

Characterization of anthropogenic and environmental pressures influencing the bacterial compartment in shallow lakes

Adélaïde Roguet

▶ **To cite this version:**

Adélaïde Roguet. Characterization of anthropogenic and environmental pressures influencing the bacterial compartment in shallow lakes. Hydrology. Université Paris-Est, 2015. English. ⟨NNT : 2015PESC1139⟩. ⟨tel-01400222⟩

HAL Id: tel-01400222

https://pastel.hal.science/tel-01400222v1

Submitted on 21 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

UNIVERSITÉ PARIS-EST

École Doctorale : Sciences, Ingénierie et Environnement

THÈSE DE DOCTORAT

Sciences et Techniques de l'Environnement

Caractérisation des pressions anthropiques et environnementales influençant le compartiment bactérien dans les lacs peu profonds

Characterization of anthropogenic and environmental pressures influencing the bacterial compartment in shallow lakes

par Adélaïde ROGUET

Thèse réalisée au Laboratoire Eau Environnement et Systèmes Urbains UMR MA 102
Dirigée par Françoise LUCAS et co-encadrée par Laurent MOULIN

Soutenue le 2 décembre 2015 à Université Paris-Est Créteil

Composition du jury :

Isabelle DOMAIZON	Directrice de Recherche, INRA	Rapporteur
Eva LINDSTRÖM	Professeur, Université d'Uppsala (Suède)	Rapporteur
Jean-François HUMBERT	Directeur de Recherche, INRA	Président
Emmanuelle CAMBAU	Professeur, CNR-MyRMA, APHP	Examinateur
Florence HULOT	Maître de Conférences, Université Paris Sud	Examinateur
Françoise LUCAS	Professeur, Université Paris Est-Créteil	Directrice de thèse
Laurent MOULIN	Docteur, Responsable R&D Biologie, Eau de Paris	Co-directeur de thèse

À Jeanne Desjouis

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to Françoise Lucas to believing in me, for her patience, her guidance and encouragements throughout the course of this thesis. These four rewarding years have made me grow. I also warmly thank Laurent Moulin for his personnal support, and this is more than anyone can expect from a supervisor.

I would like to thank the members of the jury for kindly accepting to review my work and coming to the defense.

I also would like to thank the University Paris-Est, and especially the doctoral school "Sciences, Ingénierie et Environnement" and the Laboratory Eau, Environnement et Systèmes Urbains for making this thesis possible.

Assez d'anglais... en français maintenant,

Tout d'abord, je tiens à remercier Gérard Lacroix sans qui le Projet Pulse n'aurait jamais vu le jour. Merci pour toutes tes petites phrases qui redonnent espoir, même quand il nous en reste peu. Un grand merci à Arnaud Catherine avec qui j'ai pu parler de mes résultats statistiques et des SIG. Merci à Ludwig Jardillier qui a su me dégager du temps dans son emploi du temps de ministre et pour avoir utilisé son 'joker' auprès de Ross Holland pour avoir les données de cytométrie (je pense rien que pour moi). Merci à Thomas Lerch pour avoir était mon héro d'un jour lorsque je t'ai demandé (et que tu as accepté) de m'initier à la technique de T-RFLP, car mon plan pour caractériser les communautés bactériennes par dHPLC était tombé à l'eau. Merci à Céline Bonhomme et Sophie Penisson pour m'avoir aidé à décrypter les équations du modèle neutre. Un grand merci à Pierre Legendre pour m'avoir dit que mes analyses stats étaient parfaites… mais il y a 40 ans (rassurez vous, j'ai tout refait depuis) !

Toutes ces analyses n'auraient pas été possibles sans les mesures effectuées sur le terrain. Ainsi un grand merci à ceux qui ont organisé les campagnes 49 lacs (organiser les campagnes sur un lac peut prendre du temps, alors sur 49 lacs...) et à ceux qui de près ou de loin ont participé aux campagnes 49 lacs ou Créteil. J'ai particulièrement apprécié travailler avec Jacques, pour sa beauté d'âme, avec David pour son enthousiasme et pour

son incroyable connaissance de la flore et surtout de la faune. À Philippe, Mohamed et Alex pour leur bonne humeur qu'ils ont pu apporter sur le terrain et pour tous les bons moments que nous avons partagés.

Je remercie la base de loisir de Créteil de nous avoir prêté généreusement leur bateau pendant ces deux années, la ville de Créteil pour avoir répondu favorablement à ma demande d'instrumentation du déversoir d'orage du lac de Créteil. Cette instrumentation et son suivi pendant 6 mois ont ensuite été mis en place et financés par la DSEA du conseil général du Val-de-Marne dans le cadre d'une convention. Un grand merci a Bruno, Adèle et Philippe pour le couplage du débimètre avec l'échantillonneur. Je remercie l'institut d'Aménagement et d'Urbanisme de l'Ile de France pour leurs données du MOS et plus particulièrement Manuel Pruvost-Bouvattier et Alicherif Assad. Je rends un grand merci à la ville deParis et notamment le Éric Lamelot de la Direction des Espaces Verts et de l'Environnement pour avoir accepté les prélèvements sur le lac de Daumesnil (et de nous avoir prêté gratuitement les barques pendant deux jours). Un grand merci à tous les gestionnaires et propriétaires des 49 lacs, pour avoir répondu sans réticence aux questionnaires concernant les caractéristiques et les modes de gestion de leur plan d'eau.

Évidemment cette thèse n'aurait été possible sans l'appui des membres du Leesu, qui m'ont mis dans de bonnes conditions pour accomplir ce travail et m'impliquer dans la vie du laboratoire. Je tiens tout particulièrement à remercier mon binôme microbio Claire, pour tous les fous rires que nous avons partagés, pour avoir mouillé le maillot pour moi. Merci pour ton aide quotidienne, notamment pour les filtrations, les centrifugations et les comptages et pour ton aide dans les derniers jours. Je tiens également à remercier Adèle pour nos longues discussions dans la voiture et pour tous ses petits mots d'encouragements. Un grand merci à Cécile pour sa bonne humeur, il n'y a que toi qui puisses redonner la patate aux gens comme tu le fais. Dommage que tu quittes le laboratoire... Un grand merci au reste de la cellule technique que je n'ai pas encore remercié, notamment Émilie, Lila et Damien. Enfin, merci à Catherine et Annick pour m'avoir aidé dans toutes ces tâches administratives dont vous seules connaissez les rouages.

À l'équipe R&D d'Eau de Paris, merci de m'avoir accueilli dans vos locaux comme une des vôtres.

À mes stagiaires, Grégory, Alexandre et Joeffrey, merci pour l'aide que vous m'avez apporté. Tout particulièrement Grégory pour ta bonne humeur et tous les bons moments partagés.

À tous les doctorants du laboratoire et en particulier à la dream team du LEESU, merci pour votre bonne humeur, sans vous cette thèse n'aurait pas été la même. Merci à Caroline pour toutes nos longues discussions le soir, à Romain pour tes blagues mais aussi pour ton sérieux (qui a probablement fait qu'en bas, ce ne soit pas la foire), à Mathieu pour toutes tes super théories..., à Sifax pour ton humour et ta maladresse, à Rachid pour notre complicité et à Angélique pour ta bonne humeur (et avec qui je peux parler R et couture). Je ne peux terminer ce paragraphe sans mentionner mes deux acolytes de terrain à savoir Viet et Frédéric, sans vous les campagnes n'auraient pas été les mêmes : on a franchement bien rigolé. Enfin un grand merci à Vincent et Benoit pour votre patience, notamment pour avoir pris le temps de répondre à chacune de mes questions en microbiologie et de ne pas vous être ouvertement moqué de moi lorsque j'ai réalisé ma première PCR (10 puits) en une heure. À tous, j'espère sincèrement que nous resterons en contact.

Merci pour ces deux ans d'*english therapy* passées aux côtés de Marie, Fernando, Johanna, Béatrice, Yolande, Pascal, Catherine, M'hamed, Elmiloud, Ric et bien sûr Jane.

À mes amis, et plus particulièrement à Laetitia. Merci d'avoir été là chaque semaine. Merci à Lydie, car même si nous ne nous sommes pas autant vu que je l'aurais souhaité pendant ces quatre années (comme tu l'as si bien résumé « loin des yeux, près du cœur »), la difficulté de ces dernières semaines nous ont permis de nous retrouver. Merci pour tes grands conseils de rédaction, tu devrais écrire un recueil. Un grand merci également à Chroustie, j'espère qu'une fois tout ceci terminé, nous pourrons passer d'interminables journées à prendre la faune et la flore en photo.

À toute ma famille, merci de m'avoir soutenu dans cette aventure. Merci à ma maman pour m'avoir transmit cet amour de la nature et merci à Patricia et Francis pour m'avoir accueillis chez vous pour que je puisse étudier à l'université « cette nature ». Un grand

merci à France pour avoir accepté (ou pas) que son témoin ne soit pas présent à ses côtés, mais plutôt au côté de ce présent manuscrit.

Enfin, je voudrais remercier mon compagnon, Camille Dahan, qui a toujours cru en moi, même lorsque je n'y croyais plus. Merci pour ta patience et pour ton soutien indéfectible.

TABLE OF CONTENTS

List of acronyms 12
List of figures 13
List of tables 14
Foreword 15

Bacterioplankton in shallow lakes 19

I. Shallow lakes 21
I.1 Context and characteristics 21
I.2 Shallow lakes facing anthropogenic pressures 22
I.3 Importance of the bacterial community in the ecological functioning 23

II. Biogeography of bacterioplankton in freshwater lakes 24
II.1 Context and definition 24
II.2 Conceptual biogeographical frameworks 27
II.3 Environmental, dispersal and neutral forces determining the bacterioplankton biogeography in lakes 30
II.4 Dispersal consideration 33
II.5 Biogeography depending on the taxonomic rank? 33
II.6 Spatio-temporal variability 35
II.7 Conclusion 36

III. Nontuberculous mycobacteria in freshwater habitats 36
III.1 Mycobacterial taxonomy 37
III.2 General characteristics of the mycobacteria 38
III.3 Pathogenicity of the nontuberculous mycobacteria 40
III.4 The ecology of mycobacteria in freshwater ecosystems 42
III.5 Diversity and densities of mycobacteria in freshwater 45

IV. Scope of this study 49
IV.1 Methodological considerations 51
IV.2 Chapter II and III: biogeography of the total bacterial community in shallow lakes 55
IV.3 Chapter IV and V: biogeography of the nontuberculous mycobacteria in shallow lakes 56

Neutral community model explains the bacterial community assembly in freshwater lakes 59

I. Abstract 62
II. Introduction 63
III. Materials and methods 64
III.1 Study area and sampling 64
III.2 DNA extraction 67
III.3 T-RFLP analysis 67
III.4 Bacterial enumeration 68
III.5 Environmental and meteorological parameters 69
III.6 Watershed land use assessment 69
III.7 Data analysis 70
IV. Results 72
IV.1 Spatial distribution of the BCS at the regional scale 72
IV.2 Spatial and temporal monitoring in Créteil Lake 74
V. Discussion 76
V.1 Fingerprinting methodology considerations 77

V.2 Minor influence of deterministic processes on the BCS at the regional scale 77
V.3 Bacterial assembly shape by neutral processes at the regional scale 78
V.4 Influence of deterministic and neutral processes at the local scale 80
VI. Acknowledgements ***81***
VII. References ***82***
VIII. Supplementary data ***89***

Spatial distribution processes differ between bacterial taxonomic groups in freshwater lakes **93**
I. Abstract ***96***
II. Introduction ***97***
III. Materials and methods ***98***
III.1 Study area and sampling 98
III.2 DNA extraction 99
III.3 Illumina sequencing of the V3-V4 region of the 16S rRNA gene 100
III.4 Environmental parameters 101
III.5 Watershed land use assessment 101
III.6 Data analysis 101
IV. Results ***103***
IV.1 Bacterial richness 103
IV.2 Composition of the bacterial assemblages 104
IV.3 Spatial structure of the total BCC 106
IV.4 Spatial structure of the dominant phyla 107
V. Discussion ***110***
V.1 Bacterial community composition in the 49 lakes 110
V.2 Small importance of local and regional factors on the spatial BCC 110
V.3 Stochastic processes and structure of the bacterial metacommunity 111
V.4 Local and dispersal processes effects depend on the bacterial group 112
VI. Acknowledgments ***113***
VII. References ***113***
VIII. Supplementary data ***120***

High mycobacterial diversity in recreational lakes **121**
I. Abstract ***124***
II. Introduction ***125***
III. Materials and methods ***127***
III.1 Sampling sites 127
III.2 Sample collection 127
III.3 Sample processing 127
III.4 Quantitative real-time PCR 129
III.5 Illumina sequencing of the V2-V3 region of the 16S rRNA gene 129
III.6 Data analysis 130
IV. Results ***131***
IV.1 Quantification of nontuberculous mycobacteria 131
IV.2 Richness of nontuberculous mycobacteria 133
IV.3 Composition of the mycobacterial assemblages 134
IV.4 Potential opportunistic pathogens 136
IV.5 Comparison of the mycobacterial assemblages 136
V. Discussion ***137***
V.1 Ubiquity of nontuberculous mycobacteria 137

V.2 Nontuberculous mycobacteria dominated by fast-growing species 137
V.3 Mycobacterial densities and diversity differed among compartments 139
V.4 Potential opportunistic pathogens 140
VI. Acknowledgments ***142***
VII. References ***142***
VIII. Supplementary data ***150***

Mycobacteria are everywhere, but, only the environment selects? 155
I. Abstract ***158***
II. Introduction ***159***
III. Materials and methods ***161***
III.1 Study area and sampling 161
III.2 DNA extraction and quantification of mycobacterial densities 164
III.3 PCR assay performance 165
III.4 Physicochemical and meteorological parameters 165
III.5 Data analysis 166
IV. Results ***167***
IV.1 Assay performances 167
IV.2 Mycobacterial distribution at the regional scale 167
IV.3 Spatio-temporal variations of the mycobacterial densities in Créteil Lake 168
IV.4 Rainfall impact on the bacterial densities in Créteil Lake 169
V. Discussion ***170***
V.1 Methodological considerations 170
V.2 Widespread of NTM in freshwater lakes 170
V.3 Parameters affecting the NTM densities at the regional scale 171
V.4 Local rainfall impact 172
V.5 No temporal variation in Créteil Lake 173
VI. Acknowledgments ***174***
VII. References ***174***
VIII. Supplementary data ***181***

Synthesis and general discussion 183
I.1 Bacterial biogeography at different spatial scales 185
I.2 Biogeography patterns depending on the bacterial taxonomic level 188
I.3 Impact of environmental pressures on the bacterial biogeography 191

References 195
Glossary 219
Appendix 221

Publications and communications *235*
Résumé - Abstract *236*

LIST OF ACRONYMS

ANOVA	Analysis of variance
ARISA	Automated ribosomal intergenic spacer analysis
atpE	Gene encoding for the adenosine-5'-triphosphate subunit C
BCC	Bacterial community composition
BCS	Bacterial community structure
CFU	Colony-forming unit
Chl *a*	Chlorophyll *a*
C_q	Cycle threshold
DGGE	Denaturing gradiant gel electrophoresis
DNA	Desoxyribonucleic acid
DOC	Dissolved organic carbon
DOM	Dissolved organic matter
FGM	Fast-growing mycobacteria
hsp65	Gene encoding for the 65 kDa stress-heat shock protein
ITS	Internal transcripted spacer
LM	Linear model
LMM	Linear mixte model
MAC	Mycobacterium *avium* complex
MAIS	Mycobacterium *avium-intracellulare-scrofulaceum* complex
NCM	Neutral community model
nidA	Gene encoding for the ring-hydroxylating dioxygenase α subunit
nidB	Gene encoding for the ring-hydroxylating dioxygenase β subunit
N_T	Metacommunity size
NTM	Nontuberculous mycobacteria
OTU	Operational taxonomic unit
PAH	Polycyclic aromatic hydrocarbons
PCNM	Principle coordinates of neighbour matrices
PCR	Polymerase chain reaction
POC	Particulate organic carbon
POM	Particulate organic matter
PULSE	Peri-urban lakes, society and environment
qPCR	Real-time quantitative PCR
RDA	Redundancy analysis
RDP	Ribosomal database project
RNA	Ribonucleic acid
rpoB	Gene encoding for the RNA-polymerase
rRNA	Ribosomal ribonucleic acid (16S rRNA)
rrs	Gene encoding for the component of the small subunit of ribosomal ribonucleic acid (so-called16S rRNA)
SGM	Slow-growing mycobacteria
SSO	Storm sewer outlet
STI	Space-time ANOVA models without replications
T-RF	Terminal restriction fragment
T-RFLP	Terminal restriction fragment length polymorphism
TS	Trophic status
TSS	Total suspended solids
UPGMA	Unweighted pair group method with arithmetic mean
UV	Ultraviolet
VPA	Variance partitioning analysis

LIST OF FIGURES

Figure 1 Simplified food web diagram in lacustrine pelagic zone.
Figure 2 Bacterial community composition in freshwater lake epilimnia.
Figure 3 Contributions of environmental and dispersal factors on the bacterial biogeography.
Figure 4 Relationship between the different scenarios for community assembly proposed by Martiny and colleagues (2006) and the metacommunity frameworks (Leibold *et al*. 2004).
Figure 5 Relationship between the bacterial community similarity and the geographic distance between lakes, i.e. distance-decay relationship.
Figure 6 Relationship between the observed and estimated bacterial community similarity and geographic distance between lakes.
Figure 7 Niche differentiation at the phylum and class levels.
Figure 8 Numbers of approved *Mycobacterium* species and subspecies from 1896 to 2010.
Figure 9 Schematic representation of the cell envelope of mycobacteria.
Figure 10 Comparison of phylogenetic trees based on the *rrs* gene, *hsp65* gene, and *rpoB* gene sequences of the genus *Mycobacterium*.
Figure 11 Number of publications relating the occurrence of the mycobacterial species in water samples from freshwater habitats.
Figure 12 Phylogenetic tree based on *rrs* gene of the genus *Mycobacterium*.
Figure 13 Experimental strategy followed in this study.
Figure 14 Location of the 49 lakes in the Paris area (France).
Figure 15 Sampling strategies in the Créteil Lake.
Figure 16 Relationship between the detection frequency of T-RFs and their mean relative abundance from the three years of monitoring.
Figure 17 Temporal variations of the bacterial community beta diversity, turnover and nestedness determined by comparing pairwise successive sampling campaign (n = 24).
Figure 18 Location of the 49 lakes in the Paris area (France).
Figure 19 Relative abundances per lake (percentage of total sequences) of phylogenetic phyla identified by Illumina MiSeq sequencing.
Figure 20 Relative abundances per lake (percentage of total sequences) of phylogenetic class for (a) *Proteobacteria* and (b) *Bacteroidetes* identified by Illumina MiSeq sequencing.
Figure 21 Location of the five stations in Créteil Lake (a) and Daumesnil Lake (b).
Figure 22 Comparison of the *atpE* gene copy numbers between Créteil Lake (white boxplot) and Daumesnil Lake (grey boxplot) for water (a), sediment (b) and biofilms compartments (c) (n = 10).
Figure 23 Heatmap dendrogram illustrating similarities of nontuberculous mycobacteria assemblage composition for the five compartments Créteil (C) and Daumesnil (D) Lakes.
Figure 24 Location and mycobacterial genome unit densities within the 49 lakes in Paris area (France).
Figure 25 Sampling strategies in the Créteil Lake.
Figure 26 Variations of genome unit densities along the two-year monitoring in Créteil Lake for the pooled sample M.
Figure 27 Distribution of the mycobacterial genome units per liter for each of the three stations in Créteil Lake for the six rain events.
Figure 28 Partitioning microbial community variance among meteorological, environmental and biological sets of variables for different lakes.

LIST OF TABLES

Table 1 Example of dispersal mechanisms in freshwater environment.

Table 2 Types of infections and causative agents. Compiled from mycobacteria reviews.

Table 3 Densities of mycobacteria in diverse freshwater habitats.

Table 4 Characteristics of Créteil Lake.

Table 5 Variation partitioning analysis of the bacterial community structure for the three years of sampling.

Table 6 Comparison of sequenced libraries, including the number of reads pre- and post-trimming and subsampling, the observed richness, the diversity index and the evenness index for each lake.

Table 7 Variation partitioning analysis of the bacterial community composition.

Table 8 Neutral community model parameters.

Table 9 Comparison of sequenced libraries, including the number of reads pre- and post-trimming, the observed richness of OTUs, the estimated richness, the diversity index and the evenness index for each sample.

Table 10 Characteristics of Créteil Lake.

FOREWORD

Bacteria play a pivotal role in the functioning of aquatic ecosystems owing to their involvement in biogeochemical cycles (Lindeman 1942; Torsvik *et al.* 2002). However, their investigation in natural habitats has been effective only with the recent advent of molecular tools. Such methodologies have provided a major improvement in the comprehension of the behavior of the bacterial compartment, which was considered hitherto as a black box. In the 90s, it was estimated that less than 1% of the bacterial diversity was described (Torsvik *et al.* 1990; Amann *et al.* 1995). This observation is still valid, although a tremendous effort is made to improve the microbial diversity inventory, specially using the next-generation sequencing methods. Molecular tools combined with field and experimental approaches allowed a better comprehension of the ecology of bacterial communities. These approaches revealed that bacterial community composition as well as abundance can vary spatially and temporally in response to similar processes to those developed for macroorganisms (Leibold *et al.* 2004; Martiny *et al.* 2006). Despite some differences, microorganisms and macroorganisms have several fundamental aspects in common, and ecological theories developed in vegetal and animal ecology could serve as a framework to better understand the dynamic and the role of microbial diversity in the functioning of ecosystems (Allison and Martigny 2008).

Differences among bacterial assemblages, i.e. composition and diversity, could have possible impacts on ecosystems functioning despite the important functional plasticity of bacteria (Schimel and Gulledge 1998; Bell *et al.* 2005). The link between biodiversity and ecosystem functions has emerged as a major issue in ecology research, leading to consequent debates and controversies (Loreau *et al.* 2001). The positive link between diversity and an ecosystem function was first observed in a long-term experiment performed at Cedar Creek on vegetal communities (Tilman et al. 1996). It was later confirmed in microcosm that a positive link could be found between aquatic bacterial diversity and respiration (e.g., Bell *et al.* 2005). A series of long-term experiments demonstrated that anthropogenic factors that affect the vegetal biodiversity may affect the ecosystem stability and its resilience to environmental changes (Hautier *et al.* 2015). As a consequence changes in biodiversity may determine how global environmental changes

affect the functioning of ecosystems (Hautier *et al.* 2015). Such relationship has deep consequences for the different services provided by the ecosystems to the human society.

It is crucial to improve our comprehension of the factors that govern the distribution and dynamics of bacterial species and the impact of the bacterial diversity changes over the functioning of the ecosystems. Microorganisms have played a primordial role in the co-evolution of the geosphere and the biosphere, and such interactions are still operating. Our society depends on these processes to assure the quality of water, the supply of food, and to control diseases. However, human activities have profound effects on the biodiversity and ecosystem functioning through nutrient inputs that modify the biogeochemical cycles, transformation of the land, and enhancement of biota mobility. Global environmental changes have the potential to exacerbate the ecological and societal impacts of changes in biodiversity (Chappin III *et al.* 2000). By integrating pressures on their watershed, shallow lakes are likely impacted by a large variety of global changes that act at the local (urbanization, intensification of the agriculture practices, etc.) and at the global scale (atmospheric pollution, climatic changes altering rainfall patterns, CO_2 and temperature elevations, etc.). Aquatic bacterial communities may potentially respond to these changes, which could affect their biogeographic patterns. Thus, a better understanding of the factors and processes responsible for the bacterial biogeographic patterns would improve our knowledge of these aquatic systems and how they may respond facing environmental changes.

In this context, the aim of this thesis was to gain a more comprehensive understanding of factors and processes shaping the spatial and temporal biogeography of the bacterial compartment in a set of shallow lakes. Two bacterial levels were investigated: (i) the total bacterial community and (ii) a specific bacterial group that are natural inhabitants of freshwater habitats, i.e. the nontuberculous mycobacteria.

The thesis manuscript is organized in six parts. After synthetizing in *Chapter I* the current knowledge on bacterial biogeography and nontuberculous mycobacteria in freshwater habitats, *Chapters II* and *III* will present the results on the biogeography of the total bacterial community at the local and regional scales. *Chapters IV* and *V* will present the results on the factors affecting the distribution and diversity of mycobacteria at the

local and regional scales. A general discussion (*Chapter VI*), will synthetize the results obtained at different spatial and temporal scales and at different levels of bacterial community organization. Perspectives will be discussed.

Chapter I

Bacterioplankton in shallow lakes

I. Shallow lakes

I.1 Context and characteristics

Freshwaters are essential to the establishment of civilizations. Historically, human settlements were preferentially set up close to reliable water supply, such as lakes, for their basic needs (e.g., drinking and cleaning), alimentation (e.g., irrigation or fishing) or recreational purposes (e.g., swimming, bathing) (Bolund and Hunhammar 1999; Postel and Carpenter 1997). Furthermore, among freshwater ecosystems, lakes hold an impressive hotspot of biodiversity and endemic species including macroinvertebrates or macrophytes compared with other lotic habitats, e.g., rivers or streams (Williams *et al*. 2004; Biggs *et al*. 2005).

Historically, greater efforts were performed on the study of large lakes. Yet, they represent less than 1% of the lakes with a surface area higher than 1 ha. Indeed, the majority of water bodies are small, with more than 26 million lakes on Earth with a surface area comprised from 1 to 100 ha (Downing *et al*. 2006; Verpoorter *et al*. 2014). Moreover, per unit of water area, small lakes have higher species richness compared to large lakes, notably because of their important littoral area.

Small lakes mainly associated with shallow depths (Cooke *et al*. 2005) contrast deeply with deep lakes in terms of both morphometry, but also in their ecological functioning (Scheffer 2004). Indeed, in deep lakes, upper water layer (epilimnion) is sometimes isolated from the colder deep water (hypolimnion), preventing large interactions of the upper water layer with the sediment. Therefore, exchanges of nutrients, organisms and gas diffusion are reduced between these compartments (Dodds and Whiles 2010; Padisák and Reynolds 2003). Inversely, the lack of relevant thermal stratification in shallow lakes leads to the absence of thermal barrier between the upper and the deeper waters. Consequently, the entire water column is almost always mixed and constant interactions occur between the water and the sediment compartments (Scheffer 2004). Although rare, thermal stratifications could occur during few hours to few weeks, especially in summer. Shallow lakes are thus usually considered **polymictic**[1], as a

[1] Definitions of all words in bold are found in the glossary at the end of this document.

consequence water mixing is frequent, periods of stratification are relatively shorts and not necessarily continuous. These stratified structures could vanish with wind disturbances or substantial increase of insolation (Padisák and Reynolds 2003). Exchanges between the water column and the sediment have major impact on the biogeochemical processes (by diffusion or wind-induced sediment resuspension of nutrient, e.g., nitrogen or phosphorus) (Thomas and Shallenberg 2008), and/or the vertical and horizontal distribution of organisms (e.g., **bacterioplankton** i.e. the **planktonic** bacteria that drifts in the water column) by the increase of nutrient supplies or their resuspension from sediments (Weithoff *et al*. 2000; Bai and Lung 2005).

I.2 Shallow lakes facing anthropogenic pressures

As they are linked to their watershed, lakes are good integrators of both terrestrial and atmospheric processes, thus, they are considered useful 'sentinel systems' of the environmental pressures (Williamson *et al.* 2008). These ecosystems are thus facing multiples anthropogenic pressures, especially in small and shallow lakes where pressures may be magnified due to their specific characteristics. First, shallow lakes present high surface/volume ratio compared with deeper lakes (Chapman and Bolen 2015). Second, they display intense water-sediment interplays that constantly reload large amount of nutrients in the water column. Finally underwater solar radiation can reach the bottom of the water column in shallow lakes, which could enhance the primary production due to the photodegradation of recalcitrant dissolved organic carbon into inorganic forms (de Haan 1993).

Among anthropogenic pressures, man-made eutrophication (mostly originating from agricultural practices, industrial activities and domestic wastewater discharges) may result from large **allochthonous** nutrient and micropollutant inputs in aquatic environments that could enhance algal and cyanobacterial bloom (Dodds *et al.* 2009) and bacterial production (Jansson *et al*. 2000; Kritzerg *et al*. 2006). Pesticides and herbicides may affect lake functioning through their negative effect on all **primary producers** (Seguin *et al*. 2001). Furthermore, runoff discharges from livestock, storm sewer or combined sewer outlets can introduce exogenous micropollutants, nutrients, organic matter and organisms in lacustrine ecosystems (Fig. 1). For instance, **waterborne** pathogens (e.g., the

human protozoan parasite *Cryptosporidium parvium* or the bacteria *Escherichia coli*) may be present in high numbers in rain runoffs and sewer effluents (Brookes *et al.* 2004; Nevers and Whitman 2008). All these pressures can profoundly modify the community structure and functioning and therefore, goods and services that they provide to human society could be altered (Dodds *et al.* 2009).

I.3 Importance of the bacterial community in the ecological functioning

Heterotrophic bacteria play an essential role in the ecological functioning of the lacustrine ecosystems (Lindeman 1942), especially by the regeneration and mobilization of nutrients within the microbial loop to higher trophic levels (Fig. 1) (Mostajir *et al.* 2011). Through the bacterial metabolism, dissolved and particulate organic matter originating from phytoplankton exudates or zooplankton grazing are incorporated into the bacterial biomass, and then transferred to higher trophic levels by the protist predation (Azam *et al.* 1983). Protists are then consumed by zooplankton that in turn will feed large metazoan predators. Among protists, heterotrophic nanoflagellates, **myxotrophic** flagellates and

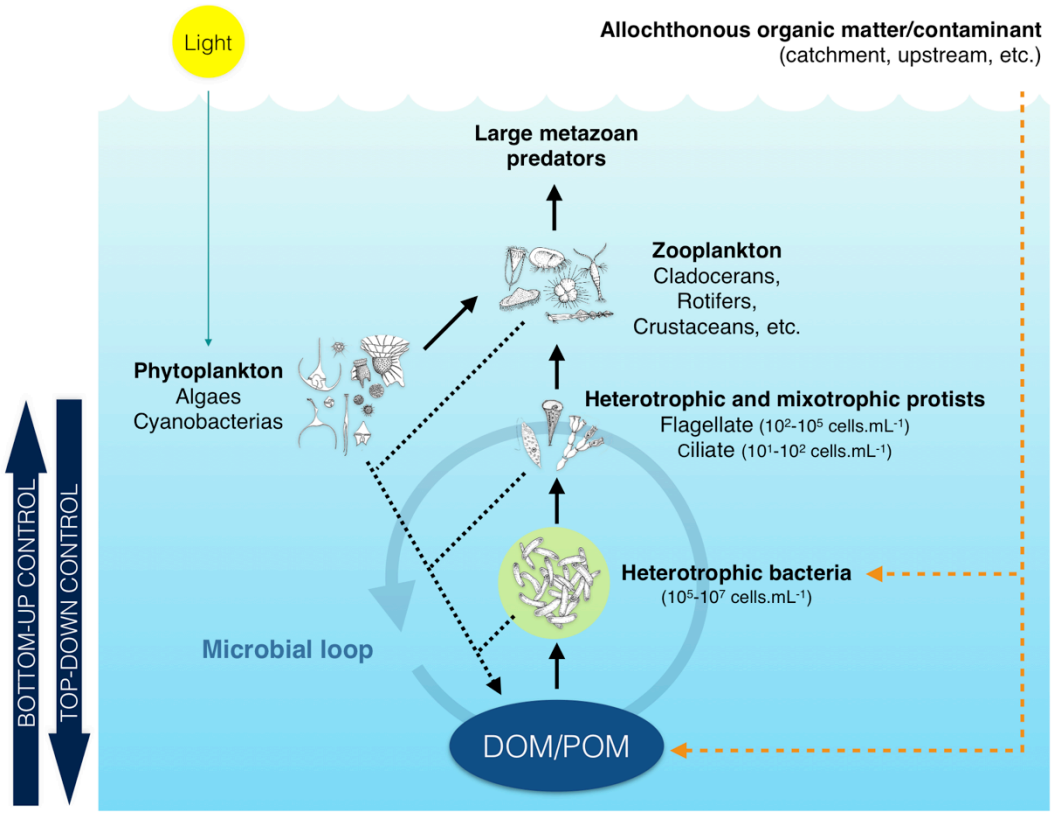


Figure 1 Simplified food web diagram in lacustrine pelagic zone. Solid black lines: predation interactions; dotted black lines: release of organic matter. DOM, dissolved organic matter; POM, particulate organic matter. Modified from Meyer (1994) and Mostajir and colleagues (2011).

ciliates (that are important grazers of heterotrophic nanoflagellates (Weisse 1991) have been observed as major consumers of the heterotrophic bacteria (Psenner and Sommaruga 1992; Domaizon *et al.* 2003; Chróst *et al.* 2009).

Several evidences indicate that the microbial loop could be regulated by both top-down and bottom-up controls (Fig. 1) (Jardillier *et al.* 2004; Muylaert *et al.* 2002). Regarding the bacterioplankton, bottom-up control refers to the limitation of bacteria by nutrients derived from allochthonous or **autochthonous** inputs, primary production and heterotrophic production. Inversely, top-down control refers to the limitation of bacteria by nano- or microplankton grazing pressures (Pace and Cole 1994). From these definitions, we could easily hypothesize that in **oligo**- and **mesotrophic** lakes, bacterial communities might be shaped by bottom-up control due to nutrient limitations whereas in **eutrophic** lakes, in absence of nutrient restrictions, top-down control might have a higher importance. However, no clear consensus supports this assumption (e.g., Muyleart *et al.* 2002; Chróst *et al.* 2009), suggesting that mechanism governing the aquatic food web may be far too complicated to be only described by top-down or bottom-up control.

II. Biogeography of bacterioplankton in freshwater lakes

II.1 Context and definition

The composition of the bacterioplankton assemblages has been extensively studied in lakes over the past decade, as noticed by the reviews of Zwart and colleagues (2002) and Newton and colleagues (2011). Opening the bacterial community black box has been possible due to the development of the culture-independent methods based on the sequences encoding for the rRNA gene (Hugenholtz *et al.* 1998). The 16S rRNA gene encoding for the component of the small subunit of prokaryotic ribosomes is commonly used for the identification and classification of bacteria due to the compulsory presence of this housekeeping gene for cell life and the slow changes over time of this sequence, permitting the detection of relatedness among species (Woese and Fox 1977). Since the appropriation of these molecular tools by microbial ecologists in the late 1990s, the number of 16S rRNA gene sequences collected from the epilimnia has steadily increased from 150 sequences to more than 11,000 sequences in 2011 (Newton *et al.* 2011). More

recently, the advent of high-throughput sequencing methods has contributed to a tremendous increase in the total number of 16S rRNA gene sequences collected from these habitats allowing a better in depth analysis of the bacterial diversity and community structure (see *Chapter I*, *section IV.1.ii*).

The first meta-analysis performed with 622 bacterial sequences collected in ten lakes from different continents revealed that most of the freshwater bacteria were distinct from bacteria in neighboring environments such as soil, sediment or marine habitats (Zwart *et al*. 2002). The same study observed that lacustrine bacterial **communities** (i.e. assemblage of taxa found together in a specific habitat at a certain time, interacting with other in this area) were dominated by five **phyla**, i.e. the *Proteobacteria* (36% of the total sequences), the *Actinobacteria* (19%), the *Cytophaga-Flavobacterium-Bacteroides* (15%) (currently known as *Bacteroidetes*), the *Cyanobacteria (9%)* and the *Verrucomicrobia* (8%). Ten years later, the same phyla were described by Newton and colleagues (2011) as dominant in the bacterioplakton within a meta-analysis carried out with data from 69 published papers (Fig. 2).

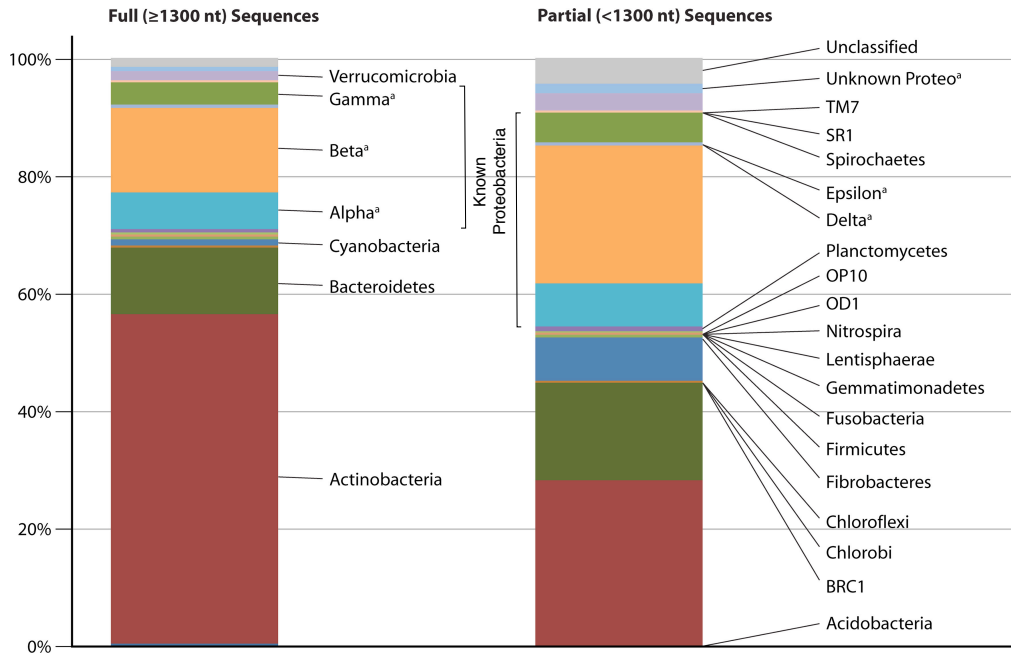


Figure 2 Bacterial community composition in freshwater lake epilimnia. Sequences (11,400 sequences) were taxonomically classified by the RDP classifier with a cutoff of 75%. Phyla listed are present in both bar plots but are listed on only one side to aid in the visualization of the plots. [a]Sequences assigned to the phylum *Proteobacteria* with a cutoff of 75%. From Newton and colleagues (2011).

Although similar global tendencies regarding the proportion of phyla were observed between these two meta-analyzes, dissimilarities (variations) among lakes in the bacterial community composition cannot be excluded at the phylum level. Indeed, conclusive supports highlighted the existence of spatial (Martiny *et al*. 2006; Hanson *et al*. 2012; Lindström and Langenheder 2012) and temporal (Boucher *et al*. 2006; Jones *et al*. 2012) variations within bacterial assemblages, among and within lakes, i.e. the study of bacterial biogeography seems to make sense.

Biogeography is the study of the distribution of organisms over space and time (Lomolino *et al.* 2006; Martiny *et al.* 2006). It aims to gain an insight into processes that cause the differences in community composition between sites, i.e. the **beta diversity**. Thus, biogeography intents to answer to "who is where, at what abundance, and why?" (Martiny *et al.* 2006), one of the fundamental questions in ecology (Sutherland *et al.* 2013).

Four fundamental ecological processes are recognized to drive the diversity and composition of the bacterial communities, i.e. the speciation, selection, ecological drift and dispersal (Vellend 2010). **Speciation** refers to the mechanisms (i.e. individual drift or selection) that create new taxa. **Selection** is the ability of taxa to survive and reproduce facing abiotic or biotic interactions (i.e. competition, predation or mutualism). **Ecological drift** refers to processes that randomly change the abundance of taxa, while the **dispersal** mechanisms (some examples are described in Table 1) may reflect the displacement of taxa to a new location. These ecological processes are at the root of the mechanisms responsible for biogeographical patterns. However, as it is difficult to evaluate separately the contribution of each of these processes (e.g., speciation) (Hanson *et al.* 2012),

Table 1 Example of dispersal mechanisms in freshwater environment. From Lindström and Langenheder (2012).

Dispersal mechanisms	Relevant scales
Dispersal via air, aerosols or particles	From meters to global scales
Dispersal between system types via terrestrial surface (storm sewer outlet) or groundwater	From meters to kilometers
Dispersal within system types with water masses (connection of lakes to river)	From meters to continents
Dispersal with living vector	From micrometers to global scales (depending on the range of the vector)
Dispersal by direct contact (contact with human host)	Meters

biogeographical frameworks (that often combine multiple ecological processes) have been established to improve our comprehension of the processes shaping the bacterial community assemblages.

II.2 Conceptual biogeographical frameworks

One early attempt to explain the microbial distribution at the regional scale was the famous statement "Everything is everywhere, but, the environment select", the so-called Baas Becking hypothesis (Beijerinck 1913; Baas Becking 1934). "Everything is everywhere" assumes that dispersal capability of microorganisms is high enough to erase historical biogeography patterns, and "but, the environment select" suggests that the bacterial assemblages are shaped by local environmental conditions. This perception that bacterial communities were governed by local interactions persisted until the early 2000s. Indeed, with the advent of DNA-based molecular technologies, substantial evidences of the impact of dispersal processes on bacterial distribution were encountered (Lindström and Bergström 2004; Reche *et al.* 2005; Yannarell and Triplett 2004).

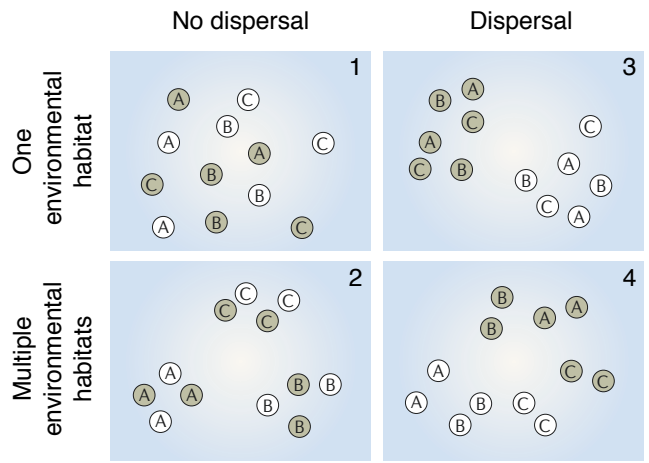


Figure 3 Contributions of environmental and dispersal factors on the bacterial biogeography. The biogeographical pattern of the four frameworks proposed by Martiny and colleagues (2006). Each bacterial community is symbolized by a circle; habitats are represented by a letter (A, B and C); different locations have a distinct color (green vs white). The closer the circles are, the more similar are the bacterial communities. Modified from Martiny and colleagues (2006).

Based on environmental and dispersal factors, Martiny and colleagues (2006) proposed four frameworks to explain the differences in the microbial assemblages at the regional scale: (i) the absence of influence of both environmental and dispersal factors on the bacterial communities, thus the bacterial assemblage appears to be randomly distributed over space, (ii) the unique influence of the local environmental conditions (i.e. Baas Becking hypothesis), (iii) the unique influence of dispersal processes often characterized with the geographic distance

between lakes and (iv) the influence of both environmental and dispersal processes. Figure 3 illustrates the biogeographical patterns observed for these four frameworks. If bacterial communities are neither influenced by environmental, nor dispersal factors, bacterial communities should be randomly distributed among habitats (scenario 1). Alternatively, if communities are clustered by local conditions, it could be concluded that bacterial communities are mainly governed by local conditions (scenario 2). If communities are clustered by geographical location, it could be assumed that bacterial communities are mainly governed by dispersal processes (scenario 3). Finally, if bacterial communities are influenced by both environmental and dispersal factors (scenario 4), communities are clustered by habitat locations and by habitat types.

The importance of dispersal mechanisms on the community composition was earlier considered and subsequently resumed into four **metacommunity** frameworks (Leibold *et al.* 2004), the metacommunity being defined as a set of local communities that are linked by dispersal (Hanski and Gilpin 1991): (i) **patch dynamic** paradigm, which assumes that diversity is determined by dispersal or local extinction and colonization among environmentally identical patches, and is not affected by the environmental conditions, (ii) **species sorting** paradigm, which assumes that dispersal is high enough to allow immigration to all patches and that community composition is governed by the local conditions of the environments/patches (iii) **mass effects** paradigm, which assumes that massive immigration can rescue species from competitive exclusion, the impact of the local environmental conditions being thus negligible, and (iv) **neutral** paradigm, developed by Hubbell (2001), assumes that all taxa are similar in their competitive and dispersal ability (i.e. they have similar fitness, as there is a functional equivalency among trophically similar taxa), and that the community composition is only govern by a stochastic balance between species loss (extinction, emigration) and gain (immigration, speciation).

Some overlaps and specificities were observed between the frameworks proposed by Martiny and colleagues (2006) and Leibold and colleagues (2004). Indeed, the second scenario defined by Martiny and colleagues (already noticed as similar of the Baas Becking hypothesis) is the equivalent of the species sorting paradigm proposed by Leibold and

colleagues (Fig. 4a). According to Lindström and Langenheder (2012), influence of species sorting must be enhanced if the bacterial communities are mainly composed of specialist taxa (Fig. 4b). Although some frameworks appeared to be similar, large differences exist among them. Indeed, in the third scenario, Martiny and colleagues (2006) only consider dispersal limitation, while Leibold and colleagues (2004) did not refer to dispersal limitation but instead described the concept of mass effects. The consequence of both mechanisms on the beta diversity is however the same: differences in the community composition at the global scale are enhanced with geographic distance at both low (dispersal limitation) and high (mass effects) dispersal rates (Lindström and Langenheder 2012) (Fig. 4a).

Connectivity between lakes could partly explain the difference of dispersal rate (Fig. 4b). In isolated water bodies (patches), dispersal should be limited, whereas in connected lakes (to a river or a storm sewer outlet) dispersal should be favored. It is also important to notice that the random scenario 1 from Martiny and colleagues (2006) is not identical to the neutral paradigm, although in both cases, bacterial community seems randomly distributed. Indeed, random scenario predicts a pure random assemblage of the bacterial community due to an absence of environmental and dispersal influences,

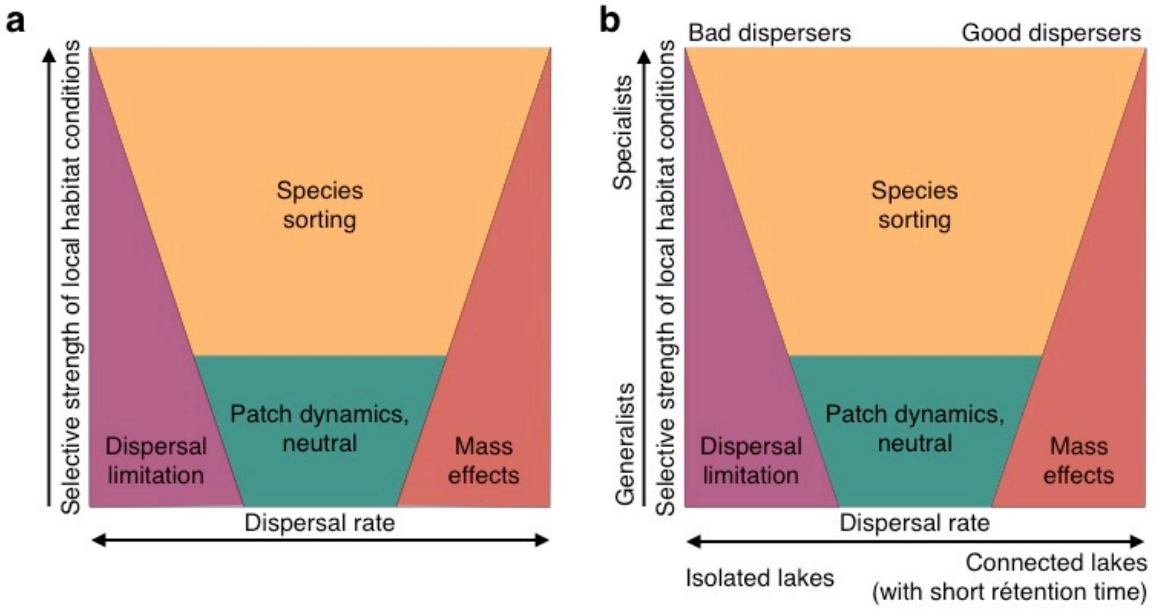


Figure 4 Relationship between the different scenarios for community assembly proposed by Martiny and colleagues (2006) and the metacommunity frameworks (Leibold *et al.* 2004). (a) Different mechanisms are assumed to act depending on the selective strength by the local environmental conditions and the rate of dispersal among communities. (b) Different mechanisms could potentially be differently important depending on type of environment and study organism. From Lindström and Langenheder (2012).

while in the neutral model patterns the beta diversity must be predicted by the occurrence frequency and the average abundances of taxa in a metacommunity (Sloan *et al.* 2006). Neutral model specificity come from the fact those taxa are considered to have the same fitness, i.e. there is a functional equivalency among trophically similar taxa (Hubbell 2001). Moreover, conclusive supports highlighted that the random framework proposed by Martiny and colleagues (2006) is obsolete as a growing number of studies found evidence of the importance of the environmental and/or dispersal factors (e.g., reviews from Tamames *et al.* 2010; Hanson *et al.* 2012; Lindström and Langenheder 2012).

Over the past decade the utilization of community assembly rules that were originally developed from the study of macroorganisms has allowed improving our knowledge in the processes governing the bacterial biogeography. From the frameworks advanced by Martiny and colleagues (2006) and Leibold and colleagues (2004), three processes appeared to be relevant in shaping the bacterial biogeography. These processes include two deterministic forces, i.e. the influence of local environmental conditions and the dispersal-related mechanisms (i.e. dispersal limitation and mass effects), and the neutral processes, a stochastic force.

II.3 Environmental, dispersal and neutral forces determining the bacterioplankton biogeography in lakes

Among local environmental conditions, numerous factors, almost all abiotic, are frequently identified as shaping the bacterial communities. These factors include water temperature (e.g., Muyleart *et al.* 2002; Wu and Hahn 2006; Shade *et al.* 2007), osmotic condition, i.e. water pH and conductivity (e.g., Kent *et al.* 2007; *Lindström et al.* 2005; Székely and Langenheder 2014), solar radiation (e.g., Warnecke *et al.* 2005; Liu *et al.* 2006), nutrients availability or primary production (e.g., Crump *et al.* 2003; Eiler *et al.* 2003; Yannarell *et al.* 2003; Yannarell and Triplett 2004; Kent *et al.* 2007) and grazing pressures (Hahn and Höfle 2001). All these factors could have direct or indirect influences on the bacterial assemblages. For instance, the water pH can directly regulate the overall growth patterns of bacterial populations (e.g., Langenheder *et al.* 2006) or indirectly influence the bacterial community through modification of the molecular structure conformation of organic compounds (Murphy *et al.* 1994).

Dispersal limitation has mostly been assessed using geographic distance between lakes (e.g., Crump *et al.* 2007; van der Gucht *et al.* 2007; Barberán and Casamayor 2010): the closer are the lakes, the more similar are the communities (Fig. 5), i.e. the distance-decay relationship (Green *et al.* 2004; Soininen *et al.* 2007). More recently, dispersal limitation is also characterized regarding the spatial structuring of the bacterial community using matrices (e.g., Drakare and Liess 2010; Székely and Langenheder 2014; Liu *et al.* 2015), such as principal coordinates of neighbor matrices (PCNM). Such matrices are composed of multiple vectors, which altogether reflect the bacterial community spatial pattern (Borcard and Legendre 2002). Figure 6 illustrates the partitioning of the observed spatial pattern of a bacterial communities (Fig. 6a) into three estimated patterns (PCNM 1, 2, and 3) (Fig. 6b,c,d). Mass effects are either directly assessed by evaluating the degree of similarities between autochthonous and allochthonous (from a river for instance) bacterial assemblages (Lindström *et al.* 2006; van der Gucht *et al.* 2007; Nelson *et al.* 2009) or indirectly by assessing the connectedness between lakes and potential sources of bacteria such as rivers, runoff or sewer discharge (van der Gucht *et al.* 2007). Neutral processes were evidenced using the model of Sloan and colleagues (2006) in various freshwater habitats (e.g., Drakare and Liess 2010; Östman *et al.* 2010; Logares *et al.* 2013). This model, modified from Hubbell's model (2001), is adapted to large population size, including bacterial communities, and allows taxa to have competitive advantage or disadvantage compare to other taxa.

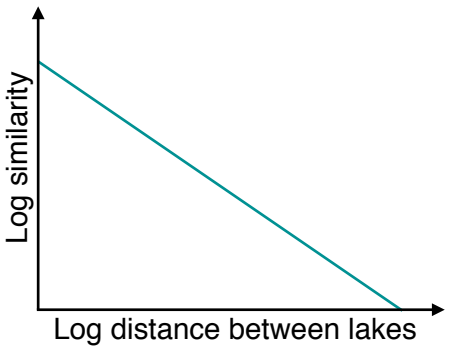


Figure 5 Relationship between the bacterial community similarity and the geographic distance between lakes, i.e. distance-decay relationship. From Soininen and colleagues (2007).

Although there are evidences in lacustrine ecosystems of the influence of all processes previously mentioned (i.e. local environmental conditions, dispersal mechanisms and neutral processes), Östman and colleagues (2010) suggested that the relative importance of environmental and neutral processes should depend on the environmental heterogeneity between lakes: greater environmental gradient should enhance the importance of local conditions. Logue and Lindström (2008) presented in

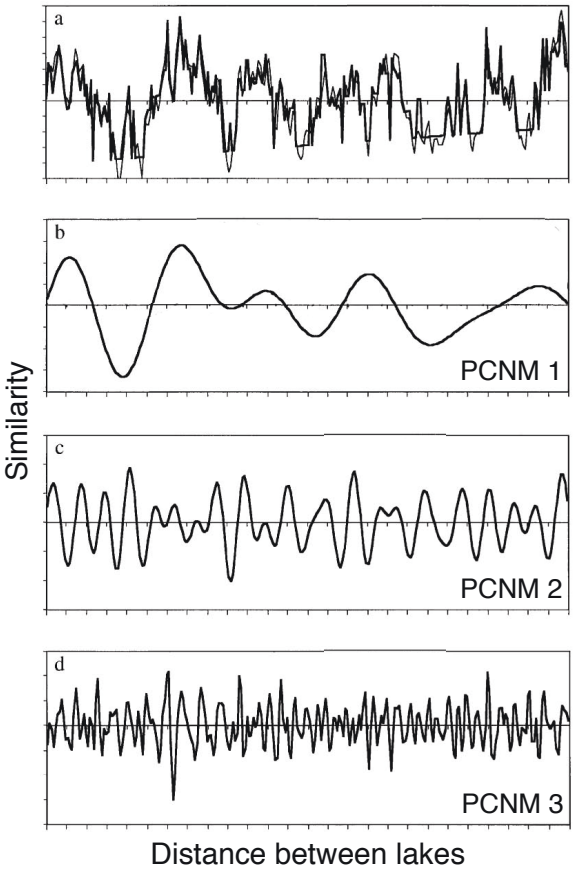


Figure 6 Relationship between the observed and estimated bacterial community similarity and geographic distance between lakes. (a) Observed spatial distribution (thin line); parsimonious estimated spatial distribution using PCNM model (combination of the three PCNM vectors 1, 2 and 3). (b, c and d) PCNM vectors representing the estimated spatial distribution at distinct spatial scale: (b) broad-scale, (c) intermediate-scale, (d) fine-scale models. Modified from Borcard and colleagues (2004).

their review a conceptual model where the importance of mass effects versus species sorting should depend on water retention time (i.e. the overall mean time that water spends in a lake), that is, in lakes characterized by long retention time (higher than 100 days (Lindström *et al.* 2006; Lindström and Bergström 2004), the influence of environmental factors may be favored contrary to lakes with a shorter retention time. However, how high dispersal rates need to be for mass effects to appear is unclear (Logue and Lindström 2010) but may for instance, depend on the trophic status of the habitat. Indeed, mass effects appear to be more efficient in oligotrophic lakes while environmental forces appear to prevail in eutrophic lakes (van der Gucht *et al.* 2007).

Determination of the influence of the processes responsible for the bacterial biogeography can be complex as biogeographical patterns may result from multiples processes. For example, distance-decay relationship or spatial clustering of the bacterial assemblages not necessarily translate the importance of dispersal limitation. Indeed, such pattern could be assessed by the spatial autocorrelation of the environmental conditions (Lindström and Langenheder 2012). Inversely, the connection of a lake to a river could

disrupt the autochthonous bacterial assemblage due to mass effects, and also change local environmental conditions (Logue and Lindström 2008), for instance, by a large supply of nutrients. Thus, it could be difficult to statistically disentangle both processes. In order to address this problem, statistical analyzes such as variation partitioning analysis (VPA) have been developed to discriminate the 'true' relative influence of both environmental and dispersal forces on the bacterial beta diversity variations (Borcard *et al.* 1992; Peres-Neto *et al*. 2006). However, the absence of a relevant environmental variable included in the model could drastically underestimate the importance of local environmental factors on the bacterial beta diversity variations (Lindström and Langenheder 2012).

II.4 Dispersal consideration

The estimation of bacterial dispersal should take into account several potential biases. First the detection of a new taxon in a lake survey does not mean that it has recently settled in the lake (after dispersal). Indeed, it is likely that this taxon was resuscitated from dormancy due to an improvement of the local conditions (Choi *et al.* 1996; Jones and Lennon 2010). Second, according to Hanson and colleagues (2012), the detection of a taxon does not imply its proper establishment in the habitat, i.e. that this taxon is metabolically and reproducibly active. The establishment of a bacterial species depends on many factors including the species competitive abilities or the presence of suitable environmental conditions. Furthermore, the detection of a bacterial taxon by PCR-based methods is adding an incertitude regarding the establishment of this taxon in the ecosystem, since it is likely to detect DNA from dead-cells and since DNA can persist several weeks in freshwaters (Nielsen *et al.* 2007; Dejean *et al.* 2011). Finally, dispersal is an integral part of the neutral processes. Indeed, dispersal is considered to be neutral when dispersal rates depend on the community size (Hubbell 2001): abundant taxa should have a greater dispersal aptitude compare to low abundance populations, regardless of taxon identity.

II.5 Biogeography depending on the taxonomic rank?

In the scientific literature, a growing number of studies support that several processes may simultaneously shape the bacterial assembly (e.g., van der Gucht *et al.*

2007; Drakare and Liess 2010; Langenheder and Székely 2011). The coexistence may result from different populations of the bacterial assemblage that are not shaped by the same forces, probably due to specific properties or traits. Green and colleagues (2008) reviewed that although bacterial community are surprisingly plastics in terms of metabolism or physiology properties, there are evidences that bacterial communities could be biogeographically distributed not regarding species composition but regarding the bacterial **functional traits**. Same observations were obtained by Burke and colleagues (2011).

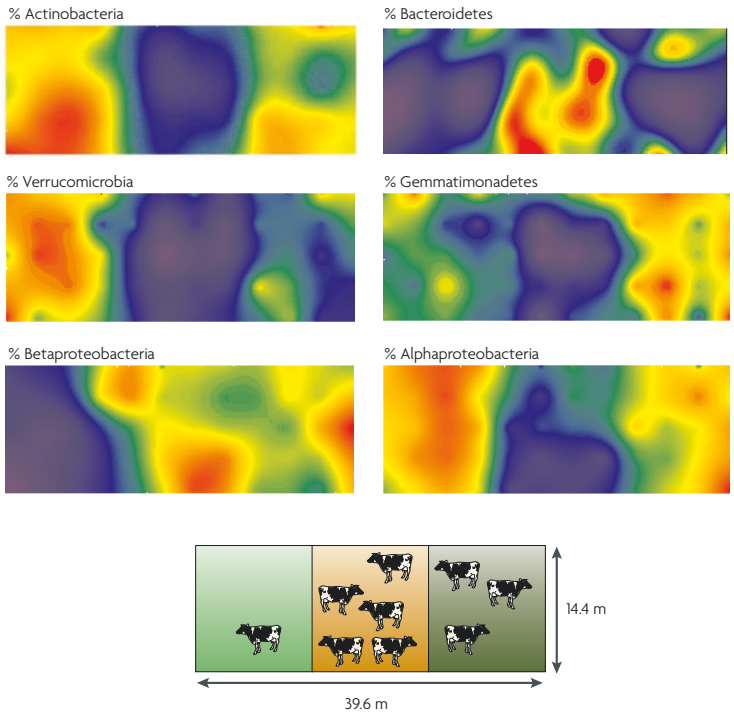


Figure 7 Niche differentiation at the phylum and class levels. The spatial distribution of the relative abundance of targeted bacterial classes or phyla within the total bacterial community of a pasture subjected to three levels of cattle impact. Field-scale variations of the relative abundance of the different taxa are indicated by the different colours, from blue (lower relative abundance) to red (higher relative abundance). From Philippot and colleagues (2009).

Independently, a recent study demonstrated that a large proportion of functional traits were not randomly distributed (83 out 89) within bacterial **lineages.** Indeed, complex traits encoded by several genes were shared by few close lineages (e.g., oxygenic photosynthesis was only encountered within the phylum *Cyanobacteria*), whereas simple genes (e.g., enzymes degrading simple carbon compounds) were widespread among bacterial lineages (Martiny *et al.* 2013). These results are congruent with those from Philippot and colleagues (2010) that observed distinct responses to a perturbation (cattle pressure) among phyla (Fig. 7). Besides, discrepancies in the importance of environmental versus dispersal-related factors among dominant phyla and classes have already been observed in the studies of Barberán and Casamayor (2010) and Székely and Langenheder (2014).

II.6 Spatio-temporal variability

Although biogeography has been previously defined as the study of the bacterial distribution over space and time, at the regional scale, a few studies have simultaneously investigated the influence of both environmental and dispersal-related factors on more than one sampling campaign (e.g., van der Gucht *et al.* 2007) and to the best of my knowledge, except in a study on an activated sludge (Ofiţeru *et al.* 2010), the importance of the neutral model has never been temporally evaluated on the bacterial assemblage. Yet, temporality appears to be an important axis of variation for aquatic microorganisms. Indeed, conclusive supports highlighted evidences of seasonal variations within lakes, related to water temperature (Shade *et al.* 2007), nutrient availability (Crump *et al.* 2003) or top-down control (Kent *et al.* 2007). Moreover, among the studies assessing the seasonality of bacterioplankton assemblages for a longer time period than a year, most of the results did not exhibit similar pattern from year to year (e.g., Lindström 1998; Yannarell *et al.* 2003; Kent *et al.* 2004; Boucher *et al.* 2006; Jones *et al.* 2012).

Intra-lake spatial variations of the beta diversity have also been poorly investigated compare with spatial surveys focusing on cross-lake comparisons. Yet, even if some papers showed conclusive supports that intra-lake variations are smaller than inter-lake variations (Yannarell and Triplett 2004; Jones *et al.* 2012), intra-lake horizontal heterogeneity has been observed in a shallow lake (that was supposed to be well-mixed) (Jones *et al.* 2012). The presence of horizontal heterogeneity in the bacterial community composition might appear when the rate of biological and ecological interactions driving bacterial community assembly occurs faster than the rate of water movement and turbulence in the lake. Besides, Jones and colleagues (2012) showed that intra-lake variations could be as important as temporal variations, i.e. the variation observed for samples separated by a single day or a meter were comparably similar. Environmental factors (de Wever *et al.* 2005), but also physical barrier have been identified as shaping the spatial heterogeneity within lakes (Yannarell and Triplett 2004). Moreover, few studies (e.g., Adams *et al.* 2014) have focused on the impact of mass effects within lake on the horizontal heterogeneity of bacterial communities. Indeed, most studies investigating the impact of natural or man-made inlets on a water body have focused on the flow and

persistence of pathogen population (McCorquodale *et al.* 2004; McLellan *et al.* 2007), but rarely on the composition of bacterial communities.

Studies investigating the bacterial dissimilarities within the water column were performed on deep lakes. They revealed large discrepancies in bacterial assemblages between the epilimnion and the hypolimnion mainly related to physicochemical water properties (e.g., Goddard *et al.* 2005; de Wever *et al.* 2005; Boucher *et al.* 2006).

II.7 Conclusion

The pertinence of biogeographical studies of bacterial assemblages have been demonstrated at both regional (among lakes) and local (within lake) scales in several published papers. Different processes were advanced to explain the bacterial community variations: two deterministic forces, i.e. local environmental conditions and dispersal-related mechanisms (i.e. dispersal limitation and mass effects), and the neutral processes. As neutral processes refer also to dispersal processes, to avoid any confusion throughout this manuscript, we will make the distinction between dispersal-related mechanisms and neutral processes by describing the dispersal-related processes as spatial processes. To date, only a few efforts have simultaneously investigated the importance of these different processes on the bacterial assembly at the regional scale, and there are even fewer studies focusing on the local scale (spatio-temporal variations within lake). Finally, at the regional scale, the factors shaping distinct bacterial groups, e.g., dominant phyla or classes, have received little attention. Since the parameters driving the bacterial assemblages seem to vary with the taxonomic rank, it appears important to investigate the bacterial biogeography not only at the total community level but also for different bacterial assemblages, from the phylum to the genus level. As a consequence this thesis explored not only the parameters that shape the whole bacterial community, but also a widespread taxonomic group: the nontuberculous mycobacteria.

III. Nontuberculous mycobacteria in freshwater habitats

Mycobacteria have been historically studied, especially because some species display pathogenicity, including the causative agents responsible for the leprosy or the

tuberculosis. To date, tuberculosis is still one of deadliest communicable diseases in the world. In 2013, 9.0 million people have contracted tuberculosis, among them, 1.5 million died of this disease (WHO 2014).

Mycobacteria are not limited to these obligate pathogens. Indeed, a large number of mycobacterial species are considered **saprophytic** (i.e. heterotrophic bacteria that are not pathogens) and natural inhabitant of environmental habitats, including freshwater lakes. Among these environmental mycobacteria, few species have also been associated with human and animal diseases as opportunistic pathogens (Falkinham III 1996). Factors affecting the biogeography of nontuberculous mycobacteria are still a broad field of research with many unknowns, although they may also play an important role in the breakdown of recalcitrant and hydrophobic organic matter.

III.1 Mycobacterial taxonomy

The term *Mycobacterium* is the contraction of two Greek roots: mukês (fungus) and baktêria (small staff), alluding the "fungus-like" morphological aspect of some colonies when they are cultured on agar plates. *Mycobacterium* genus has long been regarded as the unique genus in *Mycobacteriaceae* family (Garrity *et al.* 2004). However, the *Amycolicicoccus* genus (only represented by a single species: *A. subflavus*) has recently been classified as another representative genus of this family (Wang *et al.* 2010). This family belongs to the suborder *Corynebacterineae*, the order *Actinomycetales*, the subclass *Actinobacteridae*, the class *Actinobacteria* and the phylum *Actinobacteria*. A total of 105 official *Mycobacterium* species are described in the Bergey's manuel (Garrity *et al.* 2004). This number is increasing exponentially (Fig. 8), and to date, 172 species are identified in the List of Prokaryotic names with Standing in Nomenclature (LSPN) (Euzéby 1997).

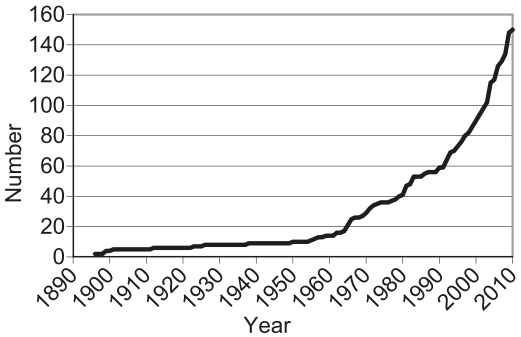


Figure 8 Numbers of approved *Mycobacterium* species and subspecies from 1896 to 2010. From Dai and colleagues (2011).

Among the large diversity within the *Mycobacterium* genus, species can be clustered in two groups according to their **virulence.** The first group includes obligate

pathogens in humans and animals i.e. the *Mycobacterium tuberculosis* complex responsible for the tuberculosis (*M. africanum, M. bovis, M. canettii, M. caprae, M. microti, M. pinnipedii,* and *M. tuberculosis*) and leprous mycobacteria responsible for leprosy (*M. leprae*). The second group, named as nontuberculous mycobacteria (NTM) or atypical mycobacteria, comprises non-obligate pathogens and represents the greatest diversity with more than 160 described species (Euzéby 1997).

III.2 General characteristics of the mycobacteria

III.2.i. Morphology

Mycobacteria are straight or slightly curved rod bacteria, sometimes branched. They measure between 0.2 to 0.6 µm wide by 1.0 to 10 µm long. Non-motile and non-spore forming, some mycobacterial species can produce carotenoid organic pigment giving them a yellow-orange color (Prescott *et al.* 2013; Runyon 1959). The main morphological characteristic of mycobacteria is their extremely complex cell envelope (Fig. 9).

The mycobacterial envelope can constitute up to 40% of the bacteria dry weight (Anderson 1943). It is composed of an inner layer and an outer layer that surround the plasma membrane (Daffé 1996). The soluble outer layer, so-called capsule, is mainly composed of proteins and lipids whereas the insoluble inner layer is constituted of three different covalently linked structures, i.e. peptidoglycan, arabinogalactan and mycolic acids (Fig. 9). This complex refers to the essential core of the mycobacterial cell wall (Hett and Rubin 2008), and it is responsible for the cell wall integrity and impermeability, notably due to the hydrophobic mycolic acid layer (Daffé 1996).

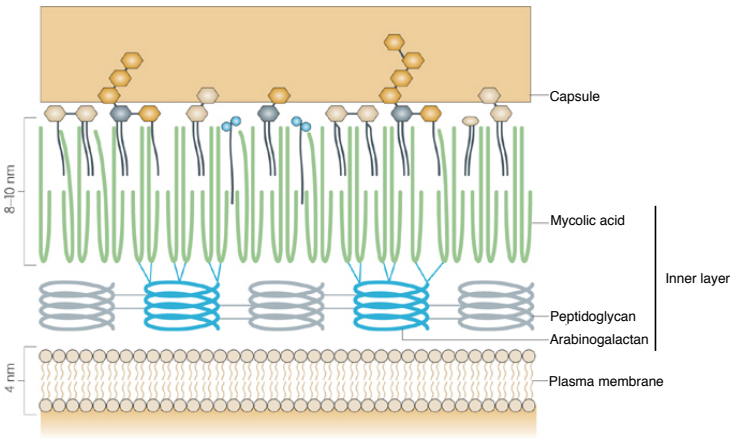


Figure 9 Schematic representation of the cell envelope of mycobacteria. Modified from Abdallah and colleagues (2007).

III.2.ii. Physiology

Mycobacteria are Gram-positive, heterotrophic, aerobic or microaerobic bacteria (Primm *et al.* 2004; Prescott *et al.* 2013). These bacteria are generally divided into two groups on the basis of their growth rate on solid medium: slow- and fast-growing mycobacteria. Slow-growing mycobacteria (SGM) are able to produce visible colonies after more than seven days contrary to rapidly or fast-growing mycobacteria (FGM) where colonies are visible in less than seven days (Shinnick and Good 1994). The limitation access of extracellular nutrients into bacteria may explain this difference of growth rate. Indeed, a growing number of studies evidenced the direct relationship between the cell envelope permeability and the growth rate, especially owing the presence of porin-like protein expressed on their membranes (Mailaender *et al.* 2004; Stephan *et al.* 2005). The phenotypic difference regarding the growth rate corresponds to a real phylogenetic division of the genus *Mycobacterium*, as observed on phylogenetic trees based on *rrs*, *hsp65* and *rpoB* gene sequences (three genes frequently used to characterized mycobacterial species) (Fig. 10).

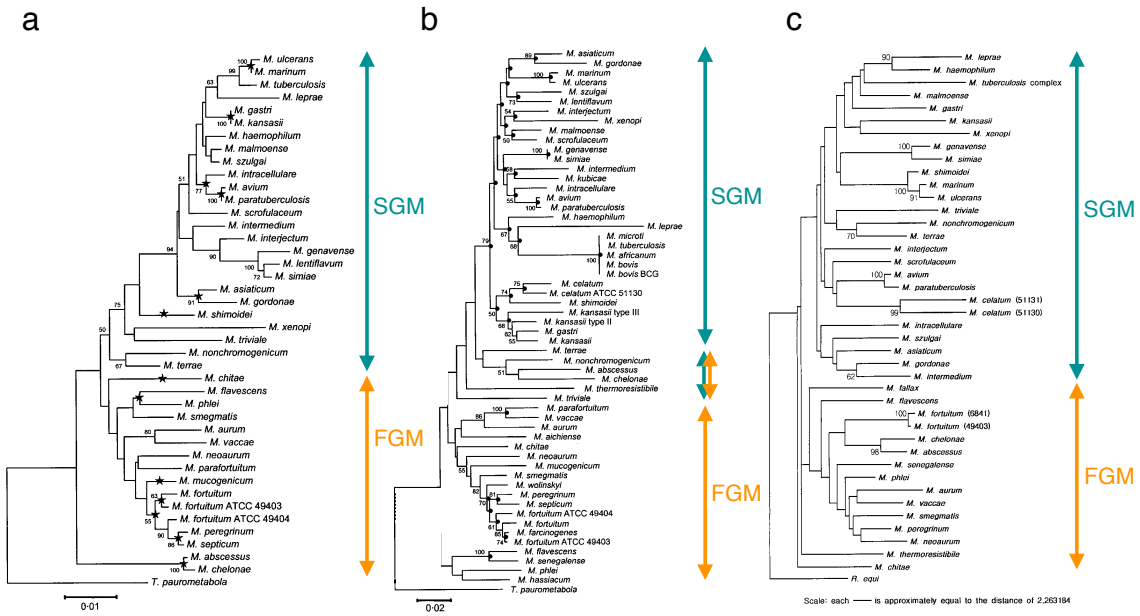


Figure 10 Comparison of phylogenetic trees based on the (a) *rrs* gene, (b) *hsp65* gene, and (c) *rpoB* gene sequences of the genus *Mycobacterium*. The *rrs, hsp65 and rpoB* sequence trees were constructed from respectively 41, 56 and 44 reference strains of mycobacteria. Trees were constructed by the neighbor-joining method. The numerical values in the tree represent bootstrap results. The distance between two strains is the sum of the branch lengths between them. SGM, slow-growing mycobacteria; FGM, fast-growing mycobacteria. Modified from Kim and colleagues (1999) and Kim and colleagues (2005).

III.3 Pathogenicity of the nontuberculous mycobacteria

Although nontuberculous mycobacteria species are primarily saprophytic, some species can behave as opportunistic pathogens and thus be responsible for infections in the presence of predisposing conditions (Falkinham III 1996; Tortoli 2009). These opportunistic pathogens may cause pulmonary, cutaneous, lumphonodal (Table 2), soft tissue, bone or joint infections (Tortoli 2009) owing to the ingestion, the inhalation or the direct physical contact with contaminated environments (Pavlik *et al.* 2009). Mycobacteriosis became particularly relevant in the last decades, in relation to the global emergence of immunocompromised (e.g., patients with AIDS) and cystic fibrosis patients who are predisposed to these infections. Indeed, from 25 to 50% and from 5 to 20% of patients with AIDS and cystic fibrosis respectively are infected with nontuberculous mycobacteria (Falkinham III 1996; Roux *et al.* 2009). In France, 500 cases are estimated per year, primarily caused by *M. avium* (CNR-MyRMA 2014).

Table 2 Type of infections and causative agents. Compiled from mycobacteria reviews (Katoch 2004; Tortoli 2009).

Mycobacterium species	Pulmonary	Cutaneous	Lymphonodal
M. abscessus	*+++*	*+++*	
M. asiaticum			
M. avium (MAC*)	+++		+++
M. bohemicum			*+*
M. chelonae		*+++*	
M. flavescens			
M. fortuitum	*+*	*+++*	
M. goodii			
M. immunogenum	*++*		
M. interjectum			*+*
M. kansasii	+++		
M. lentiflavum			*+*
M. malmoense	+++		*+*
M. marinum		*+++*	
M. scrofulaceum	*+*		+++
M. szulgai	*+*	*+++*	*+*
M. thermoresistibile			
M. ulcerans		*+++*	
M. xenopi	+++		

+++: frequent reported cases; +: seldom reported cases

* MAC (*Mycobacterium avium* complex): regroups the species *M. avium*, *M. intracellulare* and the subspecies *M. avium* spp. *avium*, *M. avium* spp. *hominissuis*, *M. avium* spp. *paratuberculosis* and *M. avium* spp. *silvaticum*.

III.3.i. Pulmonary infections

The more common NTM infections are pulmonary, representing approximately 90% of the mycobacteriosis (Kasperbauer and Huitt 2013). These infections usually occur on older adult patients suffering from pulmonary diseases, e.g., bronchiectasis, pneumoconiosis or cystic fibrosis. Pulmonary diseases are characterized by chronic cough, sputum production, fever, weight loss and weakness (Piersimoni and Scarparo 2008). The most common species associated with pulmonary infections are notably the *Mycobacterium avium* complex (MAC), the *Mycobacterium abcessus* complex (*M. abcessus*, *M. bolletii*, *M. masiliense*), *M. kansasii, M. malmoense* and *M. xenopi*.

III.3.ii. Cutaneous infections

The mycobacterial species most frequently involved in cutaneous infections are *M. marinum* and *M. ulcerans.* Mycobacteriosis causes by *M. marinum* are frequently contracted from standing water (e.g., on sinner) after infections of wound. Infections causes chronic granulomatous lesion of the skin and soft tissue, which may extend to the bone (Blackwell 1999). *M. ulcerans*, responsible for the Buruli ulcer is a common mycobacteriosis in humid area (Walsh *et al.* 2008). Buruli ulcer is a chronic infection characterized by large necrotizing ulcers due to toxins excreted by *M. ulcerans*. The infection mode is still debated; indeed, it may result from wound infection with contaminated water or bites from insects hosting *M. ulcerans* in their salivary glands (Marsollier *et al.* 2002). Other fast-growing bacteria, such as *M. abscessus*, *M. chelonae* or *M. fortuitum* are also frequently encountered in skin and soft tissues infections. These infections mostly occur among immunosuppressed patients treated with corticosteroids or chemotherapy. However, invasive gestures on immunocompetent individuals (e.g., surgery, mesotherapy, etc.) may also favor the contraction of these infections (Carbonne *et al.* 2009).

Since the mycobacteriosis transmission from human to human has never been evidenced (Wolinsky 1979), the environment appears to be the primarily source of NTM (Pavlik and Falkinham III 2009). Due to the high prevalence of mycobacteriosis contracted after water contact from drinking water distribution systems and hospital water supply networks, these habitats have been particularly well studied (e.g., du Moulin *et al.* 1988;

Fox *et al.* 1992; Covert *et al.* 1999; Falkinham III *et al.* 2001; Fujita *et al.* 2002; Vaerewijck *et al.* 2005). However, several reviews and recent studies have documented the wide range of reservoirs of mycobacteria, including soils, dust, wastewater effluents but also freshwater ecosystems (Dailloux *et al.* 1999; Falkinham III 2002; Radomski *et al.* 2011; Hruska and Kaevska 2012; Ye and Zhang 2013).

III.4 The ecology of mycobacteria in freshwater ecosystems

III.4.i. Environmental conditions promoting mycobacteria

Acidic pH is usually considered to be associated with high mycobacterial densities or recovery. Acidic conditions (up to pH 4.0) were observed to enhance the growth of *M. avium* in an experimental manipulation (Kirschner *et al.* 1999). This observation was confirmed by two biogeographical surveys performed in the United States and in Finland in brown-water swamps and in humic brooks where low water pH appeared to enhance NTM recovery (Kirschner *et al.* 1992; Iivanainen *et al.* 1993). However, NTM have also been observed in neutral and alkaline environments (Bland *et al.* 2005; Jacobs *et al.* 2009). In those habitats, water pH was not noticed as a relevant parameter shaping the NTM densities.

High nutrient concentrations were also identified to favor NTM, especially recalcitrant substrates such as humic and fulvic acid (Kirschner *et al.* 1992, 1999; Iivanainen *et al.* 1993). Combined with their ability to growth in low oxygen levels (equal to 4% of atmospheric levels (Kirschner *et al.* 1999)), mycobacteria appear to be particularly well adapted in eutrophic and hypereutrophic environments. Indeed, Jacobs and colleagues (2009) described that NTM densities were positively correlated with chlorophyll *a* and nutrients availability (total phosphorus and nitrogen) and negatively correlated with dissolved oxygen concentration and secchi depth.

Concentrations of several metals including iron, aluminum and cobalt were found to positively correlate with NTM densities (Iivanainen *et al.* 1993; Radomski 2011). Although their direct influence on NTM is still unclear (Iivanainen *et al.* 1993), the large cell wall of the mycobacteria could prevent the exchanges of metal ions from the exterior, thus a higher metal concentration in the environment may favor their uptake by mycobacteria.

III.4.ii. Biological interactions

According to the literature, commensalism (the commensal benefits from its host without benefiting or harming it) as well as parasitism (the parasite benefits from its host by causing damage to it, are two known classes of interactions between mycobacteria and other organisms. Competition is also likely to occur with other mycobacterial species as well as with other bacterial taxa, however it is rarely reported in the literature.

Amoeba appeared to be relevant hosts of NTM in the environment (see review of Salah *et al.* 2009 and thesis of Delafont 2015). *In vitro*, several studies have reported the capacity of some NTM species to colonize and multiply in these protozoan predators (e.g., Solomon *et al.* 2003; Adékambi *et al.* 2006; Mura *et al.* 2006). The intracellular association within amoeba especially with the *Centramoebida* group (comprising *Acanthamoeba* and *Protacanthamoeba*) (Delafont *et al.* 2014) appears to increase the chances of survival of NTM in unfavorable environments, particularly when amoeba are encysted (Adékambi *et al.* 2006; Salah *et al.* 2009). The interaction with the unicellular eukaryote also appears to improve NTM resistance to antibiotic treatments and promote their entry into macrophages and epithelial cells of their hosts (Cirillo *et al.* 1997; Miltner and Bermudez 2000). To date, since no evidence has demonstrated the mutual benefit of amoeba to cooperate with mycobacterial species, intracellular associations are considered commensalism interactions.

Some insects could be potential reservoirs of NTM. For example, *in vitro*, aquatic bugs (*Naucoris cimicoides*) seems to concentrate *M. ulcerans* within their salivary glands, where the bacteria appears to survive and multiply without causing any observable damage in the insect tissues (Marsollier *et al.* 2002, 2004).

NTM are responsible for parasite interactions in larger animals including reptiles, amphibians, waterfowl and terrestrial birds, ornemental and wild fishes, and livestock and wild mammals (Soldati *et al.* 2004; Biet *et al.* 2005; Chai *et al.* 2006; Gronesova *et al.* 2008; Wiinschmann *et al.* 2008). However, as for human, interactions with mycobacteria could be commensal with immunocompetent organisms and possibly pathogenic with immunocompromised organisms. Indeed, Mrlik and colleagues (2012) isolated in low prevalence several NTM species (e.g., *M. fortuitum, M. gordonae* or *M. insubricum*) mainly

from skin and gills in healthy fish population from ponds in Czech Republic (e.g., *Cyprinus carpio, Abramis brama, Tinca tinca or Pseudorasbora pava*), with no observable tissue damage.

III.4.iii. Other potential reservoirs of mycobacteria in freshwater

Although biogeographical studies targeting the mycobacterial assemblages and performed on freshwater environments mainly focused on the water column, other compartments such as biofilms could be relevant habitats for mycobacteria. Indeed, owing to their high surface hydrophobicity, mycobacteria readily attach to biofilms and can form biofilms on water-surface interface and air-water surface interface in a few days (Torvinen *et al.* 2004; Alavi *et al.* 2006). In drinking water systems for instance, water-surface interfaces are considered to be an important source of NTM (Falkinham *et al.* 2001). At the air-water interface, where aerosols are formed, Parker and colleagues (1983) observed in an experimental manipulation, those aerosols were enriched up to 15,000 times in mycobacterial cells compared with the water column, suggesting that air-water interface could represent an important source of NTM.

Sediment also appears to be a relevant habitat for NTM. For example, Debruyn and colleagues (2009) detected from 4.8×10^5 to 8.1×10^7 *rrs* gene copies per gram of dry weight sediment in the Lake Erie by specifically targeting the FGM mycobacteria. Moreover, conclusive supports report their presence in this compartment, especially since over the last decade NTM are particularly searched in sediment owing to their ability to degrade some micropollutants.

III.4.iv. Ecological function

Regarding the importance of NTM in the carbon recycling, the literature mainly focuses on their capacity to degrade polycyclic aromatic hydrocarbons (PAH). These compounds originate from incomplete combustion of organic materials such as coal or oil (Singh 2012). To date, only fast growing species have been observed as capable of mineralizing PAH, these species include *M. aromaticivorans, M. austroafricanum, M. flavescens, M. fluoranthenivorans, M. frederiksbergense, M. gilvum, M. hodleri, M. pyrenivorans* and *M. vanbaalenii* (Kleespie *et al.* 1996; Rehmann *et al.* 2001; Willumsen *et al.* 2001; Khan *et al.* 2002; Derz et al. 2004; Hormisch *et al.* 2004; Hennessee *et al.* 2009).

Isolated from soils or sediments, where PAH are concentrated owing to their hydrophobicity, these species are able to degrade light (e.g., pyrene, fluorene or anthracene) and/or heavy PAH (e.g., benzo[a]pyrene or benz[a]anthracene) (Seo *et al.* 2009). The genes responsible for PAH degradation were identified in mycobacteria to be *nidA* and *nidB,* two genes encoding for the alpha and beta subunit ring-hydroxylating dioxygenase respectively (Brezna *et al.* 2003). Besides, Debruyn and colleagues (2009) found a positive correlation between this *nidA* concentration and the fast-growing species densities.

III.5 Diversity and densities of mycobacteria in freshwater

III.5.i. Mycobacterial densities

The few studies assessing the mycobacterial densities in freshwater habitats (Table 3) allowed to evidence the wide variations of mycobacterial densities between samples (with values ranging from 1.0 to 2.7×10^8 CFU or targeted gene per liter). However, it should be noted that this overview may be biased due to the differences in the isolation methods used to count mycobacteria. Using a standardized and quantitative molecular method to compare the densities of mycobacteria among a gradient of freshwater ecosystems could help understand which environmental factors may impact the densities of mycobacteria.

Table 3 Densities of mycobacteria in diverse freshwater habitats.

Freshwater habitats	State	Densities (CFU[a] or targeted gene[b] per liter)	References
Stream	Finland	1.0×10^0–2.2×10^{3a}	(Iivanainen *et al.* 1993)
Stream	Finland	9.7×10^2–5.8×10^{3a}	(Torkko *et al.* 2001)
Stream	Finland	8.0×10^2–2.2×10^{3a}	(Torkko *et al.* 2002)
River	France	$<1.0\times10^3$–6.9×10^{5b}	(Radomski 2011)
Diverse habitats (groundwater, drinking pumps, springs, rivers and lakes)	Ethiopia	3.4×10^3–2.7×10^{8b}	(Khera 2012)

Abbreviation: CFU, colony-forming-unit.
[a]Densities determined using culture-based methods.
[b]Densities determined using PCR-based methods.

III.5.ii. Mycobacterial diversity

Although the studies investigating the composition of the NTM in freshwater remain scarce, they revealed an important diversity in these habitats. Indeed, the review of

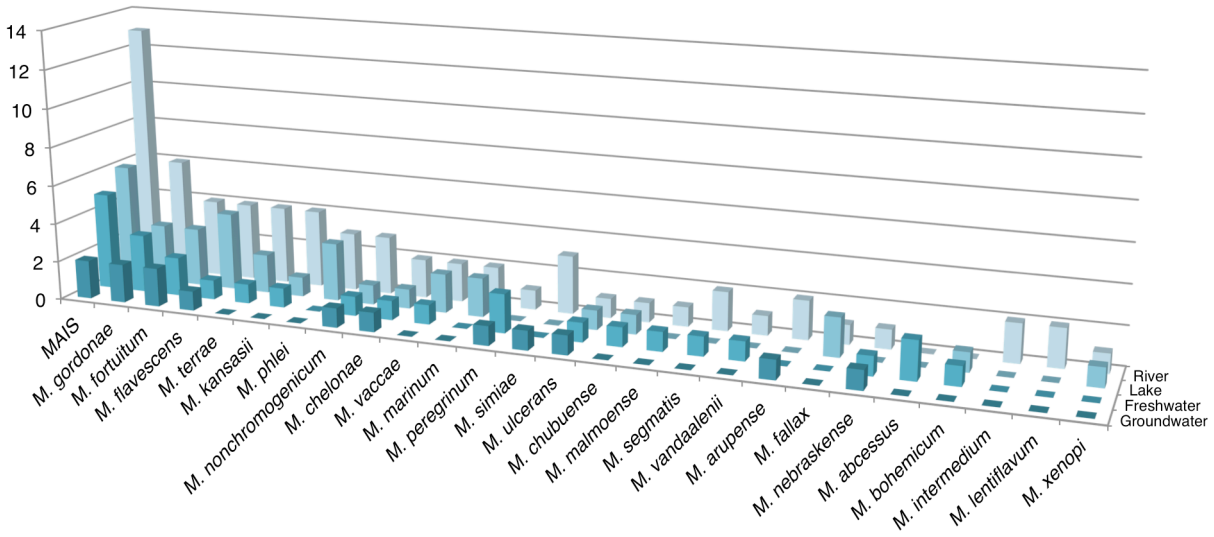


Figure 11 Number of publications relating the occurrence of the mycobacterial species in water samples from freshwater habitats. Only the species identified more than once were reported. MAIS refers to the MAIS complex (*M. avium, M. intracellulare, M. scrofulaceum*). From Viallier and Viallier (1973), Beerwerth (1973), Showalter and Wolinsky (1974), Goslee and Wolinski (1976), Falkinham *et al*. (1980), George *et al*. (1980), Wendt *et al*. (1980), Bionde *et al*. (1982), Levy-Frebault *et al*. (1983), Martin *et al*. (1987), Ichiyama *et al*. (1988), Haas and Fattal (1990), Kirschner *et al*. (1992), Iivanainen *et al*. (1993), Sabater and Zaragoza (1993), Eaton *et al*. (1995), Stinear *et al*. 2000, Torkko *et al*. (2000), Torkko *et al*. 2001, le Dantec *et al*. 2002, Torkko *et al*. 2002, Parashar *et al*. (2004), Trujillo *et al*. (2004), Bland *et al*. (2005), Pickup *et al*. (2005), Niva *et al*. (2006), Pickup *et al*. (2006), Michel *et al*. (2007), Norby *et al*. (2007), Eddyani *et al*. (2008), Eun-Sook *et al*. (2008), Drewe *et al*. (2009), Narang *et al*. (2009), Rahbar *et al*. (2010), Vandelannoote *et al*. (2010), Kankya *et al*. (2011), Radomski (2011), Khera (2012).

39 studies (Fig. 11) reveals the large diversity of the mycobacterial species found within freshwater habitat. The species from the MAIS complex (*M. avium, M. intracellulare, M. scrofulaceum*), *M. gordonae, M. fortuitum, M. flavescens, M. terrae* or *M. kansasii* seem to dominate the studied freshwater ecosystems. As previously mentioned, it should be noted that this overview may be biased either by the particular interest for species responsible for human infections or by the isolation techniques themselves. Besides, a recent study investigating the composition of the entire mycobacterial assemblage using high-throughput sequencing methodology (see *Chapter I section IV.1.ii*) evidenced that freshwater habitats could host an impressive richness and diversity of mycobacterial species with the recovery of 436 operational taxonomic units (OTU) from 42 water samples, mainly dominated by the species *M. neoaurum, M. gordonae, M. peregrinum, M. neglectum, M. vanbaalenii, M. mucogenicum, M. chubuense* or *M. vaccae* (Khera 2012). These results suggest that new sequencing techniques should unravel a new picture of mycobacterial diversity within freshwater ecosystems.

III.5.iii. Identification of mycobacterial species

To date, as for the entire bacterial community, the sequencing of the *rrs* (16S rRNA) gene is the reference for identifying mycobacteria. Other targets (e.g., *hsp65*, *rpoB* or the ITS) are frequently sequenced in combination of the *rrs* gene to obtain more robust phylogenetic trees.



Figure 12 Phylogenetic tree based on *rrs* gene of the genus *Mycobacterium*. Species having identical sequences are shown in bold. The numbers on the abscissa represent the percent distance between different isolates. From Harmsen and colleagues (2003).

The *rrs* gene is an approximately 1,500-nucleotides sequence (position 1,487,542 to 1,489,058, *M. avium* 104 GeneBank NC_008595) that encodes for the component of the small subunit of prokaryotic ribosomes. Overall, within the *Mycobacterium* genus, slow-growing mycobacteria possess a single rRNA operon, whereas fast-growing species possess two rRNA operons (Bercovier *et al.* 1986). As for the entire bacterial community, the sequencing of the *rrs* gene is still the primary target of molecular taxonomic studies (Tortoli 2003). Thus, the *rrs* gene available database is relatively well documented compared with other genes. However, the *rrs* gene has the disadvantage to not fully discriminate all the mycobacterial species (Fig. 12). Indeed, the near full length sequencing of the *rrs* gene does not permit differentiation between certain SGM and FGM species (Kim *et al.* 1999; Harmsen *et al.* 2003; Adékambi and Drancourt 2004; Gomila *et al.* 2007). For instance, the species *Mycobacterium chelonae*, *M. massiliense*, *M.*

abscessus and *M. bolletii* have similar *rrs* gene sequences, as well as the species *M. mucogenicum* and *M. phocaicum*, the species *M. peregrinum* and *M. septicum*, the species *M. neworleansense* and *M. porcinum*, or the species *M. kansasii and M. gastrii*.

The 65-kDa heat shock protein gene (*hsp65*) has a total length sequence about 1,600-nucletotides (position 4,835,762 to 4,837,387, *M. avium* 104 GeneBank NC_008595). As the rrs gene, hsp65 gene is also highly conserved among mycobacterial species. However, it presents two hypervariable regions whose sequences may be used for identification purposes (Ringuet *et al.* 1999). The sequencing of a 440-bp portion targeting these two hypervariable regions allows to distinguish mycobacterial species including *M. chelonae* and *M. immunogenum*, from *M. abscessus*, *M. bolletii* and *M. massiliense*, which are difficult to differentiate with the complete sequence of the *rrs* gene (Ringuet *et al.* 1999; Adékambi and Drancourt 2004; Gomila *et al.* 2007).

Encoding the beta subunit of RNA polymerase, the *rpoB* gene has a total length about 3,400-nucleotides (position 4,631,379 to 4,634,813, *M. avium* 104 GeneBank NC_008595). The partial sequencing of the *rpoB* gene allows to distinguished *M. kansasii* from *M. gastri*, which are difficult to discriminate with the complete sequence of the *rrs* gene (Kim *et al.* 1999). If phylogenetic trees obtained from the *rrs* and *rpoB* are almost similar, it should be underlined that trees based on *rpoB* gene sequence display higher bootstrap values and more divergence than the *rrs* gene-based trees (Adékambi and Drancourt 2004).

III.5.iv.A limited overview

Most of the studies exploring the mycobacterial diversity or densities were performed using culture-based methods, yet this methodology may introduce several biases. First, not all bacteria are able to growth on culture media. Second, because of the slow growth of many mycobacterial species, culture from environmental samples necessitates a decontamination step, which can drastically reduce both the density and diversity of the recovered isolates. Indeed, le Dantec and colleagues (2002) and Radomski and colleagues (2010) estimated that more than 99% of the mycobacterial isolates were removed by this step. Thirdly, despite the decontamination step, culture media could be contaminated by interfering flora owing to the slow-growing rate of the mycobacteria

compare with other organisms, making the counting and the identification difficult or even impossible (e.g., Bland *et al.* 2005).

Furthermore, the advent of the PCR-based methods has greatly enhanced the ability to detect NTM in the environment, especially due to a broader sensitivity. However, PCR methods do have some serious limitations for environmental analysis, especially the presence of PCR-inhibitory substances. Moreover, total bacterial DNA extraction also appears to be a critical step. Indeed, the improvement of the extraction protocols and especially the use of a beat-beating step (that disrupt the cell wall) significantly improve the recovery of the *Actinobacteria* DNA (Guo and Zhang 2013). Such amelioration could explain why the *Mycobacterium* genus has been recently found to be a dominant group in wastewater treatment plant (Ye and Zhang 2013), whereas this genus was usually not detected in those environments in the previous studies (e.g., Moura *et al.* 2009; Hu *et al.* 2012).

As a consequence, all these biases could partly justify the absence of NTM recovery in several bacterial community studies or in studies focusing on mycobacteria (e.g., Viallier and Viallier 1973; Bland *et al.* 2005; Drewe *et al.* 2009; Radomski 2011) favoring the limited and patchy overview of these bacteria in the freshwater ecosystems.

IV. Scope of this study

The aim of this study was to investigate the factors and processes shaping the spatial and temporal biogeography of the bacterial metacommunity in a set of shallow lakes. Two bacterial levels were investigated: (i) the total bacterial community and (ii) the nontuberculous mycobacteria. Among the nontuberculous mycobacteria, some species are considered opportunistic pathogens and several species have been described as organic pollutant degraders, so this bacterial group may represent a useful sentinel of anthropogenic pressures.

Regarding literature several hypotheses were advanced. First, we assumed that the impact of environmental and dispersal processes on bacterial assemblages depend on the taxonomic rank that is considered (Philippot *et al.* 2010). As a consequence we

expected dissimilar responses between the total bacterial assembly and the nontuberculous mycobacteria. Second, according to Martiny and colleagues (2006), we hypothesized that at the regional scale (from 1 to 150 km between lakes) bacterial assemblages would be shaped by both environmental factors and dispersal processes. However, owing to mass effect we expected that lakes linked to rivers or storm sewer outlet may have dissimilar bacterial community structure compared with isolated lakes.

At the local scale (within lake), we expected that nutrient and bacterial inputs from storm sewer effluents would have a local impact that on the bacterial communities in the lake. However, due to the importance of water column mixing in shallow lakes, we did not compulsorily expect a spatial structuring of the bacterial communities. Temporal variation of bacterial assemblage may be mostly driven by seasonal fluctuation of local factors driven by climatic variations. Finally, neutral processes have recently been observed to drive both spatial (e.g., Drakare and Liess 2010; Östman *et al.* 2010) and temporal (Ofiţeru *et al.* 2010) bacterial assembly in diverse aquatic habitats. So, we would expect same tendencies in our study.

To test these hypotheses, two complementary approaches were undertaken (Fig. 13) in the context of the research project PULSE (Peri-Urban Lakes, Society and Environment) supported by the French National Research Agency. First, 49 shallow lakes with surface area ranging from 5 to 120 ha were sampled once a year in 2011, 2012 and 2013 summers in Paris area. Representative in terms of abundance and properties (i.e. land use, hydrology, altitude and surface) of the water bodies in Paris area with surface area higher than 5 ha, these 49 shallow lakes constitute a useful model to evaluate the relative importance of the factors and processes shaping the spatial distribution of the bacterial assemblages at the regional scale since they display large environmental gradients and strong variations in variables that could impact the bacterial dispersion (e.g., river or storm sewer inputs).

Second, in order to assess the spatio-temporal dynamic of the bacterial compartment, a finer approach was performed on a pilot lake, Créteil Lake (Val-de-Marne, France) by a two-year monthly monitoring of a horizontal and a vertical transects. Additionally, the impact of the stormwater runoff into Créteil Lake was evaluated on the

nontuberculous mycobacteria densities through six important rain events that occurred from June to November 2013. Their favorable habitats in lacustrine environments were explored in two recreational shallow lakes

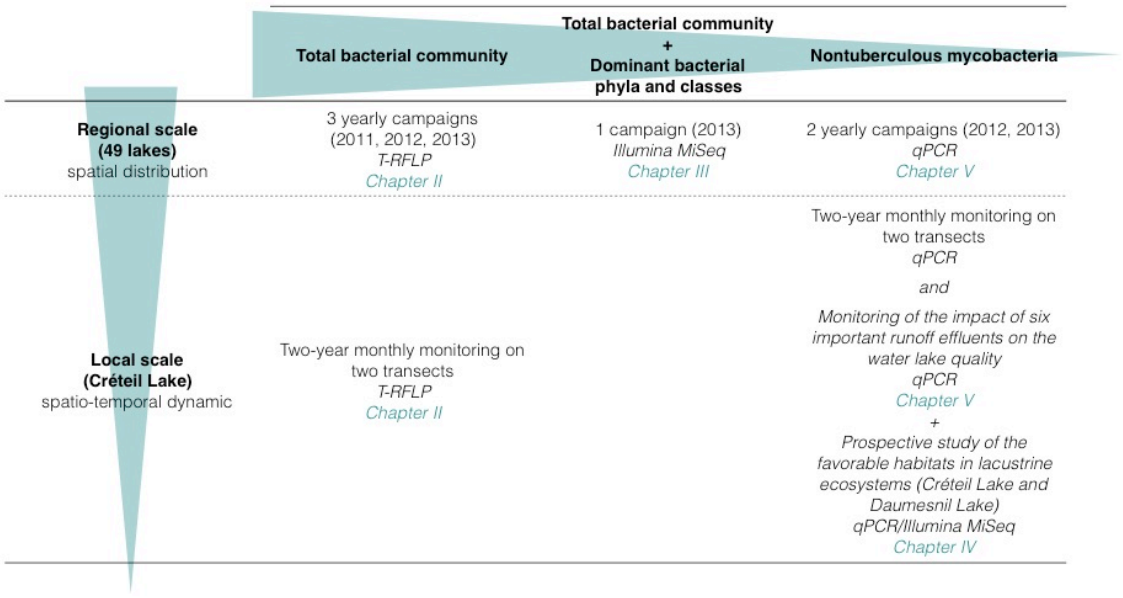


Figure 13 Experimental strategy followed in this study.

This work combines molecular approaches and multivariate statistical analysis to delineate the relative importance of environmental factors (physicochemical parameters, watershed land use type, trophic status), spatial and neutral processes that could shape the bacterial and mycobacterial communities. Using innovative molecular tools, such as high-throughput sequencing and more conventional molecular approaches (T-RFLP profiles), the diversity of the bacterial communities and the mycobacteria was evaluated. Densities of mycobacteria were evaluated using a real-time quantitative PCR method. Physicochemical parameters were measured by the partners of the PULSE project during the different campaigns (Appendix 1).

IV.1 Methodological considerations

In the last three decades, major advents have been made to investigate the microbial community. Limitations and pitfalls associated with traditional culture-based methods (see *Chapter I* section III.5.iv) were in part solved with molecular methods that constitute valuable tools to investigate the diversity and the abundance of the bacterial community (Dorigo *et al.* 2005). Among the different molecular methods available, we

chose two different approaches for this study: a fingerprinting and a next-generation sequencing method to analyze the bacterial diversity and a real-time quantitative PCR to assess the bacterial densities. These methods are PCR-based tools that hold advantages and limitations, which orientated our choice.

IV.1.i. Advantages and limitations associated with PCR-based tools

The primarily advantage of molecular methods is to detect both cultivable but also non-cultivable organisms. Compared to culture approaches, molecular methods also appear to be faster, more specific, more sensitive and more accurate (Justé *et al.* 2008). Despite the fact that molecular tools offer extensible advantages compare to culture-based methods, it should be kept in mind when performing and interpreting community analyses that these methods also provide limitations and pitfalls that could bias the overview of the bacterial community.

Overall, all processing steps between sampling and PCR amplification could potentially bias the bacterial community (Justé *et al.* 2008), i.e. sampling procedure, storage and nucleic acid extraction could have consequences of the bacterial community. Indeed, according the extraction protocol used, large discrepancies in the bacterial community structure could be observed, some taxa being often underestimated such as the *Actinobacteria* phylum (Terrat *et al.* 2011; Guo and Zhang 2013).

Regarding the amplification step, the choice of the primer set has major consequences of the bacterial community retrieved, the specificity and the sensitivity of the primer set being two key properties. Yet, as illustrated in Appendix 2b, large discrepancies of sensibilities could be observed from one universal bacterial probe to another. During the PCR reaction, dominant DNA strains due to probability distribution may predominantly be amplified. Moreover, inhibitors could also reduce the number of reactions, and thus, bias the analysis. Finally, in absence of any pretreatment with intercalating molecules, such as the live/dead® protocol (Fittipaldi *et al.* 2012), DNA of sufficient quality is amplified even if it comes from dead cells or free-DNA, which could bias the representation of the bacterial composition too.

IV.1.ii. Bacterial diversity analysis

To date, two semi-quantitative molecular approaches are widely used to characterize the bacterial community assemblages: fingerprinting and high-throughput sequencing methods. For both approaches, suitable primers to assess the bacterial diversity should have both variables/hypervariables and conserved regions, as the *rrs* genes (Appendix 2a). Variables and hypervariables regions offer the power of discrimination over a wide range of taxonomic level.

Fingerprinting methodologies

Fingerprinting methods have widely been used to assess the bacterial community structure (BCS). Especially since these low-cost tools are extremely suitable to compare BCS between a large number of sample at a sufficient taxonomic resolution to conclude on changes in the environment (Fierer *et al.* 2007). Considering the number of samples generated by the sampling campaigns on the 49 lakes during three regional campaigns, and the two-year monthly monitoring of the Créteil Lake, we considered that the fingerprinting approach was an acceptable choice. Among the different genetic fingerprinting methods available for community analysis, we chose to use the terminal restriction fragment length polymorphism (T-RFLP) analysis. T-RFLP analysis allows the identification of DNA fragments labeled with a fluorochrome at the terminal ends of the *rrs* genes after digestion of DNA sequences by one or more restriction enzymes (Liu *et al.* 1997). In order to be separated, amplicon are then separate by their length using capillary electrophoresis. This technique is one of the most used in genetic footprint with denaturating gradient gel electrophoresis (DGGE) and automated ribosomal intergenic spacer analysis (ARISA). However, DGGE compared to ARISA and T-RFLP exhibits a lower sensitivity, moreover, the difficulty in comparing samples from different gels (Nocker *et al.* 2007) may constitute an important flaw. Furthermore, the high sensitivity and the good inter-run comparability of the T-RFLP analysis (Nocker *et al.* 2007) convinced us to use this method. We decided to use a protocol already validated (by Thomas Lerch). *In silico* the primers used in this protocol (63F-1389R) does not seem to be a useful universal bacterial primers (Appendix 2b). However, similar limitations could be observed *in silico* with the

other primer sets usually employed in freshwater studies (27F-1492R) (e.g., Jardillier *et al.* 2004; Anderson-Glenna *et al.* 2008; Besemer *et al.* 2009).

As any fingerprinting method, T-RFLP only detects dominant taxa, one peak can represent more than a bacterial species and no taxonomic identification could be performed from T-RFLP profiles (Frakruddin *et al.* 2013; Douterelo *et al.* 2014). Moreover fingerprinting methods assess only the structure of the bacterial community, and they should be coupled with cloning and sequencing in order to identify the taxa that compose the bacterial community. As a consequence we decided to use also a high-throughput methods to study the bacterial communities in the 49 lakes, but for a single regional campaign.

High-throughput sequencing methodologies

Contrary to fingerprinting methods, next-generation sequencing (NGS) are more resolutive to perform phylogenetic comparison among samples (Liu *et al.* 2012). The most frequently used NGS platforms are Roche 454 and Illumina MiSeq. The recent advances made by Illumina MiSeq technology mean that today both technologies are comparable regarding the length of the sequenced fragments (NGS Field Guide 2014, *http://www.molecularecologist.com/next-gen-table-2-2014/* consulted on October 2015) and regarding the sequencing error (Luo *et al.* 2012). However, since Illumina is cheaper than Roche, we decided to assess the bacterial and mycobacterial community composition using Illumina MiSeq technology.

Bacterial communities were assessed using primer sets targeting the hypervariable regions V3 and V4 of the *rrs* genes. *In silico*, this primer set offers the advantage to well amplify bacterial DNA without amplification of chloroplast DNA (Appendix 2b) that could avoid the massive chloroplast DNA recovery from lakes with important algal bloom. Mycobacterial communities were assessed using a primer set targeting the hypervariable regions V2 and V3 of the *rrs* genes (338F-797R). Although it is not high enough resolutive to characterize all NTM at the species level (Kim *et al.* 2005), these primers appeared to be more sensitive (Pontiroli *et al.* 2013) than primers targeting other genes. Indeed, van der Wielen and colleagues (2013) amplified 69% of non-targeted sequences using a primer set targeting the *hsp65* gene. Moreover, data analysis from high-throughput sequencing

database is easier to perform when focusing on the *rrs* gene due to the amount of reference deposited in the public databases.

IV.1.iii. Bacterial quantification

To quantify NTM densities, real-time quantitative PCR (qPCR) was used. The qPCR is a sensitive, specific and reproducible tool to detect and quantify bacterial populations from environmental samples by estimating the number of targeted copies present in a sample (Arya *et al.* 2005; Douterelo *et al.* 2014). Briefly, this method can measure the amount of PCR product obtained during the exponential phase of the PCR reaction by quantifying a fluorescent reporter. The amount of detected reporter is then correlated with the initial amount of target template allowing the quantification of the target population (Kubista *et al.* 2006).

To quantify *Mycobacterium* spp., we used a recent protocol developed in our laboratory (Radomski *et al.* 2013). Targeting the *atpE* genes, this primer set offers the advantage to be highly sensitive and specific to the *Mycobacterium* genus. Moreover, contrary to the *rrs* gene, only one copy of the *atpE* gene is present in mycobacterial species.

These three methods (T-RFLP, Illumina MiSeq platform, and qPCR) allowed the acquisition of data from the bacterial communities in the 49 Lakes, Créteil Lake and Daumesnil Lake. These results are presented in *Chapter II, III, IV* and *V*.

IV.2 Chapter II and III: biogeography of the total bacterial community in shallow lakes

Chapter II reports the metacommunity study at both regional and local scales that evaluated factors and processes (i.e. environmental conditions, spatial and neutral processes) shaping the spatial and temporal bacterial community assembly within the lacustrine water column. The bacterial community structure was assessed using T-RFLP method targeting the 16S rRNA genes.

This study has been published in *FEMS Ecology Microbiology*.

Chapter III As T-RFLP method only detects the dominant bacterial taxa, I also evaluated for the year 2013 the factors and processes governing the biogeography patterns of the lacustrine bacterial community composition using a highly resolutive method, i.e. high-throughput sequencing (Illumina MiSeq).

This paper is in preparation.

The key results are:

At the regional scale

- The neutral community model predicted well the spatial community structure (assessed by T-RFLP) compared to the environmental conditions and the spatial processes that both explained a small fraction of the bacterial community structure total variance.
- Similar tendencies were observed in 2013 using high-throughput sequencing method.
- Similar trends were noticed within dominant phyla and classes in the processes involved in their biogeographical patterns.

At the local scale

- Intra-lake spatial variations of the bacterial community structure appeared to be mainly shaped by neutral processes.
- The temporal dynamic of the bacterial assembly was primarily governed by the water temperature (an environmental factor driven by meteorological processes).

IV.3 Chapter IV and V: biogeography of the nontuberculous mycobacteria in shallow lakes

Chapter IV describes a prospective study that aimed to identify the potential habitats of the nontuberculous mycobacteria in five compartments within lacustrine ecosystems, i.e. water column, air-water interface, sediment and biofilms from plants and rocks. For that, two contrasted lakes were sampled: the mesotrophic Créteil Lake (Val-de-Marne, France) and the eutrophic Daumesnil Lake (Paris, France). In this study, the mycobacterial density

as well as their diversity were assessed using respectively a new designed real-time quantitative PCR and high-throughput sequencing method (Illumina MiSeq).

This paper is accepted for publication in *Antonie van Leeuwenhoek*.

Chapter V details the biogeographical study of the mycobacterial densities at both regional and local scales in lacustrine environments. At the regional scale, the importance of the environmental conditions as well as factors that could enhance the mycobacterial dispersion was assessed for the years 2012 and 2013. Within Créteil Lake, the influence of the physicochemical and meteorological parameters on the spatial variation of mycobacterial densities was evaluated. Physicochemical and meteorological parameters as well as neutral processes were also investigated to understand the temporal dynamics of the mycobacterial densities.

This paper is in preparation.

<u>The key results are:</u>

- Ubiquity of the nontuberculous mycobacteria in lacustrine environments. Indeed, this bacterial group was quantified in all the collected samples in high densities (in all the five compartments and 49 lakes)
- In Créteil and Daumesnil lakes, nontuberculous mycobacteria displayed a remarkable diversity, dominated by fast-growing species.
- Lakes and compartments shaped the mycobacterial assemblage composition as well as their densities.
- At the regional scale, density variations were significantly explained by environmental factors including water pH, labile iron concentration, but also spatial factors (lakes connected to a river displayed significantly higher mycobacterial densities).
- At the local scale, although stormwater runoffs locally affected mycobacterial densities after important rainfalls (in front of the storm sewer outlet), no significant spatial variation was observed across the Créteil Lake and from the surface to the bottom of the lake.

- No significant seasonal pattern was observed in the mycobacterial densities over the two years of survey. However, the temporal dynamic of the mycobacteria was predicted up to 87% by neutral processes.

Chapter II

Neutral community model explains the bacterial community assembly in freshwater lakes

Neutral community model explains the bacterial community assembly in freshwater lakes

Adélaïde Roguet[a], Grégory S. Laigle[a], Claire Therial[a], Adèle Bressy[a], Frédéric Soulignac[a], Arnaud Catherine[b], Gérard Lacroix[c,d], Ludwig Jardillier[e], Céline Bonhomme[a], Thomas Z. Lerch[f] and Françoise S. Lucas[a].

[a]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.

[b]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.

[c]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.

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

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

[f]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.

key words: bacterioplankton, community structure, deterministic factors, neutral community model, lake, T-RFLP.

Accepted in FEMS Microbiology Ecology: DOI: 10.1093/femsec/fiv125

I. 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 scales 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 scale, 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.

II. 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 scales (e.g., Muylaert *et al.* 2002; Yannarell and Triplett 2004, 2005; Lindström *et al.* 2005; Kent *et al.* 2007; Shade *et al.* 2007; Jones *et al.* 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) the dispersal limitation, i.e. the extent to which taxa to 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 local community (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 *et al.* 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, 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 a monthly monitoring over two 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.

III. Materials and methods

III.1 Study area and sampling

This study was conducted in the Paris area (Fig. 14), 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 3.0® (IGN, Paris, France), 49 lakes (Fig. 14 and Supplementary data 1) were selected using a random and stratified sampling strategy (Catherine *et al.* 2008). Briefly, the 49 lakes were chosen to represent an unbiased set of water 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 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 prefiltration through 50 µm pore-size nylon mesh. Sterivex filters were kept at 4°C during transport and then stored at – 20°C.

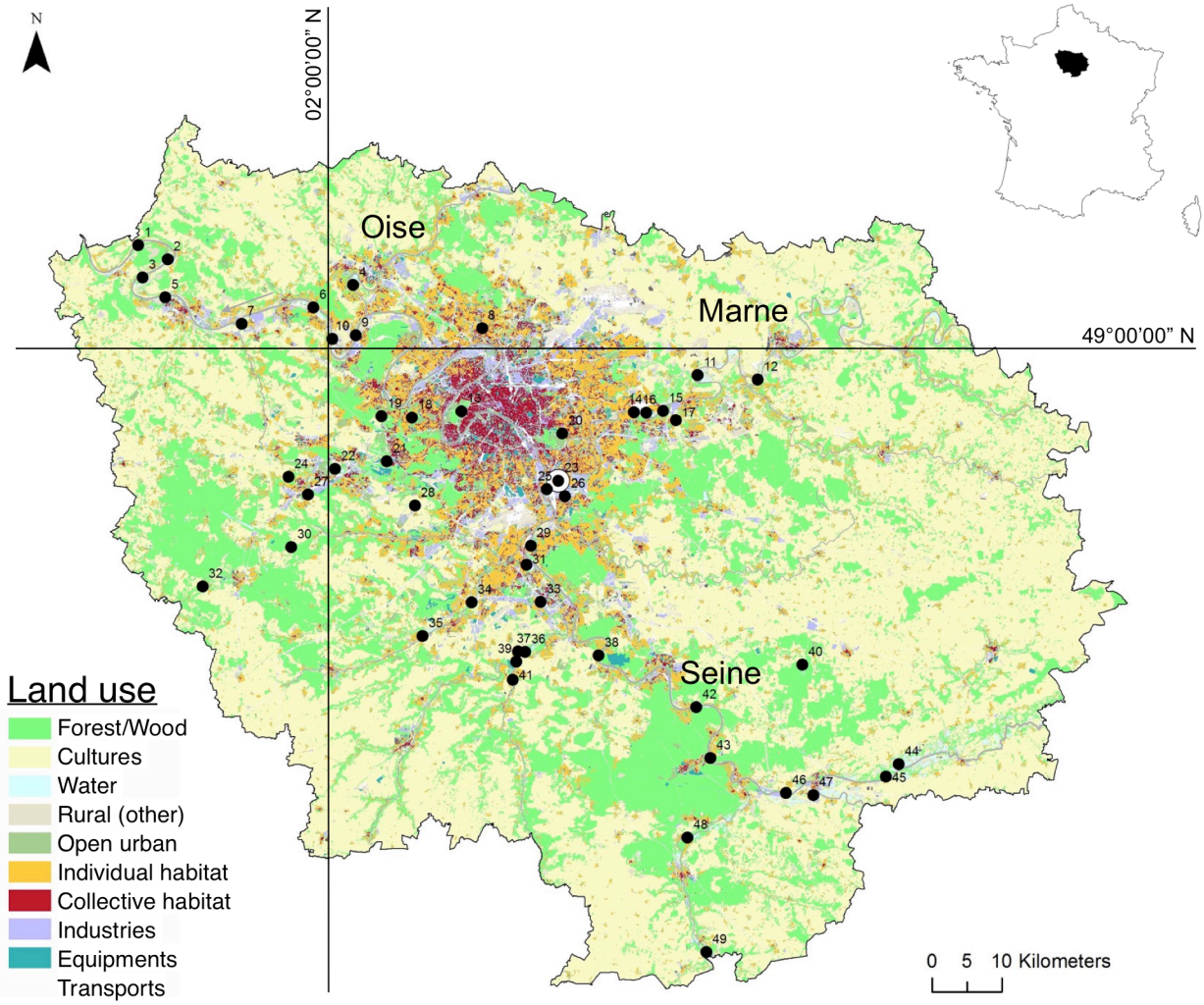


Figure 14 Location of the 49 lakes in the Paris area (France). Créteil Lake is indicated by an open circle. Numbers correspond to the ID lake.

Among the 49 lakes, Créteil Lake (Fig. 14 and 15) was monthly monitored 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 4) 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.

A horizontal and a vertical transect were studied (Fig. 15). 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 outflow (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 date, an integrated sample (M) was obtained as previously described. This sample 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, one liter of lake water was filtered through a Sterivex GP-filter cartridge as previously described and then stored at – 20°C prior to analysis.

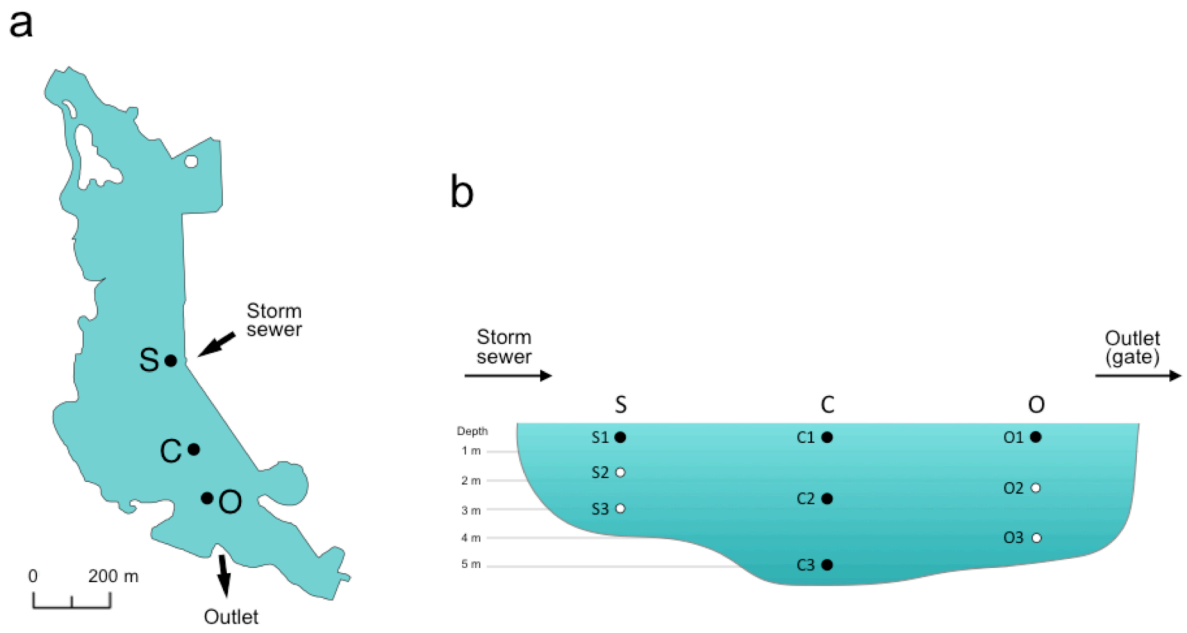


Figure 15 Sampling strategies in the Créteil Lake. (a) Location of the three stations. (b) Representation of the two transects: the horizontal transect (S1, C1 and O1) and the vertical transect (C1, C2 and C3).

Table 4 Characteristics of Créteil Lake. Values are the averages (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)
DOC ($mgC.L^{-1}$)	6.1 (0.6)
POC ($mgC.L^{-1}$)	0.8 (0.3)
TSS ($mg.L^{-1}$)	4.1 (1.5)
pH	7.8 (0.5)
Conductivity ($\mu S.cm^{-1}$)	1509 (63)

III.2 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 ms^{-1} and an additional wash was performed on the SPIN filters.

III.3 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-carboxyfluorescein) and 1392R (5'-ACGGGCGGTGTGTACAAG-3') targeting the bacterial domain (Osborn *et al.* 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 X of GoTaq® Colorless buffer and 0.5 U of GoTaq® DNA polymerase (Promega, Madison (WI), USA). Reactions were carried out in a T1 thermocycler (Biometra, Gottingen, 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 *et al.* 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. T-RFLP data were then processed and analyzed 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.

III.4 Bacterial enumeration

Bacterial counts (heterotrophic bacteria and cyanobacteria) from the Créteil Lake samples were assessed by flow cytometry using a Becton Dickinson FACScan (BD Biosciences) instrument equipped with a 15 mW 488 nm laser. Pigmented microorganisms were identified based on the autofluorescence 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 0.5 and 1 micron microspheres (Polysciences, Eppelheim, Germany) according to Zubkov and Burkill (2006).

III.5 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 and colleagues (2006) and the French standard (AFNOR NF T 90-023).

Secchi depth and vertical profiles of physico-chemical 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).

Concentration of total suspended solids (TSS) were quantified after filtration of one liter 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).

III.6 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 for 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).

III.7 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, i.e. 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 the 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: (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 and colleagues (2011) and Legendre and Gauthier (2014).

To assess the neutral assembly of the bacterial communities, we used the method developed by Sloan and colleagues (2006) to fit the regional relative abundance of the T-RFs and their observed detection frequency. Contrary to the 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 two 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 and colleagues (2010). Included in the 'STI' package (Legendre *et al.* 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 two years 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 and colleagues (2010) to evaluate to what extent the neutral model could explain the temporal variations of the relative abundance for each T-RF.

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

IV. Results

IV.1 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 replacement that explained respectively 76, 70 and 67% of the beta diversity, whereas the random shift of T-RF 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 (Table 5). On average over the three 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 5).

The neutral model explained a large fraction of the relationship between the occurrence frequency of the T-RFs and their relative abundance variations (Fig. 16), with

Table 5 Variation partitioning analysis of the bacterial community structure for the three years of sampling.

	Relative variance explained (R^2_{adj})						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
2012	0.08	0.11	0.03*	0.06*	0.05	0.86	TS, DOC	Hydro, PCNM
2013	0.06	0.03	0.04*	0.01^{NS}	0.02	0.93	TS	Hydro, PCNM
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)		

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

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 three years of sampling. Indeed, the N_Tm value 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 also performed. The same values of N_Tm with 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.

IV.2 Spatial and temporal monitoring in Créteil Lake

In Créteil Lake, the two-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 was used for the statistical analysis with a median 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 two years of survey (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

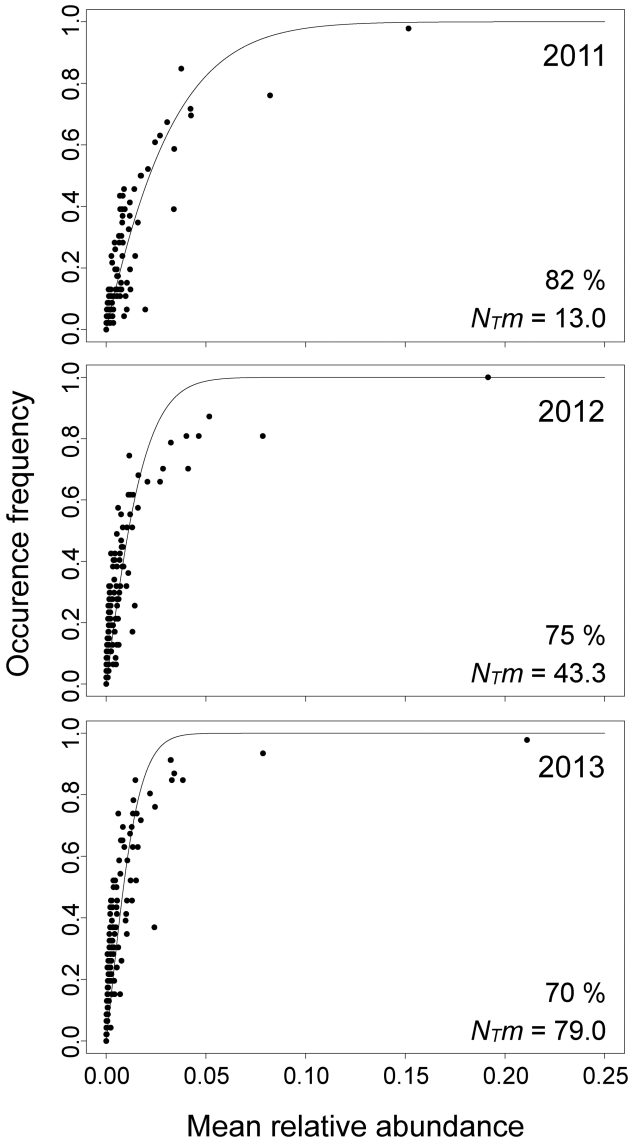


Figure 16 Relationship between the detection frequency of T-RFs and their mean relative abundance from the three years of monitoring. Each point represents a T-RF. The lines show the best fit to the neutral model with random immigration from a common source pool as described by Sloan and colleagues (2006). N_Tm values represent the meta-community size multiplied by the immigration rate. Percentage represents R^2 of fit.

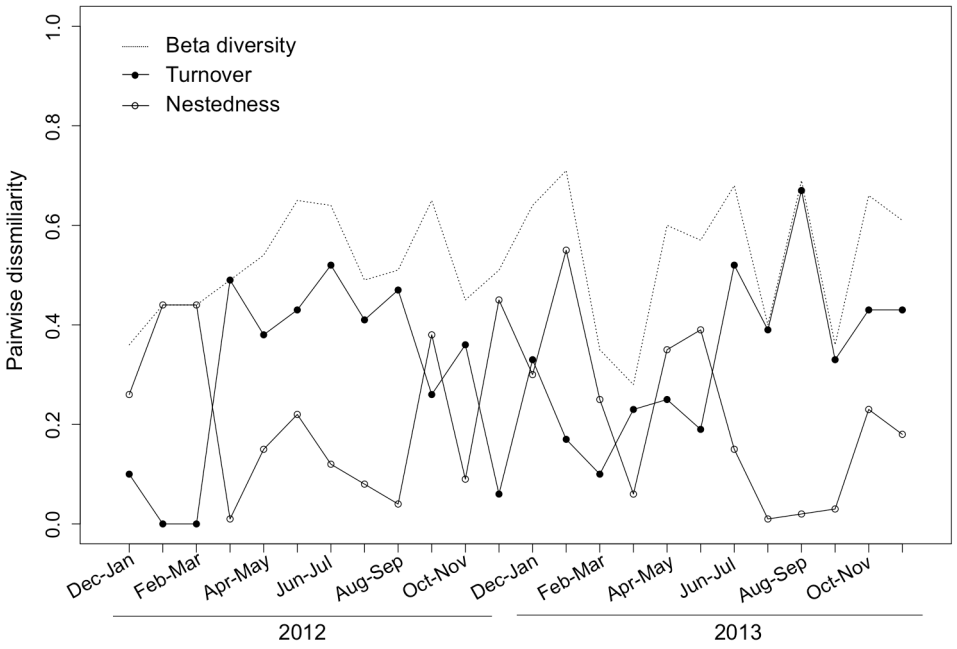


Figure 17 Temporal variations of the bacterial community beta diversity, turnover and nestedness determined by comparing pairwise successive sampling campaign (n = 24).

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 analyze 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. 17). No clear temporal pattern in the beta diversity fluctuations could be reported (Fig. 17). However, when decomposing the beta diversity into its turnover and nestedness components, we observed distinct temporal variations (Fig. 17) 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 data 2), indicating that T-RF replacement was higher in summer. In contrast, the negative relationship between temperature and nestedness (Supplementary

data 3) suggested that the composition of the T-RFs was strongly influenced by richness variations in winter.

At the local scale, the neutral community model significantly explained the relationship between the occurrence frequency of the T-RFs and their relative abundance variations in 22 out of the 25 campaigns (Supplementary data 3). 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 data 3). Regarding the temporal dynamics of the abundance of the T-RFs on the average sample M, the neutral time-series model proposed by Ofiţeru and colleagues (2010) could explain the relative abundance variations of only one T-RF (R^2_{adj} = 80%, 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 two 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 adjusted R-squared of 34% (data not shown).

V. 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 two-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.

V.1 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 T-RFLP (e.g., Castro-Carrera *et al*. 2014; van Dorst *et al*. 2014; Elsayed *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.

V.2 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 and colleagues (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 as significant over the three 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 PCNM variables). Furthermore, in this analysis, no significant direct influence of the presence of storm sewer effluents 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 *et al*. 2006; McLellan *et al*. 2007; Lear and Lewis 2009).

V.3 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 agreement with the study performed by Drakare and Liess (2010) on thirteen 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 influence on BCS (Chisholm and Pacala 2010), especially as the fraction of variance unexplained by the variance partitioning analysis could not be attributed to stochastic processes (Vellend *et al*. 2014). Moreover, as discussed in Anderson and colleagues (2011), the neutral and the deterministic mechanisms might coincidentally mirror the same patterns. Therefor, 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 three 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 and colleagues (2006) and Nelson and colleagues (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, the 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 the identification of the 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.

V.4 Influence of deterministic and neutral processes at the local scale

In Créteil Lake, we expected a spatial structure between the storm sewer and the lake outlet since the effluents from the sewer 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 and the inlet (Lindström *et al.* 2006; Nelson *et al.* 2009). However, our two-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 *et al.* 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 two-year 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.

VI. Acknowledgements

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].

The authors thank Pierre Legendre for his helpful advice 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 two 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 three years.

VII. 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.

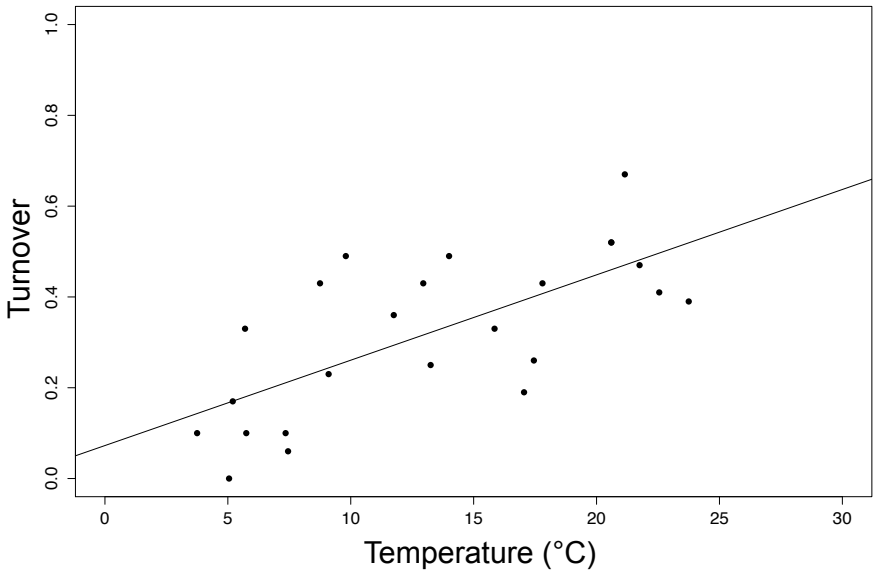
VIII. Supplementary data

Supplementary data 1 Localization, morphological and physicochemical characteristics of the 49 lakes

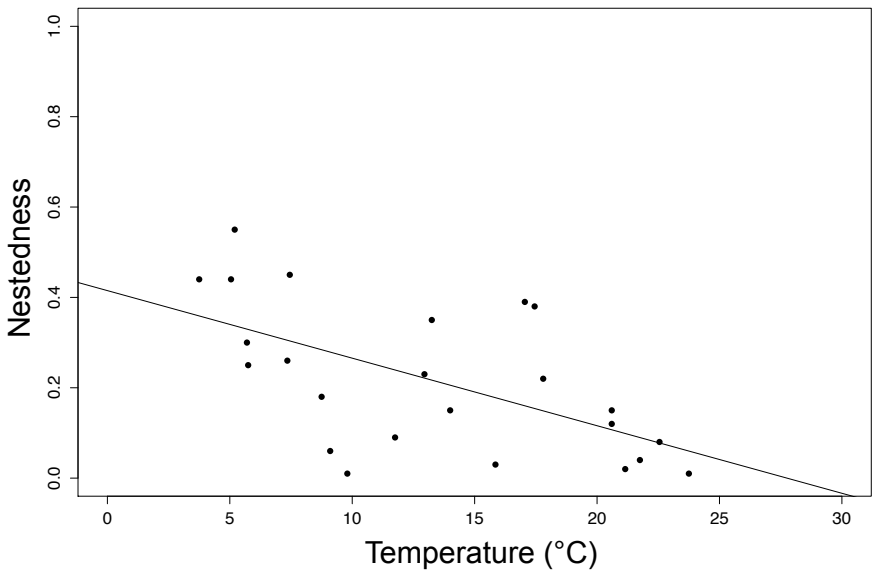
ID lake	Name	Localization	LUI	Trophic status	Hydro	SSO	Surface area (km^2)	Mean depth (m)	pH	Temp (°C)	Cond ($\mu S.cm^{-1}$)	DOC ($mgC.L^{-1}$)
1	Freneuse Cove	49°04′ N 01°35′ E	N	H	C	+	0.14	2.4	7.9 (0.8)	22.0 (2.8)	473 (110)	4.3 (0.4)
2	Grand Lavacourt Pond	49°03′ N 01°41′ E	N	E	I	-	1.12	4.3	8.8 (0.4)	21.4 (2.7)	276 (25)	4.3 (0.3)
3	Ilon Cove	49°01′ N 01°37′ E	N	E	C	+	0.66	3.8	8.2 (0.1)	21.3 (2.3)	503 (43)	3.6 (0.2)
4	Les Galets Pond	49°01′ N 02°03′ E	OU	E	I	-	0.11	2.2	9.1 (0.3)	22.6 (3.9)	277 (18)	7.2 (0.2)
5	ASM Club Cove	49°00′ N 01°40′ E	N	H	C	+	0.37	2.5	8.0 (0.1)	21.8 (2.1)	491 (12)	3.9 (0.3)
6	Grosse Pierre Pond	49°00′ N 01°58′ E	N	H	I	+	0.46	4.3	9.1 (0.6)	21.2 (2.2)	417 (46)	7.6 (0.6)
7	Bout du Monde Pond	48°58′ N 01°50′ E	N	H	I	-	0.20	1.2	8.6 (0.5)	21.1 (3.7)	738 (68)	6.4 (0.2)
8	Enghien Lake	48°58′ N 02°18′ E	U	H	Cr	+	0.34	1.0	7.7 (0.7)	23.6 (1.0)	689 (132)	6.8 (1.4)
9	Gaule Achéroise Pond	48°58′ N 02°03′ E	A	M	I	-	0.05	6.2	7.7 (0.6)	23.2 (1.7)	775 (16)	3.6 (0.4)
10	Triel Pond	48°57′ N 02°00′ E	N	H	I	-	0.34	3.9	8.8 (0.2)	22.0 (3.4)	1069 (93)	11.4 (0.7)
11	Jabline-Anet Lake	48°54′ N 02°43′ E	OU	O	I	-	0.77	7.0	8.5 (0.1)	21.0 (1.6)	1196 (48)	4.3 (0.2)
12	Isles-les Villenoy Pond	48°54′ N 02°50′ E	A	M	I	-	0.42	2.6	8.6 (0.1)	22.3 (2.7)	324 (12)	4.4 (0.4)
13	Lake Inférieur	48°51′ N 02°16′ E	OU	E	Cr	+	0.11	1.2	8.0 (0.9)	23.7 (3.5)	312 (57)	4.1 (0.6)
14	Les Pâtis Pond	48°52′ N 02°36′ E	OU	H	I	-	0.10	2.6	8.9 (0.3)	21.2 (2.8)	591 (43)	7.9 (1.9)
15	Torcy Lake	48°52' N 02°39'E	OU	M	I	-	0.25	5.8	8.1 (0.0)	22.5 (3.3)	1185 (47)	5.2 (0.1)
16	UCPA Centre Pond	48°52′ N 02°37′ E	N	O	I	-	0.92	5.3	8.7 (0.1)	21.6 (2.4)	455 (5)	4.2 (0.2)
17	Loy Pond	48°51′ N 02°41′ E	N	H	Cr	+	0.07	1.3	8.1 (0.6)	20.6 (3.6)	427 (98)	5.7 (0.9)
18	Saint-Cucufa Pond	48°51′ N 02°10′ E	N	H	I	-	0.02	1.8	7.6 (0.6)	21.0 (3.1)	733 (26)	7.2 (0.7)
19	UTE Louveciennes	48°51′ N 02°06′ E	OU	O	I	-	0.10	2.5	7.3 (0.3)	18.6 (2.8)	695 (41)	1.6 (0.3)
20	Lake Minimes	48°50′ N 02°27′ E	N	H	Cr	+	0.06	1.2	8.0 (0.9)	22.5 (3.8)	520 (38)	5.0 (0.2)
21	Swiss Pond	48°48′ N 02°07′ E	OU	M	I	-	0.14	1.9	8.9 (0.8)	25.6 (3.0)	1031 (75)	6.1 (0.5)
22	Saint-Quentin Reservoir	48°47′ N 02°01′ E	A	E	I	-	1.20	1.6	9.4 (0.4)	23.7 (3.5)	420 (25)	6.4 (0.7)
23	Créteil Lake	48°46′ N 02°27′ E	U	M	I	+	0.41	4.5	7.9 (0.4)	22.6 (2.7)	1494 (78)	6.8 (0.5)
24	Maurepas Reservoir	48°46′ N 01°55′ E	U	H	Cr	+	0.08	3.4	8.4 (0.3)	22.5 (4.3)	387 (77)	6.0 (0.9)
25	Choisy Pond	48°46′ N 02°25′ E	OU	M	I	-	0.34	2.6	7.7 (0.6)	23.3 (3.4)	1228 (84)	5.2 (0.8)
26	Plage Bleue Pond	48°45′ N 02°28′ E	OU	M	I	-	0.09	3.1	7.8 (0.7)	23.1 (2.8)	2279 (230)	6.5 (1.1)
27	Noés Reservoir	48°45′ N 01°58′ E	U	H	I	+	0.24	0.7	8.2 (0.1)	24.0 (5.5)	713 (169)	8.9 (3.4)
28	Saclay Reservoir	48°44′ N 02°10′ E	A	H	I	+	0.30	1.5	9.3 (0.2)	24.3 (2.7)	540 (50)	7.2 (1.3)
29	La Veyssière Pond	48°41′ N 02°24′ E	OU	H	C	+	0.25	1.6	8.3 (0.4)	23.5 (3.1)	310 (16)	4.1 (0.2)
30	Vaux-de-Cernay Pond	48°41′ N 01°56′ E	N	H	Cr	+	0.06	1.4	8.5 (0.5)	21.7 (4.4)	464 (15)	5.7 (0.7)
31	Noues de Seine Pond	48°40′ N 02°23′ E	U	E	C	+	0.97	3.4	8.9 (0.2)	23.8 (3.6)	760 (19)	4.8 (0.1)
32	Gazeran Pond	48°38′ N 01°46′ E	N	H	Cr	+	0.08	0.7	7.6 (0.3)	22.9 (4.8)	766 (78)	7.4 (1.0)
33	Courcouronnes Pond	48°37′ N 02°25′ E	U	H	I	+	0.07	1.4	9.1 (0.2)	25.3 (4.7)	255 (102)	4.6 (1.5)
34	Epinoche Pond	48°37′ N 02°17′ E	A	H	C	+	0.06	2.4	8.2 (0.3)	22.0 (2.6)	558 (27)	4.6 (0.8)
35	Port-Sud Pond	48°34′ N 02°11′ E	A	H	Cr	+	0.07	1.7	8.3 (0.4)	23.5 (3.4)	494 (58)	4.5 (1.8)
36	Fontenay/Vic Pond	48°33′ N 02°23′ E	A	H	C	+	0.14	1.2	8.7 (0.2)	24.6 (2.5)	380 (27)	4.3 (0.1)
37	Vert-le-Petit Pond	48°33′ N 02°22′ E	A	E	C	+	0.01	2.0	8.1 (0.3)	22.4 (3.0)	575 (48)	7.0 (0.6)
38	Seine Port Pond	48°33′ N 02°32′ E	N	E	C	+	0.06	2.9	8.0 (0.4)	21.7 (1.7)	280 (34)	4.1 (1.0)
39	Fleuri Pond	48°32′ N 02°22′ E	A	H	C	+	0.08	1.3	8.1 (0.2)	20.8 (1.8)	516 (33)	2.7 (0.3)
40	Villefermoy Reservoir	48°32′ N 02°56′ E	N	H	Cr	+	0.32	2.1	8.4 (0.2)	22.2 (3.2)	289 (41)	9.4 (0.7)
41	Saint-Blaise Pond	48°31′ N 02°22′ E	A	E	C	+	0.05	1.4	8.0 (0.2)	20.5 (2.3)	525 (22)	2.9 (0.2)
42	Bois-le-Roi pond	48°29′ N 02°43′ E	OU	E	I	-	0.08	2.6	8.2 (0.2)	23.3 (3.6)	306 (18)	4.1 (0.1)
43	Samoreau Pond	48°25′ N 02°45′ E	OU	O	I	-	0.10	2.4	8.9 (0.3)	23.3 (4.8)	215 (30)	2.7 (0.4)
44	Delomez Pond	48°24′ N 03°07′ E	N	O	I	-	0.07	4.2	8.2 (0.2)	23.0 (2.5)	225 (28)	3.3 (0.2)
45	Leclerc Pond	48°24′ N 03°05′ E	N	M	I	-	0.05	2.8	8.0 (0.0)	22.9 (3.0)	320 (14)	2.8 (0.4)
46	Grande Paroisse Pond	48°22′ N 02°54′ E	A	E	C	+	0.53	3.7	8.3 (0.2)	22.6 (2.2)	319 (15)	4.5 (0.7)
47	Grand Marais Pond	48°22′ N 02°57′ E	N	H	I	-	0.17	2.4	9.5 (0.6)	23.2 (2.2)	368 (7)	10.4 (0.6)
48	Clarette Pond	48°19′ N 02°42′ E	N	E	I	-	0.14	1.4	8.8 (0.7)	23.2 (4.1)	279 (42)	7.4 (0.2)
49	Souppes/Loing Pond	48°10′ N 02°45′ E	A	M	C	+	0.21	2.9	8.1 (0.1)	23.3 (2.8)	401 (9)	4.2 (1.4)

Values represent the means (standard deviations) of the three yearly campaigns.
Abbreviations: LUI, dominant watershed land use indice; Hydro, type of connection to the hydrological network; DOC, dissolved organic carbon concentration; SSO, presence of a storm sewer outlet; Temp, temperature; Cond, conductivity; N, natural; A, agricultural; OU, open urban; U, urban; O, oligotrophic; M, mesotrophic; E, eutrophic; H, hypereutrophic; I, isolated; C, connected; Cr, crossed.

Supplementary data 2 Relationship between the water temperature and the bacterial community turnover from Créteil Lake. Values were calculated by comparing a sampling campaign with the following (n = 24).

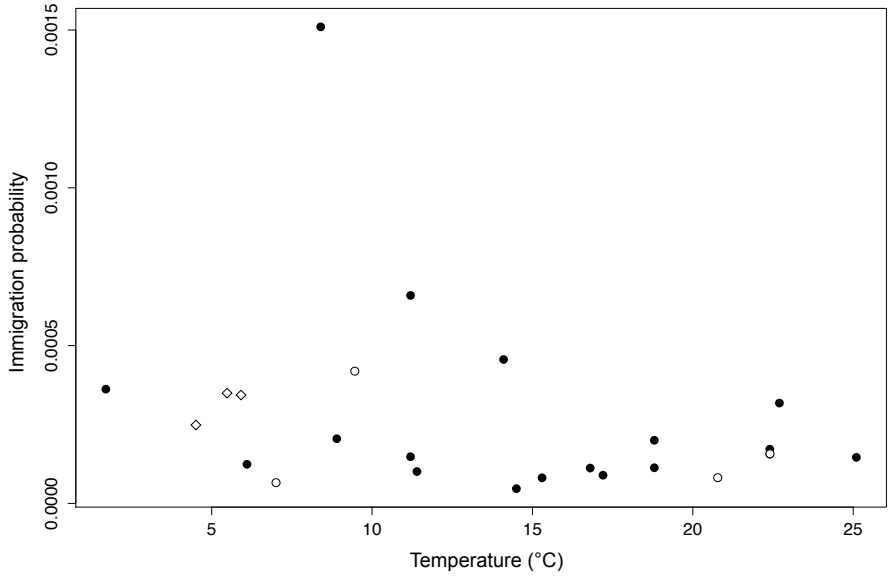


Supplementary data 3 Relationship between the water temperature and the bacterial community nestedness from Créteil Lake. Values were calculated by comparing a sampling campaign with the following (n = 24).



Supplementary data 4 Relationships between the immigration probability *m* and the water temperature over the two-year monthly monitoring of Créteil Lake.

The immigration probability was calculated as the relation between the output parameter $N_T m$ of the neutral community model (NCM) and the mean abundance N_T of the heterotrophic bacteria and cyanobacteria from the points S1, C1, C2, C3 and R1 determined by flow cytometry. Solid symbols represent campaigns where the NCM explained more than 50% of the bacterial assembly variance. Diamond-shaped represents campaigns where the NCM did not significantly shaped the BCS ($P > 0.05$).



Chapter III

Spatial distribution processes differ between bacterial taxonomic groups in freshwater lakes

Spatial distribution processes differ between bacterial taxonomic groups in freshwater lakes

Adélaïde Roguet[a], Claire Therial[a], Adèle Bressy[a], Arnaud Catherine[b], Gérard Lacroix[c,d], Ludwig Jardillier[e] and Françoise S. Lucas[a]

[a]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 94010 Créteil, France.

[b]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.

[c]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.

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

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

Key words: bacterioplankton, community composition, neutral community model, deterministic factors, lake, high-throughput sequencing

I. Abstract

A growing number of studies highlight the importance of neutral processes on bacterial community assemblages, in addition to other dispersal-related processes and environmental conditions. However, no clear consensus has emerged regarding the relative importance of these processes in shaping the entire bacterial communities and the different taxonomic groups. In a metacommunity analysis, we investigated the effects of these mechanisms on the bacterial community composition (BCC) within 49 contrasted lakes, located in the same hydrographical basin. BCC was assessed using high-throughput sequencing (Illumina MiSeq) targeting a portion of the 16S rRNA genes. The bacterial composition of these freshwaters was dominated by the *Actinobacteria*, *Proteobacteria* and *Bacteroidetes*. Overall, this BCC was better predicted by the neutral community model (R^2 = 0.64) than by local environmental conditions and spatial factors (16%). However, the relative importance of the processes involved in the spatial distribution of the bacteria differed among the dominant phyla and classes, suggesting that, although not categorical, some ecological traits were shared among the species that belong to the same phylum or class.

II. Introduction

Processes that shape the spatial distribution of bacterial communities in aquatic habitats remain a central issue in microbial ecology. Over the past decades, several studies showed that freshwater bacterial communities display biogeographical patterns governed by local environmental conditions and dispersal-related processes (Martiny *et al.* 2006; Hanson *et al*. 2012; Lindström and Langenheder 2012). Local environmental factors, also called species sorting, generally include resource availability, osmotic conditions or water temperature (Muylaert *et al.* 2002; Lindström *et al*. 2005; Kent *et al*. 2007; Shade *et al*. 2007; Jones *et al*. 2009). Dispersal-related processes enclose the dispersal limitation of bacterial taxa to reach another location (Hubbell 2001), and mass effect implies a massive immigration of exogenous taxa disturbing the local community composition (Leibold *et al*. 2004). Dispersal-related processes could be impacted by the geographic distance between lakes, geographic barriers, or the links between lakes and the hydrological network (Beisner *et al*. 2006; van der Gucht *et al*. 2007; Nelson *et al*. 2009; Lindström *et al*. 2010). Additionally to the dispersal-related mechanisms, a growing number of recent studies (Sloan *et al*. 2006, Woodcock *et al*. 2007; Drakare and Liess 2010; Ofiţeru *et al*. 2010; Langenheder and Székely 2011) demonstrated that bacterial communities could also be neutrally assembled in various aquatic environments. These results are in agreement with the theory proposed by Hubbell (2001), suggesting that bacterial species with similar fitness may be driven by a stochastic balance between immigration, speciation, emigration and extinction.

Several studies reported the joint action of these processes although not always apprehended with the same parameters (e.g., Drakare and Liess 2010; Ofiţeru *et al*. 2010; Langenheder and Székely 2011; Stegen *et al*. 2012). These observations may potentially result from different bacterial groups linked to specific intrinsic properties or traits (Green *et al*. 2008). Besides, a growing number of studies highlighted that distinct bacterial lineages, even at the phylum rank, could have specific functional traits that could explain unique spatial patterns (e.g., Lindström *et al*. 2005; Holmfeldt *et al*. 2009; Philippot *et al*. 2009; Skékely and Langenheder 2014).

To date, only a few studies have conjointly quantified the relative importance of the local environmental conditions and the spatial factors involved in the dispersal limitation of the dominant bacterial phylum using metagenomic analysis (e.g., Langenheder and Székely 2011; Székely and Langenheder 2014). Similarly, studies integrating the neutral assembly theory from metagenomic data are still scarce (e.g., Logares *et al*. 2013; Langenheder and Székely 2011), although the neutral model can correctly explain on its own the bacterial community composition (BCC) in diverse aquatic environments (e.g., Sloan *et al*. 2006; Woodcock *et al*. 2007; Drakare and Liess 2010; Ofiţeru *et al*. 2010).

Using molecular analysis, we investigated the spatial distribution of freshwater bacterial communities in a large set of shallow lakes located in Paris area (France). Three specific aims were held. First, we sought to determine to what extent local and spatial factors could shape the bacterial community at the regional scale. Then, we evaluated the importance of stochastic processes on the bacterial community assemblage. Third, we explored how these processes acted at different taxonomic ranks. Freshwater from 49 lakes was collected and bacterial community composition was assessed using high-throughput sequencing (Illumina MiSeq) targeting a fragment of the 16S rRNA genes.

III. Materials and methods

III.1 Study area and sampling

This study was conducted in Paris area (Fig. 18), the most populated area in France that 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% are still covered by forests (INSEE 2012). Among the 248 waterbodies larger than 5 ha referenced in the hydrological database Carthage 3.0® (IGN, Paris, France), 49 lakes (Fig. 18 and Supplementary data 1) were selected using a random and stratified sampling strategy (Catherine *et al.* 2008). All the 49 lakes have an artificial anthropogenic origin and can be considered shallow according to the definition of Scheffer (2004). Lakes were sampled in less than 15 days (from 16 to 30 July 2013) to reduce the variability caused by short-term changes in meteorological conditions and nutrient inputs.

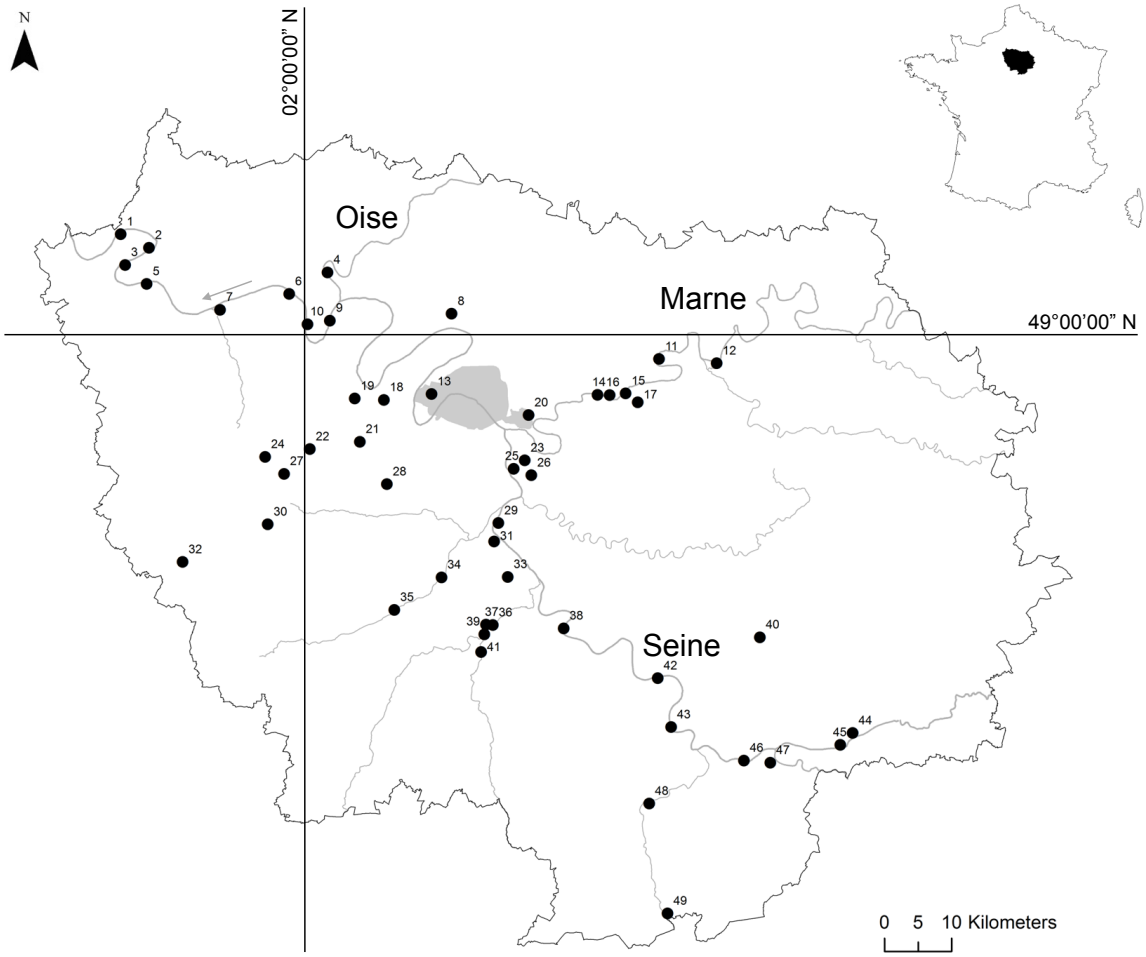


Figure 18 Location of the 49 lakes in the Paris area (France). Numbers correspond to ID lakes.

For each lake, three equidistant sampling stations were selected. At each station, water samples were collected at three depths (depending on the depth of the water column) using a Niskin bottle (General Oceanics Inc., Miami (FL), USA). All samples were then pooled together to obtain an integrated sample. Immediately after collection, samples were kept at 4 °C until their processing.

Successive centrifugations (7,500 × g, 4°C) were performed to concentrate one liter into a 2-mL sterile tube. In order to facilitate pellet resuspension, 1 mL of Tween 80 (final concentration, 0.01% vol/vol) was added before centrifugation to each sample. All samples were stored at – 20°C until DNA extraction.

III.2 DNA extraction

Total DNA was extracted using the FastDNA® SPIN Kit (QBiogene, Carlsbad (CA), USA) according to the manufacturer's instructions. Two modifications to this protocol were

applied, cells were lysed in a FastPrep bead beater three times for 30 s at 4.0 ms^{-1} and the SPIN filters were washed twice. All samples were eluted in 50 µL of deionized sterile water. The quality and quantity of extracted DNA were analyzed at 230, 260, and 280 nm by spectrophotometry before storage at -20°C.

III.3 Illumina sequencing of the V3-V4 region of the 16S rRNA gene

PCR libraries were created by amplifying a fragment including two hypervariable regions V3-V4 of the 16S rRNA genes using the bacterial primers 338F (5'-TCCTACGGGAGGCAGCAGT -3') and 797R (5'- GGACTACCAGGGTATCTAATCCTGTT -3') (Nadkarni *et al*. 2002; Santelli *et al*. 2008). The efficiency of this primer set was tested *in silico* using the probeBase online resource (http://www.microbial-ecology.net/probebase/search.asp, data not shown, pers. comm.). The in silico analysis showed that this set of primers offers the advantage to well amplify bacterial DNA (72% of the total bacterial sequences in SILVA SSU Ref 108 NR dataset) without amplification of chloroplast DNA (5% of the total sequences of chloroplasts). PCR and sequencing were performed by Research and Testing Laboratory (Lubbock (TX), USA), using the Illumina MiSeq platform (Illumina Inc., San Diego (CA), USA). All DNA extracts were adjusted to 120 $ng.\mu L^{-1}$ prior to PCR reactions. Quality control was processed for each sequence using the software QIIME v. 1.8.0-20140103 (Caporaso *et al*. 2010). Sequences of poor quality were removed using the default parameters (quality score < 25, length < 100 nt and at least one ambiguous base or seven successive homopolymers). Chimeric sequences were then identified with USEARCH 6.0 using the UCHIME algorithm in *de novo* mode (Edgar *et al*. 2011) implemented in the online service FunGene pipeline and removed from the dataset. Sequences were clustered into 97% similarity-based operational taxonomic units (OTUs). Since the number of reads differed between samples, a subsampling strategy was performed to normalize the dataset as suggested by Aguirre de Cárcer and colleagues (2011). Briefly, subsampling consisted of a randomly resampling (100 times) of the whole reads until the median (16326 of reads) was reached and then all singletons were deleted. Rarefaction curves of the OTUs are presented in Supplementary data 2. Sequences corresponding to OTUs were deposited in the GenBank database with accession number

(the submission on NCBI is in progress). OTUs were assigned to the genus level with the non-redundant SILVA 111 database (Quast *et al.* 2013) using UCLUST (Edgar 2010).

III.4 Environmental parameters

Total nitrogen and phosphorus concentrations were assayed by colorimetry using a Cary 50 Scan spectrophotometer (Varian Inc., Palo Alto (CA), USA) according to Rogora and colleagues (2006) and the French standard (AFNOR NF T 90-023) respectively. Secchi depth and vertical profiles of physicochemical parameters were determined for each lake at the three stations. 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 (Sea-Bird Electronics Inc., Bellevue (WA), USA). Total suspended solids (TSS) concentration was quantified after filtration of one liter on precombusted tarred Whatman GF/F filter. Dissolved organic carbon (DOC) concentrations were measured using a TOC-VCSN carbon analyzer (Shimadzu, Columbia (MD), USA).

III.5 Watershed land use assessment

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 for physical barriers that potentially modify the water flow (e.g., roads or underground drainages).

III.6 Data analysis

Diversity index and evenness determination. The degree of similarity was analyzed among the bacterial assemblages. Evenness was calculated using the Simpson's evenness index (Smith and Wilson 1996). Results of evenness ranged from 0 (when one OTU was predominant) to 1 (when all OTUs were equally abundant). The alpha diversity was estimated using the exponential form of Shannon entropy. This index gives intuitive measures that correspond to the number of OTUs when OTUs are equally abundant in a sample (Whittaker 1972; Jost 2006). To evaluate the influence of environmental factors on the exponential form of Shannon entropy, a linear model was performed. Normality and homoscedasticity of residuals were checked according to Zuur and colleagues (2009). The

dissimilarity of NTM assemblages among the 49 lakes (beta diversity) was evaluated using the Sørenson dissimilarity index. The beta diversity was then partitioned using the *beta.pair* function from the 'betapart' package (Baselga 2010; Baselga *et al.* 2013) to quantify the fraction of dissimilarity explained by OTUs replacement (based on the Simpson's dissimilarity index) and from pure random richness variation of the OTUs (Azeria *et al.* 2011).

Drivers of the spatial distribution of the bacterial community The relative importance of local environmental and spatial factors was quantified by decomposing the total bacterial community variation into unique environmental (E) and spatial (S) components using variance partitioning analysis (VPA) (Peres-Neto *et al.* 2006). Six local environmental variables were injected in VPAs, 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 reflecting the dispersal and the structure at the regional scale. The first component included variables reflecting the dispersal of bacterial taxa, i.e. the presence of direct storm sewer discharge into the lake and the type of link with the hydrographic network. The type of linkage to the hydrographic network was decomposed into three categories (that 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. The second component was represented by variables integrating the spatial patterns of the communities at the regional scale. This variable was evaluated using the eigenvectors derived from the principal components of neighbor matrices (PCNM) of spatial coordinates (Borcard and Legendre 2002; Legendre and Gauthier 2014). PCNM variables reflect the spectral decomposition of the spatial relationships among the study sites. Variance partitioning analyses estimates the proportion of BCC variation (adjusted R-squared) that can be attributed to local environmental variation [E], spatial variation [S], local variation without

regional component [E|S], spatial variation without environmental component [S|E], variation explained by their interaction [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). The method was implemented using the *varpart* function from the 'vegan' package (Oksanen *et al.* 2013). Prior to VPA, forward selections were performed for both components according to Borcard and colleagues (2011) and Legendre and Gauthier (2014).

To assess the neutral assembly of the bacterial communities, we used the method developed by Sloan and colleagues (2006) to fit the regional relative abundance of OTUs and their observed detection frequency. Contrary to the Hubbell's model (Hubbell 2001), this continuous model is particularly suited for bacterial communities (i.e. with large population size) detected using molecular tools (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 bacterial community and "m" to the immigration rate (the probability that a dead individual is replaced by an immigrant). This parameter was estimated using the best fit between detection frequency of OTUs and their regional relative abundance, by minimizing the sum of squares of errors. The detection limit was fixed to 1.10^{-5}, which corresponds approximately to the threshold of the sequencing method. The goodness of fit was evaluated using the determination coefficient R^2.

All statistical analyses and indices computing were conducted using the statistical environment R version 3.1.1 (R Development Core Team 2014). The default statistical significance was based on a $P \leq 0.05$ level.

IV. Results

IV.1 Bacterial richness

A total of 886,505 reads were recovered from the 49 lakes with an average length of 410 bp. Trimming, chimera check and subsampling removed 193,627 sequences (Table 7). Finally 691,739 sequences clustered in 5,290 OTUs were identified as belonging to the bacteria kingdom (Table 7). Only three percent of the OTUs (180/5,290) had a relative

abundance higher than 0.5% in at least one of the 49 lakes samples. The bacterial assemblages were dominated by few OTUs and composed by a majority of rare OTUs, as underlined by the low values of the evenness index (Table 7). The alpha diversity ranged from 14 to 69 OTUs per lake (if all OTUs were equally distributed), with a median of 26 (Table 6). This diversity index was significantly driven by the pH and the trophic status of lakes, with a bacterial diversity significantly higher in eutrophic lakes (linear model, $F_{8,40} = 8.16$, $R^2 = 0.62$, $P < 0.001$ for both parameters).

IV.2 Composition of the bacterial assemblages

Overall, the bacterial composition was dominated by the *Actinobacteria* with 62% of the total sequences, followed by the *Proteobacteria* (29%) and the *Bacteroidetes* (6%) (Fig. 19). *Actinobacteria* phylum was dominated by the *Actinobacteria* subclass (Supplementary data 3), whereas *Proteobacteria* and *Bacteroidetes* were dominated by three main classes, respectively the *Alpha-*, *Beta-* and *Gammaproteobacteria,* and *Sphingobacteriia*, *Cytophagia* and *Flavobacteria* (Fig. 20). Dominant families and genus are listed in Table 7.

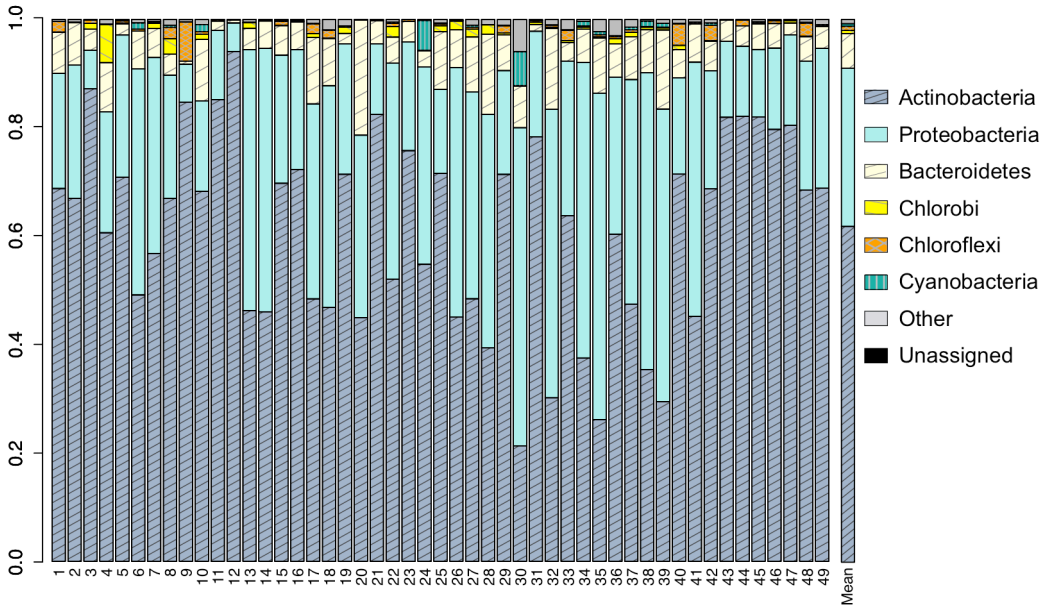


Figure 19 Relative abundances per lake (percentage of total sequences) of phylogenetic phyla identified by Illumina MiSeq sequencing. The group "Other" includes the phyla (in decreasing order) *Verrucomicrobia, Gemmatimonadetes, Firmicutes, Planctomycetes, Armatimonadetes, Acidobacteria,* Candidate_division_TM7, *Nitrospirae, Spirochaetes, Fusobacteria, Deinococcus-Thermus,* WCHB1-60, *Lentisphaerae,* TM6, *Elusimicrobia,* Candidate_division_OP3, *Tenericutes, Caldiserica*, OC31, *Chlamydiae,* Candidate_division_WS3 and BD1-5.

Table 6 Comparison of sequenced libraries, including the number of reads pre- and post-trimming and subsampling, the observed richness, the diversity index and the evenness index for each lake.

ID lake	No. of raw sequences	No. of filtered sequences	Subsampling		No. sequences belonging to the bacteria	No. OTUs belonging to the bacteria	Diversity index (Exp, form of Shannon)	Evenness index % (Simpson evenness)
			No. sequences	No. OTUs				
1	13133	12289	11990	316	11788	309	34.9	0.3
2	16337	13795	13451	300	13437	294	24.4	0.3
3	25716	23312	16326	632	15955	627	22.0	0.14
4	13560	12347	12053	288	12043	285	30.8	0.32
5	18256	16326	15855	378	15840	374	28.4	0.24
6	16423	15309	15030	268	15028	267	28.7	0.35
7	16377	15160	14734	378	14732	377	39.0	0.25
8	19290	17330	16323	985	15424	958	69.4	0.1
9	32688	29553	16323	708	15980	698	13.9	0.12
10	17233	15684	15256	347	15223	342	32.0	0.27
11	21965	19132	16326	608	15887	600	21.6	0.15
12	23837	20360	16326	579	15947	578	17.1	0.15
13	10680	9973	9628	258	9628	258	26.5	0.35
14	11310	10625	10379	217	10360	215	25.8	0.43
15	25031	22625	16326	751	15832	730	28.2	0.12
16	19418	17802	16326	583	15858	575	26.9	0.15
17	12995	11805	11292	493	11275	487	76.8	0.2
18	13201	12124	11780	324	11757	318	37.8	0.29
19	13472	12312	11875	349	11859	344	36.9	0.27
20	14938	13428	13177	220	13164	219	25.5	0.42
21	8945	7860	7662	153	7633	148	12.0	0.55
22	12745	11744	11425	294	11407	289	28.7	0.32
23	26913	24168	16326	579	16001	569	21.5	0.16
24	21005	19436	16326	675	15853	668	36.4	0.14
25	23932	21630	16326	801	15812	782	37.6	0.12
26	23440	21897	16326	648	15932	638	31.0	0.14
27	18524	16928	16320	1068	15368	1045	88.9	0.09
28	23208	21517	16326	562	15977	555	32.0	0.16
29	23406	21353	16326	1059	15627	1027	46.3	0.09
30	18604	17907	16326	579	15869	569	33.6	0.16
31	7349	6735	6517	172	6515	171	18.1	0.52
32	14603	13126	12623	569	12615	566	92.3	0.17
33	14967	13684	13324	395	13307	388	52.8	0.24
34	14177	12822	12282	509	12262	504	93.2	0.19
35	15521	14060	13513	611	13490	605	112.6	0.16
36	21037	19006	16326	1076	15487	1043	62.1	0.09
37	20303	18682	16326	1094	15491	1061	82.8	0.09
38	14676	13492	13179	342	13160	337	52.8	0.28
39	16284	14702	14034	640	14017	633	98.2	0.15
40	19340	17339	16325	879	15585	856	62.0	0.11
41	20671	18792	16326	1146	15480	1130	63.6	0.08
42	14946	13575	13153	351	13121	342	32.8	0.26
43	25583	22537	16326	543	16011	535	16.9	0.16
44	20255	17600	16326	622	15811	614	18.5	0.13
45	21239	18027	16326	594	15851	581	22.9	0.15
46	17393	15588	15264	270	15250	266	17.7	0.32
47	28460	25432	16326	705	15938	702	24.2	0.13
48	19481	17318	16326	863	15166	839	43.4	0.11
49	16771	15083	14724	350	14686	340	32.9	0.27

OTUs, operational taxonomic units

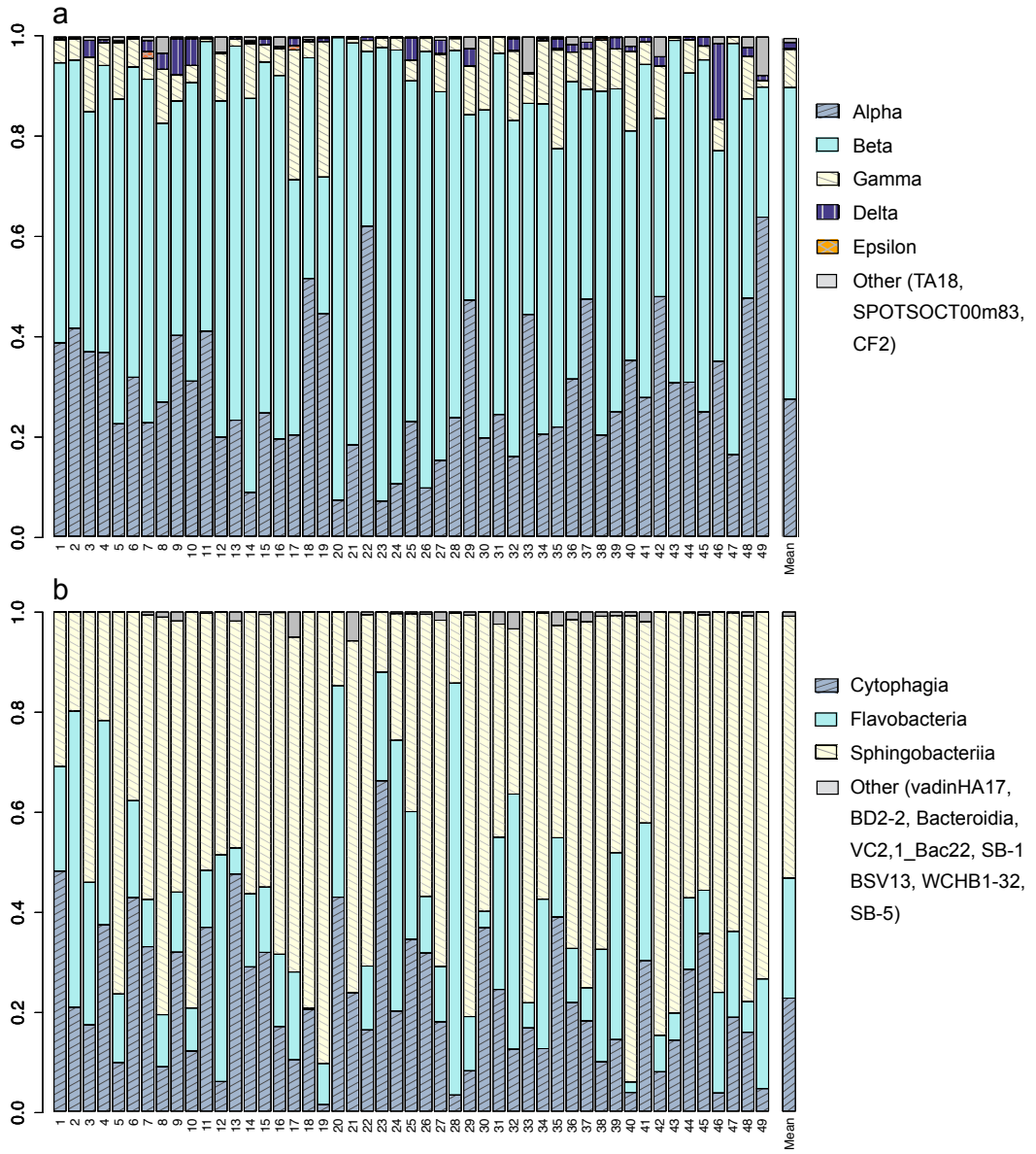


Figure 20 Relative abundances per lake (percentage of total sequences) of phylogenetic class for (a) *Proteobacteria* and (b) *Bacteroidetes* identified by Illumina MiSeq sequencing.

IV.3 Spatial structure of the total BCC

The beta diversity between the 49 lakes was of 67% of dissimilarity. When only considering the dominant OTUs (≥ 0.5%), the beta diversity felt to 30% whereas rare OTUs displayed a value of 76%. By partitioning the beta diversity of the bacterial community, we observed that OTUs replacements explained 79% of the beta diversity, against 21% explained by a random shift of OTUs. Local conditions (mainly trophic status and water pH) and spatial factors alone (type of link to the hydrological network and the PCNM

variables) and their interaction explained only 16% of the BCC dissimilarity among the 49 lakes, as evaluated by the variance partitioning analysis (VPA) analysis (Table 7). Thus, a large amount of the variance, i.e. 84%, remained unexplained by these factors (Table 7). Regarding the dominant and rare OTUs, VPA explained 19% of the variability within the dominant fraction against 9% for the rare OTUs (Table 7).

Contrary to local and spatial factors, the neutral model explained a large fraction in the relationship between the OTUs occurrence frequency and their relative abundance variations with an R^2 of 64% (Table 8). The neutral community model also properly fit the dominant taxa and the rare OTUs with a R^2 of 68 and 73% respectively (Table 8).

IV.4 Spatial structure of the dominant phyla

We explored whether the patterns and processes observed at the bacterial community level were similar inside specific taxonomic groups from phyla to classes. From the class level, only the three dominant bacterial groups with a regional relative abundance higher than 0.5% were considered. Although no strong relation between the phylogeny and the processes involved in the spatial structure of bacterial communities was identified (UPGMA clustered analysis based on Euclidian distance, data not shown), there were similar trends within different dominant phyla and classes in the relative importance of the processes shaping the bacterial communities.

Large discrepancies within the three dominant bacterial phyla (Actinobacteria, Proteobacteria, Bacteroidetes) were observed in the total variance explained by environmental and spatial factors (Table 7). Within the *Actinobacteria* phylum, the relative importance of the spatial factors differed from one taxonomic group to another. *Acidimicrobiia* assembly was mainly explained by the factor 'type of linkage to the hydrological network' (up to 15%), while the dominant *Actinobacteria* subclass had a smaller total variance explained by the VPA compared to the whole bacterial community. In addition, *Alpha- and Betaproteobacteria* were significantly sensitive to the environmental conditions (up to 13%) and not to the spatial factors, contrary to the *Gammaproteobacteria*. Finally, *Bacteroidetes* appeared to be poorly shaped by both the environmental conditions and the selected spatial variables.

Table 7 Variation partitioning analysis of the bacterial community composition.

	Variance partitioning analysis (R^2_{adj})						Variables selected by forward selection	
	Env. factors [E]	Spatial factors [S]	Env. factors alone [E|S]	Spatial factors alone [S|E]	Interaction [E∩S]	Residuals	Environmental factors	Spatial factors
Total bacterial community	**0.10**	**0.12**	**0.04***	**0.06***	**0.06**	**0.84**	**TS, pH**	**Hydro, PCNM**
Dominant community	0.11	0.15	0.05*	0.08*	0.06	0.81	TS, pH	Hydro, PCNM
Rare community	0.06	0.06	0.03*	0.03*	0.04	0.91	TS, pH, DOC	Hydro, PCNM
Actinobacteria (phylum)	**0.09**	**0.07**	**0.04^{NS}**	**0.03***	**0.04**	**0.88**	**TS, pH**	**Hydro, PCNM**
Acidimicrobiia (subclass)$^{+}$	**0.10**	**0.23**	**0.02^{NS}**	**0.15***	**0.08**	**0.75**	**TS**	**Hydro, PCNM**
Actinobacteria (subclass)	**0.10**	**0.09**	**0.05***	**0.04***	**0.05**	**0.86**	**TS, pH**	**Hydro, PCNM**
Frankiales (suborder)$^{+}$	0.06	0.14	0.00^{NS}	0.07*	0.06	0.87	TS	Hydro, SSO, PCNM
Micrococcales (suborder)$^{+}$	0.12	0.04	0.09*	0.01^{NS}	0.03	0.86	TS, pH, Cond	Hydro
PeM15 (suborder)$^{+}$	0.09	0.03	0.05*	0.00^{NS}	0.03	0.91	TS	Hydro
Proteobacteria (phylum)	**0.13**	**0.07**	**0.07***	**0.01^{NS}**	**0.06**	**0.85**	**TS, pH, Cond, Temp**	**Hydro, PCNM**
Alphaproteobacteria (class)	**0.13**	**0.11**	**0.07***	**0.05***	**0.06**	**0.82**	**TS, pH, Cond**	**Hydro, PCNM**
Rhizobiales (order)$^{+}$	0.16	0.05	0.11*	0.00^{NS}	0.05	0.85	TS, pH, Cond	Hydro
Rhodobacterales (order)$^{+}$	0.17	0.06	0.12*	0.00^{NS}	0.05	0.82	TS, pH, Cond	Hydro
Sphingomonadales (order)$^{+}$	0.07	0.09	0.01^{NS}	0.04^{NS}	0.06	0.90	pH	Hydro, PCNM
Betaproteobacteria (class)	**0.15**	**0.06**	**0.10***	**0.02^{NS}**	**0.04**	**0.84**	**TS, pH, Cond, Temp**	**Hydro, SSO, PCNM**
Burkholderiales (order)	0.16	0.07	0.10*	0.01^{NS}	0.06	0.83	TS, pH, Cond, Temp	Hydro, PCNM
Burkholderiaceae (family)$^{+}$	0.15	0.05	0.12*	0.02^{NS}	0.03	0.83	TS, pH, Cond	Hydro
Comamonadaceae (family)$^{+}$	0.19	0.07	0.13*	0.01^{NS}	0.06	0.80	TS, pH, Cond, Temp	Hydro, SSO, PCNM
Methylophilales (order)$^{+}$	0.15	0.10	0.07*	0.02^{NS}	0.08	0.83	TS	Hydro
Gammaproteobacteria (class)$^{+}$	**0.09**	**0.07**	**0.05***	**0.04***	**0.04**	**0.88**	**TS, pH, Cond**	**Hydro, PCNM**
Bacteroidetes (phylum)	**0.05**	**0.03**	**0.03***	**0.01^{NS}**	**0.02**	**0.94**	**TS**	**Hydro, PCNM**
Cytophagia (class)$^{+}$	**0.08**	**0.04**	**0.06***	**0.02^{NS}**	**0.02**	**0.91**	**TS**	**Hydro**
Flavobacteria (class)$^{+}$	**0.04**	**0.07**	**0.02^{NS}**	**0.05***	**0.02**	**0.91**	**TS, DOC**	**Hydro, PCNM**
Sphingobacteriia (class)$^{+}$	**0.08**	**0.05**	**0.04***	**0.01^{NS}**	**0.04**	**0.91**	**TS, pH**	**Hydro, PCNM**

Abbreviations: TS, trophic status; DOC, dissolved organic carbon concentration; Cond, water conductivity; Temp, water temperature; PCNM, reflect the community structuring using PCNM variables; Hydro, reflects the dispersal limitation by the type of linkage to the hydrographical network; SSO, storm sewer outlet discharges.

Symbols associated to the adjusted R squared correspond to the significance (symbols: *, $P<0.05$; NS, non significant) of the partial contribution of factors tested alone in the presence of the others.

$^{+}$, indicates that the bacterial class/subclass, order/suborder or family is dominated by a single genus.

Table 8 Neutral community model parameters.

	Regional relative abundance	$N_T m$	m'	R^2	p-value
Total bacterial community	**1.00**	**559.2**	**5.6**	**0.64**	**<0.001**
Dominant community	0.89	111.6	1.3	0.68	<0.001
Rare community	0.11	63.4	5.7	0.73	<0.001
Actinobacteria (phylum)	**0.62**	**741.7**	**12.0**	**0.77**	**<0.001**
Acidimicrobiia (subclass)$^+$	**0.06**	**19.7**	**3.1**	**0.65**	**<0.001**
Actinobacteria (subclass)	**0.55**	**770.2**	**13.9**	**0.81**	**<0.001**
Frankiales (suborder)$^+$	0.50	765.6	15.3	0.82	<0.001
Micrococcales (suborder)$^+$	0.04	26.6	6.0	0.86	<0.001
PeM15 (suborder)$^+$	0.01	0.6	0.9	0.81	<0.001
Proteobacteria (phylum)	**0.29**	**99.6**	**3.4**	**0.69**	**<0.001**
Alphaproteobacteria (class)	**0.08**	**19.4**	**2.4**	**0.76**	**<0.001**
Rhizobiales (order)$^+$	0.02	2.0	0.9	0.62	<0.001
Rhodobacterales (order)$^+$	0.02	4.1	2.5	0.81	<0.001
Sphingomonadales (order)$^+$	0.03	4.6	1.4	0.84	<0.001
Betaproteobacteria (class)	**0.18**	**65.9**	**3.6**	**0.73**	**<0.001**
Burkholderiales (order)	0.16	58.2	3.6	0.73	<0.001
Burkholderiaceae (family)$^+$	0.04	24.0	5.3	0.71	<0.001
Comamonadaceae (family)$^+$	0.03	31.0	11.4	0.76	<0.001
Methylophilales (order)$^+$	0.01	1.8	1.3	0.81	<0.001
Gammaproteobacteria (class)	**0.02**	**4.0**	**1.8**	**0.61**	**<0.001**
Bacteroidetes (phylum)	**0.06**	**8.6**	**1.4**	**0.60**	**<0.001**
Cytophagia (class)$^+$	**0.01**	**0.9**	**0.6**	**0.81**	**<0.001**
Flavobacteria (class)$^+$	**0.02**	**1.9**	**1.2**	**0.81**	**<0.001**
Sphingobacteriia (class)$^+$	**0.03**	**4.1**	**1.2**	**0.59**	**<0.001**

$N_T m$ parameter was estimated using the model developed by Sloan and colleagues (2006). It represents the meta-community size multiplied by the immigration probability *m*. Since we did not know the real abundance of the bacterial community, the immigration probability was estimated *m'* using the regional relative abundance of the bacterial groups.

Whatever the bacterial group, the neutral community models well explained the relationship between the OTUs occurrence frequency and their relative abundance variations with R^2 ranging from 59 to 86% (74% on average). However, assuming that the parameter N_T was proportional to the relative abundance of the phylum among the 49 lakes, variations in the estimated immigration probability m' were observed between the three dominant phyla. The immigration rate was relatively higher for the *Actinobacteria* compared to the *Proteobacteria* or the *Bacteroidetes* (Table 8).

V. Discussion

This regional-scale study not only explored the relative influence of the local environmental conditions and the spatial factors, but also the neutral processes that governed the spatial variations in the bacterial community composition and the spatial distribution of the dominant taxonomic groups. The bacterial communities from a large set of lakes were assessed using a high-throughput sequencing methodology that allowed an in-depth analysis of the bacterial diversity. Major results suggested that the entire bacterial community composition (BCC) was shaped by stochastic processes, whereas at specific taxonomic ranks (phylum to classes), local and spatial processes were also found as important mechanisms governing the bacterial composition, with global tendencies within each phyla or classes.

V.1 Bacterial community composition in the 49 lakes

Overall, *Actinobacteria* dominated the bacterial metacommunity. A similar tendency was previously recorded in freshwater habitats, where the *Actinobacteria* phylum represented a large fraction (20-70%) of the total bacterial assemblage, specifically the *Actinobacteria* subclass (e.g., Warnecke *et al*. 2005; Allgaier and Gossart 2006; Holmfeldt *et al*. 2009). *Proteobacteria* (especially the *Betaproteobacteria*) and the *Bacteroidetes* (*Sphingobacteria*) were also found to be abundant in the 49 lakes. In the literature, these phyla are often identified as abundant in lakes epilimnia (e.g., Glöckner *et al*. 1999; Mueller-Spitz *et al*. 2009; Barberán and Casamayor 2010).

Despite the dominance of few phyla, large inter-lake variations in the BCC were observed at the OTU level between the 49 lakes. This variability among lakes was not the pure consequence of a random shift in the OTUs richness, but was rather influenced by species turnover. Main results evidenced the importance of stochastic processes compared with the local and regional factors on the total BCC.

V.2 Small importance of local and regional factors on the spatial BCC

Our results indicate that local conditions and spatial factors have a small influence in structuring the BCC at the regional scale, as suggested by the low amount of total variance explained by these two factors. A similar tendency was encountered in a three-

year survey performed on the same set of lakes using the T-RFLP method (Roguet *et al*. 2015). Some environmental and spatial factors were identified as having a significant effect on the BCC with both approaches, such as the trophic status and the link to the hydrological network. The fact that both approaches gave similar conclusions over the three-year survey brings some credit in the analysis and results. The trophic status, nutrients availability or the osmotic conditions has been previously reported to shape the bacterial community (Lindström 2000; Kent *et al*. 2007; Shade *et al*. 2007). Among the spatial factors, the potential ability of bacteria to disperse (characterized by the type of linkage to the hydrographical network) and the spatial structuring were mainly identified as shaping the BCC. Moreover, in this analysis, no relevant major influence of the presence of storm sewer discharge into the lakes was identified.

Although in this study the deterministic factors (environmental and spatial) seemed to marginally explain the regional variation of beta-diversity, the interpretation of these results should be taken with caution since several methodological and statistical biases could overestimate the relative importance of the unexplained variance or underestimate the contribution of the environmental and spatial factors to the variance partitioning analysis (Gilbert and Bennett 2010, Lindström and Langenheder 2012, Székely and Langenhender 2014, Roguet *et al*. 2015). Moreover, the relevant variables may be absent among the tested variables, in particular the organic matter quality (Judd *et al*. 2006) or the biotic factors reflecting the top-down pressures exerted on the bacterial community (van der Gucht *et al*. 2007).

V.3 Stochastic processes and structure of the bacterial metacommunity

The neutral community model was relatively powerful (R^2 = 64%) for explaining the OTUs proportion observed in the 49 lakes, suggesting that the total BCC was driven by a stochastic equilibrium between taxa loss and gain (Hubbell 2001; Sloan *et al*. 2006). Similar observations have already been reported in previous experimental and limnological studies, where stochastic processes also displayed a higher influence than local factors and/or regional processes (Langenheder and Székely 2011; Lee *et al*. 2013; Drakare and Liess 2010). The prevalence of neutral processes on the bacterial assemblage could be explained by the apparent reduced sensitivity of bacterial communities to local

factors, due to their strong adaptability and plasticity to environmental perturbations and gradients (Östman *et al*. 2010) or to the dominance of generalists in the bacterial community (Langenheder and Székely 2011). However it should be kept in mind that the analyzed dataset represent a snapshot that may give a biased view of the taxa loss and gain analysis.

Although the total bacterial community assemblage seemed to be driven by stochastic processes, we could not infer the absence of local and spatial factors influence on BCC (Chisholm and Pacala 2010). Indeed, within the different phyla, local and spatial factors could have a higher influence on the species assemblage.

V.4 Local and dispersal processes effects depend on the bacterial group

Although not categorical, there are evidences of specific processes associated with phyla or classes. First, *Actinobacteria* and *Bacteroidetes* are generally considered to be generalist groups in freshwater habitats (Tamames *et al*. 2010; Langenheder and Székely 2011; Newton *et al*. 2011). For both groups their generalist nature could explain the low impact of the environmental and spatial factors on their spatial structuration. However the two groups seemed to differ in their dispersal capacity. The spatial distribution of the *Acidimicrobiia* (*Actinobacteria*) was strongly related to the type of link to the hydrological network. This result may suggest the existence of mass effect overriding the influence of the environmental conditions (Hubbell 2001) and affecting the spatial distribution of *Actinobacteria*, as a massive load of taxa could be expected in lakes crossed by a river. The widespread distribution of *Actinobacteria* may be the consequence of the absence of a significant influence of the local environmental conditions on two dominant bacterial groups *Acidimicrobiia* and *Frankiales*. Finally, the *Proteobacteria* was the unique phylum to be strongly influenced by local water conditions and, in a lesser extent by spatial factors or immigration rate.

Although some patterns were observed within phyla or classes, evidencing that the spatial distribution of bacterial taxa may reflect their ecological traits (Fierer *et al*. 2007; Philippot *et al*. 2010; Lennon *et al*. 2012), the absence of categorical relationship between the phylogeny and the importance of the different processes suggested a random association between phylogenetic constraints and some functional traits (Boucher *et al.*

2003), likely due to gene loss, convergent evolution or lateral transfer (Snel *et al.* 2002; Martiny *et al*. 2013).

In this study, we observed that the bacterial community assemblages in 49 lakes located in Paris area were mainly shaped by stochastic processes and poorly dependent on local conditions and spatial factors. Although this study only provided a snapshot of the bacterial community, as we did not take temporal variation into account, these tendencies (regarding the total and the dominant bacterial community) are in accordance with those obtained from a fingerprinting study performed between 2011 and 2012 on the identical set of lake. However, we also attested that similar tendencies were found between close bacterial taxonomic lineages in the processes involved in their spatial distribution. Considering the importance of dispersal-related processes, notably of the type of linkage to the hydrographic on the bacterial community composition, it would be interesting to take into account this parameters in further analyzes.

VI. Acknowledgments

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]. The authors thank Sophie Penisson for her support in decrypting the equation of the neutral model. We are grateful to the stakeholders and owners for their authorization to collect samples in each lake. Finally we warmly thank all the colleagues who helped in sampling and quantifying the ecological parameters in the 49 lakes.

VII. References

Aguirre de Cárcer D, Denman SE, McSweeney C and Morrison M (2011) Evaluation of subsampling-based normalization strategies for tagged high-throughput sequencing data sets from gut microbiomes. *Appl Environ Microbiol* 77:8795–8798.

Allgaier M and Grossart H (2006) Diversity and seasonal dynamics of Actinobacteria populations in four lakes in northeastern Germany. *Appl Environ Microbiol* 72:3489–3497.

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.

Boucher Y, Douady CJ, Papke RT, Walsh DA, Boudreau MER, Nesbø CL, Case RJ and Doolittle WF (2003). Lateral gene transfer and the origins of prokaryotic groups. *Annu Rev Genet* 37: 283–328.

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.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, Mcdonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J and Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336.

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.

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.

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.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST.

Bioinformatics 26:2460–2461.

Edgar RC, Haas BJ, Clemente JC, Quince C and Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200.

Fierer N, Morse JL, Berthrong ST, Bernhardt ES and Jackson RB (2007) Environmental controls on the landscape-scale biogeography of stream bacterial communities. *Ecology* 88:2162–2173.

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

Glöckner FO, Fuchs BM and Amann R (1999) Bacterioplankton composition in lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Appl Environ Microbiol* 65:3721–3726.

Green JL, Bohannan BJM and Whitaker RJ (2008) Microbial biogeography: from taxonomy to traits. *Science* 320:1039–1043.

Hanson CA, Fuhrman JA, Horner-Devine MC and Martiny JBH (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* 10:497–506.

Holmfeldt K, Dziallas C, Titelman J, Pohlmann K, Grossart HP and Riemann L (2009) Diversity and abundance of freshwater Actinobacteria along environmental gradients in the brackish northern Baltic Sea. *Environ Microbiol* 11:2042–2054.

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.

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.

Jost L (2006) Entropy and diversity. *Oikos* 113:363–375.

Judd KE, Crump BC and Kling GW (2006) Variation in dissolved organic matter controls bacterial production and community composition. *Ecology* 87:2068–2079.

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 Székely AJ (2011) Species sorting and neutral processes are both

important during the initial assembly of bacterial communities. *ISME J* 5:1086–1094.

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

Lennon JT, Aanderud ZT, Lehmkuhl BK and Schoolmaster DRJ (2012) Mapping the niche space of soil microorganisms using taxonomy and traits. *Ecology* 93:1867–1879.

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

Lindström ES, Feng XM, Grane W and Kritzberg ES (2010) The interplay between bacterial community composition and the environment determining function of inland water bacteria. *Limnol Oceanogr* 55:2052–2060.

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.

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.

Martiny AC, Treseder K and Pusch G (2013) Phylogenetic conservatism of functional traits in microorganisms. *ISME J* 7:830–838.

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.

Mueller-Spitz SR, Goetz GW and McLellan SL (2009) Temporal and spatial variability in nearshore bacterioplankton communities of Lake Michigan. *FEMS Microbiol Ecol* 67:511–522.

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.

Nadkarni MA, Martin FE, Jacques NA and Hunter N (2002) Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* 148:257–266.

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(4):1292–1305.

Newton RJ, Jones SE, Eiler A, McMahon KD and Bertilsson S (2011) A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* 75:14–49.

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 (2013) Package "vegan". Community ecology package, version, 2(9).

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

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.

Philippot L, Bru D, Saby NPA, Čuhel J, Arrouays D, Šimek M and Hallin S (2009) Spatial patterns of bacterial taxa in nature reflect ecological traits of deep branches of the 16S rRNA. *Environ Microbiol* 11:3096–3104.

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.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J and Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:590–596.

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.

Roguet A, Laigle GS, Therial C, Bressy A, Soulignac F, Catherine A, Lacroix G, Jardillier L, Bonhomme C, Lerch TZ and Lucas FS. Neutral community model explains the bacterial community assembly in freshwater lakes. FEMS Microbiol Ecol 91(11), fiv125.

Santelli CM, Orcutt BN, Banning E, Bach W, Moyer CL, Sogin ML, Staudigel H and Edwards KJ (2008) Abundance and diversity of microbial life in ocean crust. *Nature* 453:653–656.

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

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* 8:732–440.

Smith B and Wilson JB (1996) A consumer's guide to evenness indices. *Oikos* 76:70–82.

Snel B, Bork P and Huynen MA (2002) Genomes in flux: the evolution of Archaeal and Proteobacterial gene content. *Genome Res* 12:17–25.

Stegen JC, Lin X, Konopka AE and Fredrickson JK (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J* 6:1653–1664.

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.

Tamames J, Abellán JJ, Pignatelli M, Camacho A and Moya A (2010) Environmental distribution of prokaryotic taxa. *BMC Microbiol* 10:85.

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.

Warnecke F, Sommaruga R, Sekar R, Hofer JS and Pernthaler J (2005) Abundances, identity, and growth state of *Actinobacteria* in mountain lakes of different UV transparency. *Appl Environ Microbiol* 71:5551–5559.

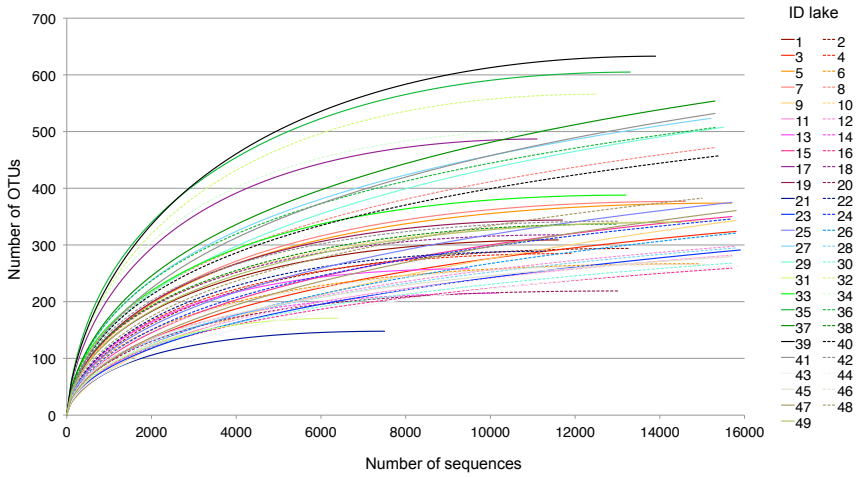
Whittaker RH (1972) Evolution and measurement of species diversity. *Taxon* 21:213–251.

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.

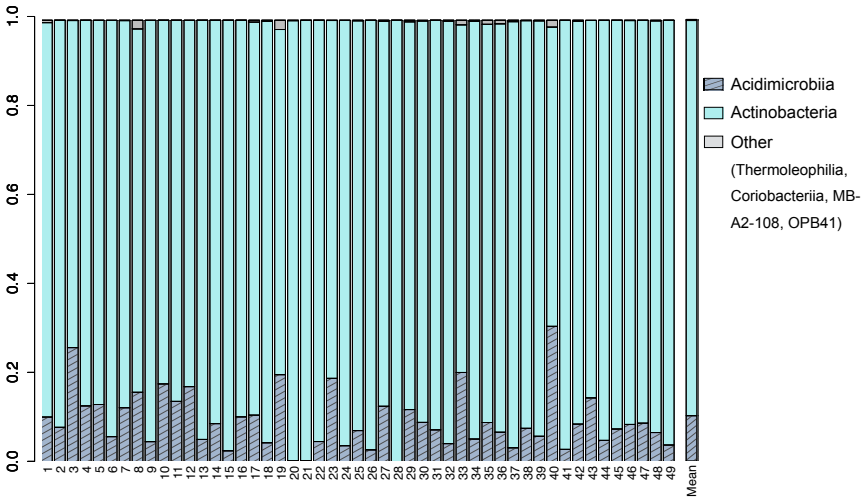
Zuur A, Ieno EN, Walker N, Saveliev AA and Smith GM (2009). *Mixed effects models and extensions in ecology with R*. Springer Science and Business Media.

VIII. Supplementary data

Supplementary data 1 Rarefaction curves obtained for the forty-nine water samples collected in the 49 lakes in Paris area (at a cutoff level of 3%). The correspondence between the ID lake and the name of lakes are listed in Table S1. Abbreviation: OTU, operational taxonomic unit.



Supplementary data 2 Relative abundances per lake (percentage of total sequences) of the *Actinobacteria* identified by Illumina MiSeq sequencing.



Chapter IV

High mycobacterial diversity in recreational lakes

High mycobacterial diversity in recreational lakes

Adélaïde Roguet [a], Claire Therial[a], Mohamed Saad[a], Lila Boudahmane[a], Laurent Moulin [b] and. Françoise S. Lucas[a]

[a]Laboratoire Eau Environnement et Systèmes Urbains (LEESU UMR MA 102), Université Paris-Est, AgroParisTech, Faculté des Sciences et Technologie, 61 avenue du Général de Gaulle, FR 94000 Créteil, France.
[b]Eau de Paris, Direction Recherche et Développement Qualité de l'Eau (DRDQE), 33 avenue Jean Jaurès, FR 94200 Ivry-sur-Seine, France.

Key words: Nontuberculous mycobacteria, Pathogens, Lake, Real-time quantitative PCR, High-throughput sequencing

Accepted in Antonie van Leeuwenhoek, Journal of Microbiology. DOI: 10.1007/s10482-016-0665-x

I. Abstract

Although nontuberculous mycobacteria (NTM) are natural inhabitant of freshwater ecosystems, few studies have focused on their distribution in these habitats. Thus, the knowledge about the abundance as well as the composition of NTM remains limited and patchy in these environments. In this context, a prospective study was performed to identify favorable habitats for mycobacteria in two recreational lakes. Mycobacterial density and diversity were measured using real-time quantitative PCR and the MiSeq Illumina patform.

For both lakes, five compartments were investigated, i.e. water column, air-water interface, sediment, epilithon and epiphyton biofilms. Nontuberculous mycobacteria were detected in all compartments in large densities and displayed a remarkable diversity. NTM were dominated by fast-growing species. Lakes and compartments appeared to shape mycobacteria assemblage composition as well as their densities. In both lakes, some OTUs assigned to the species level, a few were identified as related to known opportunistic pathogens.

II. Introduction

The genus *Mycobacterium* contains more than 170 species, mostly described as nontuberculous mycobacteria (NTM) (Euzéby 1997). Although some species are recognized as opportunistic pathogens, many of these bacteria are saprophytic and therefore are natural inhabitants of terrestrial and aquatic environments (Collins *et al.* 1984; Falkinham III 2002; Hruska and Kaevska 2012). Yet, studies focusing on NTM in natural ecosystems remain rare in comparison with drinking water distribution systems (e.g., Covert *et al.* 1999; Falkinham III *et al*. 2001; Vaerewijck *et al.* 2005) or hospital water supply networks (e.g., du Moulin *et al.* 1998; Fox *et al.* 1992; Fujita *et al.* 2002). These two types of habitats have been particularly well studied (Kazda *et al.* 1999), owing to the increased susceptibility of immunodeficient persons to mycobacteriosis following contact with waters harboring mycobacteria.

The lack of studies exploring the environmental habitats of mycobacteria could partly be explained by the difficulties in isolating slow-growing bacteria from natural environments using classical culture methods (Falkinham III 2002). Cultivation approaches also provide a limited view of NTM density and diversity. However, improvements in extracting DNA from natural samples (Guo and Zhang 2013) as well as PCR-based methods, which have been developed over the last decades, should solve some of these biases. Nevertheless, studies describing the entire mycobacterial assemblage in ecosystems using molecular tools are still scarce. These few published papers mainly focused on soil habitats (e.g., Niva *et al.* 2006; Pontiroli *et al.* 2013). Moreover, the majority of studies investigating freshwater habitats, either focused on particular species (e.g., Stinear *et al.* 2000; Pickup *et al.* 2005, 2006; Gauthier *et al.* 2010) or on cultivable species (e.g., Viallier and Viallier 1973; Kirschner *et al.* 1992; Iivanainen *et al.* 1993; Bland *et al.* 2005). Only few studies considering the whole NTM assemblage have been performed in aquatic ecosystems (Niva *et al.* 2006; Parashar *et al.* 2009; Debruyn *et al.* 2009; Khera 2012). Consequently, it appears necessary to improve our knowledge in terms of abundance, diversity and potential niches of mycobacterial species in surface freshwater ecosystems, especially since NTM probably acquired their virulence traits

under selective pressures in their natural habitat, as have other environmental pathogens (Dyble *et al.* 2008; Abida *et al.* 2010).

Among freshwater ecosystems, lakes appear to be interesting models to investigate the distribution of NTM. Indeed lakes are common ecosystems in urban area, where they attract a high number of visitors due to the cultural and recreational services they provide. Moreover, according to the literature, lakes could provide relevant ecological niches for mycobacteria. In boreal lakes, Niva *et al.* (2006) observed that the mycobacterial assemblages constituted a large part of the *Actinobacteria* phyla, one of the dominant phylum in these ecosystems (Newton *et al.* 2011). Moreover, numerous NTM species (e.g., *M. avium*, *M. chelonae*, *M. xenopi*) are able to persist or develop in aquatic environments within biofilms (Hall-Stoodley and Lappin-Scott 1998; Dailloux *et al.* 2003; Williams *et al.* 2009) or at water interfaces (Alavi *et al.* 2006) due to their hydrophobic cell wall. Sediment seems also to constitute a suitable habitat for NTM (Pickup *et al.* 2005; Debruyn *et al.* 2009; Gauthier *et al.* 2010).

This prospective study investigated the spatial distribution of the mycobacterial assemblages in two shallow lakes with different trophic status (eutrophic and mesotrophic) in order to evaluate the density and diversity of NTM in different compartments. First, we sought to determine the main reservoirs of NTM among a set of different compartments (water column, air-water interface, sediment and biofilms). Second, we evaluated to what extent mycobacterial density could be impacted by spatial parameters. We hypothesized that NTM would be preferentially found at the air-water interface and in biofilms, and that the species composition would differ among the different compartments in both lakes. For that purpose, an approach combining a quantitative and a compositional analysis of NTM based on recent molecular tools (Pontiroli *et al.* 2013; Radomski *et al.* 2013) was used. Mycobacterial densities were quantified using real-time quantitative PCR and NTM diversity was assessed using MiSeq Illumina high-throughput sequencing.

III. Materials and methods

III.1 Sampling sites

Créteil Lake and Daumesnil Lake are two shallow lakes separated by less than 10 km. They are located in the Paris area (France), near the confluence of Seine and Marne Rivers. Créteil Lake (Supplementary data 1) is a 0.40 km^2 mesotrophic lake that was originally a sandpit. This lake is mainly supplied by an alluvial groundwater, however the water quality of Créteil Lake can be affected by the presence of a storm sewer outlet, which drains 1 km^2 of a residential area. Daumesnil Lake (Supplementary data 1) is a 0.12 km^2 eutrophic lake that was dug for recreational purposes in the 1860s in the Bois de Vincennes (Paris). Daumesnil Lake is supplied by water pumped in the Ourcq Channel.

III.2 Sample collection

Lakes were sampled in August and October 2012. For both lakes, five stations (C1 to C5 and D1 to D5 for Créteil and Daumesnil Lake respectively) were surveyed (Fig. 21). For each station, five compartments were collected: water column, surface microlayer (air-water interface), sediment, epilithic and epiphytic biofilms. For each station, water column samples were collected using a Niskin bottle. The surface microlayer was collected at three depths (depending on the depth of the water column) using a Niskin bottle (General Oceanics Inc., Miami, USA). The surface microlayer was collected using a metal screen as described in detail by Agogué *et al.* (2004). Pooled sediment samples were constituted with the aerobic top sediment layer (~ 1 cm) from three cores. Epilithic biofilms (epilithon) were removed from rock surface by scraping 20 cm^2 biofilm area with a sterile syringe-toothbrush sampler inspired from Steinman *et al.* (2006). Epiphytic biofilms (epiphyton) were collected by harvesting submerged leaves sheath of phragmites (*Phragmites australis*) and milfoil (*Myriophyllum spicatum)* within Créteil and Daumesnil lakes respectively. All compartments were stored in sterile containers and placed at 4°C until return to the laboratory in less than ten hours.

III.3 Sample processing

One liter of water sample (water column and surface microlayer) was centrifuged (7,500 × g, 4°C) and the resulting cell pellet was collected into a 2-mL sterile tube. In order

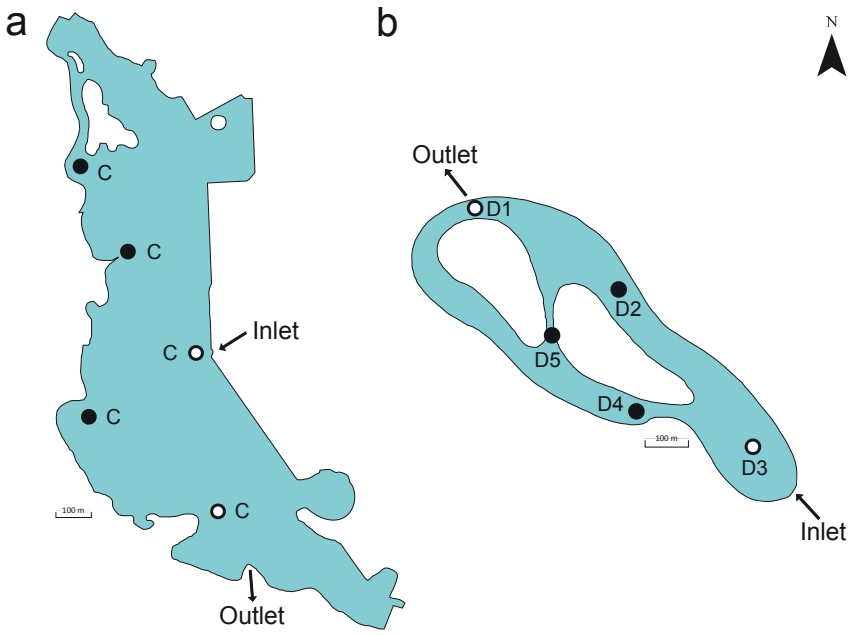


Figure 21 Location of the five stations in Créteil Lake (a) and Daumesnil Lake (b). Nontuberculous mycobacteria composition was analyzed in the stations symbolized by open circles.

to facilitate pellet resuspension, before centrifugations 1 mL of Tween 80 (final concentration, 0.01% vol/vol) was added to each sample. Sediments and biofilms were immediately frozen, then lyophilized. All samples were stored at – 20°C until DNA extraction.

For water samples, pellets were resuspended in 400 µL of sterile water. Extractions were performed using a MagNA Pure Compact system (Roche Applied Science, Bâle, Switzerland) and the MagNA Pure Compact Nucleic Acid Isolation Kit I using the Bacteria V3 protocol, according to the manufacturer's instructions. Prior to extraction, samples underwent freeze-thaw cycles that consisted in three cycles of 1 min in liquid nitrogen and 5 min at 90°C. Approximately 250 mg of sediment, 50 mg of dry epilithic biofilms and 40 mg of dry leaves were extracted using the FastDNA® SPIN Kit (Qbiogene, Carlbad, CA, USA) according to the manufacturer's instructions. Two modifications to this protocol were applied: cells were lysed in a FastPrep bead beater three times for 30 sec at 4.0 ms^{-1} and the SPIN filters were washed twice with 500 µL of SEWS-M buffer. DNA was recovered in 50 µL eluent buffer and DNA quality and quantity were analyzed at 230, 260, and 280 nm by spectrophotometry before storage at – 20°C.

III.4 Quantitative real-time PCR

To quantify the abundance of *Mycobacterium* in the five compartments, TaqMan® real-time PCR assays targeting the *atpE* gene were carried out as previously described by Radomski and colleagues (2013). The assay was performed using the forward primer F*atpE* 5'-CGGYGCCGGTATCGGYGA-3', the reverse primer R*atpE* 5'-CGAAGACGAACARSGCCAT-3' and the probe P*atpE* 5'-ACSGTGATGAAGAACGGBGTRAA-3' labeled with the fluorescent dyes 6-carboxyfluorescein (5' end) and Black Hole Quencher (3' end). Contaminations in PCR mix were checked using negative controls. Moreover, the presence of PCR inhibitors in DNA templates was verified using a non-competitive exogenous internal control that was included in the PCR buffer. This internal control was made of a partial sequence of the *human β-actin* gene cloned in pGEM-T-easy vector (Promega, Madison, WI, USA) (Wurtzer *et al.* 2014). Absence of significant PCR inhibition of the *atpE* assay was confirmed based on the average C_q values (± standard deviation) obtained from 100 repeated PCR reactions containing 1,000 copies of *human β-actin*.

The *atpE* copy numbers concentration was estimated from crude extracts using a *Mycobacterium chelonae* standard curve from 1.0×10^1 to 1.0×10^6 copies.μL^{-1}. The PCR method presented a high amplification efficiency (82%) and the standard curve had a good linearity ($R^2 = 0.998$). Results were expressed as *aptE* gene copies per liter for water samples, per gram (dry weight) for sediment samples and per square centimeter for biofilms. Relationship between biofilm area and mass are presented in the Supplementary data 2.

III.5 Illumina sequencing of the V2-V3 region of the 16S rRNA gene

For the study of mycobacterial diversity in each compartment, twenty samples from the summer campaign were analyzed. The samples were selected from two stations corresponding to the inlet and the outlet of both lakes: C2 and C3 for Créteil Lake; D1 and D3 for Daumesnil Lake (Fig. 21).

PCR libraries were created by amplifying a fragment of the 16S rRNA gene (*rrs*) including the V2-V3 hypervariable regions using mycobacterial primers JSY16SF 5'-TGGGAAACTGGGAAACTGGGTCTAATA-3' and JSY16SR 5'-

CCCGCACGCCCAAGTTAAGCTGTGAG-3' (Pontiroli *et al.* 2013). PCR and sequencing were performed by Research and Testing Laboratory (Lubbock (TX), USA), using the Illumina MiSeq platform (Illumina, Inc., San Diego (CA), USA). All DNA extracts were adjusted to 120 ng.μL^{-1} prior to PCR reactions.

Quality control and sample processing were performed using QIIME v. 1.8.0-20140103 (Caporaso *et al.* 2010). Sequences of poor quality were removed using the default parameters (quality score < 25, length < 200 nt and the presence of at least six ambiguous bases or six successive homopolymers). Chimeric sequences were then identified with USEARCH 6.0 using the UCHIME algorithm in *de novo* mode (Edgar *et al.* 2011) implemented in the online service FunGene pipeline and ignored for further analysis. Filtered sequences were clustered into classical 97% similarity-based operational taxonomic units (OTU). A clustering cutoff at 99% rather than 97% would lead to an six-fold increase of the OTU number (data not shown). Each OTU represented by a single sequence was removed from the analysis. Rarefaction curves are presented in the Supplementary data 3. OTUs were assigned to the genus level with the SILVA 111 database (Quast *et al.* 2013) using UCLUST (Edgar 2010). Assignation to the species level was performed using blastn 2.2.31 search (Morgulis *et al.* 2008) provided by GenBank (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Query coverage was set to at least 97%. A 97% threshold was chosen based on Pontiroli and colleagues (2013) who did not notice a change of the most preponderant species identities or the overall numbers of sequences affiliated to these species using the more stringent cutoff of 99%. Sequence data have been deposited in the NCBI Sequence Read Archive (SRA) and can be accessed through accession number SRP061716.

III.6 Data analysis

All statistical analyses and indices computing were conducted using the statistical environment R version 3.1.1 (R Development Core Team 2014) and the 'vegan' package (Oksanen *et al.* 2013). The default statistical significance was based on a $P \leq 0.05$ level.

Quantitative data analysis. Paired Wilcoxon or t-tests were used to compare the data between August and October for each compartment and each lake. An intra-lake analysis was conducted to identify significant differences between the water

compartments and between the biofilms. For that purpose, linear models were performed with stations as covariables. An inter-lake analysis was also conducted to characterize differences between the two lakes for all of the five compartments using linear models. All statistical analyses were performed with log-transformed data.

Mycobacterial composition analysis. First, OTU richness, diversity and evenness indices were calculated. Theoretical richness was estimated according to the non-parametric model of Chao-1 (Chao 1987). Alpha diversity was estimated using the exponential form of Shannon entropy. Evenness was calculated using the Simpson's evenness index (Smith and Wilson 1996). Results of evenness ranged from 0 (when one OTU is predominant) to 1 (when all OTUs are equally abundant). Second, variations in the NTM community assembly were explored. For each lake, a Mann-Whitney test was conducted with Bray-Curtis dissimilarities to determine if the variability in the community structure within each compartment was significantly different from the variability between compartments. The dissimilarity of NTM assemblages among compartments (beta diversity) was evaluated using the Sørenson dissimilarity index. The beta diversity was then partitioned following the framework proposed by Baselga using the 'betapart' package (Baselga 2010; Baselga *et al.* 2013) to quantify the fraction of dissimilarity explained by OTUs replacement (based on the Simpson's dissimilarity index) and from pure random richness variation of the OTUs (Azeria *et al.* 2011). To evaluate the influence of lakes and compartments in the composition of the NTM communities, a redundancy analysis was performed on Hellinger-transformed data (Legendre and Gallagher 2001). Finally, samples were associated depending of the NTM community composition similarity using the unweighted-pair group method with UPGMA clustering (bootstrap = 999).

IV. Results

IV.1 Quantification of nontuberculous mycobacteria

The *atpE* gene was successfully amplified from all the samples with a copy numbers of *atpE* gene ranging from 5.0×10^2 to 1.1×10^5 L^{-1} in water samples, from 2.2×10^3 to 5.3×10^5 mg^{-1} in dry sediment and from 1.0×10^3 to 6.9×10^5 cm^{-2} in biofilms (Fig. 22). No significant difference in the *atpE* gene copy number was found between August and

October for each compartment and each lake, except for the epiphytic biofilm from Créteil. For this compartment, the *atpE* gene copy number was significantly higher in October than in August (paired t-test, P = 0.006), with values of 4.4 ± 2.4 × 10^4 copies.cm^{-2} in summer against 1.3 ± 0.9 × 10^5 copies.cm^{-2} in autumn. Despite this unique difference, the remaining analysis was performed by combining data from the two sampling dates for all the compartments.

Intra-lake analysis indicated that for both lakes, the *atpE* gene copy numbers were significantly higher in the surface microlayer compared to the water column (linear models, $P = 0.014$ for Créteil Lake and $P = 0.007$ for Daumesnil Lake) (Fig. 22a). The copy numbers of *atpE* gene in epilithic biofilms were significantly higher than the copy number in epiphytic biofilms (linear model, $P < 0.001$ within Créteil and Daumesnil Lake) (Fig. 22c).

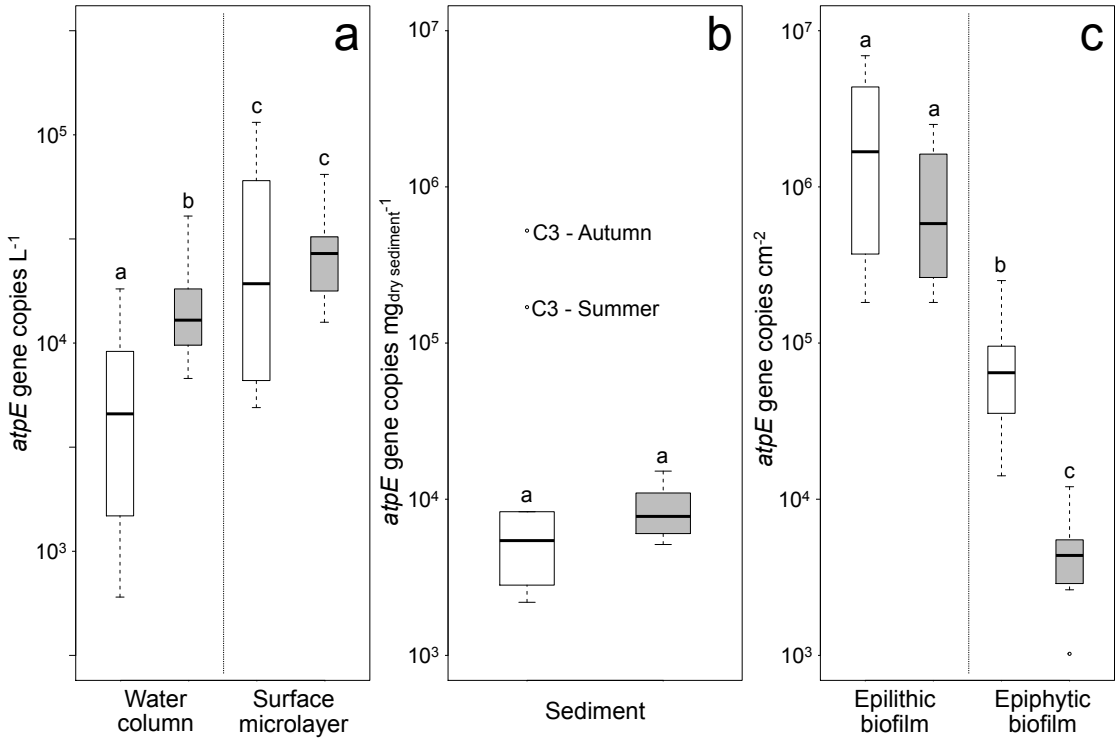


Figure 22 Comparison of the *atpE* gene copy numbers between Créteil Lake (white boxplot) and Daumesnil Lake (grey boxplot) for water (a), sediment (b) and biofilms compartments (c) (n = 10). Different letters indicate a significant difference between lakes or compartments (linear models). Box- and-whisker plots represent the median (inner box line), 25th and 75th percentiles (lower and upper outer box lines), whiskers cover 10th and 90th percentiles and outliers are represented by open circles.

For each compartment, an inter-lake comparison revealed that the *atpE* gene copy numbers in the water column was significantly lower in Créteil compared to Daumesnil Lake (linear model, $P = 0.010$), whereas for the epiphytic biofilm the copy number was significantly higher in Créteil Lake (linear model, $P < 0.001$). No significant difference was observed between the two lakes for the NTM densities in the surface microlayer, the epilithic biofilm and the sediment (linear models, $P = 0.555$, $P = 0.191$ and $P = 0.748$ respectively). Interestingly, the Figure 22b showed high NTM densities for the two sediment samples collected in front of the storm sewer outlet of the Créteil Lake (Fig. 21). After removal of these two outliers from the analysis, the *atpE* gene copy numbers were significantly higher in Daumesnil than in Créteil Lake (linear model, $P < 0.001$).

IV.2 Richness of nontuberculous mycobacteria

For each lake, ten samples collected in summer at two stations were sequenced, which produced a total of 503,164 sequences with an average length of 418 bp. Trimming and chimera check removed 39,939 sequences (Table 9). For Créteil Lake, two samples from different compartments were removed from the analysis due to a problem in amplification and sequencing (Table 9), thus the rest of the analysis was performed on 18 samples. After removal of the singleton sequences, 317,923 sequences were identified as belonging to the *Mycobacterium* genus (Table 9). These sequences were clustered into 658 OTUs. The observed richness covered on average 73% of the estimated richness (Chao-1 index). A median of 125 OTUs was encountered in the different samples. Only six percent of these OTUs (37/658), presented in Figure 23, had a relative abundance higher than 1% in at least one of the 18 samples. This result suggests that the mycobacterial assemblages were dominated by few OTUs and mainly composed of rare OTUs, as underscored by the low values of the evenness index (Table 9).

Although Créteil and Daumesnil Lakes have different trophic status, no significant difference was encountered between these lakes regarding the observed and the estimated richness (t-tests, $P = 0.966$ and $P = 0.435$, respectively). However, the exponential form of Shannon entropy was significantly higher (t-test, $P = 0.002$) within Créteil Lake (11.6 ± 9.3 OTUs) compared to Daumesnil samples (3.3 ± 2.3 OTUs). Epilithic biofilms from Créteil possessed the highest diversity (Table 9).

Table 9 Comparison of sequenced libraries, including the number of reads pre- and post-trimming, the observed richness of OTUs, the estimated richness, the diversity index and the evenness index for each sample.

			Belonging to the *Mycobacterium* genus				
Sample name [a]	No. of raw sequences	No. of filtered sequences	No. of sequences[b]	No. of OTUs[b]	Richness estimator (Chao-1)	Diversity index (Exp. form of Shannon)	Evenness index (Simpson evenness)
C2 W	3548	3128	2708	87	108.7	10.69	0.07
C3 W	7961	6862	6616	105	181.6	9.11	0.06
C2 SML[c]	135	99	19	3	3.0	2.76	0.85
C3 SML	6963	5504	5466	64	77.6	7.21	0.07
C2 S	14052	13182	13092	85	114.3	3.84	0.02
C3 S[c]	159736	136815	116827	402	435.5	19.47	0.02
C2 El	6992	5681	5630	167	226.6	26.92	0.09
C3 El	20789	17377	16699	268	332.5	26.60	0.05
C2 Ep	14817	14097	13997	139	160.1	4.20	0.01
C3 Ep	10589	10010	9816	101	117.0	5.17	0.02
D1 W	3079	2969	2556	51	61.9	4.23	0.05
D3 W	16007	15354	12842	110	129.4	5.15	0.03
D1 SML	2564	2490	2056	59	92.0	7.57	0.07
D3 SML	3764	3684	2744	78	109.2	5.67	0.03
D1 S	126217	123967	123587	147	184.1	1.10	0.01
D3 S	16974	16682	16578	61	204.5	1.09	0.02
D1 El	22332	20907	20422	222	315.8	3.04	0.01
D3 El	40176	38979	38359	158	246.4	1.89	0.01
D1 Ep	14808	13917	13257	79	102.4	1.92	0.02
D3 Ep	11661	11521	11498	14	23.3	1.23	0.08

Abbreviations: OTUs, operational taxonomic units; C, Créteil Lake; D, Daumesnil Lake; SML, surface microlayer; W, water column; S, sediment; Ep, epiphytic biofilm; El, epilithic biofilm.
[a]Numbers correspond to the stations numbers.
[b]Without singleton sequences.
[c]Samples removed from the analysis due to a problem of amplification (C2 SML) and sequencing (C3 S).

IV.3 Composition of the mycobacterial assemblages

The blastn algorithm allowed a putative assignment of 364 out of 658 total OTUs to *Mycobacterium* species, each belonging to the nontuberculous mycobacteria (see Supplementary data 4). The remaining OTUs were not assigned or were assigned to unclassified mycobacteria. Among the 364 identified OTUs, 94 were assigned to several mycobacterial species without any possible discrimination. Among the 270 remaining OTUs, 141 were assigned to a single species, 71 to two species and 58 OTUs to three species. When only considering these 270 OTUs, a large part (68 OTUs) was affiliated to *M. moriokaense*, 35 to the undifferentiated species *M. neglectum/tusciae*, 34 to the

undifferentiated species *M. bacteremicum/frederiksbergense/sacrum*, 12 to *M. rhodesiae* and 9 to *M. asiaticum*.

The majority of all the sequences (87%) were assigned to fast-growing mycobacteria. These fast-growing species were mostly identified as *M. moriokaense/barrassiae* (41%), *M. frederiksbergense* (11%) or *M. austraoafricanum* (5%). Some OTUs related to fast-growing mycobacterial species dominated some of the compartments. Such was the case for the otu1026 that represented up to 99% of all the sequences retrieved from the Daumesnil sediments (Fig. 23).

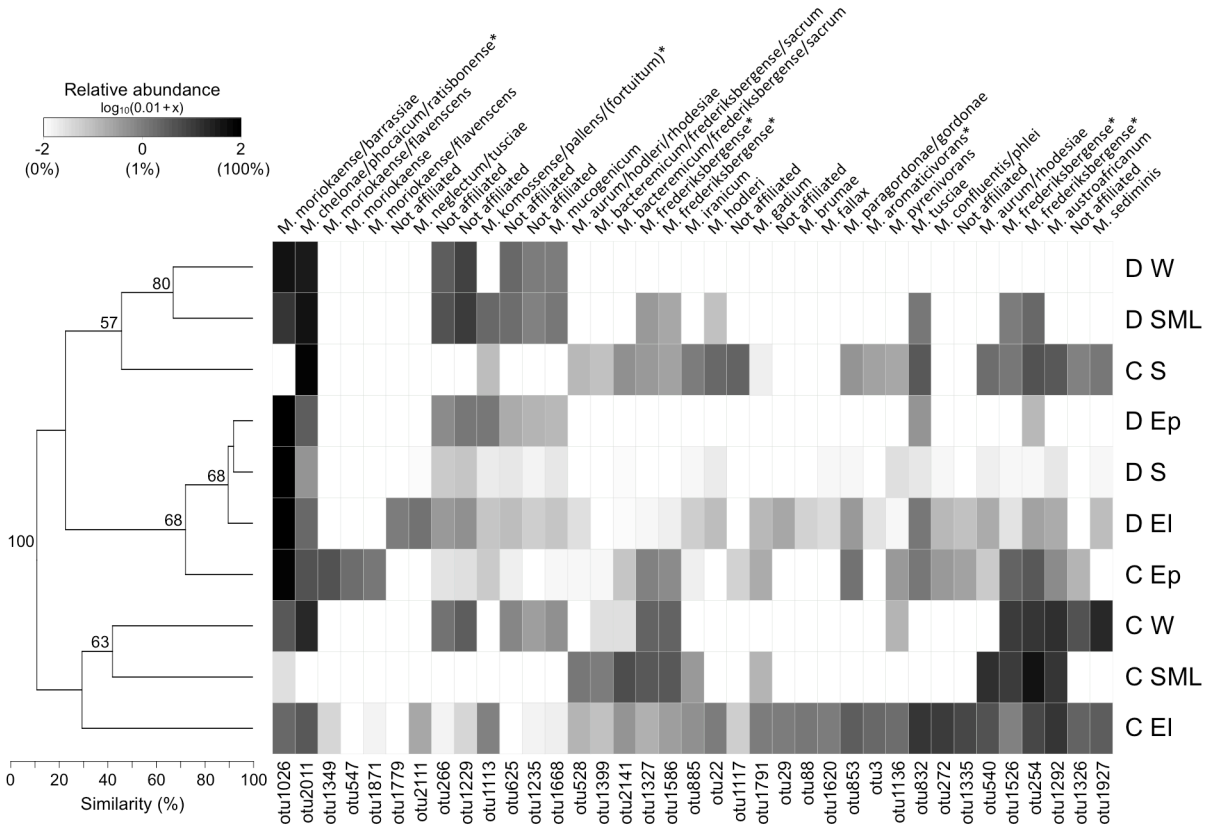


Figure 23 Heatmap dendrogram illustrating similarities of nontuberculous mycobacteria assemblage composition for the five compartments (W, water; SML, surface microlayer; S, sediment; El, epilithic and Ep, epiphytic biofilms) in Créteil (C) and Daumesnil (D) Lakes. The OTU ID as well as species to which were assigned are presented in abscissa (only OTUs with relative abundances > 1% in at least one compartment were represented). Asterisk indicates an assignation for a single OTU to more than three species.

IV.4 Potential opportunistic pathogens

When considering the 270 OTUs (Supplementary data 4), 93 were affiliated with at least one potential pathogen species listed by Katoch (2004) and Tortoli (2009). However, few of these 93 OTUs (16 OTUs) were exclusively assigned to opportunistic pathogen species. Among these 16 OTUs a large fraction (9 OTUs) were identified as *M. asiaticum*. The remaining OTUs were affiliated to *M. fortuitum*, *M. immunogenum*, *M. kansasii, M. lentiflavum, M. smegmatis, M. thermoresistibile* and *M. vaccae*. All the OTUs were assigned with an identity close to 97.0% except for *M. vaccae* and *M. thermoresistibile* with identities of 98.1 and 99.4% respectively. *M. asiaticum* was mainly encountered in the epilithic biofilm for both lakes and also in the sediment of Créteil Lake. These 16 OTUs were detected in both lakes, in almost all compartments except in waters. However, for all these OTUs characterized as potential opportunistic pathogens, the prevalence in each sample never exceeded 0.34%.

IV.5 Comparison of the mycobacterial assemblages

For both lakes, the intra-compartment variations of the NTM assemblages, evaluated using the Bray-Curtis distances, were significantly lower than the inter-compartment variations (Mann-Whitney, $P = 0.006$). Overall, mycobacterial community structure variations between compartments, i.e. beta diversity, was about 72% of dissimilarity. The beta diversity partitioning showed that the variation in the mycobacterial assembly was mainly due to lake or compartment particularities (83%) and only 17% was explained by pure random fluctuation of the OTUs richness. Besides, a redundancy analysis performed on the 18 samples with a goodness-of-fit of 0.73 (adjusted R-squared) revealed that the NTM assemblages were significantly shaped by the lake ($P < 0.001$) and the type of compartment (Fig. 23).

Finally, we observed that the majority of the 37 dominant OTUs were present in both lakes and in almost all the compartments (Fig. 23). Inversely, two OTUs were only detected in a single sample: otu547 in epilithon from Créteil Lake and otu1779 in epiphyton from Daumesnil Lake. Otu29 and otu88 were identified in both lakes but exclusively in epilithic biofilms. Only one OTU (otu1399) was present in all the compartments but only in Créteil Lake.

V. Discussion

This study aimed to prospect the distribution of nontuberculous mycobacteria and identify their natural reservoirs in two recreational lakes. For this purpose, NTM assemblages were conjointly quantified and characterized within five compartments (water column, surface microlayer, sediment, epilithic and epiphytic biofilms) using real-time quantitative PCR and high-throughput sequencing. To date, this sequencing approach has already been applied in soils (Pontiroli *et al.* 2013) or in drinking water systems (van der Wielen *et al.* 2013), but as far as we are aware it is the first time that these analyzes have been performed in recreational lakes and in different compartments within the same lake. The results reveal that NTM are ubiquitous and diverse.

V.1 Ubiquity of nontuberculous mycobacteria

In natural habitats, NTM have been more frequently identified in extreme environmental conditions, such as in alkaline or acid brown-water and in highly polluted soils and sediments or in environments with low oxygen concentration (Brooks *et al.* 1984; Kirschner *et al.* 1992; Iivanainen *et al.* 1993; Bland *et al.* 2005; Leys *et al.* 2005). However, in this study, NTM were detected at high levels in both lakes, in all the collected samples. This result suggests that even in non-extreme environments, mycobacteria are able to persist and/or grow in the five investigated compartments at relatively high diversity and abundance. However, it is difficult to compare the densities measured in this study with the literature, since no other study has used the same extraction and amplification methodologies. Yet, some studies using real-time quantitative PCR targeting the entire NTM assemblages reported similar values in a coastal lagoon (Jacobs *et al.* 2009), in freshwater samples in Ethiopia (Khera 2012) or in sediments from Lake Erie (Debruyn *et al.* 2009).

V.2 Nontuberculous mycobacteria dominated by fast-growing species

High-throughput sequencing of the hypervariable regions V2-V3 from the 16S rRNA gene allowed an in depth-analysis of the NTM composition. With 41% of the OTUs assigned to one, two or three *Mycobacterium* species, the percent of unclassified mycobacteria was comparable to that reported by Pontiroli *et al.* (2013) who used the

same primers to characterize soil samples. The large proportion of unassigned sequences suggests, as already known, that the *rrs* gene does not have enough information to properly identify NTM to the species level (Kim *et al.* 2005) and/or that many NTM species are still unknown or poorly described. Indeed, partly due to the improvement of the isolation and identification techniques, the number of described species belonging to the *Mycobacterium* genus is steadily growing, with 50 identified species in 1997 and up to 170 presently (Euzéby 1997).

For all the analyzed samples, NTM assemblages were mainly composed of rare OTUs and were dominated by fast-growing species. Similar results were encountered in studies focusing on water and soil samples (Khera 2012; Pontiroli *et al.* 2013), which also identified NTM by targeting the *rrs* gene with the same primer set. This prevalence could suggest that fast-growing species possess more plasticity to environmental changes and/or are more competitive compared to slow-growing species. It could also be a PCR artifact due to a difference in *rrs* gene copy numbers between fast and slow-growers (Bercovier *et al.* 1986) and/or it may also be due to preferential binding of the primer set to the *rrs* gene in fast-growing mycobacteria. Besides, two previous culture-based studies that investigated freshwater ecosystems (Viallier and Viallier 1973; Bland *et al.* 2005) did not find any noticeable dominance of the fast-growing species.

Among the fast-growing OTUs identified in the present study, numerous species have already been identified for their potential role in polycyclic aromatic hydrocarbon degradation. Indeed, *M. austroafricanum* or *M. frederiksbergense* have been characterized for their abilities to degrade anthracene or pyrene (Willumsen *et al.* 2001; Wick *et al.* 2003; Leys *et al.* 2005; Uyttebroek *et al.* 2006). The dominant OTU, identified as *M. moriokaense* and *M. barrassiae* (up to 99% of the identified OTUs for the sediment in Daumesnil Lake), have been previously observed in Japanese and Ethiopian soils (Tsukamura *et al.* 1986; Pontiroli *et al.* 2013), in German hospital tap water (Hussein *et al.* 2009) or in a French water distribution system (Dubrou *et al.* 2013), but this is the first time they were reported in lakes.

V.3 Mycobacterial densities and diversity differed among compartments

The abundance and diversity of mycobacterial assemblage was significantly shaped by the compartments, potentially due to differences in physicochemical properties and type of substrates.

In water compartments, NTM densities were significantly higher in the surface microlayer compared to the water column. This result is consistent with Parker and colleagues (1983) who experimentally found a concentration of *M. intracellulare* up to 15,000 higher in aerosols (formed from the surface microlayer) compared with the bulk water. Even if the surface microlayer is exposed to high levels of UV, the enrichment of mycobacteria in the lipid microlayer could be due to the hydrophobicity of the mycobacteria, which can be concentrated at the surface by preferential binding to air bubbles rising in the water column (Blanchard 1964). Moreover the high NTM concentration in this compartment could be due to high concentrations of hydrophobic compounds such as hydrocarbons (Cincinelli *et al.* 2001; Wurl and Obbard 2004; Manodori *et al.* 2006) that could constitute a substrate for mycobacteria.

In the sediment, large differences among assemblages were observed between the different samples. This result could be due to differences of the quantity and/or the quality of the nutrients available for the bacteria. The high densities of NTM in the sediment near the storm sewer outlet in the Créteil Lake were in agreement with Pickup and colleagues (2006), who found that densities of *M. avium* subsp. *paratuberculosis* in river were well predicted by rainfall events. This result suggests that significant densities of NTM are transported by runoff effluents and settle in the lake sediment close to the storm sewer outlet.

Mycobacteria are well known to colonize biofilms, as they tend to easily stick to surfaces and to clump together. Indeed, high mycobacterial densities and diversity were found in the epilithic and epiphytic biofilms. Moreover, the density and composition of NTM differed between the epilithic and epiphytic biofilms, which may be related to the difference of biofilm ages. Epilithic biofilms were established since several decades, while phragmites and milfoils regrew every year. It may also be due to the colonization of specific algal species in the epilithon or epiphyton (Danilov and Ekelund 2000) that could

interact differently with the bacteria. The mycobacterial densities were significantly higher in the phragmite biofilms compared to the milfoil biofilms, although the mycobacterial assemblages showed similar OTU compositions. This result indicates that the plant species did not exhibit a strong influence on the NTM assemblage composition, contrary to what have already been found for the total bacterial community (Hempel *et al.* 2008), potentially due to the microaerophile nature of NTM. However, plant biofilms had lower *Mycobacterium* density compared with epilithic biofilms. One potential explanation could rely in the capacity of *Myriophyllum spicatum* to produce polyphenols that can affect the bacterial growth (Walenciak *et al.* 2002; Hempel *et al.* 2009).

V.4 Potential opportunistic pathogens

Several OTUs were identified as potential opportunistic pathogens (*M. asiaticum, M. fortuitum, M. immunogenum, M. kansasii, M. lentiflavum, M. smegmatis, M. thermoresistibile and M. vaccae*), known to be responsible for a large range of diseases including pulmonary, cutaneous and soft tissue infections (Katoch 2004; Griffith *et al.* 2007; Tortoli 2009). All these species have previously been isolated in terrestrial or aquatic habitats (Viallier and Viallier 1973; Engel and Berwald 1980; Covert *et al.* 1999; Narang *et al.* 2009; Pontiroli *et al.* 2013; Klanicova *et al.* 2013). Although most of these pathogenic species were detected in both lakes in almost all compartments except in waters, some species were preferentially present in particular compartments. For instance, *M. asiaticum* was only identified in epilithon and sediment.

Due to methodological aspects, it is possible that some species could be either underestimated or overestimated. On one hand, it is likely that the number of potential pathogens was underestimated owing to the highly conservative sequence of the *rrs* gene. On the other hand, it is possible that this number was overestimated due to the poor discrimination of the 16S rDNA sequences at the species levels using the 97% similarity threshold. Drancourt and colleagues (2000) recommended using 99% as identity cutoff to assign OTU sequences to bacterial species. When we applied this recommendation, only one OTU could be identified as an opportunistic pathogen species: *M. thermoresistibile*.

Sequencing genes encoding for *hsp65* and *rpoB* are necessary for a proper identification of mycobacterial species (Adékambi and Drancourt 2004; Kim *et al.* 2005). Although *rrs* sequencing may not be a suitable tool to reliably identify mycobacterial species in environmental samples, it could be an initial screen for dominant NTM groups that could later be studied in detail using quantitative real-time PCR or DNA microarrays.

To conclude, our results showed that a molecular approach combining quantification and characterization of the bacterial composition is suitable to screen for NTM in complex aquatic ecosystems. High-throughput sequencing of the *rrs* gene offered sufficient depth to investigate variations in the mycobacterial assemblages between lakes and compartments. Although the 16S rRNA gene fragment had a too low resolution to assign all the OTUs at the species level, this approach represents a useful tool to prospect for potential pathogenic species that should be studied in detail with appropriate methods.

This study emphases the ubiquity of NTM in natural aquatic environments and the high diversity of mycobacterial assemblages. Yet, although they were detected in all five compartments, large variations shaped by compartments and lakes were observed in the mycobacterial densities and diversity. For water and epilithic compartments, the mycobacterial composition was clearly distinct between the two lakes. Moreover, the density of NTM in the water column and sediment was significantly greater in the eutrophic Daumesnil Lake. These results could be due to the difference of quantity and/or quality of nutrients available for bacteria (Supplementary data 1). The difference in epilithic mycobacteria between the two lakes could be also due to differences in the substratum nature (rocks in Créteil and concrete in Daumesnil Lake). However, it is difficult to conclude about the environmental parameters that structure the NTM community since these two lakes were highly contrasted. To establish a clear relationship between limnological properties and mycobacteria density and diversity, it would be necessary to perform a study on a larger number of lakes in order to perform a robust statistical analysis. Despite all these discrepancies, NTM assemblages were mainly

composed of fast-growing species, regardless of compartment. Finally, it would be interesting to take into account biotic reservoirs, such as amoebae (Delafont *et al.* 2014), fishes (Mrlik *et al.* 2012) or insects (Marsollier *et al.* 2002), which could represent relevant vectors of mycobacteria and favor the survival of these bacteria in environment.

VI. Acknowledgments

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] and by Eau de Paris. The authors are grateful to the Créteil leisure base and the Green Space and Environment Department of Paris for their authorization to collect samples in both lakes and for lending their boats. We also thank Fabien Joux for showing us how to use and manufacture the metal screen, Sylvain Huon, Cécile Bernard and Brigitte Vinçon-Leite for respectively lending us the corer, the fluoroprobe and the submersive data logger. Finally, we warmly thank Gérard Lacroix for sediment C/N ratio analysis.

VII. References

Adékambi T and Drancourt M (2004) Dissection of phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, *hsp65*, *sodA*, *recA* and *rpoB* gene sequencing. *Int J Syst Evol Microbiol* 54:2095–2105.

Adiba S, Nizak C, van Baalen M, Denamur E and Depaulis F (2010) From grazing resistance to pathogenesis: the coincidental evolution of virulence factors. *PLoS One* 5:e11882.

Agogué H, Casamayor EO, Joux F, Obernosterer I, Dupuy C, Lantoine F, Catala P, Weinbauer MG, Reinthaler T, Herndl GJ and Lebaron P (2004) Comparison of samplers for the biological characterization of the sea surface microlayer. *Limnol Oceanogr Methods 2*:213–225.

Alavi MR, Shukla HD, Whitaker B, Arnold J and Shahamat M (2006) Attachment and biofilm formation of *Mycobacterium marinum* on a hydrophobic surface at the air interface. *World J Microbiol Biotechnol* 23:93–101.

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.

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.

Bercovier H, Kafri O and Sela S (1986) Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. *Biochem Biophys Res Commun* 136:1136–1141.

Blancard DC (1964) Sea-to-air transport of surface active material. *Science* 146:396–397.

Bland CS, Ireland JM, Lozano E, Alvarez ME and Primm TP (2005) Mycobacterial ecology of the Rio Grande. *Appl Environ Microbiol* 71:5719–5727.

Brooks RW, George KL, Parker BC, Falkinham III JO and Gruff H (1984) Recovery and survival of nontuberculous mycobacteria under various growth and decontamination conditions. *Can J Microbiol* 30:1112–1117.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, Mcdonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J and Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336.

Chao A (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43:783–791.

Cincinelli A, Stortini AM, Perugini M, Checchini L and Lepri L (2001) Organic pollutants in sea-surface microlayer and aerosol in the coastal environment of Leghorn - (Tyrrhenian Sea). *Mar Chem* 76:77–98.

Collins CH, Grange JM and Yates MD (1984) Mycobacteria in water. *J Appl Bacteriol* 57:193–211.

Covert TC, Rodgers MR, Reyes AL and Stelma GN (1999) Occurrence of nontuberculous mycobacteria in environmental samples. *Appl Environ Microbiol* 65:2492–2496.

Dailloux M, Albert M, Laurain C, Andolfatto S, Lozniewski A, Hartemann P and Mathieu L (2003) *Mycobacterium xenopi* and drinking water biofilms. *Appl Environ Microbiol* 69:6946–6948.

Danilov RA and Ekelund NGA (2000) The use of epiphyton and epilithon data as a base for calculating ecological indices in monitoring of eutrophication in lakes in central Sweden. *Sci Total Environ* 248:63–70.

Debruyn JM, Mead TJ, Wilhelm SW and Sayler GS (2009) PAH biodegradative genotypes in Lake Erie sediments: evidence for broad geographical distribution of pyrene-

degrading mycobacteria. *Environ Sci Technol* 43:3467–3473.

Delafont V, Mougari F, Cambau E, Joyeux M, Bouchon D, Yann H and Moulin L (2014) First evidence of Amoebae−Mycobacteria association in drinking water network. *Environ Sci Technol* 48:11872–11882.

Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP and Raoult D (2000) 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *J Clin Microbiol* 38:3623–3630.

Dubrou S, Konjek J, Macheras E, Welté B, Guidicelli L, Chignon E, Joyeux M, Gaillard JL, Heym B, Tully T and Sapriel G (2013) Diversity, community composition, and dynamics of nonpigmented and late-pigmenting rapidly growing mycobacteria in an urban tap water production and distribution system. *Appl Environ Microbiol* 79:1–12.

du Moulin G, Stottmeier K, Pelletier P, Tsang A and Hedley-Whyte J (1988) Concentration of *Mycobacterium avium* by hospital hot water systems. *JAMA* 260:1599–1601.

Dyble J, Bienfang P, Dusek E, Hitchcock G, Holland F, Laws E, Lerczak J, McGillicuddy Jr DJ, Minnett P, Moore SK, O'Kelly C, Solo-Gabriele H and Wang JD (2008) Environmental controls, oceanography and population dynamics of pathogens and harmful algal blooms: connecting sources to human exposure. *Environ Heal* 7:S5.

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

Edgar RC, Haas BJ, Clemente JC, Quince C and Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200.

Engel H and Berwald L (1980) The occurence of *Mycobacterium kansasii* in tapwater. *Tubercle* 61:21–26.

Euzéby JP (1997). List of bacterial names standing in nomenclature: a folder available on the internet. *Int J Syst Bacteriol* 47:590–592. (Consulted on August 2015, URL: http://www.bacterio.net/).

Falkinham III JO (2002) Nontuberculous mycobacteria in the environment. *Clin Chest Med* 23:529–551.

Falkinham III JO, Norton CD and LeChevalier MW (2001) Factors influencing numbers of *Mycobacterium avium*, *Mycobacterium intracellulare*, and other mycobacteria in drinking water distribution systems. *Appl Environ Microbiol* 67:1225–1231.

Fox GE, Wisotzkey JD and Jurtshuk P (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* 42:166–170.

Fujita J, Nanki N, Negayama K, Tsutsui S, Taminato T and Ishida T (2002) Nosocomial contamination by *Mycobacterium gordonae* in hospital water supply and super-

oxidized water. *J Hosp Infect* 51:65–68.

Gauthier DT, Reece KS, Xiao J, Rhodes MW, Kator HI, Latour RJ, Bonzek CF, Hoenig JM and Vogelbein WK (2010) Quantitative PCR assay for *Mycobacterium pseudoshottsii* and *Mycobacterium shottsii* and application to environmental samples and fishes from the Chesapeake Bay. *Appl* 76:6171–6179.

Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, Holland SM, Horsburgh R, Huitt G, Iademarco MF, Iseman M, Olivier K, Ruoss S, von Reyn CF, Wallace Jr RJ and Winthrop K (2007) An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am J Respir Crit Care Med* 175:367–416.

Guo F and Zhang T (2013) Biases during DNA extraction of activated sludge samples revealed by high throughput sequencing. *Appl Microbiol Biotechnol* 97:4607–4616.

Hall-Stoodley L and Lappin-Scott H (1998) Biofilm formation by the rapidly growing mycobacterial species *Mycobacterium fortuitum*. *FEMS Microbiol Lett* 168:77–84.

Hempel M, Blume M, Blindow I and Gross EM (2008) Epiphytic bacterial community composition on two common submerged macrophytes in brackish water and freshwater. *BMC Microbiol* 8.

Hempel M, Grossart HP and Gross EM (2009) Community composition of bacterial biofilms on two submerged macrophytes and an artificial substrate in a pre-alpine Lake. *Aquat Microb Ecol* 58:79–94.

Hruska K and Kaevska M (2012) Mycobacteria in water, soil, plants and air: a review. *Vet Med (Praha)* 57:623–679.

Hussein Z, Landt O, Wirths B and Wellinghausen N (2009) Detection of non-tuberculous mycobacteria in hospital water by culture and molecular methods. *Int J Med Microbiol* 299:281–290.

Iivanainen EK, Martikainen PJ, Väänänen PK and Katila ML (1993) Environmental factors affecting the occurrence of mycobacteria in brook waters. *Appl Environ Microbiol* 59:398–404.

Jacobs J, Rhodes M, Sturgis B and Wood B (2009) Influence of environmental gradients on the abundance and distribution of *Mycobacterium* spp. in a coastal lagoon estuary. *Appl Environ Microbiol* 75:7378–7384.

Katoch VM (2004) Infections due to non-tuberculous mycobacteria (NTM). *Indian J Med Res* 120:290–304.

Kazda J, Pavlik I, Falkinham III JO and Hruska K (1999) *The ecology of mycobacteria: impact on animal's and human's health*. Springer Science and Business Media.

Khera T (2012) The diversity and distribution of *Mycobacterium* species in varying

ecological and climatic environments.

Kim H, Kim SH, Shim TS, Kim M, Bai GH, Park YG, Lee SH, Chae GT, Cha CY, Kook YH and Kim BJ (2005) Differentiation of *Mycobacterium* species by analysis of the heat-shock protein 65 gene (hsp65). *Int J Syst Evol Microbiol* 55:1649–1656.

Kirschner RA, Parker BC and Falkinham III JO (1992) Epidemiology of infection by nontuberculous mycobacteria. *Mycobacterium avium*, *Mycobacterium intracellulare*, and *Mycobacterium scrofulaceum* in acid, brown-water swamps of southeastern United States and their association with environmental variables. *Am Rev Respir Dis* 145:271–275.

Klanicova B, Seda J, Slana I, Slany M and Pavlik I (2013) The tracing of mycobacteria in drinking water supply systems by culture, conventional, and real-time PCRs. *Curr Microbiol* 67:725–731.

Legendre P and Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271–280.

Leys NM, Ryngaert A, Bastiaens L, Wattiau P, Top EM, Verstraete W and Springael D (2005) Occurrence and community composition of fast-growing *Mycobacterium* in soils contaminated with polycyclic aromatic hydrocarbons. *FEMS Microbiol Ecol* 51:375–388.

Manodori L, Gambaro A, Piazza R, Ferrari S, Stortini AM, Moret I and Capodaglio G (2006) PCBs and PAHs in sea-surface microlayer and sub-surface water samples of the Venice Lagoon (Italy). *Mar Pollut Bull* 52:184–192.

Marsollier L, Robert R, Aubry J, Saint Andre JP, Kouakou H, Legras P, Manceau A, Mahaza C and Carbonnelle B (2002) Aquatic insects as a vector for *Mycobacterium ulcerans*. *Appl Environ Microbiol* 68:4623–4628.

Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R and Schäffer AA (2008) Database indexing for production MegaBLAST searches. *Bioinformatics* 24:1757–1764.

Mrlik V, Slany M, Kubecka J, Seda J, Necas A, Babak V, Slana I, Kriz P and Pavlik I (2012) A low prevalence of mycobacteria in freshwater fish from water reservoirs, ponds and farms. *J Fish Dis* 35:497–504.

Narang R, Narang P and Mendiratta DK (2009) Isolation and identification of nontuberculous mycobacteria from water and soil in central India. *Indian J Med Microbiol* 27:247–250.

Newton RJ, Jones SE, Eiler A, McMahon KD and Bertilsson S (2011) A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* 75:14–49.

Niva M, Hernesmaa A, Haahtela K, Salkinoja-Salonen M, Sivonen K and Haukka K (2006)

Actinobacterial communities of boreal forest soil and lake water are rich in mycobacteria. *Boreal Environ Res* 11:45–53.

Oksanen J, Blanchet G, Kindt R, Legendre P, Minchin P, O'Hara R, Simpson G, Solymos P, Stevens M and Wagner H (2013) Package "vegan". Community ecology package, version, 2(9).

Parashar D, Das R, Chauhan DS, Sharma VD, Lavania M, Yadav VS, Chauhan SVS and Katoch VM (2009) Identification of environmental mycobacteria isolated from Agra, north India by conventional and molecular approaches. *Indian J Med Res* 129:424–431.

Parker BC, Ford MA, Gruft H and Falkinham III JO (1983) Epidemiology of infection by nontuberculous mycobacteria: IV. preferential aerosolization of *Mycobacterium intracellulare* from natural waters. *Am Rev Respir Dis* 128:652–656.

Pickup RW, Rhodes G, Arnott S, Bull TJ, Weightman A, Hurley M and Hermon-Taylor J (2005) *Mycobacterium avium* subsp. *paratuberculosis* in the catchment area and water of the river taff in South Wales, United Kingdom, and its potential relationship to clustering of Crohn' s disease cases in the city of Cardiff. *Appl Environ Microbiol* 71:2130–2139.

Pickup RW, Rhodes G, Bull TJ, Arnott S, Sidi-Boumedine K, Hurley M and Hermon-Taylor J (2006) *Mycobacterium avium* subsp. *paratuberculosis* in lake catchments, in river water abstracted for domestic use, and in effluent from domestic sewage treatment works: diverse opportunities for environmental cycling and human exposure. *Appl Environ Microbiol* 72:4067–4077.

Pontiroli A, Khera TT, Oakley BB, Mason S, Dowd SE, Travis ER, Erenso G, Aseffa A, Courtenay O and Wellington EMH (2013) Prospecting environmental mycobacteria: combined molecular approaches reveal unprecedented diversity. *PLoS One* 8:e68648.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J and Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:590–596.

R Development Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Radomski N, Roguet A, Lucas FS, Veyrier FJ, Cambau E, Accrombessi H, Moilleron R, Behr MA and Moulin L (2013) *atpE* gene as a new useful specific molecular target to quantify *Mycobacterium* in environmental samples. *BMC Microbiol* 13:277.

Smith B and Wilson JB (1996) A consumer's guide to evenness indices. *Oikos* 76:70–82.

Steinman AD, Lamberti GA and Leavitt PR (2006) Biomass and Pigments of Benthic Algae. *Methods Stream Ecol* 1st:295–313.

Stinear T, Davies JK, Jenkin GA, Hayman JA, Oppedisano F and Johnson PDR (2000) Identification of *Mycobacterium ulcerans* in the environment from regions in Southeast Australia in which it is endemic with sequence capture-PCR. *Appl Environ Microbiol* 66:3206–3213.

Tortoli E (2009) Clinical manifestations of nontuberculous mycobacteria infections. *Clin Microbiol Infect* 15:906–910.

Tsukamura M, Yano I and Imaeda T (1986) *Mycobacterium moriokaense* sp. nov., a rapidly growing, nonphotochromogenic *Mycobacterium*. *Int J Syst Bacteriol* 36:333–338.

Uyttebroek M, Breugelmans P, Janssen M, Wattiau P, Joffe B, Karlson U, Ortega-Calvo JJ, Bastiaens L, Ryngaert A, Hausner M and Springael D (2006) Distribution of the *Mycobacterium* community and polycyclic aromatic hydrocarbons (PAHs) among different size fractions of a long-term PAH-contaminated soil. *Environ Microbiol* 8:836–847.

Vaerewijck MJM, Huys G, Palomino JC, Swings J and Portaels F (2005) Mycobacteria in drinking water distribution systems: ecology and significance for human health. *FEMS Microbiol Rev* 29:911–934.

van der Wielen PWJJ, Heijnen L and van der Kooij D (2013) Pyrosequence analysis of the hsp65 genes of nontuberculous mycobacterium communities in unchlorinated drinking water in the Netherlands. *Appl Environ Microbiol* 79:6160–6166.

Viallier J and Viallier G (1973) Inventaire des mycobacteries de la nature. *Ann Soc Belg Med Trop (1920)* 53:361–371.

Walenciak O, Zwisler W and Gross EM (2002) Influence of *Myriophyllum spicatum*–derived tannins on gut microbiota of its herbivore *Acentria ephemerella*. *J Chem Ecol* 28:2045–2056.

Wick LY, Pasche N, Bernasconi SM, Pelz O and Harms H (2003) Characterization of multiple-substrate utilization by anthracene-degrading *Mycobacterium frederiksbergense* LB501T. *Appl Environ Microbiol* 69:6133–6142.

Williams MM, Yakrus M a, Arduino MJ, Cooksey RC, Crane CB, Banerjee SN, Hilborn ED and Donlan RM (2009) Structural analysis of biofilm formation by rapidly and slowly growing nontuberculous mycobacteria. *Appl Environ Microbiol* 75:2091–2098.

Willumsen P, Karlson U, Stackebrandt E and Kroppenstedt RM (2001) *Mycobacterium frederiksbergense* sp. nov., a novel polycyclic aromatic hydrocarbon-degrading *Mycobacterium* species. *Int J Syst Evol Microbiol* 51:1715–1722.

Wurl O and Obbard JP (2004) A review of pollutants in the sea-surface microlayer (SML): a unique habitat for marine organisms. *Mar Pollut Bull* 48:1016–1030.

Wurtzer S, Prevost B, Lucas FS and Moulin L (2014) Detection of enterovirus in

environmental waters: A new optimized method compared to commercial real-time RT-qPCR kits. *J Virol Methods* 209:47–54.

VIII. Supplementary data

Supplementary data 1 Physicochemical characteristics of Créteil and Daumesnil Lakes (water columns parameters and sediment C/N ratios). Values represent arithmetic means of the two sampling campaigns. Standard deviations are expressed in brackets.

Parameter	Créteil	Daumesnil
Location	48°46.50'N 2°27.10'E	48° 49.80'N 2° 25.00'E
Surface area (km^2)	0.40	0.12
Mean depth (m)	4.50	1.10
Max depth (m)	6.00	2.00
Trophic status	Mesotrophic	Eutrophic
Secchi depth (m)	1.9 (1.1)	0.6 (0.1)
Chlorophyll a ($\mu g.L^{-1}$)	19.6 (6.7)	43.8 (10.1)
PO_4 ($mg.L^{-1}$)	0.08 (0.07)	0.03 (0.03)
NO_3 ($mg.L^{-1}$)	0.22 (0.18)	0.69 (0.52)
NH_4 ($mg.L^{-1}$)	0.17 (0.25)	0.06 (0.05)
DOC ($mgC.L^{-1}$)	6.1 (0.3)	4.1 (0.6)
POC ($mgC.L^{-1}$)	1.7 (0.6)	5.0 (1.7)
TSS ($mg.L^{-1}$)	8.1 (1.8)	23.4 (15.7)
pH	8.4 (0.2)	8.6 (0.1)
Conductivity ($\mu S.cm^{-1}$)	1530 (19)	455 (41)
Sediment C/N ratio	36.9 (14.2)	19.6 (2.6)

Abbreviations: DOC, dissolved organic carbon concentration; POC, particulate organic carbon concentration. C/N, carbon to nitrogen ratio.

Supplementary data 2 Relationship between the dry weight and the biofilm surface.

Sampling collection

For both Créteil and Daumesnil Lake, epilithic biofilms were removed from rock surface by scraping 20 cm^2 biofilm area with a sterile syringe-toothbrush sampler. Epilithic biofilms were stored in a sterile container that was previously weighed.

Epiphytic biofilms were collected by harvesting submerged leaves of *Phragmites australis* (the base attached to the stem) in Créteil Lake and *Myriophyllum spicatum* within Daumesnil Lake. Epiphytic biofilms were stored in sterile containers. After collection, biofilms samples were placed at 4°C until return to the laboratory, where they were immediately frozen at – 20°C in order to be lyophilized.

Surface determination

Dry epilithic biofilms were weighed to quantifying the biofilm mass harvested by scraping 20 cm^2 of rocks.

For both *P. australis* and *M. spicatum,* six replicates of 40 mg of dry leaves (mass used for DNA extraction) were placed between two glass plates with a size marker. Pictures taken from these replicates were analyzed using the area calculation tool in ImageJ (Schneider *et al.* 2012). Regarding *P. australis*, we assumed that environmental mycobacteria can only grow on the outer part of the leaves. Thus the area measured by ImageJ corresponds to biofilm area where mycobacteria can grow (Fig. S1). In contrast with *M. spicatum,* we considered that leaves are composed of a multitude of small cylindric objects (Fig. S1). Consequently, to assess the area of cylinders we used the formula: $Area_{Cylinder}$ = pi × diameter × height. However, only a limited area of the cylinders was visible (Fig. S1), this area (in green) corresponding to the parameter: diameter × height. Thus, to determine the whole surface cylinders, representing the biofilm area, we multiplied the visible surface area by pi.

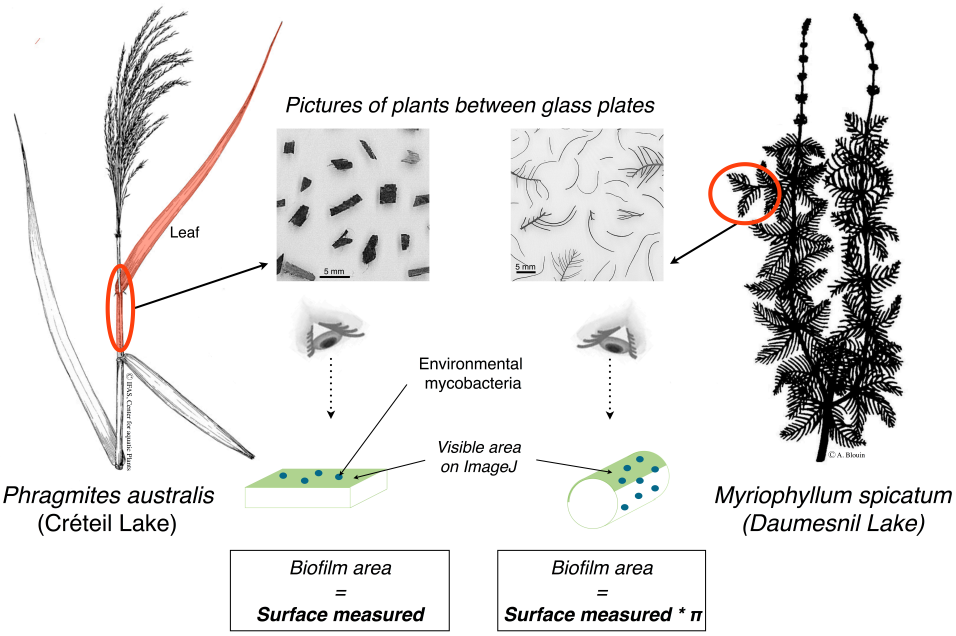


Figure S1 Diagram illustrating the methodology to determine biofilm areas on *P. australis* and *M. spicatum.*

Correlation between dry mass and surface

To determine the relationship between the dry mass of epilithic biofilms and the surface where mycobacteria could proliferate, the following equation was used for each sample:

$$Surf_{Epilithic} = \frac{M \times Surf_{Scraped}}{(m_1 - m_0)} \quad (1)$$

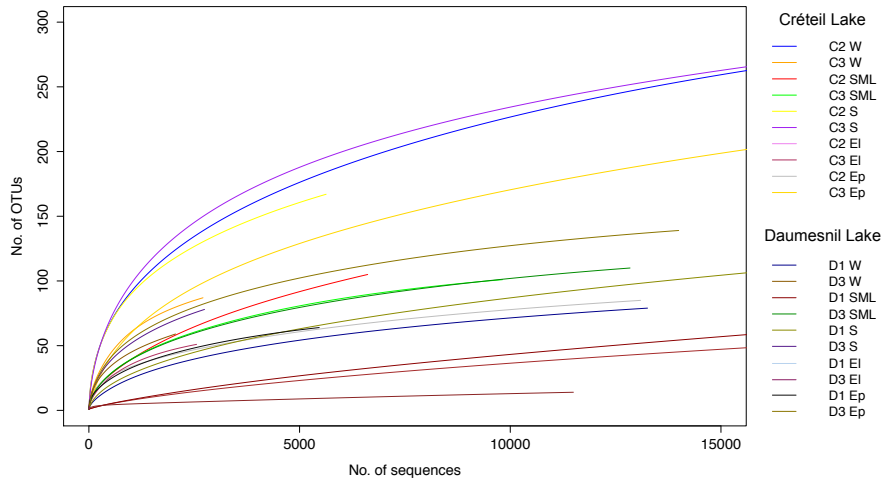
where $Surf_{Epilithic}$ is the surface; M is the dry mass of epilithic biofilm used for the DNA extraction (50 mg); $Surf_{Scraped}$ is the total surface scrapped during the sample collection, m_0 is the mass of the empty container and m_1 the mass of the container with dry biofilm. Measures of biofilms area were compiled in the Table S1. Thus, we admitted that 40 mg of dry *P. australis* leaves from this plant corresponds to 9.5 cm^2 of biofilm surface whereas 40 mg of dry *M. spicatum* leaves corresponds to 37.6 cm^2 of biofilms surface.

Table S1 Biofilm area (in cm^2) measured from 40 mg of dry *P. australis* and *M. spicatum* leaves

Replicates	*Phragmites*	*Myriophyllum*
1	10.27	35.17
2	8.42	37.22
3	9.88	37.34
4	8.39	38.17
5	9.66	39.07
6	10.11	38.45
Mean (cm^2)	9.46	37.57
Standard deviation (cm^2)	0.77	1.25
Coefficient of variation (%)	8.11	3.31

Schneider CA, Rasband WS and Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9: 671-675.

Supplementary data 3 Rarefaction curves obtained for the twenty samples collected in Créteil and Daumesnil lakes at a cutoff level of 3%. C2, C3, D1 and D3 correspond to the stations numbers within the two lakes. Abbreviations of the five compartments: W, water column; SML, surface microlayer; S, sediment; El, epilithic biofilm; Ep, epiphytic biofilm.



Supplementary data 4 OTUs assignation to the species level (with a cutoff >97% identity) using the blastn algorithm on GenBank website (*http://blast.ncbi.nlm.nih.gov/Blast.cgi*). Available at: http://dx.doi.org/10.1007/s10482-016-0665-x

Chapter V

Mycobacteria are everywhere, but, only the environment selects?

Mycobacteria are everywhere, but, only the environment selects?

Adélaïde Roguet[a], Claire Therial[a], Adèle Bressy[a], Viet Tran[a], Bruno Lemaire[a], Philippe Dubois[a], Arnaud Catherine[b], Gérard Lacroix[c,d], Ludwig Jardillier[e], Laurent Moulin[f] and Françoise S. Lucas[a].

[a]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 94010 Créteil, France.

[b]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.

[c]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.

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

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

[f]Eau de Paris, Direction Recherche et Développement Qualité de l'Eau (DRDQE), 33 avenue Jean Jaurès, FR 94200 Ivry-sur-Seine, France.

Key words: mycobacteria, community composition, density, lake, real-time quantitative PCR

I. Abstract

Although nontuberculous mycobacteria (NTM) are considered to be widespread in aquatic ecosystems, only few studies investigated their distribution in surface freshwaters especially using molecular tools. The aim of this study was to discriminate the factors affecting the spatial and temporal variations of mycobacterial densities in freshwater lakes at the regional scale (49 lakes in the Paris area) and at the local scale (two-year monitoring of Créteil Lake). *Mycobacterium* spp. was quantified from water samples by real-time quantitative PCR targeting the *atpE* gene.

At the regional scale, mycobacteria were detected in all the sampled lakes, with densities ranging from 8.6×10^3 to 9.4×10^6 genome units per liter. Spatial density variations were significantly explained by local environmental factors including the water pH, the labile iron concentration, but they were also explained by spatial factors. Lakes connected to a river displayed significantly higher mycobacterial densities compared to isolated lakes. At the local scale, storm sewer effluents locally affected NTM densities in the Créteil Lake immediately after rainfalls. However, no significant spatial or temporal pattern was observed over the two-year monthly survey. Nevertheless, the temporal dynamic of the NTM densities was well predicted by the neutral community model, suggesting a random balance between loss and gain of the mycobacterial taxa. This study highlights the ubiquitous character of NTM in freshwater lakes, which appeared to be relevant habitats for mycobacteria.

II. Introduction

Among the *Mycobacterium* genus, about 170 species have been described, most of them being defined as nontuberculous mycobacteria (Hruska and Kaevska 2012). Some of these species has been recognized to be opportunistic pathogens of humans and animals (Falkinham III 1996; Mrlik *et al.* 2012), and it is widely assumed that nontuberculous mycobacteria (NTM) infecting humans have an environmental origin (Marras and Daley 2002). Historically most environmental studies of mycobacteria have focused on the drinking water distribution systems (e.g., Covert *et al.* 1999; Falkinham III *et al.* 2001; Vaerewijck *et al.* 2005) or hospital water supply networks (e.g., du Moulin *et al.* 1988; Fox *et al.* 1992; Fujita *et al.* 2002) due to the increase number of mycobacteriosis contracted after water contact.

Recent molecular studies of the bacterial communities suggested that mycobacteria may be widespread and represent a dominant taxon in different aquatic and terrestrial ecosystems (e.g., Kwon *et al.* 2010; Nacke *et al.* 2011; Liu *et al.* 2012; Cai and Zhang 2013; Ye and Zhang 2013; Kim *et al.* 2014; Lührig *et al.* 2015; Unno *et al.* 2015; Yang *et al.* 2015). Although mycobacteria are common inhabitants of terrestrial and aquatic ecosystems (Collins *et al.* 1984; Falkinham III 2002), their favorable habitats in natural environment remain poorly documented. This gap is probably enhanced by the difficulties in isolating NTM from the environment using culture methods (Falkinham III 2002) and by the resistance of their wall cell to DNA extraction procedures. It could also be due to a lack of interest in ecosystems that do not represent the main infection routes in most mycobacteriosis cases.

Environmental studies focusing on NTM were preferentially performed in soil or sediment, notably for their role as polycyclic aromatic hydrocarbon degraders (e.g., Leys *et al.* 2005; Niva *et al.* 2006; Pontiroli *et al.* 2013). Among the few studies exploring freshwater habitats, most of them focused on one specific or cultivable species (e.g., Viallier and Viallier 1973; Stinear *et al.* 2000; Pickup *et al.* 2005, 2006; Gauthier *et al.* 2010; Bland *et al.* 2005; Rahbar *et al.* 2010), providing a limited and patchy overview of the NTM assemblages in these environments.

Among freshwater ecosystems, lakes appear to be useful models to survey mycobacteria. In boreal water lakes for example, mycobacterial assemblages constituted a large part of the *Actinobacteria* phylum (Niva *et al.* 2006), generally one of the dominant phylum in lacustrine ecosystems (Newton *et al.* 2011). Moreover, lakes may constitute an important source of infection as they attract a high number of visitors due to the numerous cultural, provisional, and recreational services that they provide. Conclusive supports from experimental and environmental monitoring performed in diverse habitats indicated the prevalence of mycobacteria in extreme environments, i.e., habitats with a low amount of oxygen, high metal concentration, acid brown-water (e.g., Kirschner *et al.* 1999; Iivanainen *et al.* 1993). However, there is little knowledge about the prevalence of NTM in less extreme aquatic environments that may represent important water bodies. Thus it is essential to better understand NTM reservoirs or factors that could affect their occurrence, in aquatic environments that are widely frequented by humans.

Finally, temporal survey of NTM in freshwater ecosystems revealed important shifts in mycobacterial diversity or variations in the densities of some species such as *M. avium* (e.g., Bland *et al.* 2005; Pickup *et al.* 2005), especially after rainfall events (Pickup *et al.* 2006). However, as far as we are aware, no such monitoring has already been performed on the entire NTM assemblage using quantitative molecular tools. As a consequence the comprehension of the factors and processes that shape the mycobacterial communities is still poor.

NTM assemblages should be subjected to ecological processes that shape their spatial and temporal distribution, i.e. dispersal, drift, selection and speciation (Vellend 2010). Some ecological processes being difficult to assess, e.g., the ecological drift (Hanson *et al.* 2012), the spatial and the temporal distribution of bacterial communities are often studied by investigating the local environmental conditions, the processes affecting the species dispersal, and the stochastic processes (neutral processes). The neutral processes correspond to the forces that randomly shape a bacterial population by the stochastic balance between dispersal, apparition and death of taxa from a set of habitats (see Leibold *et al.* 2004; Martiny et al. 2006).

In this study, we aimed to determine the environmental and spatial factors that shape the distribution of NTM at the regional and local scale in freshwater lakes. At the regional scale, a total of 49 freshwater lakes located in Paris area (France) were sampled once a year for two consecutive years. This set of shallow lakes, some being used for recreational activities, displays a large gradient of environmental parameters and spatial features that could enhance the bacterial dispersion (e.g., river or watershed connections) (Catherine *et al.* 2008, 2010). At the local scale, a two-year monthly monitoring combined with the survey of six runoff discharges into the water column of Créteil Lake allowed to characterize the spatio-temporal dynamic of mycobacterial densities and to discriminate the local factors (i.e. physicochemical parameters and storm sewer effluents). We also evaluated at the local scale the influence of the neutral processes on the temporal dynamic of the NTM density variations. NTM were assessed by quantifying their densities using real-time quantitative PCR (Radomski *et al.* 2013).

III. Materials and methods

III.1 Study area and sampling

The 49 lakes This study was conducted in Paris area, the most populated area in France (with 18% of the metropolitan French population). Among the 248 waterbodies larger than 5 ha referenced in the hydrological database Carthage 3.0® (IGN, Paris, France), 49 lakes were selected (Fig. 24 and Supplementary data 1) to obtain a representative and unbiased set of water bodies reflecting the whole range of environmental conditions in the Paris area (Catherine *et al.* 2008). Trophic status was determined for each of the 49 lakes according to OECD (1982).

The lakes were sampled twice, in late July in 2012 and 2013. Both sampling campaigns were 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 depths using a Niskin bottle (General Oceanics Inc., Miami (FL), USA). All samples were then pooled together to obtain an integrated sample. Immediately after collection, samples were kept at 4°C for a maximum of 24h. Successive centrifugations

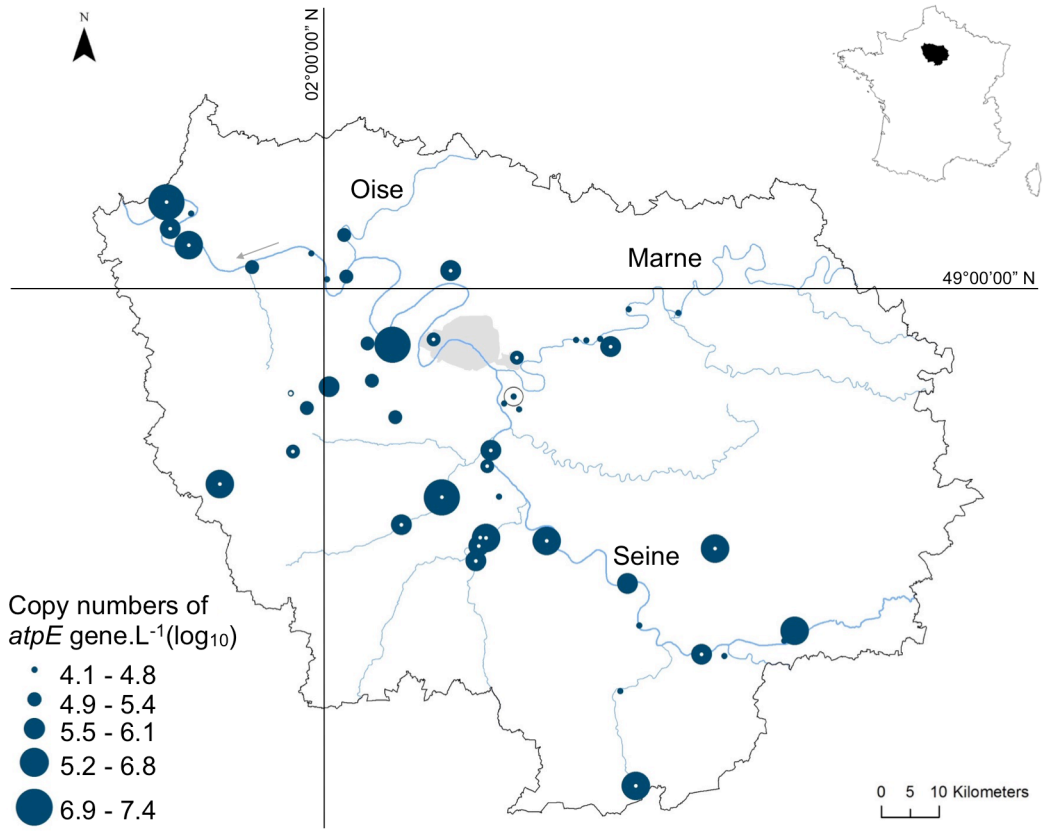


Figure 24 Location and mycobacterial genome unit densities within the 49 lakes in Paris area (France). Lakes linked to the hydrological network are symbolized by white dots. Cérteil Lake is indicated by an open circle.

(7,500 × g, 4°C) were performed to concentrate one liter into a 2-mL sterile tube. In order to facilitate pellet resuspension, 1 mL of Tween 80 (final concentration, 0.01% vol/vol) was added before centrifugation to each sample. All samples were stored at – 20°C until DNA extraction.

Créteil Lake Among the 49 lakes, Créteil Lake (Fig. 24 and 25) was monthly monitored at three different stations from December 2011 to December 2013 (25 sampling dates). This mesotrophic lake (Table 10) covers 40 ha in an urbanized area (Val-de-Marne, France) and is mainly supplied by alluvial groundwater. However, Créteil Lake also receives effluents from a storm sewer that collects runoff from an urban area of 100 ha.

A horizontal and a vertical transects were studied (Fig. 25). Sub-surface samples were obtained at three stations along the inlet-outlet axis of the lake (storm sewer outlet (S1), center of the lake (C1), and lake outflow (O1)). At the central station, the vertical

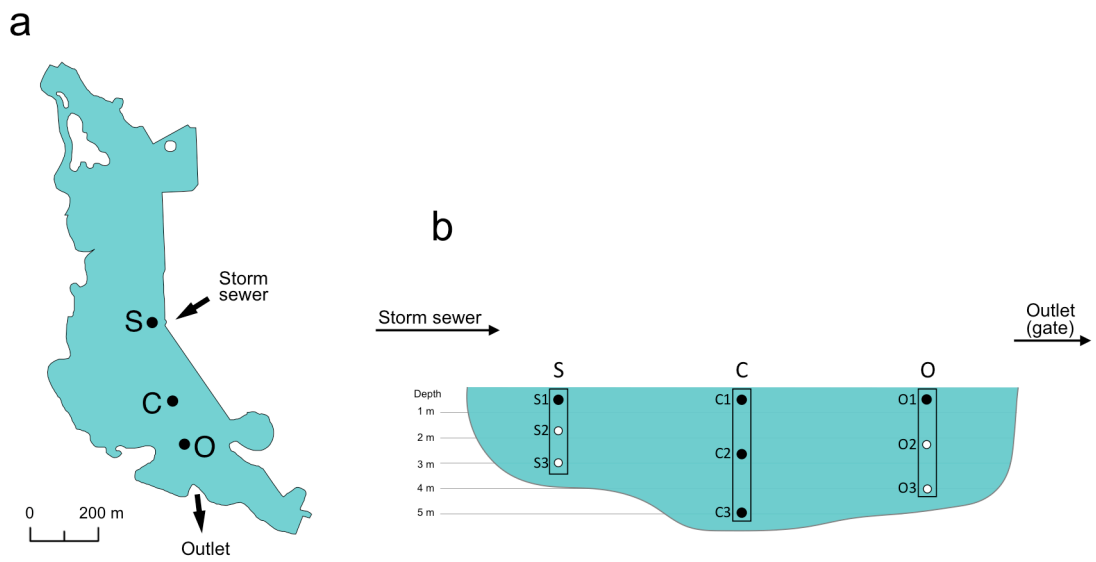


Figure 25 Sampling strategies in the Créteil Lake. (a) Location of the three stations. (b) Representation of the samples collected over the two-year monitoring (solid circles) and the rainfall monitoring (open rectangles).

axis was sampled at three depths (C1, C2 and C3). In addition, for all sampling dates, an average sample (M) was obtained by mixing the three stations at the three depths (S1, S2, S3, C1, C2, C3, O1, O2 and O3). All samples were stored at 4°C up to 3-4 hours. One liter was filtered through a 0.22 µm pore-size Sterivex GP-filter (Millipore, Billerica (MA), USA) after prefiltration through 50 µm pore-size nylon mesh. Filtered samples were kept at 4°C during transport and then stored at – 20°C. Prior to DNA extraction, the membrane inside

Table 10 Characteristics of Créteil Lake. Average values (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)
DOC ($mgC.L^{-1}$)	6.1 (0.6)
POC ($mgC.L^{-1}$)	0.8 (0.3)
TSS ($mg.L^{-1}$)	4.1 (1.5)
pH	7.8 (0.5)
Conductivity ($\mu S.cm^{-1}$)	1509 (63)

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 2-mL sterile tube.

Rainfall events The impact of the storm sewer effluents on the mycobacterial densities in Créteil Lake was evaluated from six rain events from June to November 2013. For that purpose, the storm sewer was equipped with the flowmeter Mainstream IV (Hydreka Inc., France) and a piezometric sensor upstream the outfall and above the high-water line of the lake to avoid back flow and sample dilution. Only isolated rainfalls higher than 4 mm that occurred after 4 PM were collected. A representative effluent sample was collected in a 10 L sterile glass bottle using an automatic Sigma sampler (Hach Inc., Hach Loveland, (CO), USA) that was triggered to the flow. The next early morning, immediately after recovering the effluent sample, a sampling campaign was performed on Créteil Lake as describe above (S1, C1, C2, C3, O1). Aside the average point M, an integrating sample of the water column was collected for each of the three stations (Fig. 25). One liter of lake samples and from 100 to 800 mL of the effluent samples were filtered as previously described.

The choice of Sterivex filtration was taken as it allows a fast concentration of the water samples. Indeed, it appeared logistically difficult to centrifuge all Créteil samples in a short time after each sampling campaign. However, we assume that the difference of protocol used to concentrate the bacteria between the 49 lakes (centrifugation) and the Créteil Lake (filtration) did not introduce a considerable bias in NTM densities, since no significant difference of DNA recovery was observed between these two protocols over preliminary experiments and during conjoint campaigns (data not shown).

III.2 DNA extraction and quantification of mycobacterial densities

Total DNA was extracted using the FastDNA® SPIN Kit (QBiogene, Carlsbad (CA), USA) according to the manufacturer's instructions. Two modifications to this protocol were applied, cells were lysed in a FastPrep bead beater three times for 30 s at 4.0 ms^{-1} and the SPIN filters were washed twice. All samples were eluted in 50 µL of deionized sterile water. The quality and quantity of extracted DNA were analyzed at 230, 260, and 280 nm by spectrophotometry before storage at – 20°C.

To quantify the abundance of *Mycobacterium* genus in the water samples, TaqMan® real-time quantitative PCR assays targeting the *atpE* gene were carried out in duplicate as previously described by Radomski and colleagues (2013). The *atpE* copy numbers concentration was estimated from crude extracts using a *Mycobacterium chelonae* standard curve from 1.0×10^1 to 1.0×10^6 copies.μL^{-1}. Contaminations in PCR mix were checked using negative controls. Presence of PCR inhibitors in DNA templates was verified using a non-competitive exogenous internal control that was included in the PCR buffer at 1,000 copies. This internal control was made of a partial sequence of the *human β-actin* gene cloned in pGEM-T-easy vector (Promega, Madison (WI), USA) (Wurtzer *et al.* 2014). Absence of significant PCR inhibition of the atpE assay was done by comparing results with the average C_q value (± standard deviation) obtained from 100 repeated PCR reactions containing 1,000 copies of *human β-acti*n gene fragment.

III.3 PCR assay performance

The performance of the modified extraction protocol was evaluated by centrifuging and extracting in triplicate one liter of water from Créteil Lake and one liter inoculated with a known concentration of *Mycobacterium chelonae* (determined using culture based methods). Real-time quantitative PCR efficiency reached 94.6% with a R^2 of 0.999 based on a standard curve of *M. chelonae*. With a single copy of *atpE* gene per mycobacterial genome, results were expressed as genome unit per liter, based on the *M. chelonae* standard curve.

III.4 Physicochemical and meteorological parameters

For the 49 lakes, total nitrogen and phosphorus concentrations in the water samples were measured by colorimetry using a Cary 50 Scan spectrophotometer (Varian Inc., Palo Alto (CA), USA) according to Rogora and colleagues (2006) and the French standard (AFNOR NF T 90-023) respectively. Secchi depth and vertical profiles of physicochemical parameters were determined for each lake at the three stations (49 lakes and Créteil Lake), and 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 profiler (Sea-Bird Electronics Inc., Bellevue (WA), USA). Total suspended

solids (TSS) concentration was quantified after filtration of one liter on precombusted tarred Whatman GF/F filter. Dissolved organic carbon (DOC) concentrations were measured using a TOC-VCSN carbon analyzer (Shimadzu, Columbia (MD), USA).

For the 49 lakes, the polycyclic aromatic hydrocarbons (PAHs) concentrations in the water samples were quantified both in dissolved and particulate fractions by internal calibration using gas chromatography coupled with mass spectrometry GC/MS Focus DSQ (Thermo Fisher Scientific, Waltham (MA), USA) (Bressy *et al.* 2011). Results are displayed as the sum of 13 PAHs (Fluorene, phenanthrene, anthracene, fluoranthene, pyrene, benzo[a]anthracene, chrysene, benzo[b]fluoranthene, benzo[k]fluoranthene, benzo[a]pyrene, indeno[1,2,3]pyrene, di-benzo[a,h]anthracene, and benzo[g,h,i]perylene). In 2012, the labile fraction of trace metals (i.e., Ti, Mn, Fe, Co, Ni, Cu, Zn, Cd, Ba and Pb) was estimated by ICP-MS at the LSCE laboratory (Gif-sur-Yvette, France) after water filtration on chelating disk cartridge (Chaminda *et al.* 2008).

During the Créteil Lake campaigns, cumulative precipitation data of each rain event from 1 to 15 days before the campaign were collected using a weather real-time transmitter WXT520 (Vaisala Inc., Boulder (CO), USA) placed on a sensing platform LakeESP (PME Inc., Vista (CA), USA).

III.5 Data analysis

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

Inter-lake variation analysis Since no significant difference was observed between the two years of sampling for the 49 lakes dataset (data not shown), the statistical analysis was performed on the entire dataset regardless the year of sampling. A linear model (LM) was performed to evaluate the influence of the environmental and dispersal factors on the mycobacterial densities. Due to important colinearities between variables, only seven quantitative (water pH, conductivity and temperature, DOC, TSS, total PAHs, total nitrogen and total phosphorous concentrations) and two qualitative variables (the presence of a storm sewer discharge and the presence of a connection to the hydrological network) were injected into a variable selection. Furthermore, for the year 2012 only, the labile fraction of trace metals was added to the LM.

Spatial and temporal variability within Créteil Lake NTM densities variation along the two spatial transects were assessed with linear mixed-effect models (LMM) using the *lme* function from the 'nlme' package (Pinheiro *et al*. 2015) with the sampling dates as mixed effect. The temporal dynamic of the mycobacteria over the two-year survey was evaluated using a linear model. We also evaluated if the neutral model could explain the temporal variations of the NTM densities as previously described by Ofiţeru and colleagues (2010). Prior to analysis, non-log data densities were transformed into relative abundance, considering the highest abundance as equal to one. The goodness of fit was evaluated using the determination coefficient R^2.

Rainfall events The impact of effluents on the horizontal transect between the storm sewer (S in Figure 25) and the outlet (O in Figure 25) was assessed using a LMM with the sampling dates as mixed effect, following by a post-hoc test using the *glht* function from the 'multcomp' package (Hothorn *et al*. 2015). *P*-values were corrected with the Bonferroni correction to reveal significant differences.

Data exploration and model validation (normality, homoscedasticity and independence of residuals) of all analyzes were assessed according to Zuur and colleagues (2009). Variable selections were performed prior to each analysis.

IV. Results

IV.1 Assay performances

After centrifugation and nucleic acid extraction by the modified FastDNA® SPIN kit, the recovery of *atpE* gene copies from environmental spiked water samples averaged 124 ± 29% (standard deviation; $n = 3$). The *atpE* gene was successfully amplified in all the 49 lakes and Créteil Lake samples. No inhibition was identified in any of these samples.

IV.2 Mycobacterial distribution at the regional scale

Over the two years, from 8.6×10^3 to 9.4×10^6 genome units per liter (Fig. 24) were quantified in the 49 lakes. A linear model ($F_{3,89} = 18.30$, $R^2 = 0.38$; $P < 0.001$) showed that lakes connected to the hydrological network (river or stream) had significantly higher numbers of genome units per liter compared to non-connected lakes ($F = 22.52$,

$P < 0.001$). Water pH appeared to be weakly but significantly correlated (F = 7.74, $P = 0.007$) with the genome unit concentration (Spearman correlation, rho = -0.37; $P < 0.001$). No significant influence of nutrient, polycyclic aromatic hydrocarbon, total suspended solids concentrations or the presence of a storm sewer outlet was identified. Furthermore, in 2012 a linear model ($F_{2,39} = 19.12$, $R^2 = 0.50$; $P < 0.001$) revealed that the labile iron fraction also positively correlated with genome unit densities (linear model, F = 8.99, $P < 0.001$; spearman correlation, rho = 0.61, $P < 0.001$) in addition to the connection to the hydrological network (F = 14.84, $P < 0.001$).

IV.3 Spatio-temporal variations of the mycobacterial densities in Créteil Lake

The two-year monitoring of the Créteil Lake revealed no significant spatial variability of NTM densities along the vertical transect located at the central station (LMM, F(46) = 0.58, $P = 0.561$, degrees of freedom in brackets), even when the lake was stratified according to the temperature gradient (data not shown). Similarly, no significant impact of the storm sewer was visible along the horizontal transect (LMM, F(48) = 2.67, $P = 0.079$). Since no spatial variability was found, temporal variation analysis was carried out on the pooled samples (M).

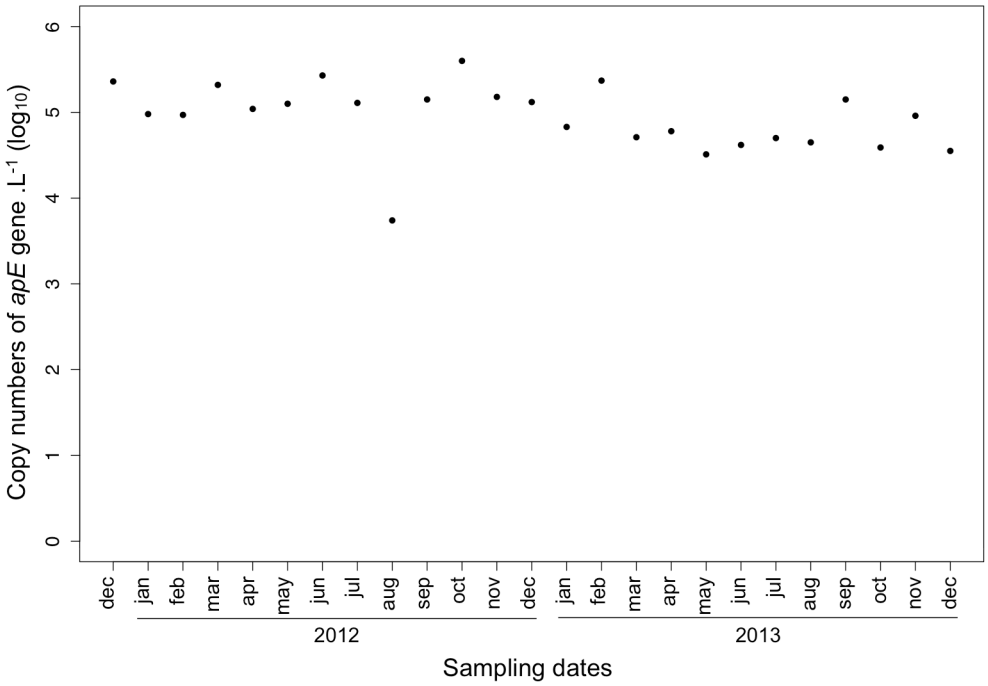


Figure 26 Variations of genome unit densities along the two-year monitoring in Créteil Lake for the pooled sample M.

Without considering an extreme value in August 2012 (due to a sampling issue), no significant temporal variation of the genome unit densities was assessed at the average point M over the two-year survey (LM, $F_{1,22} = 2.00$, $R^2 = 0.08$; $P = 0.172$) (Fig. 26). The neutral time-series model proposed by Ofiţeru and colleagues (2010) well predicted the temporal dynamic of the NTM densities ($R^2_{adj} = 87\%$, $P < 0.001$).

IV.4 Rainfall impact on the mycobacterial densities in Créteil Lake

Regarding the six rain events that were monitored, the concentrations of NTM in the sewer effluent ranged from 2.2×10^7 to 2.2×10^8 genome units per liter, which represented a median input of $1.7 \times 10^{14} \pm 1.5 \times 10^{14}$ genome units in the lake per rain event. This input significantly impacted the *atpE* gene densities into Créteil Lake (Fig. 27) within 5 to 20h following the rainfall, time corresponding to the extreme duration between the rainfall events and the lake sampling campaigns. Indeed shortly after the rain events, NTM concentrations were significantly higher at the station (S) in front of the storm sewer outlet compared to the two other stations (Post-hoc test, stations S-C: $P = 0.0337$; S-O: $P < 0.001$). No impact of the storm sewer effluent was observed on the average point M (Post-hoc test, $P > 0.05$).

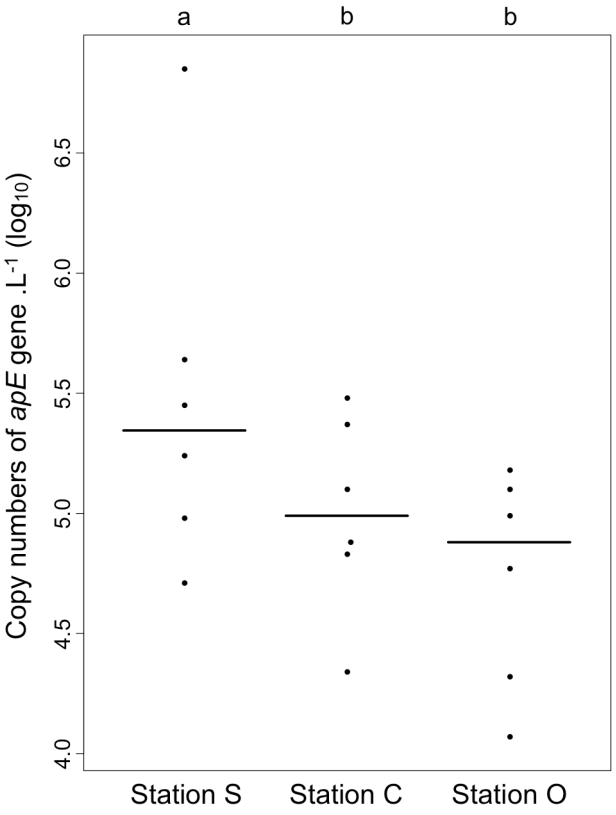


Figure 27 Distribution of the mycobacterial genome units per liter for each of the three stations in Créteil Lake for the six rain events. Bars represent the median of the six observations. Distributions connected by the same letter are not significantly different (Post-hoc tests corrected by Bonferroni correction, $P < 0.05$).

V. Discussion

The primary objective of this study was to examine the abundances and distributions of the nontuberculous mycobacteria not only at the regional scale but also at the local scale to discriminate factors that could affect the spatial and/or temporal dynamics of the mycobacterial assemblages. Such an original investigation, using a recently developed real-time quantitative PCR provides a new insight of NTM ecology in freshwater ecosystems.

V.1 Methodological considerations

Assessment of the mycobacterial abundance in freshwater using exclusively the molecular approach remains rare (Radomski 2011; Khera 2012). Quantitative PCR offer the advantage to avoid the culture biases that prevent the estimation of the slow-growing mycobacteria in the environmental samples (Radomski *et al.* 2010a,b). However, PCR-based methods also present the inconvenience to detect DNA from both live and dead cells. Thus, our results may overestimate the abundance of the NTM live-cells, especially due to the persistence of DNA after cell death (Josephson *et al.* 1993; Masters *et al.* 1994).

In this assay, the DNA recovery from raw lake water samples was close to 100 % using centrifugation and the modified extraction protocol. This high efficiency was probably due to the extraction protocol improvements. Indeed, Guo and Zhang (2013) noticed the high extraction yields obtained for actinobacterial DNA using the FastDNA® SPIN Kit, notably due to the beat-beating step.

V.2 Widespread of NTM in freshwater lakes

Environmental studies exploring the ecological preferences of mycobacteria usually found higher abundances of mycobacteria in a restricted range of aquatic habitats, especially in humic rich waters (Kirschner *et al.* 1999; Iivanainen *et al.* 1993; Jacobs *et al.* 2009). Nonetheless, in this study, NTM were detected in all the 49 lakes for all years at relatively high abundances (with values ranging from 8.6×10^3 to 9.4×10^6 genome units per liter). In spite of a relatively important variability, the estimated mycobacterial densities are in agreement with reported values in diverse freshwater habitats or coastal lakes (Jacobs *et al.* 2009; Khera 2012). This finding highlights the ubiquity of NTM and their

ability to persist and/or grow in non-extreme conditions. As previously mentioned, the successful recovery of NTM in all the collected water samples could potentially result from the use of a more efficient DNA extraction protocol, a lower limit of detection of the quantitative PCR method and/or the analysis of a larger volume of water compared to the previous studies (Jacobs *et al*. 2009; Khera 2012).

V.3 Parameters affecting the NTM densities at the regional scale

Although NTM were encountered in all freshwater lakes, relevant dispersal and environmental parameters were identified as affecting the mycobacterial densities. The connection of the lakes to a river appeared to strongly increase the NTM densities, suggesting that rivers could represent important potential reservoirs and sources of NTM. Indeed, a survey performed in Paris area, reported high densities of NTM in the Marne River with a mean of $2.2 \times 10^5 \pm 2.4 \times 10^5$ 16S rRNA gene copy numbers per liter (Radomski 2011). Alternatively, this relationship could also be indirect, through the modification of the water properties and nutrients availability within lakes by the river inflow. Such changes could improve the persistence and/or the competitiveness of NTM in lakes connected to hydrological network. However, this relationship can also be explained by the dependency between the type of connection and the trophic status. Indeed, isolated lakes are mainly oligotrophic and mesotrophic while connected lakes are mostly eutrophic and hypertrophic, which seems to favor NTM.

Mycobacteria are reported to occur in high abundances in aquatic ecosystems with acidic waters, low oxygen levels, high nutrient availability or high levels of micropollutants (Iivanainen *et al*. 1993; Kirschner *et al*. 1999; Jacobs *et al*. 2009). The significant negative relationship, even weak, between the NTM densities and water pH values that was found in the 49 lakes dataset supports previous observations in freshwater habitats about the importance of water pH for mycobacteria (Iivanainen *et al.* 1993; Kirschner *et al.* 1999). Moreover, as previously reported in various terrestrial or aquatic habitats (Iivanainen *et al.* 1993; Norby *et al.* 2007), NTM were positively correlated with iron content. Iron is an important biocatalyzer of redox reaction in many living cells. Its deficiency in lacustrine ecosystems (Mckay *et al.* 2004) could drastically impact the biomass of mycobacteria that is dependent upon the production of iron-binding siderophores to acquire this compound

(Barclay and Ratledge 1983; Chan 2009). This dependence to siderophores is particularly important in waters with high pH values, when iron is insoluble, and thus, non-bioavailable (Emmenegger *et al.* 2001).

Considering the fact that NTM are hydrophobic and are therefore preferentially associated to hydrophobic surfaces (e.g., suspended particles), surprisingly no relationship between NTM abundances and total suspended solids concentrations. Similarly, no significant influence of the nutrient levels (total nitrogen, total phosphorus or DOC concentrations) or the lake trophic status was observed in this study. Yet, previous experiments reported positive correlations between bacterial growth and nutrients concentration (Elser *et al.* 1995; Eiler *et al.* 2003). Thus, this result could suggest that in the sampled lakes, (i) mycobacteria were not limited by the supply of nutrients, (ii) relevant organic matter variables were not measured, e.g., acid humic or fulvic content (Iivanainen *et al.* 1993; Kirschner *et al.* 1999), (iii) mycobacterial growth was too small to be detected or (iv) mycobacterial populations were not active. Moreover, although a growing number of studies demonstrate the role of some NTM species as polycyclic aromatic hydrocarbon degraders (Uyttebroek *et al.* 2006, 2007; Debruyn *et al.* 2009), no direct relationship was noticed in this survey regarding the PAH concentration, despite the large variability of PAH concentration among the lakes.

V.4 Local rainfall impact

In Créteil Lake, runoff effluents carried large amount of mycobacteria that the may result from the erosion of sedimented matters and biofilms in the storm sewer. In fact, biofilms are well-known reservoirs of mycobacteria in diverse habitats, such as drinking water systems (Dailloux *et al.* 2003; Williams *et al.* 2009). These large densities of mycobacteria in the runoff discharges appeared to locally enhance the NTM abundance in Créteil Lake the day following important rainfalls. However, this impact on the surface water was temporary, as revealed by the absence of relevant spatial heterogeneity in the surface water of the lake over the two-year monitoring. No vertical variability was observed along the vertical transect, suggesting that the water column from the polymictic Créteil Lake was sufficiently well mixed to avoid significant spatial heterogeneity. Discharge plumes were probably rapidly diluted in the lake. Besides, even after important rainfalls,

no significant mycobacterial densities rise was observed on the average sample (M), which could explain the absence of any significant direct impact of storm sewer effluents among the 49 lakes. Similar results were found by Pickup and colleagues (2006), who did not detected a positive correlation between the precipitation and the *Mycobacterium avium* subsp. *paratuberculosis* recovery after nine days. Inversely, Iivanainen and colleagues (1993) observed a positive correlation between the abundance of *Mycobacterium* spp. and the cumulative precipitation in 53 brook waters, but only the second and the third weeks after rainfalls.

V.5 No temporal variation in Créteil Lake

Mycobacterial densities appeared to be roughly stable over the two years monitoring. This temporal dynamics, apparently stable, was instead particularly well predicted by the neutral community model, suggesting that the variations of the *Mycobacterium* densities were governed by a stochastic balance between loss and gain of taxa (Hubbell 2001) as previously demonstrated for many other bacterial groups (Sloan *et al.* 2006). Although deterministic parameters such as the temperature fluctuation did not appeared to significantly affect the *Mycobacterium* densities, we cannot exclude a temporal dynamic of the mycobacterial species. Indeed, in the Rio Grande, Bland and colleagues (2005) observed important shifts in the NTM assemblage composition over a one-year monthly monitoring. Similar tendency was also noticed with the non-systematic recovery of *M. avium* over a one-year survey in the river Taff (Pickup *et al.* 2005, 2006).

In summary, this study emphasis the ubiquity of mycobacteria in non-extreme freshwater habitats, even if some forces could temporally or spatially enhance their abundance (i.e. neutral processes, connection to rivers or storm sewers discharges). As previously described, NTM densities were related to pH and iron concentrations. Freshwater lakes appeared thus as relevant niche of NTM in urban lakes. The famous citation 'everything is everywhere, but, the environment selects' seems to apply to the ubiquitous mycobacteria (Beijerinck 1913; Baas Becking 1934). This citation suggests that no spatial biogeography occur due to dispersal limitation, and that the spatial structure is only governed by environmental conditions. However, our results highlight the

importance of dispersal and neutral processes influencing the mycobacterial densities. The monitoring of water physicochemical parameters alone does not appear to be sufficient to predict mycobacterial densities. Nevertheless an additional study assessing the species composition of the NTM communities should help to reveal the existence of a biogeographical structure in mycobacterial communities at local and regional scale. Especially, it would be interesting to assess the influence of neutral processes on the spatial variations of the NTM densities and diversity.

VI. Acknowledgments

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], Eau de Paris, that financed the mycobacterial analysis and the DSEA (Direction des Services de l'Environnement et de l'Assainissement), the DGST (Direction Générale des Services Techniques), which belongs to the General Council of the Val-de-Marne district that financed the equipment of the storm sewer of Créteil Lake. The Direction Générale des Services Techniques of Créteil City provided the authorizations and the street signalization equipment that were necessary to conduct the storm sewer study in Créteil Lake. We like to thank more specifically Mrs Chamayou, Mrs Vernin, Mrs Berdoulay and Mrs Butel-Gomis. We are grateful to the stakeholders, owners and municipalities for their authorizations to collect samples in each lake, and we especially thank the city of Créteil for their support. We also thank the leisure base of Créteil Lake for lending us their boats over the two 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 during these two years, and Gilles Varrault for his constructive advices.

VII. References

Baas Becking LGM (1934) *Geobiologie of Inleiding Tot de Milieukunde*. Van Stockkum and Zoon, The Hague.

Barclay R and Ratledge C (1983) Iron-binding compounds of *Mycobacterium avium, M. intracellulare, M. scrofulaceum*, and mycobactin-dependent *M. paratuberculosis* and

M. avium. *J Bacteriol* 153:1138–1146.

Beijerinck MW (1913) *De infusies en de ontdekking der backteriën*. In Jaarboek van de Koninklijke Akademie van Wetenschappen. Müller, Amsterdam.

Bland CS, Ireland JM, Lozano E, Alvarez ME and Primm TP (2005) Mycobacterial ecology of the Rio Grande. *Appl Environ Microbiol* 71:5719–5727.

Bressy A, Gromaire MC, Lorgeoux C and Chebbo G (2011) Alkylphenols in atmospheric depositions and urban runoff. *Water Sci Technol* 63.4:671–679.

Cai L and Zhang T (2013) Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. *Environ Sci Technol* 47:5433–5441.

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.

Chaminda GGT, Nakajima F and Furumai H (2008). Heavy metal (Zn and Cu) complexation and molecular size distribution in wastewater treatment plant effluent. *Water Science and Technology* 58:1207-1213

Chan K (2009) Exochelin production in Mycobacterium neoaurum. *Int J Mol Sci* 10:345–353.

Collins CH, Grange JM and Yates MD (1984) Mycobacteria in water. *J Appl Bacteriol* 57:193–211.

Covert TC, Rodgers MR, Reyes AL and Stelma GN (1999) Occurrence of nontuberculous mycobacteria in environmental samples. *Appl Environ Microbiol* 65:2492–2496.

Dailloux M, Albert M, Laurain C, Andolfatto S, Lozniewski A, Hartemann P and Mathieu L (2003) *Mycobacterium xenopi* and drinking water biofilms. *Appl Environ Microbiol* 69:6946–6948.

Debruyn JM, Mead TJ, Wilhelm SW and Sayler GS (2009) PAH biodegradative genotypes in Lake Erie sediments: evidence for broad geographical distribution of pyrene-degrading mycobacteria. *Environ Sci Technol* 43:3467–3473.

du Moulin G, Stottmeier K, Pelletier P, Tsang A and Hedley-Whyte J (1988) Concentration of *Mycobacterium avium* by hospital hot water systems. *JAMA* 260:1599–1601.

Eiler A, Langenheder S, Bertisson S and Tranvik LJ (2003) Heterotrophic bacterial growth efficiency and community structure at different natural organic carbon

concentrations. *Appl Environ Microbiol* 69:3701–3709.

Elser JJ, Stabler LB and Hassett RP (1995) Nutrient limitation of bacterial growth and rates of bacterivory in lakes and oceans: a comparative study. *Aquat Microb Ecol* 9:105–110.

Emmenegger L, Sigg L and Sulzberger B (2001) Light-induced redox cycling of iron in circumneutral lakes. *Limnol Oceanogr* 46:49–61.

Falkinham III JO (1996) Epidemiology of infection by nontuberculous mycobacteria. *Clin Microbiol Rev* 9:177–215.

Falkinham III JO (2002) Nontuberculous mycobacteria in the environment. *Clin Chest Med* 23:529–551.

Falkinham III JO, Norton CD and LeChevalier MW (2001) Factors influencing numbers of *Mycobacterium avium, Mycobacterium intracellulare*, and other mycobacteria in drinking water distribution systems. *Appl Environ Microbiol* 67:1225–1231.

Fox GE, Wisotzkey JD and Jurtshuk P (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* 42:166–170.

Fujita J, Nanki N, Negayama K, Tsutsui S, Taminato T and Ishida T (2002) Nosocomial contamination by *Mycobacterium gordonae* in hospital water supply and super-oxidized water. *J Hosp Infect* 51:65–68.

Gauthier DT, Reece KS, Xiao J, Rhodes MW, Kator HI, Latour RJ, Bonzek CF, Hoenig JM and Vogelbein WK (2010) Quantitative PCR assay for *Mycobacterium pseudoshottsii* and *Mycobacterium shottsii* and application to environmental samples and fishes from the Chesapeake Bay. *Appl* 76:6171–6179.

Guo F and Zhang T (2013) Biases during DNA extraction of activated sludge samples revealed by high throughput sequencing. *Appl Microbiol Biotechnol* 97:4607–4616.

Hanson CA, Fuhrman JA, Horner-Devine MC and Martiny JBH (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* 10:497–506.

Hothorn T, Bretz F, Westfall P, Heiberger RM, Schuetzenmeister A and Scheibe S (2015) Package "multcomp". *Simultaneous Inference in General Parametric Models*, version 1.4-0.

Hruska K and Kaevska M (2012) Mycobacteria in water, soil, plants and air: a review. *Vet Med (Praha)* 57:623–679.

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

Iivanainen EK, Martikainen PJ, Väänänen PK and Katila ML (1993) Environmental factors affecting the occurrence of mycobacteria in brook waters. *Appl Environ Microbiol* 59:398–404.

Jacobs J, Rhodes M, Sturgis B and Wood B (2009) Influence of environmental gradients on the abundance and distribution of *Mycobacterium spp*. in a coastal lagoon estuary. *Appl Environ Microbiol* 75:7378–7384.

Josephon KL, Gerba CP and Pepper IL (1993) Polymerase chain reaction detection of nonviable. *Appl Environ Microbiol* 59:3513–3515.

Khera TT (2012) *The diversity and distribution of* Mycobacterium *species in varying ecological and climatic environments*. Doctoral dissertation, University of Warwick.

Kim YE, Yoon H, Kim M, Nam YJ, Kim H, Seo Y, Lee GM, Ja Kim Y, Kong WS, Kim JG and Seu YB (2014) Metagenomic analysis of bacterial communities on Dokdo Island. *J Gen Appl Microbiol* 60:65–74.

Kirschner RA, Parker BC and Falkinham III JO (1999) Humic and fulvic acids stimulate the growth of *Mycobacterium avium*. *FEMS Microbiol Ecol* 30:327–332.

Kwon S, Kim TS, Yu GH, Jung JH and Park HD (2010) Bacterial community composition and diversity of a full-scale integrated fixed-film activated sludge system as investigated by pyrosequencing. *J Microbiol Biotechnol* 20:1717–1723.

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.

Leys NM, Ryngaert A, Bastiaens L, Wattiau P, Top EM, Verstraete W and Springael D (2005) Occurrence and community composition of fast-growing Mycobacterium in soils contaminated with polycyclic aromatic hydrocarbons. *FEMS Microbiol Ecol* 51:375–388.

Liu R, Yu Z, Guo H, Liu M, Zhang H and Yang M (2012) Pyrosequencing analysis of eukaryotic and bacterial communities in faucet biofilms. *Sci Total Environ* 435-436:124–131.

Lührig K, Canbäck B, Paul CJ, Johansson T, Persson KM and Rådström P (2015) Bacterial community analysis of drinking water biofilms in Southern Sweden. *Microbes Environ* 30:99–107.

Marras TK and Daley CL (2002) Epidemiology of human pulmonary infection with nontuberculous mycobacteria. *Clin Chest Med* 23:553–567.

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.

Masters CI, Shallcross JA and Mackey BM (1994) Effect of stress treatments on the detection of *Listeria monocytogenes* and enterotoxigenic *Escherichia coli* by the polymerase chain reaction. *J Appl Bacteriol* 77:73–79.

Mckay RML, Bullerjahn GS, Porta D, Brown ET, Sherrell RM, Smutka TM, Sterner RW, Twiss MR and Wilhelm SW (2004) Consideration of the bioavailability of iron in the North American Great Lakes: development of novel approaches toward understanding iron biogeochemistry. *Aquat Ecosyst Heal Manag* 7:475–490.

Mrlik V, Slany M, Kubecka J, Seda J, Necas a, Babak V, Slana I, Kriz P and Pavlik I (2012) A low prevalence of mycobacteria in freshwater fish from water reservoirs, ponds and farms. *J Fish Dis* 35:497–504.

Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, Herold N, Schöning I, Schrumpf M and Daniel R (2011) Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLoS One* 6:e17000.

Newton RJ, Jones SE, Eiler A, McMahon KD and Bertilsson S (2011) A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* 75:14–49.

Niva M, Hernesmaa A, Haahtela K, Salkinoja-Salonen M, Sivonen K and Haukka K (2006) Actinobacterial communities of boreal forest soil and lake water are rich in mycobacteria. *Boreal Environ Res* 11:45–53.

Norby B, Fosgate GT, Manning EJB, Collins MT and Roussel AJ (2007) Environmental mycobacteria in soil and water on beef ranches: association between presence of cultivable mycobacteria and soil and water physicochemical characteristics. *Vet Microbiol* 124:153–159.

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 U S A* 107:15345–15350.

Pickup RW, Rhodes G, Arnott S, Bull TJ, Weightman A, Hurley M and Hermon-Taylor J (2005) *Mycobacterium avium* subsp. *paratuberculosis* in the catchment area and water of the river taff in South Wales, United Kingdom, and its potential relationship to clustering of Crohn' s disease cases in the city of Cardiff. *Appl Environ Microbiol* 71:2130–2139.

Pickup RW, Rhodes G, Bull TJ, Arnott S, Sidi-Boumedine K, Hurley M and Hermon-Taylor J (2006) *Mycobacterium avium* subsp. *paratuberculosis* in lake catchments, in river water abstracted for domestic use, and in effluent from domestic sewage treatment

works: diverse opportunities for environmental cycling and human exposure. *Appl Environ Microbiol* 72:4067–4077.

Pinheiro J, Bates D, DebRoy S and Sarkar D (2015) Package "nlme". *Fit and compare Gaussian linear and nonlinear mixed-effects models*, version 3.1-120.

Pontiroli A, Khera TT, Oakley BB, Mason S, Dowd SE, Travis ER, Erenso G, Aseffa A, Courtenay O and Wellington EMH (2013) Prospecting environmental mycobacteria: combined molecular approaches reveal unprecedented diversity. *PLoS One* 8:e68648.

R Development Core Team (2014) *R: A Language and Environment for Statistical Computing*. Vienna, Autria: R Foundation for Statistical Computing.

Radomski N (2011) *Sources of nontuberculous mycobacteria in watersheds*. Doctoral dissertation, Université Paris-Est École des Ponts ParisTech.

Radomski N, Cambau E, Moulin L, Haenn S, Moilleron R and Lucas FS (2010a) Comparison of culture methods for isolation of nontuberculous mycobacteria from surface waters. *Appl Environ Microbiol* 76:3514–3520.

Radomski N, Lucas FS, Moilleron R, Cambau E, Haenn S and Moulin L (2010b) Development of a real-time qPCR method for detection and enumeration of Mycobacterium spp. in surface water. *Appl Environ Microbiol* 76:7348–7351.

Radomski N, Roguet A, Lucas FS, Veyrier FJ, Cambau E, Accrombessi H, Moilleron R, Behr MA and Moulin L (2013) *atpE* gene as a new useful specific molecular target to quantify *Mycobacterium* in environmental samples. *BMC Microbiol* 13:277.

Rahbar M, Lamei A, Babazadeh H and Yavari SA (2010) Isolation of rapid growing mycobacteria from soil and water in Iran. *African J Biotechnol* 9:3618–3621.

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.

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.

Stinear T, Davies JK, Jenkin GA, Hayman JA, Oppedisano F and Johnson PDR (2000) Identification of *Mycobacterium ulcerans* in the environment from regions in Southeast Australia in which it is endemic with sequence capture-PCR. *Appl Environ Microbiol* 66:3206–3213.

Unno T, Kim J, Kim Y, Nguyen SG, Guevarra RB, Kim GP, Lee JH and Sadowsky MJ (2015) Influence of seawater intrusion on microbial communities in groundwater. *Sci Total Environ* 532:337–343.

Uyttebroek M, Breugelmans P, Janssen M, Wattiau P, Joffe B, Karlson U, Ortega-Calvo JJ, Bastiaens L, Ryngaert A, Hausner M and Springael D (2006) Distribution of the *Mycobacterium* community and polycyclic aromatic hydrocarbons (PAHs) among different size fractions of a long-term PAH-contaminated soil. *Environ Microbiol* 8:836–847.

Uyttebroek M, Spoden A, Ortega-Calvo JJ, Wouters K, Wattiau P, Bastiaens L and Springael D (2007) Differential responses of eubacterial, *Mycobacterium*, and *Sphingomonas* communities in polycyclic aromatic hydrocarbon (PAH)-contaminated soil to artificially induced changes in PAH profile. *Appl Environ Microbiol* 36:1403–1411.

Vaerewijck MJM, Huys G, Palomino JC, Swings J and Portaels F (2005) Mycobacteria in drinking water distribution systems: ecology and significance for human health. *FEMS Microbiol Rev* 29:911–934.

Vellend M (2010) Conceptual synthesis in community ecology. *Q Rev Biol* 85:183–206.

Viallier J and Viallier G (1973) Inventaire des mycobacteries de la nature. *Ann Soc Belg Med Trop (1920)* 53:361–371.

Williams MM, Yakrus MA, Arduino MJ, Cooksey RC, Crane CB, Banerjee SN, Hilborn ED and Donlan RM (2009) Structural analysis of biofilm formation by rapidly and slowly growing nontuberculous mycobacteria. *Appl Environ Microbiol* 75:2091–2098.

Wurtzer S, Prevost B, Lucas FS and Moulin L (2014) Detection of enterovirus in environmental waters: a new optimized method compared to commercial real-time RT-qPCR kits. *J Virol Methods* 209:47–54.

Yang C, Hamel C and Gan Y (2015) Incongruous variation of denitrifying bacterial communities as soil N level rises in Canadian canola fields. *Appl Soil Ecol* 89:93–101.

Ye L and Zhang T (2013) Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. *Appl Microbiol Biotechnol* 97:2681–2690.

Zuur A, Ieno EN, Walker N, Saveliev AA and Smith GM (2009). Mixed effects models and extensions in ecology with R. *Springer Science and Business Media*.

VIII. Supplementary data

Supplementary data 1 Location, morphological and physicochemical characteristics of the 49 lakes

ID lake	Name	LUI	Trophic status	Hydro	SSO	Surface area (km^2)	Mean depth (m)	pH	Temp (°C)	Cond ($\mu S.cm^{-1}$)	DOC ($mgC.L^{-1}$)	TSS ($mg.L^{-1}$)
1	Freneuse Cove	N	H	C	+	0.14	2.4	7.8 (1.1)	23.3 (2.6)	430 (113)	22.7 (3.6)	4.5 (0.3)
2	Grand Lavacourt Pond	N	E	I	-	1.12	4.3	9 (0.2)	22.5 (2.7)	264 (7)	2.3 (0.4)	4.3 (0.4)
3	Ilon Cove	N	E	C	+	0.66	3.8	8.3 (0)	22.1 (2.7)	478 (11)	8.4 (1.3)	3.5 (0.1)
4	Les Galets Pond	OU	E	I	-	0.11	2.2	9.3 (0.2)	24.4 (3.3)	272 (22)	7.0 (3.1)	7.2 (0.2)
5	ASM Club Cove	N	H	C	+	0.37	2.5	8 (0)	21.7 (3)	491 (17)	20.3 (7.9)	3.9 (0.3)
6	Grosse Pierre Pond	N	H	I	+	0.46	4.3	8.9 (0.5)	22.3 (1.8)	421 (65)	8.9 (6.5)	7.3 (0.5)
7	Bout du Monde Pond	N	H	I	-	0.20	1.2	8.4 (0.1)	22.8 (3.1)	776 (28)	10.2 (5.4)	6.3 (0.1)
8	Enghien Lake	U	H	Cr	+	0.34	1.0	7.8 (0.9)	23.3 (1.2)	689 (186)	36.4 (13.2)	6.3 (1.4)
9	Gaule Achéroise Pond	A	M	I	-	0.05	6.2	7.5 (0.7)	24 (1.4)	769 (16)	7.1 (0.4)	3.4 (0.4)
10	Triel Pond	N	H	I	-	0.34	3.9	8.9 (0.2)	23.6 (2.7)	1031 (94)	8.9 (6.1)	11.1 (0.7)
11	Jabline-Anet Lake	OU	O	I	-	0.77	7.0	8.6 (0.1)	21.6 (1.7)	1173 (40)	1.3 (0.0)	4.2 (0.2)
12	Isles-les Villenoy Pond	A	M	I	-	0.42	2.6	8.6 (0.1)	23.4 (2.9)	324 (17)	5.0 (3.2)	4.6 (0.5)
13	Lake Inférieur	OU	E	Cr	+	0.11	1.2	8 (1.3)	24.8 (4.1)	312 (80)	10.2 (0.1)	4.3 (0.7)
14	Les Pâtis Pond	OU	H	I	-	0.10	2.6	8.8 (0.2)	22.2 (3.2)	615 (22)	11.9 (12.3)	6.8 (0.4)
15	Torcy Lake	OU	M	I	-	0.25	5.8	8.1 (0)	22.5 (3.2)	1185 (47)	2.5 (0.2)	5.2 (0.1)
16	UCPA Centre Pond	N	O	I	-	0.92	5.3	8.8 (0)	22.4 (2.8)	457 (3)	1.3 (0.5)	4.2 (0.3)
17	Loy Pond	N	H	Cr	+	0.07	1.3	7.8 (0.4)	22.2 (3.5)	476 (69)	58.9 (34.7)	6.3 (0.1)
18	Saint-Cucufa Pond	N	H	I	-	0.02	1.8	7.5 (0.6)	22.4 (3)	718 (4)	11.9 (3.3)	7.4 (0.8)
19	UTE Louveciennes	OU	O	I	-	0.10	2.5	7.3 (0.4)	20 (2.1)	672 (17)	1.3 (0.1)	1.6 (0.4)
20	Lake Minimes	N	H	Cr	+	0.06	1.2	7.7 (0.9)	24 (4)	511 (50)	39.3 (5.6)	5.1 (0)
21	Swiss Pond	OU	M	I	-	0.14	1.9	8.8 (1.1)	27.2 (1.6)	988 (9)	3.0 (1.3)	6.1 (0.7)
22	Saint-Quentin Reservoir	A	E	I	-	1.20	1.6	9.2 (0.3)	25.7 (1.5)	406 (6)	10.7 (0.4)	6 (0.3)
23	Créteil Lake	U	M	I	+	0.41	4.5	7.8 (0.4)	24 (1.8)	1454 (48)	2.6 (0.2)	6.9 (0.7)
24	Maurepas Reservoir	U	H	Cr	+	0.08	3.4	8.5 (0.4)	24.7 (3)	410 (94)	9.8 (7.9)	6 (1.2)
25	Choisy Pond	OU	M	I	-	0.34	2.6	7.4 (0.6)	25.1 (1.6)	1228 (119)	3.7 (0.6)	4.9 (0.8)
26	Plage Bleue Pond	OU	M	I	-	0.09	3.1	7.6 (0.8)	24.6 (1.6)	2168 (178)	3.0 (2.0)	5.9 (0.7)
27	Noés Reservoir	U	H	I	+	0.24	0.7	8.3 (0.1)	27.1 (1.1)	619 (70)	37.6 (4.6)	9.5 (4.6)
28	Saclay Reservoir	A	H	I	+	0.30	1.5	9.3 (0.2)	25.8 (1.5)	518 (47)	14 (14.4)	7.6 (1.5)
29	La Veyssière Pond	OU	H	C	+	0.25	1.6	8.3 (0.5)	25.2 (1.6)	305 (19)	14.9 (0.2)	4 (0.2)
30	Vaux-de-Cernay Pond	N	H	Cr	+	0.06	1.4	8.8 (0.2)	24.2 (1)	458 (15)	12.9 (1.6)	6 (0.6)
31	Noues de Seine Pond	U	E	C	+	0.97	3.4	8.9 (0.3)	25.8 (1.9)	758 (26)	2.6 (0)	4.7 (0.2)
32	Gazeran Pond	N	H	Cr	+	0.08	0.7	7.8 (0.2)	25.7 (0.6)	807 (46)	42.3 (35.4)	8 (0.2)
33	Courcouronnes Pond	U	H	I	+	0.07	1.4	9.1 (0.1)	27.7 (3.3)	298 (102)	22.5 (1.2)	5.2 (1.7)
34	Epinoche Pond	A	H	C	+	0.06	2.4	8.1 (0.4)	23.4 (1.6)	565 (34)	17 (0.4)	4.9 (0.6)
35	Port-Sud Pond	A	H	Cr	+	0.07	1.7	8.2 (0.4)	25.4 (1.3)	511 (70)	23.8 (16.6)	5.1 (2.1)
36	Fontenay/Vic Pond	A	H	C	+	0.14	1.2	8.7 (0.1)	25.9 (1.4)	395 (9)	26.6 (2.7)	4.3 (0.1)
37	Vert-le-Petit Pond	A	E	C	+	0.01	2.0	8 (0.1)	23.7 (3)	601 (22)	14.5 (2.8)	7.2 (0.5)
38	Seine Port Pond	N	E	C	+	0.06	2.9	8.1 (0.4)	22.6 (1)	263 (26)	6.0 (1.5)	3.6 (0.4)
39	Fleuri Pond	A	H	C	+	0.08	1.3	8 (0.3)	21.7 (1.3)	509 (43)	15.6 (0.4)	2.7 (0.5)
40	Villefermoy Reservoir	N	H	Cr	+	0.32	2.1	8.4 (0.3)	23.4 (3.6)	273 (42)	28.0 (0.8)	9.8 (0.1)
41	Saint-Blaise Pond	A	E	C	+	0.05	1.4	7.9 (0.1)	21.5 (2.1)	530 (29)	11.2 (1)	2.9 (0.3)
42	Bois-le-Roi pond	OU	E	I	-	0.08	2.6	8.2 (0.2)	24.6 (3.8)	300 (21)	11.9 (1.5)	4.2 (0)
43	Samoreau Pond	OU	O	I	-	0.10	2.4	9 (0.3)	25.5 (4.1)	200 (22)	2.1 (0.5)	2.6 (0.6)
44	Delomez Pond	N	O	I	-	0.07	4.2	8.3 (0.2)	23.8 (2.9)	220 (39)	2.8 (0.9)	3.4 (0.2)
45	Leclerc Pond	N	M	I	-	0.05	2.8	8 (0)	24 (3.2)	309 (15)	4.1 (1.3)	2.8 (0.6)
46	Grande Paroisse Pond	A	E	C	+	0.53	3.7	8.3 (0.1)	23.4 (2.4)	315 (20)	9.7 (2.4)	4.1 (0.2)
47	Grand Marais Pond	N	H	I	-	0.17	2.4	9.3 (0.6)	23.7 (2.9)	364 (1)	9.1 (5.1)	10.1 (0.6)
48	Clarette Pond	N	E	I	-	0.14	1.4	8.5 (0.4)	24.9 (4)	300 (29)	11.3 (3.4)	7.4 (0.2)
49	Souppes/Loing Pond	A	M	C	+	0.21	2.9	8.1 (0.1)	23.3 (2.8)	401 (9)	5.8 (1.6)	4.2 (1.4)

Values represent the means (standard deviations) of the three yearly campaigns.
Abbreviations: LUI, dominant watershed land use indice; Hydro, type of connection to the hydrological network; DOC, dissolved organic carbon concentration; SSO, presence of a storm sewer outlet; Temp, temperature; Cond, conductivity; N, natural; A, agricultural; OU, open urban; U, urban; O, oligotrophic; M, mesotrophic; E, eutrophic; H, hypereutrophic; I, isolated; C, connected; Cr, crossed.

Chapter VI

Synthesis and general conclusion

This study contributes to the comprehension of the factors and processes shaping the biogeography of the bacterial compartment in lacustrine habitats, which remains a major issue for microbial ecologists. One of the main contributions of this investigation was to consider in the same study the effect of environmental, spatial and stochastic processes at several spatial scales (regional and local) focusing on different taxonomic levels.

The manuscript comprises four scientific articles or manuscripts that present the main results. These papers are based on a consequent database that was explored in depth using multivariate statistical analysis. The database was generated with the measurement of physicochemical parameters and molecular analysis of the bacterial community diversity. The samples were collected during an intense field work, with three-year sampling of 49 lakes, two-year monitoring of Créteil Lake (every month and six rain events), and with the investigation within Créteil and Daumesnil lakes of the favorable habitats for the NTM.

In this last section, the main results will be discussed in a broader context in light of their contributions and limitations, as well as the perspective of this work. The processes that structure the bacterial community will be commented at different spatial scales, then at different bacterial taxonomic levels. Finally, the perspectives, and the contribution of this study to a better comprehension of the response of the shallow lakes facing anthropogenic pressures will be addressed in the context of global change impact on aquatic ecosystems.

I.1 Bacterial biogeography at different spatial scales

For both regional and local scales, the spatial distribution of the bacterial community was well predicted by the neutral community model (*Chapter II* and *III*). Knowing that these tendencies were observed for each of the 49 lakes campaigns (with both T-RFLP fingerprinting and next-generation sequencing approaches) and for 88% of the Créteil Lake campaigns, it gives us confidence in the conclusions drawn from this work. However, it would be worthwhile to evaluate the co-trends between these methods using co-inertia analysis (Dray *et al.* 2003). Thus at both spatial scales, results indicated that bacterial assembly seemed to be primarily driven by stochastic balance between

births, deaths and immigration (Hubbell 2001), suggesting that taxa dispersal was high enough to allow a random immigration among the 49 lakes and within Créteil Lake.

The identification at both regional and local spatial scales of the same forces structuring the bacterial assembly may potentially reflect that regional patterns result from the coupling of local processes with dispersal processes (Chave 2013). As a consequence the beta diversity that we observed among the different lakes would be explained by the spatio-temporal dynamic of the bacterial community within each single lake, linked together through dispersal limitation. Within Créteil Lake, neutral forces explained up to 91% of the spatial structure of the bacterial assembly. This high percentage could notably be promoted by the regular mixing of the entire water column, enhancing dispersal of taxa and avoiding any spatial discrimination along the two transects that we investigated (longitudinal and vertical). However, the temporal dynamic of the bacterial assemblage over the two-year monitoring was slightly explained by the neutral model (probably due to statistical consideration, c.f. *Chapter II*). Indeed, the bacterial community was mainly driven by a deterministic factor: the water temperature. This result is not surprising since water temperature is an integrative seasonal parameter that is often associated with microbial composition variations (Yannarell and Triplett 2004; Shade *et al.* 2007).

Nevertheless we cannot assume that the Créteil Lake is representative of the 48 other lakes that were considered in this project. Further studies focusing on shallow lakes are therefore needed since the survey of a single shallow lake is not sufficient to provide conclusive support regarding the forces responsible for the spatio-temporal dynamics of the bacterial assemblages observed at the regional scale. It would also be important to consider the temporal variation at the regional scale, which could not be really achieved with only three summer campaigns.

Furthermore, except for the temporal variations of the bacterial assembly within Créteil Lake, deterministic forces (i.e. environmental factors and spatial processes) were not identified to have a relevant impact on the bacterial beta diversity, which is surprising regarding the literature. Indeed, a recent review (Hanson *et al.* 2012) showed that environmental selection and dispersal processes were frequently observed as shaping microbial composition (in more than 90% and 50% of the aquatic studies respectively).

However most of the studies used for this review did not consider the stochastic processes in their evaluation of the processes shaping the aquatic bacterial communities (Hanson *et al.* 2012).

The low percentage of variation explained by environmental factors that we found in our study could either be due to real low contribution of the deterministic processes to the bacterial assembly with respect to stochastic forces, or it could also reveal, as mentioned in the 'discussion' of the *Chapter II,* the absence of relevant parameters in the statistical model (Lindström and Langenheder 2012). Indeed, among the environmental factors integrated in our analysis, we only considered abiotic parameters. Therefore, we only investigated bottom-up control neglecting the top-down control. Yet, biotic interactions could also play an important role in shaping bacterial communities, especially due to protist predation pressures or viral lysis (Jürgens and Matz 2002; Weinbauer and Rassoulzadegan 2004). Such influence is illustrated by the Figure 28, where biotic explanatory variables (phytoplankton and heterotrophic nanoflagellate densities) coupled with interactions with other factors explained more than half of the spatial variation of the bacterial structure in six bogs (Kent *et al.* 2007). Although biotic factors were not currently included in our analysis, this issue could be resolved regarding some of the biotic factors. Indeed during the sampling campaigns, samples of zooplankton and phytoplankton have been collected by our scientific collaborators (ESE, iEES Paris, MCAM). However, their analysis is still in progress and could not be included in the database that was used for this thesis.

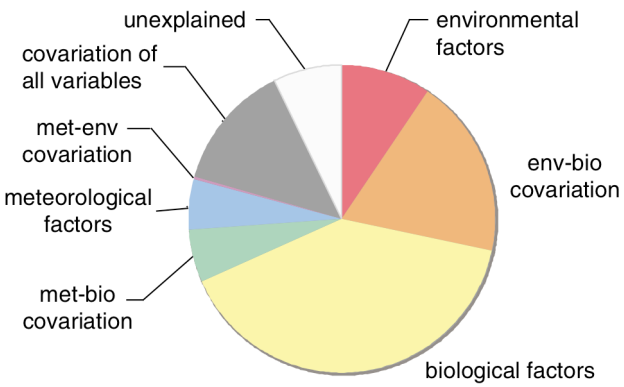


Figure 28 Partitioning microbial community variance among meteorological (met), environmental (env) and biological (bio) sets of variables for different lakes. Variance partitioning analysis results are averaged among six bogs. From Kent *et al.* 2007.

Furthermore, although historical processes may play a major role in shaping the bacterial communities, they were not evaluated in our study. Indeed, past events (even just a few days) may leave a contemporary signature in the observed patterns of spatial structure at the regional and local scale (Andersson *et al.* 2014). The incorporation in our statistical analysis of integrative parameters could potentially enhance the explanatory power of statistical models. For instance, we could use some parameters taking into accounts historical fluctuations prior sampling such as the cumulative rainfalls or water temperatures, or drastic changes of nutrients availability from one year to the other.

Moreover, it should be kept in mind that the sampling strategy that was used for the PULSE research project may have produced a bias that could influence the results of the biogeographical study. Indeed, the selection of the 49 lakes was initially performed to evaluate the regional distribution of cyanobacteria in a representative set of Paris area lakes (Catherine *et al.* 2008). These lakes were not selected to address the influence of local environmental, dispersal-related mechanisms and neutral processes on the bacterial assemblages. For some categorical variables related to dispersal processes the sampling design is unbalanced. For instance, a majority of the sampled lakes are isolated (n = 27), while few are connected to a river (n = 13), and even fewer are crossed by a river (n = 9). Similarly, a majority of the 49 lakes are not impacted by storm sewer effluents (n = 30). In spite of this drawback, this sampling design still allowed to extract conclusive tendencies regarding the relevant processes shaping the bacterial communities. However, in the prospective of improving our comprehension of relative importance of stochastic and deterministic factors for the bacterial assembly, it would be interesting if possible to design a totally balanced sampling strategy.

I.2 Biogeography patterns depending on the bacterial taxonomic level

One challenge of microbial geography is to determine at what taxonomic levels these forces operate. The choice of a target organism is often dictated by practical issues (Ramette and Tiedje 2007). Although a growing number of recent studies investigate the biogeography of the total bacterial community (see reviews of Martiny et al. 2006; Hanson et al. 2012, Lindström and Langenheder 2012), historically the species is the most commonly found unit in microbial biogeographic studies, probably as an inheritance of

botany and zoology. However it may be questionable to extrapolate results based on few selected taxa to the total community. Indeed, our study revealed large discrepancies in the forces responsible for the biogeography patterns depending on the bacterial taxonomic level or group that was considered.

At the regional scale, the influence of both deterministic forces (i.e. environmental factors and spatial processes) never exceeded 16% on the entire bacterial structure (*Chapter II* and *III*), whereas the same parameters explained up to 25% of the *Acidimicrobiia* subclass (*Actinobacteria*) and only 9% of the *Shingobacteriia* (*Bacteroidetes*) composition (*Chapter III*). This result is not surprising since distinct taxa may respond differently to evolutionary and ecological processes, even if recent supports highlighted the ecological convergences within close lineages (Philippot *et al.* 2010), especially when shared characters are complex (Martiny *et al.* 2013). These observations may explain why we noticed similar tendencies inside the dominant bacterial phyla or classes regarding the importance of the biogeographical processes. It would it be interesting to use the next-generation sequencing dataset (*Chapter III*) to investigate if some particular taxa deviated from a predictive random model (Šizling *et al.* 2009), and if it is the case, which ones?

Multivariate comparisons of microbial assemblages require decisions to be made about the taxonomic resolution to be used and whether the contributions of common, intermediate and rare taxa should be emphasized (Lasiak 2003). Rare biosphere is often described as containing most of the bacterial diversity and cosmopolitan taxa owing to their high persistence and dispersal (Pedrós-Alio 2006; Caron and Countway 2009). This so-called 'seed bank' could potentially represents a source of taxa waiting for favorable environmental conditions (Simon *et al.* 2015). Exchanges between rare to dominant taxa could explain part of the results of Galand and colleagues (2009). These authors found similar composition between dominant and rare members of bacterial communities among sites in the Arctic Ocean, implying that these communities were shaped by equivalent biogeographical processes (Galand *et al.* 2009). In our study, we noticed that the composition of rare and dominant taxa in the bacterial communities was shaped by similar factors but not with the same relative importance (*Chapter III*). Indeed, in the 49

lakes, deterministic factors explained 19% of the composition of dominant biosphere, while together with DOC they explained only 9% of the rare biosphere. Same conclusions were drawn with the influence of neutral processes that well predicted the taxa distribution variations for both biospheres, but up to 68% for dominant taxa against 73% for rare members.

The advent of high-throughput technologies has radically changed the microbial biogeography. Diversity and dynamics of rare taxa in microbial communities were rarely explored before the next-generation sequencing (Crump *et al.* 2012). Low resolutive methodologies (fingerprinting) could potentially not detect finer biogeographical patterns since these methods revealed only a small part of the diversity (Ramette and Tiedje 2007). In our case, this could explain the slight but higher importance of both deterministic forces on the total bacterial assembly structure when assessed with high-throughput sequencing compared to the T-RFLP approach (*Chapter II* and *III*). Although high-throughput sequencing using universal bacterial ribosomal RNA gene probes offers the opportunity to detect rare taxa, this methodology is still inefficient to properly assign sequences at taxonomic levels below the genus level (Hanson *et al.* 2012). Indeed, in our study mycobacteria were retrieved in only 71% of the 49 lakes, with a relative abundance higher than 5% in one lake (Freneuse Cove), whereas the NTM population was quantified in all these lakes by real-time quantitative PCR (*Chapter III* and *V*). Furthermore, as suggested by Hanson and colleagues (2012), the likelihood to detect distance effect will increase with the increasing resolution of the diversity analysis. This assumption was illustrated by Martiny and colleagues (2009) that observed in marine environment, that the *Prochlorococcus* community composition assessed by pyrosequencing correlated with the dispersal rate only at the highest taxonomic resolution (from a cutoff of 99%). Besides, in our study, although it is difficult to compare the densities and the bacterial community structure together (notably since densities do not reflect ecological processes such as selection, dispersal or drift), we observed at the regional scale a difference in the total variance explained by the deterministic forces that was of 38% for the *Mycobacterium* genus densities and lesser for broader taxonomic levels (total bacteria, dominant phyla and classes) (*Chapter II, III* and *V*).

The focus on mycobacteria gave the opportunity to investigate the impact of environmental and spatial factors at finer taxonomic level (genus). Although we were able to show variations among lakes and within lakes of the mycobacterial densities and diversity (*Chapter IV* and *V*), the methods we used did not allow a fine biogeography study (fragment of the 16S rRNA gene sequence and real-time quantitative PCR at the genus level). It would be interesting to pursue the study of the mycobacterial biogeography using gene targets that would be less conservative than the 16S rRNA gene, such as the hsp65 gene (van der Wielen *et al.* 2013). It would also be interesting to investigate the biogeography of mycobacterial community at both local (over the two year monitoring of Créteil Lake) and regional scales using taxonomic diversity tools.

I.3 Impact of environmental pressures on the bacterial biogeography

The biogeography structure of bacterial communities may be disturbed by anthropogenic factors, which could have important impact on the functioning of the aquatic ecosystems and on their ecosystem services. For instance, many pathogens display well-defined biogeographical distributions that may be modified when their dispersal limitation is modified by human wastewater management or human population displacement or other activities.

In our study, pressures resulting from anthropogenic activities significantly impacted each of the taxonomic levels investigated. Among these pressures, the trophic status and the connection of the lake to a river were almost always identified as shaping the bacterial assemblages (*Chapter II, III* and *V*). Regarding the connection of lakes to river, it would have been interesting in collect samples upstream the river-lake connection to compare the similarity between allochthonous and autochthonous communities. Such approach would allow to evaluate the importance of mass effects on local communities (Nelson et al. 2009). The trophic status is related to autotrophic production and may reflect pressures related to the input of allochthonous nutrients resulting from the agricultural, industrial and domestic human activities. The connection of the lake to the hydrographic network reflects pressures resulting from watershed managements and water policies. Discharges of punctual storm sewer effluents into the lakes were not observed to affect the bacterial community structure, except for the mycobacterial

densities that appeared to be locally and temporally affected by the runoffs of the Créteil Lake storm sewer (*Chapter V*). It would be interesting to assess the bacterial community structure of the samples collected during the 6 rain events. For instance, using the T-RFLP method, we could verify the sewer effluents contribute to the input and establishment of allochthonous taxa in the bacterial community of the lake water column. In order to check whether the mycobacteria discharged by the storm sewer come from road runoffs or from the sewer itself (pipe biofilms and retention chamber), it would have been interesting to sample the runoff directly from the streets and inside the sewer.

The dominant land use present on the watershed was not observed as governing the bacterial communities. This result appears surprising, because Scopel and colleagues (2006), McLellan and colleagues (2007) and Lear and Lewis (2009) found that bacterial river biofilms were directly driven by the watershed land use. However, in the context of these 49 artificial lakes, watersheds were not representative of the proximal environment that surrounds the lakes. Indeed, for most of the lakes, their watersheds are drained resulting in a low contribution of the watershed runoff to the lake. Consequently, only a small stripe around lakes may impact the autochthonous communities.

Although bacterial biogeographic variations have predominantly been studied focusing on taxonomic diversity, a vivid interest is resurging regarding the biogeographical patterns of functional traits among the bacterial communities, especially to better understand how microorganisms will respond to environmental pressures (Green *et al.* 2008). Biogeochemical processes that are primarily driven by microbial activities constitute an important part of the ecosystem services offered by lacustrine ecosystems in terms of regulation of water quality, regulation of pathogens and regulation of the climate (Downing 2010). It would be interesting to characterize whether differences in trophic status or the fact those lakes are connected to the river system have an impact on the functional traits of the bacterial community, despite the functional redundancy that is often observed in bacterial assemblages (Langenheder *et al.* 2006). Bacterial functional redundancy is probably due to the rare biosphere that encloses metabolically similar populations that are able to contend with changing environmental conditions (Lennon and Jones 2011). The functional diversity of a bacterial community could be indirectly assessed

using microplates supplemented with different substrates such as EcoPlates (Biolog) or Microresp™ (Preston-Mafham *et al.* 2002; Campbell *et al.* 2003). It could also be directly assessed by transcriptomic approaches. As part of this study, I unfortunately only touched upon this issue by predicting the functional profiles from our NGS database using the 'Tax4Fun' package in R project (Aßhauer *et al.* 2015). Nevertheless I gave up on this idea due to the difficulty in analyzing the data obtained from the Tax4Fun model that provides information for more than 6,400 enzymatic reactions. However, there are other models that may be easier to interpret such as PICRUSt (Langille *et al.* 2013) or consenTRAIT (Martiny *et al.* 2013).

References

Abdallah M, van Pittius NCG, DiGiuseppe Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CMJE, Appelmelk BJ and Bitter W (2007) Type VII secretion—mycobacteria show the way. Nat Rev Microbiol 5:883–891.

Adams HE, Crump BC and Kling GW (2014) Metacommunity dynamics of bacteria in an arctic lake: the impact of species sorting and mass effects on bacterial production and biogeography. *Front Microbiol* 5:1–10.

Adékambi T and Drancourt M (2004) Dissection of phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, *hsp65*, *sodA*, *recA* and *rpoB* gene sequencing. *Int J Syst Evol Microbiol* 54:2095–2105.

Adékambi T, Ben Salah S, Khlif M, Raoult D and Drancourt M (2006) Survival of environmental mycobacteria in *Acanthamoeba polyphaga*. *Appl Environ Microbiol* 72:5974–5981.

Alavi MR, Shukla HD, Whitaker B, Arnold J and Shahamat M (2006) Attachment and biofilm formation of *Mycobacterium marinum* on a hydrophobic surface at the air interface. *World J Microbiol Biotechnol* 23:93–101.

Allison SD and Martiny JBH (2008) Colloquium paper: resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci USA* 105(Suppl 1):11512–11519.

Amann RI, Ludwig W and Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59:143–169.

Anderson RJ (1943) The chemistry of the lipids of the tubercle bacillus. *Yale J Biol Med 15*:311–345.

Anderson-Glenna MJ, Bakkestuen V and Clipson NJW (2008) Spatial and temporal variability in epilithic biofilm bacterial communities along an upland river gradient. *FEMS Microbiol Ecol* 64:407–418.

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

Arya M, Shergill IS, Williamson M, Gommersall L, Arya N and Patel HRH (2005) Basic principles of real-time quantitative PCR. *Mol Biotechnol* 15:209–219.

Aßhauer KP, Wemheuer B, Daniel R and Meinicke P (2015) Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinforma Adv* btv287.

Azam F, Fenchel T, Field JG, Gray JS, Meyer-Reil LA and Thingstad F (1983) The ecological role of water-column microbes in the sea. *Mar Ecol* 10:257–263.

Baas Becking LGM (1934) *Geobiologie of Inleiding Tot de Milieukunde*. Van Stockkum and Zoon, The Hague.

Bai S and Lung WS (2005) Modeling sediment impact on the transport of fecal bacteria. *Water Res* 39:5232–5240.

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

Beerwerth W (1973) Über die verwertbarkeit natürlicher substrate als nahrungsquellen der mycobakterien. *Ann Soc Belg Med Trop* 54:355–360.

Beijerinck MW (1913) *De infusies en de ontdekking der backteriën*. In Jaarboek van de Koninklijke Akademie van Wetenschappen. Müller, Amsterdam.

Bell T, Newman JA, Silverman BW, Turner SL and Lilley AK (2005) The contribution of species richness and composition to bacterial services. *Nature* 436:1157–1160.

Bercovier H, Kafri O and Sela S (1986) Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. *Biochem Biophys Res Commun* 136:1136–1141.

Besemer K, Singer G, Hödl I and Battin TJ (2009) Bacterial community composition of stream biofilms in spatially variable-flow environments. *Appl Environ Microbiol* 75:7189–7195.

Biet F, Boschiroli ML, Thorel MF and Guilloteau LA (2005) Zoonotic aspects of *Mycobacterium bovis* and *Mycobacterium avium-intracellulare* complex (MAC). *Vet Res* 36:411–436.

Biggs J, Williams P, Whitfield M, Nicolet P and Weatherby A (2005) 15 years of pond assessment in Britain: results and lessons learned from the work of pond conservation. *Aquat Conserv Mar Freshw Ecosyst* 15:693–714.

Bionde M., Marranzano M, Allegra A, di Fazio, Fraro G and Ferrante M (1982) Micobatteri atipici isolati dal suolo e dalle acque. *Annali Sclavo* 24:496–502.

Blackwell V (1999) *Mycobacterium marinum* infections. *Curr Opin Infect Dis* 12:181–184.

Bland CS, Ireland JM, Lozano E, Alvarez ME and Primm TP (2005) Mycobacterial ecology of the Rio Grande. *Appl Environ Microbiol* 71:5719–5727.

Bolund P and Hunhammar S (1999) Ecosystem services in urban areas. *Ecol Econ* 29:293–301.

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, Avois-Jacquet C and Tuomisto H (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85:1826–1832.

Boucher D, Jardillier L, Debroas D and Pascal B (2006) Succession of bacterial community composition over two consecutive years in two aquatic systems: a natural lake and a lake-reservoir. *FEMS Microbiol Ecol* 55:79–97.

Brezna B, Khan AA and Cerniglia CE (2003) Molecular characterization of dioxygenases from polycyclic aromatic hydrocarbon-degrading *Mycobacterium* spp.. *FEMS Microbiol Lett* 223:177–183.

Brookes JD, Antenucci J, Hipsey M, Burch MD, Ashbolt NJ and Ferguson C (2004) Fate and transport of pathogens in lakes and reservoirs. *Environ Int* 30:741–759.

Burke C, Steinberg P, Rusch D, Kjelleberg S and Thomas T (2011) Bacterial community assembly based on functional genes rather than species. *Proc Natl Acad Sci USA* 108:14288–14293.

Campbell M, Howard K, Le K, Shriver J and Wan L (2013) Linking land use to inland lake ecosystem service values. p. 32.

Carbonne A, Brossier F, Arnaud I, Bougmiza I, Caumes E, Meningaud JP, Dubrou S, Jarlier V, Cambau E and Astagneau P (2009) Outbreak of nontuberculous mycobacterial subcutaneous infections related to multiple mesotherapy injections. *J Clin Microbiol* 47:1961–1964.

Caron DA and Countway PD (2009) Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquat Microb Ecol* 57:227–238.

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.

Centre National de Référence des Mycobactéries et de la Résistance des Mycobactéries aux Antituberculeux (CNR-MyRMA) (2014*) Rapport d'activité pour l'année 2013*. http://cnrmyctb.free.fr/IMG/pdf/Rapport_CNR_2013_web.pdf.

Chai N, Deforges L, Sougakoff W, Truffot-Pernot C, De Luze A, Demeneix B, Clément M and Bomsel MC (2006) *Mycobacterium szulgai* infection in a captive population of African clawed frogs (*Xenopus tropicalis*). *J Zoo Wildl Med* 37:55–58.

Chapin III FS, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM, Reynolds HL, Hooper DU, Lavorel S, Sala OE, Hobbie SE, Mack MC and Díaz S (2000) Consequences of changing biodiversity. *Nature* 405:234–242.

Chapman BR and Bolen EG (2015) *Ecology of North America*. 2nd edition. Wiley Blackwell.

Chave J (2013) The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecol Lett* 16:4–16.

Chróst RJ, Adamczewki T, Kalinowska K and Skowrońska A (2009) Abundance and structure of microbial loop components (bacteria and protists) in lakes of different

trophic status. *J Microbiol Biotechnol* 19:858–868.

Cirillo JD, Falkow S, Tompkins LS and Bermudez LE (1997) Interaction of *Mycobacterium avium* with environmental amoebae enhances virulence. *Infect Immun* 65:3759–3767.

Cooke GD, Welch EB, Peterson S and Nichols SA (2005) *Restoration and management of lakes and reservoirs*, CRC press, p. 28.

Covert TC, Rodgers MR, Reyes AL and Stelma GN (1999) Occurrence of nontuberculous mycobacteria in environmental samples. *Appl Environ Microbiol* 65:2492–2496.

Crump BC, Adams HE, Hobbie JE and Kling GW (2007) Biogeography of bacterioplankton in lakes and streams of an artic tundra catchment. *Ecology* 88:1365–1378.

Crump BC, Kling GW, Bahr M and Hobbie JE (2003) Bacterioplankton community shifts in an arctic lake correlate with seasonal changes in organic matter source. *Appl Environ Microbiol* 69:2253–2268.

Daffé M (1996) Structure de l'enveloppe de *Mycobacterium tuberculosis*. *Méd Mal Infect* 26:891–897.

Dailloux M, Laurain C, Weber M and Hartemann P (1999) Water and nontuberculous mycobacteria. *Water Res* 33:2219–2228.

De Haan H (1993) Solar UV-light penetration and photodegradation of humic substances in peaty lake water. *Limnol Oceanogr* 38:1072–1076.

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.

Debruyn JM, Mead TJ, Wilhelm SW and Sayler GS (2009) PAH biodegradative genotypes in Lake Erie sediments: evidence for broad geographical distribution of pyrene-degrading mycobacteria. *Environ Sci Technol* 43:3467–3473.

Dejean T, Valentini A, Duparc A, Pellier-Cuit S, Pompanon F, Taberlet P and Miaud C (2011) Persistence of environmental DNA in freshwater ecosystems. *PLoS One* 6:8–11.

Delafont V (2015) Diversité et implication des amibes libres dans la survie et la persistance des mycobactéries non tuberculeuses au sein d'un réseau d'eau potable. Doctoral dissertation, Université de Poitiers.

Delafont V, Mougari F, Cambau E, Joyeux M, Bouchon D, Yann H and Moulin L (2014) First evidence of Amoebae−Mycobacteria association in drinking water network. *Environ Sci Technol* 48:11872–11882.

Derz K. Klinner U, Schuphan I, Stackebrandt E and Kroppenstedt RM (2004) *Mycobacterium pyrenivorans* sp. nov., a novel polycyclic-aromatic-hydrocarbon-

degrading species. *Int J Syst Evol Microbiol* 54:2313–2317.

Dodds WK and Whiles MR (2010). *Freshwater ecology: concepts and environmental applications*. Academic press. p. 160.

Dodds WK, Bouska WW, Eitzmann JL, Pilger TJ, Pitts KL, Riley AJ, Schloesser JT and Thornbrugh DJ (2009) Eutrophication of U.S. freshwaters: analysis of potential economic damages. *Environ Sci Technol* 43:12–19.

Domaizon I, Viboud S and Fontvieille D (2003) Taxon-specific and seasonal variations in flagellates grazing on heterotrophic bacteria in the oligotrophic Lake Annecy – Importance of mixotrophy. *FEMS Microbiol Ecol* 46:317–329.

Dorigo U, Volatier L and Humbert JF (2005) Molecular approaches to the assessment of biodiversity in aquatic microbial communities. *Water Res* 39:2207–2218.

Douterelo I, Boxall JB, Deines P, Sekar R, Fish KE and Biggs CA (2014) Methodological approaches for studying the microbial ecology of drinking water distribution systems. *Water Res* 65:134–156.

Downing JA (2010) Emerging global role of small lakes and ponds: little things mean a lot. *Limnetica* 29:9–24.

Downing JA, Prairie YT, Cole JJ, Duarte CM, Tranvik LJ, Striegl RG, McDowell WH, Kortelainen P, Caraco NF, Melack JM and Middelburg JJ (2006) The global abundance and size distribution of lakes, ponds and impoundments. *Limnol Oceanogr* 51:2388–2397.

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.

Dray S, Chessel D and Thioulouse J (2003) Co-inertia analysis and the linking of ecological data tables. *Ecology* 84:3078–3089.

Drewe JA, Mwangi D, Donoghue HD and Cromie RL (2009) PCR analysis of the presence and location of *Mycobacterium avium* in a constructed reed bed, with implications for avian tuberculosis control. *FEMS Microbiol Ecol* 67:320–328.

du Moulin G, Stottmeier K, Pelletier P, Tsang A and Hedley-Whyte J (1988) Concentration of *Mycobacterium avium* by hospital hot water systems. *JAMA* 260:1599–1601.

Eaton T, Falkinham III JO, Aisu TO and Daniel TM (1995) Isolation and characteristics of *Mycobacterium avium* complex from water and soil samples in Uganda. *Tuber Lung Dis* 76:570–574.

Eddyani M, de Jonckheere JF, Durnez L, Suykerbuyk P, Leirs H and Portaels F (2008) Occurrence of free-living amoebae in communities of low and high endemicity for Buruli Ulcer in Southern Benin. *Appl Environ Microbiol* 74:6547–6553.

Eiler A, Langenheder S, Bertisson S and Tranvik LJ (2003) Heterotrophic bacterial growth efficiency and community structure at different natural organic carbon concentrations. *Appl Environ Microbiol* 69:3701–3709.

Eun-Sook L, Lee MY, Han SH and Ka JO (2008) Occurrence and molecular differentiation of environmental mycobacteria in surface waters. *J Microbiol Biotechnol* 18:1207–1215.

Euzéby JP (1997). List of bacterial names standing in nomenclature: a folder available on the internet. *Int J Syst Bacteriol* 47:590–592. (Consulted on September 2015, URL: http://www.bacterio.net/).

Fakruddin M and Mannan KS Bin (2013) Methods for analyzing diversity of microbial communities in natural environments. *Ceylon J Sci (Biological Sci* 42:19–33.

Falkinham III JO (1996) Epidemiology of infection by nontuberculous mycobacteria. *Clin Microbiol Rev* 9:177–215.

Falkinham III JO (2002) Nontuberculous mycobacteria in the environment. *Clin Chest Med* 23:529–551.

Falkinham III JO, Norton CD and LeChevalier MW (2001) Factors influencing numbers of *Mycobacterium avium*, *Mycobacterium intracellulare*, and other mycobacteria in drinking water distribution systems. *Appl Environ Microbiol* 67:1225–1231.

Falkinham III JO, Parker BC and Gruff H (1980) Epidemiology of infection by nontuberculous mycobacteria. *Am Rev Respir Dis* 121: 931–937.

Fierer N, Morse JL, Berthrong ST, Bernhardt ES and Jackson RB (2007) Environmental controls on the landscape-scale biogeography of stream bacterial communities. *Ecology* 88:2162–2173.

Fittipaldi M, Nocker A and Codony F (2012) Progress in understanding preferential detection of live cells using viability dyes in combination with DNA amplification. *J Microbiol Methods* 91:276–289.

Fox GE, Wisotzkey JD and Jurtshuk P (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* 42:166–170.

Fujita J, Nanki N, Negayama K, Tsutsui S, Taminato T and Ishida T (2002) Nosocomial contamination by Mycobacterium gordonae in hospital water supply and super-oxidized water. *J Hosp Infect* 51:65–68.

Galand PE, Casamayor EO, Kirchman DL and Lovejoy C (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci USA* 106:22427–22432.

Garrity GM, Bell JA and Lilburn TG (2004). *Taxonomic outline of the prokaryotes*. *Bergey's manual of systematic bacteriology*. Springer, New York, Berlin, Heidelberg.

George KL, Parker BC, Gruft G and Falkinham III JO (1980) Epidemiology of infection by nontuberculous mycobacteria. Growth and survival in natural waters. *Am Rev Respir Dis* 122:89–94.

Goddard V, Baker A, Davy J, Adams D, de Ville M, Thackeray S, Maberly S and Wilson W (2005) Temporal distribution of viruses, bacteria and phytoplankton throughout the water column in a freshwater hypereutrophic lake. *Aquat Microb Ecol* 39:211–223.

Gomila M, Ramirez A and Lalucat J (2007) Diversity of environmental *Mycobacterium* isolates from hemodialysis water as shown by a multigene sequencing approach. *Appl Environ Microbiol* 73:3787–3797.

Goslee S and Wolinsky E (1976) Water as a source of potentially pathogenic mycobacteria. *Am Rev Respir Dis* 113:287–291.

Green JL, Bohannan BJM and Whitaker RJ (2008) Microbial biogeography: from taxonomy to traits. *Science* 320:1039–1043.

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M, Gillings M and Beattie AJ (2004) Spatial scaling of microbial eukaryote diversity. *Nature* 432:747–750.

Griffith GW (2012) Do we need a global strategy for microbial conservation? *Trends Ecol Evol* 27:1–2.

Gronesova P, Ficova M, Mizakova A, Kabat P, Trnka A and Betakova T (2008) Prevalence of avian influenza viruses, *Borrelia garinii*, *Mycobacterium avium*, and *Mycobacterium avium* subsp. *paratuberculosis* in waterfowl and terrestrial birds in Slovakia, 2006. *Avian Pathol* 37:537–543.

Guo F and Zhang T (2013) Biases during DNA extraction of activated sludge samples revealed by high throughput sequencing. *Appl Microbiol Biotechnol* 97:4607–4616.

Haas H and Fattal B (1990) Distribution of mycobacteria in different types of water in Israel. *Water Res* 24:1233–1235.

Hahn MW and Höfle MG (2001) Grazing of protozoa and its effect on populations of aquatic bacteria. *FEMS Microbiol Ecol* 35:113–121.

Hanski I and Gilpin M (1991). Metapopulation dynamics: brief history and conceptual domain. *Biol J Linn Soc* 42:3–16.

Hanson CA, Fuhrman JA, Horner-Devine MC and Martiny JBH (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* 10:497–506.

Harmsen D, Dostal S, Roth A, Niemann S, Rothgänger J, Sammeth M, Albert J, Frosch M and Richter E (2003) RIDOM: comprehensive and public sequence database for identification of *Mycobacterium* species. *BMC Infect Dis* 3:26.

Hautier Y, Tilman D, Isbell F, Seabloom EW, Borer ET and Reich PB (2015) Anthropogenic environmental changes affect ecosystem stability via biodiversity. *Science* 348:336–340.

Hennessee CT, Seo JS, Alvarez AM and Li QX (2009) Polycyclic aromatic hydrocarbon-degrading species isolated from Hawaiian soils: *Mycobacterium crocinum* sp. nov., *Mycobacterium pallens* sp. nov., *Mycobacterium rutilum* sp. nov., *Mycobacterium rufum* sp. nov. and *Mycobacterium aromaticivorans* sp. nov. *Int J Syst Evol Microbiol* 59:378–387.

Hett EC and Rubin EJ (2008) Bacterial growth and cell division: a mycobacterial perspective. *Microbiol Mol Biol Rev* 72:126–156.

HormischD, Brost I, KohringGW, Giffhorn F, Kroppenstedt RM, Stackebrandt E, Farber P and Holzapfel WH (2004) *Mycobacterium fluoranthenivorans* sp. nov., a fluoranthene and aflatoxin B1 degrading bacterium from contaminated soil of a former coal gas plant. *Syst Appl Microbiol* 27:653–660.

Hruska K and Kaevska M (2012) Mycobacteria in water, soil, plants and air: a review. *Vet Med* 57:623–679.

Hu M, Wang X, Wen X and Xia Y (2012) Microbial community structures in different wastewater treatment plants as revealed by 454-pyrosequencing analysis. *Bioresour Technol* 117:72–79.

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

Hugenholtz P, Goebel BM and Pace NR (1998) Impact of culture independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 180:4765–4774.

Ichiyama S, Shimokata K and Tsukamura M (1988) Isolation from of *Mycobacterium avium* complex from soil, water and dusts. *Microbiol Immunol* 32:733–739.

Iivanainen EK, Martikainen PJ, Väänänen PK and Katila ML (1993) Environmental factors affecting the occurrence of mycobacteria in brook waters. *Appl Environ Microbiol* 59:398–404.

Jacobs J, Rhodes M, Sturgis B and Wood B (2009) Influence of environmental gradients on the abundance and distribution of *Mycobacterium* spp. in a coastal lagoon estuary. *Appl Environ Microbiol* 75:7378–7384.

Jansson M, Bergstrom A, Blomqvist P and Drakare S (2000) Allochthonous organic carbon and phytoplankton/bacterioplankton production relationships in lakes. *Ecology* 81:3250–3255.

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 and Lennon JT (2010) Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci USA* 107:5881–5886.

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

Jürgens K and Matz C (2002) Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. *Antonie Van Leeuwenhoek* 81:413–434.

Justé A, Thomma BPHJ and Lievens B (2008) Recent advances in molecular techniques to study microbial communities in food-associated matrices and processes. *Food Microbiol* 25:745–761.

Kankya C, Muwonge A, Djønne B, Munyeme M, Opuda-Asibo J, Skjerve E, Oloya J and Johanson TB (2011). Isolation of non-tuberculous mycobacteria from pastoral ecosystems of Uganda: public health significance. BMC Public Health, 11, 1–9

Kasperbauer S and Huitt G (2013). Management of extrapulmonary nontuberculous mycobacterial infections. *Semin Respir Crit Care Med* 34:143–150.

Katoch VM (2004) Infections due to non-tuberculous mycobacteria (NTM). *Indian J Med Res* 120:290–304.

Kent AD, Jones SE, Yannarell AC, Graham JM, Lauster GH, Kratz TK and Triplett EW (2004) Annual patterns in bacterioplankton community variability in a humic lake. *Microb Ecol* 48:550–560.

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

Khan AA, Kim S, Paine DD and Cerniglia CE (2002) Classification of a polycyclic aromatic hydrocarbon-metabolizing bacterium, *Mycobacterium* sp. strain PYR-1, as *Mycobacterium vanbaalenii* sp. nov. *Int J Syst Evol Microbiol* 52:1997–2002.

Khera TT (2012) The diversity and distribution of Mycobacterium species in varying ecological and climatic environments. Doctoral dissertation, University of Warwick.

Kim BJ, Lee SH, Lyu MA, Bai GH, Kim SJ, Chae GT, Kim EC, Cha C and Kook YH (1999) Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). *J Clin Microbiol* 37:1714–1720.

Kim H, Kim SH, Shim TS, Kim M, Bai GH, Park YG, Lee SH, Chae GT, Cha CY, Kook YH and Kim BJ (2005) Differentiation of *Mycobacterium* species by analysis of the heat-shock protein 65 gene (*hsp65*). *Int J Syst Evol Microbiol* 55:1649–1656.

Kirschner RA, Parker BC and Falkinham III JO (1992) Epidemiology of infection by

nontuberculous mycobacteria. *Clin Microbiol Rev* 9:177–215.

Kirschner RA, Parker BC and Falkinham III JO (1999) Humic and fulvic acids stimulate the growth of *Mycobacterium avium*. *FEMS Microbiol Ecol* 30:327–332.

Kleespies M, Kroppenstedt RM, Rainey FA, Webb LE and Stackebrandt E (1996). *Mycobacterium hodleri* sp. nov., a new member of the fast-growing mycobacteria capable of degrading polycyclic aromatic hydrocarbons. *Int J Syst Bacteriol* 46:683–687.

Kritzberg ES, Cole JJ, Pace MM and Granéli W (2006) Bacterial growth on allochthonous carbon in humic and nutrient-enriched lakes: results from whole-lake 13C addition experiments. *Ecosystems* 9:489–499.

Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonák J, Lind K, Sindelka R, Sjöback R, Sjögreen B, Strömbom L, Ståhlberg A and Zoric N (2006) The real-time polymerase chain reaction. *Mol Aspects Med* 27:95–125.

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.

Langenheder S, Lindström ES and Tranvik LJ (2006) Structure and function of bacterial communities emerging from different sources under identical conditions. *Appl Environ Microbiol* 72:212–220.

Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG and Huttenhower C (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 31:814–821.

Lasiak T (2003) Influence of taxonomic resolution, biological attributes and data transformations on multivariate comparisons of rocky macrofaunal assemblages. *Mar Ecol Prog Ser* 250:29–34.

le Dantec C, Duguet J, Montiel A, Dumoutier N, Dubrou S and Vincent V (2002) Chlorine disinfection of atypical mycobacteria isolated from a water distribution system. *Appl Environ Microbiol* 68:1025–1032.

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

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.

Lennon JT and Jones SE (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol* 9:119–130.

Lévy-Frébault V, Rafidinarivo E, Promé JC, Grandry J, Boisvert H and David HL (1983)

Mycobacterium fallax sp. nov. *Int J Syst Bacteriol* 33:336–343.

Lindeman RL (1942) The trophic-dynamic aspect of ecology. *Ecology* 23:399–417.

Lindström ES (1998) Bacterioplankton community composition in a boreal forest lake. *FEMS Microbiol Ecol* 27:163–174.

Lindström ES and Bergström AK (2004) Influence of inlet bacteria on bacterioplankton assemblage composition in lakes of different hydraulic retention time. *Limnol Oceanogr* 49:125–136.

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

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.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L and Law M (2012) Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012:251364.

Liu L, Phanikumar MS, Molloy SL, Whitman RL, Shively D a, Nevers MB, Schwab DJ and Rose JB (2006) Modeling the transport and inactivation of *E. coli* and enterococci in the near-shore region of Lake Michigan. *Environ Sci Technol* 40:5022–5028.

Liu L, Yang J, Yu Z and Wilkinson DM (2015) The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J* 9:2068–2077.

Liu WT, Marsh TL, Cheng H and Forney LJ (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S. *Appl Environ Microbiol* 63:4516–4522.

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 (2008) Biogeography of bacterioplankton in inland waters. *Freshw Rev* 1:99–114.

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.

Lomolino MV, Riddle BR and Brown JH (2006) *Biogeography*. 3th edition. Sunderland, MA, USA:Sinauer Associates.

Loreau M, Naeem S, Inchausti P, Bengtsson J, Grime JP, Hector A, Hooper DU, Huston MA, Raffaelli D, Schmid B, Tilman D and Wardle DA (2001) Biodiversity and ccosystem functioning: current knowledge and future challenges. *Science (80)* 294:804–808.

Luo C, Tsementzi D, Kyrpides N, Read T and Konstantinidis KT (2012) Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One* 7:e30087.

Mailaender C, Reiling N, Engelhardt H, Bossmann S, Ehlers S and Niederweis M (2004) The *MspA* porin promotes growth and increases antibiotic susceptibility of both *Mycobacterium bovis* BCG and *Mycobacterium tuberculosis*. *Microbiology* 150:853–864.

Marsollier L, Robert R, Aubry J, Saint Andre JP, Kouakou H, Legras P, Manceau A, Mahaza C and Carbonnelle B (2002) Aquatic insects as a vector for *Mycobacterium ulcerans*. *Appl Environ Microbiol* 68:4623–4628.

Marsollier L, Sévérin T, Aubry J, Merritt RW, Saint André JP, Legras P, Manceau AL, Chauty A, Carbonnelle B and Cole ST (2004) Aquatic snails, passive hosts of *Mycobacterium ulcerans*. *Appl Environ Microbiol* 70:6296–6298.

Martin EC, Parker BC and Falkinham III JO (1987) Epidemiology of infection by nontuberculous mycobacteria. VII. Absence of mycobacteria in southeastern groundwaters. *Am Rev Respir Dis* 136:344–348.

Martiny AC, Treseder K and Pusch G (2013) Phylogenetic conservatism of functional traits in microorganisms. *ISME J* 7:830–838.

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.

McCorquodale JA, Georgiou I, Carnelos S and Englande AJ (2004) Modeling coliforms in storm water plumes. *J Environ Eng Sci* 3:419–431.

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.

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.

Meyer JL (1994) The microbial loop in flowing waters. *Microb Ecol* 28:195–199.

Michel AL, de Klerk L, van Pittius NCG, Warren RM and van Helden PD (2007) Bovine

tuberculosis in African buffaloes: observations regarding *Mycobacterium bovis* shedding into water and exposure to environmental mycobacteria. *BMC Vet Res* 3:23.

Miltner EC and Bermudez LE (2000) *Mycobacterium avium* grown in *Acanthamoeba castellanii* is protected from the effects of antimicrobials. *Antimicrob Agents Chemother* 44:1990–1994.

Morris CE, Bardin M, Berge O, Frey-Klett P, Fromin N, Girardin H, Guinebretière MH, Lebaron P, Thiéry JM and Troussellier M (2002) Microbial biodiversity: approaches to experimental design and hypothesis testing in primary scientific literature from 1975 to 1999. *Microbiol Bolecular Biol Rev* 66:592–616.

Mostajir B, Amblard C, Buffan-Dubau E, de Wit R, Lensi R and Sime-Ngando T (2011) Les réseaux trophiques microbiens des milieux aquatiques et terrestre, in *Ecologie Microbienne : Microbiologie des milieux naturels et anthropisés*. Presses universitaires de Pau et des Pays de l'Adour.

Moura A, Tacão M, Henriques I, Dias J, Ferreira P and Correia A (2009) Characterization of bacterial diversity in two aerated lagoons of a wastewater treatment plant using PCR–DGGE analysis. *Microbiol Res* 164:560–569.

Mrlik V, Slany M, Kubecka J, Seda J, Necas a, Babak V, Slana I, Kriz P and Pavlik I (2012) A low prevalence of mycobacteria in freshwater fish from water reservoirs, ponds and farms. *J Fish Dis* 35:497–504.

Mura M, Bull TJ, Evans H, Sidi-Boumedine K, McMinn L, Rhodes G, Pickup RW and Hermon-Taylor J (2006) Replication and long-term persistence of bovine and human strains of *Mycobacterium avium* subsp. *paratuberculosis* within *Acanthamoeba polyphaga*. *Appl Environ Microbiol* 72:854–859.

Murphy EM, Zachara JM, Smith SC, Phillips JL and Wletsma TW (1994) Hydrophobic organic compounds with mineral-bound humic substances. *Environ Sci Technol* 28:1291–1299.

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.

Narang R, Narang P and Mendiratta DK (2009) Isolation and identification of nontuberculous mycobacteria from water and soil in central India. *Indian J Med Microbiol* 27:247–250.

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.

Nevers MB and Whitman RL (2008) Coastal strategies to predict *Escherichia coli*

concentrations for beaches along a 35 km stretch of Southern Lake Michigan. *Environ Sci Technol* 42:4454–4460.

Newton RJ, Jones SE, Eiler A, McMahon KD and Bertilsson S (2011) A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* 75:14–49.

Nielsen KE, Johnsen PA, Bensasson D and Daffonchio D (2007) Release and persistence of extracellular DNA in the environment. *Environ Biosafety Res* 6:37–53.

Niva M, Hernesmaa A, Haahtela K, Salkinoja-Salonen M, Sivonen K and Haukka K (2006) Actinobacterial communities of boreal forest soil and lake water are rich in mycobacteria. *Boreal Environ Res* 11:45–53.

Nocker A, Burr M and Camper AK (2007) Genotypic microbial community profiling: a critical technical review. *Microb Ecol* 54:276–289.

Norby B, Fosgate GT, Manning EJB, Collins MT and Roussel AJ (2007) Environmental mycobacteria in soil and water on beef ranches: association between presence of cultivable mycobacteria and soil and water physicochemical characteristics. *Vet Microbiol* 124:153–159.

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.

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

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

Pace M and Cole J (1994) Comparative and experimental approaches to top-down and bottom-up regulation of bacteria. *Microb Ecol* 28:181–193.

Padisák J and Reynolds CS (2003) Shalow lakes: the absolute, the relative, the funcional and the pragmatic. *Hydrobiologia* 506:1–11.

Parashar D, Chauhan DS, Sharma VD, Chauhan A, Chauhan SVS and Katoch V (2004) Optimization of procedures for isolation of mycobacteria from soil and water samples obtained in northern India. *Appl Environ Microbiol* 70:3751–3753.

Parker BC, Ford MA, Gruft H and Falkinham III JO (1983) Epidemiology of infection by nontuberculous mycobacteria: IV. preferential aerosolization of *Mycobacterium intracellulare* from natural waters. *Am Rev Respir Dis* 128:652–656.

Pavlik I and Falkinham III JO (2009) The occurrence of pathogenic and potentially pathogenic mycobacteria in animals and the role of the environment in the spread of infection, in *The ecology of mycobacteria: impact on animal's and human's health*. Springer Science and Business Media.

Pavlik I, Falkinham III JO and Kazda J (2009) Transmission of mycobacteria from the environment to susceptible hosts, in *The ecology of mycobacteria: impact on animal's and human's health*. Springer Science and Business Media.

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.

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.

Pickup RW, Rhodes G, Arnott S, Bull TJ, Weightman A, Hurley M and Hermon-Taylor J (2005) *Mycobacterium avium* subsp. *paratuberculosis* in the catchment area and water of the river taff in South Wales, United Kingdom, and its potential relationship to clustering of Crohn's disease cases in the city of Cardiff. *Appl Environ Microbiol* 71:2130–2139.

Pickup RW, Rhodes G, Bull TJ, Arnott S, Sidi-Boumedine K, Hurley M and Hermon-Taylor J (2006) *Mycobacterium avium* subsp. *paratuberculosis* in lake catchments, in river water abstracted for domestic use, and in effluent from domestic sewage treatment works: diverse opportunities for environmental cycling and human exposure. *Appl Environ Microbiol* 72:4067–4077.

Piersimoni C and Scarparo C (2008) Pulmonary infections associated with non-tuberculous mycobacteria in immunocompetent patients. *Lancet Infect Dis* 8:323–334.

Pontiroli A, Khera TT, Oakley BB, Mason S, Dowd SE, Travis ER, Erenso G, Aseffa A, Courtenay O and Wellington EMH (2013) Prospecting environmental mycobacteria: combined molecular approaches reveal unprecedented diversity. *PLoS One* 8:e68648.

Postel S and Carpenter S (1997) Freshwater ecosystem services. *Nature's Serv Soc Depend Nat Ecosyst* 195, Island Press, Washington, DC.

Prescott LM, Wiley J, Sherwood L and Woolverton C (2013) *Microbiologie*. 4ème édition. De Boeck Supérieur (page 574).

Preston-Mafham J, Boddy L and Randerson PF (2002) Analysis of microbial community functional diversity using sole-carbon-source utilisation profiles - a critique. *FEMS Microbiol Ecol* 42:1–14.

Primm TP, Lucero CA and Falkinham III JO (2004) Health impacts of environmental mycobacteria. *Clin Microbiol Rev* 17:98–106.

Psenner R and Sommaruga R (1992) Are rapid changes in bacterial biomass caused by

shifts from top-down to bottom-up control? *Limnol Oceanogr* 37:1092–1100.

Radomski N (2011) *Sources of nontuberculous mycobacteria in watersheds*. Doctoral dissertation, Université Paris-Est École des Ponts ParisTech.

Radomski N, Betelli L, Moilleron R, Haenn S, Moulin L, Cambau E, Rocher V, Gonçalves A and Lucas FS (2011) *Mycobacterium* behavior in wastewater treatment plant, a bacterial model distinct from *Escherichia coli* and enterococci. *Environ Sci and Technol* 45: 5380–5386

Radomski N, Roguet A, Lucas FS, Veyrier FJ, Cambau E, Accrombessi H, Moilleron R, Behr MA and Moulin L (2013) *atpE* gene as a new useful specific molecular target to quantify Mycobacterium in environmental samples. *BMC Microbiol* 13:277.

Rahbar M, Lamei A, Babazadeh H and Yavari SA (2010) Isolation of rapid growing mycobacteria from soil and water in Iran. *African J Biotechnol* 9:3618–3621.

Ramette A and Tiedje JM (2007) Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb Ecol* 53:197–207.

Reche I, Pulido-Villena E, Morales-Baquero R and Casamayor EO (2005) Does ecosystem size determine aquatic bacterial richness? *Ecology* 86:1715–1722.

Rehmann K, Hertkorn N and Kettrup AA (2001) Fluoranthene metabolism in *Mycobacterium* sp. strain KR20 : identity of pathway intermediates during degradation and growth. *Microbiology* 147:2783–2794.

Ringuet H, Akoua-Koffi C, Honore S, Varnerot A, Vincent V, Berche P, Gaillard JL and Pierre-Audigier C (1999) *hsp65* sequencing for identification of rapidly growing mycobacteria. *J Clin Microbiol* 37:852–857.

Roux AL, Catherinot E, Ripoll F, Soismier N, Macheras E, Ravilly S, Bellis G, Vibet MA, Le Roux E, Lemonnier L, Gutierrez C, Vincent V, Fauroux B, Rottman M, Guillemot D and Gaillard JL (2009) Multicenter study of prevalence of nontuberculous mycobacteria in patients with cystic fibrosis in France. *J Clin Microbiol* 47:4124–4128.

Runyon EH (1959) Anonymous mycobacteria in pulmonary disease. *Med Clin N Amer* 43:273–290.

Sabater JF and Zaragoza JM (1993) A simple identification system for slowly growing mycobacteria. II. Identification of 25 strains isolated from surface water in Valencia (Spain). *Acta Microbiol Hung* 40:343-349.

Salah IB, Ghigo E and Drancourt M (2009) Free-living amoebae, a training field for macrophage resistance of mycobacteria. *Clin Microbiol Infect* 15:894–905.

Scheffer M (2004). *Ecology of shallow lakes*. Springer Science and Business Media. Originally published by Kluwer Academic Publishers.

Schimel J and Gulledge J (1998) Microbial community structure and global trace gases. *Glob Chang Biol* 4:745–758.

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.

Seguin F, Leboulanger C, Rimet F, Druart JC and Bérard A (2001) Effects of atrazine and nicosulfuron on phytoplankton in systems of increasing complexity. *Arch Environ Contam Toxicol* 40:198–208.

Seo JS, Keum YS and Li QX (2009) Bacterial degradation of aromatic compounds. *Int J Environ Res Public Health* 6:278–309.

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.

Shinnick T and Good R (1994) Mycobacterial taxonomy. *Eur J Clin Microbiol Infect Dis* 13:884–901.

Showalter SG and Wolinsky E (1974). Water as a source of pathogenic mycobacteria. *Am Rev Respir Dis* 109:710.

Simon M, López-García P, Deschamps P, Moreira D, Restoux G, Bertolino P and Jardillier L (2015) Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. *ISME J* 9:1941–1953.

Singh SN (2012). *Microbial degradation of xenobiotics*. Springer-Verlag Berlin Heidelberg, p.189.

Šizling AL, Šizlingová E, Storch D, Reif J and Gaston KJ (2009) Rarity, commonness, and the contribution of individual species to species richness patterns. *Am Nat* 174:82–93.

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.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM and Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proc Natl Acad Sci USA* 103:12115–12120.

Soininen J, McDonald R and Hillebrand H (2007) The distance decay of similarity in ecological communities. *Ecography* 30:3–12.

Soldati G, Lu ZH, Vaughan L, Polkinghorne A, Zimmermann DR, Huder JB and Pospischil A (2004) Detection of mycobacteria and chlamydiae in granulomatous inflammation of reptiles: a retrospective study. *Vet Pathol* 41:388–397.

Solomon JM, Leung GS and Isberg RR (2003) Intracellular replication of *Mycobacterium marinum* within *Dictyostelium discoideum*: efficient replication in the absence of host coronin. *Infect Immun* 71:3578–3586.

Stephan J, Bender J, Wolschendorf F, Hoffmann C, Roth E, Mailänder C, Engelhardt H and Niederweis M (2005) The growth rate of *Mycobacterium smegmatis* depends on sufficient porin-mediated influx of nutrients. *Mol Microbiol* 58:714–730.

Stinear T, Davies JK, Jenkin GA, Hayman JA, Oppedisano F and Johnson PDR (2000) Identification of *Mycobacterium ulcerans* in the environment from regions in Southeast Australia in which it is endemic with sequence capture-PCR. *Appl Environ Microbiol* 66:3206–3213.

Sutherland WJ, Freckleton RP, Godfray HC, Beissinger SR, Benton T, Cameron DD, Carmel Y, Coomes DA, Coulson T, Emmerson MC, Hails RS, Hays GC, Hodgson DJ, Hutchings MJ, Johnson D, Jones JPG, Keeling MJ, Kokko H, Kunin WE, Lambin X, Lewis OT, Malhi Y, Mieszkowska N, Milner-Gulland EJ, Norris K, Phillimore AB, Purves DW, Reid JM, Reuman DC, Thompson K, Travis JMJ, Turnbull LA, Wardle DA and Wiegand T (2013) Identification of 100 fundamental ecological questions. *J Ecol* 101:58–67.

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.

Tamames J, Abellán JJ, Pignatelli M, Camacho A and Moya A (2010) Environmental distribution of prokaryotic taxa. *BMC Microbiol* 10:85.

Terrat S, Christen R, Dequiedt S, Lelièvre M, Nowak V, Regnier T, Bachar D, Plassart P, Wincker P, Jolivet C, Bispo A, Lemanceau P, Maron PA, Mougel C and Ranjard L (2012) Molecular biomass and MetaTaxogenomic assessment of soil microbial communities as influenced by soil DNA extraction procedure. *Microb Biotechnol* 5:135–141.

Thomas DB and Schallenberg M (2008) Benthic shear stress gradient defines three mutually exclusive modes of non-biological internal nutrient loading in shallow lakes. *Hydrobiologia* 610:1–11.

Tilman D, Wedin D and Knops J (1996) Productivity and sustainability influences by biodiversity in grassland ecosystems. *Nature* 379:718–720.

Torkko P, Suomalainen S, Iivanainen E, Suutari M, Paulin L, Rudbäck E, Tortoli E, Vincent V, Mattila R and Katila ML (2001) Characterization of Mycobacterium bohemicum isolated from human, veterinary, and environmental sources. *J Clin Microbiol* 39:207–211.

Torkko P, Suomalainen S, Iivanainen E, Suutari M, Tortoli E, Paulin L and Katila ML (2000) *Mycobacterium xenopi* and related organisms isolated from stream waters in Finland and description of *Mycobacterium botniense* sp. nov. *Int J Syst Evol Microbiol*

50:283–289.

Torkko P, Suomalainen S, Iivanainen E, Tortoli E, Suutari M, Seppänen J, Paulin L and Katila ML (2002) *Mycobacterium palustre* sp. nov., a potentially pathogenic, slowly growing mycobacterium isolated from clinical and veterinary specimens and from Finnish stream waters. *Int J Syst Evol Microbiol* 52:1519–1525.

Torsvik V, Goksøyr J and Daae FL (1990) High diversity in DNA of soil bacteria. *Appl Environ Microbiol* 56:782–787.

Torsvik V, Ovreas L and Thingstad TF (2002) Prokaryotic diversity: magnitude, dynamics, and controlling factors. *Science* 296:1064–1066.

Tortoli E (2003) Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. *Clin Microbiol Rev* 16:319–354.

Tortoli E (2009) Clinical manifestations of nontuberculous mycobacteria infections. *Clin Microbiol Infect* 15:906–910.

Torvinen E, Suomalainen S, Lehtola MJ, Miettinen IT, Zacheus O, Paulin L, Katila M and Martikainen PJ (2004) Mycobacteria in water and loose deposits of drinking water distribution systems in Finland. *Appl Environ Microbiol* 70:1973–1981.

Trujillo ME, Velázquez E, Kroppenstedt RM, Schumann P, Rivas R, Mateos PF and Eustoquio MM (2004) *Mycobacterium psychrotolerans* sp. nov., isolated from pond water near a uranium mine. *Int J Syst Evol Microbiol* 54:1459–1463.

Vaerewijck MJM, Huys G, Palomino JC, Swings J and Portaels F (2005) Mycobacteria in drinking water distribution systems: ecology and significance for human health. *FEMS Microbiol Rev* 29:911–934.

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 der Wielen PWJJ and van der Kooij D (2013) Nontuberculous mycobacteria, fungi, and opportunistic pathogens in unchlorinated drinking water in The Netherlands. *Appl Environ Microbiol* 79:825–834.

Vandelannoote K, Durnez L, Amissah D, Gryseels S, Dodoo A, Addo P, Eddyani M, Leirs H, Ablordey A and Portaels F (2010) Application of real-time PCR in Ghana, a Buruli ulcer-endemic country, confirms the presence of *Mycobacterium ulcerans* in the environment. *FEMS Microbiol Lett* 304:191–194.

Vellend M (2010) Conceptual synthesis in community ecology. *Q Rev Biol* 85:183–206.

Verpoorter C, Kutser T, Seekell DA and Tranvik LJ (2014) A global inventory of lakes based

on high-resolution satellite imagery. *Geophys Res Lett* 6396–6402.

Viallier J and Viallier G (1973) Inventaire des mycobacteries de la nature. *Ann Soc Belg Med Trop* 53:361–371.

Walsh DS, Portaels F and Meyers WM (2008) Buruli ulcer (*Mycobacterium ulcerans* infection). *Trans R Soc Trop Med Hyg* 102: 969–978.

Wang Y, Chi C, Cai M, Lou Z, Tang Y, Zhi X, Li W, Wu X and Du X (2010) *Amycolicicoccus subflavus* gen. nov., sp. nov., an actinomycete isolated from a saline soil contaminated by crude oil. *Int J Syst Evol Microbiol* 60:638–643.

Warnecke F, Sommaruga R, Sekar R, Hofer JS and Pernthaler J (2005) Abundances, identity, and growth state of *Actinobacteria* in mountain lakes of different UV transparency. *Appl Environ Microbiol* 71:5551–5559.

Weinbauer MG and Rassoulzadegan F (2004) Are viruses driving microbial diversification and diversity? *Environ Microbiol* 6:1–11.

Weisse T (1991) The annual cycle of heterotrophic freshwater nanoflagellates: role of bottom-up versus top-down control. *J Plankton Res* 13:167–185.

Weithoff G, Lorke A and Walz N (2000) Effects of water-column mixing on bacteria, phytoplankton, and rotifers under different levels of herbivory in a shallow eutrophic lake. *Oecologia* 125:91–100.

Wendt SL, George KL, Parker BC, Gruft H and Falkinham III JO (1980) Epidemiology of infection by nontuberculous mycobacteria: isolation of potentially pathogenic mycobacteria from aerosols. *Am Rev Respir Dis* 122: 259–263.

WHO (2014) Global tuberculosis report. Geneva, World Health Organization.

Wiinschmann A, Armien A, Harris NB, Brown-Elliott BA, Wallace RJ Jr, Rasmussen J, Willette M and Wolf T (2008) Disseminated panniculitis in a bottlenose dolphin (*Tursiops truncatus*) due to *Mycobacterium chelonae* infection. *J Zoo Wildl Med* 39:412–420.

Williams P, Whitfield M, Biggs J, Bray S, Fox G, Nicolet P and Sear D (2004) Comparative biodiversity of rivers, streams, ditches and ponds in an agricultural landscape in Southern England. *Biol Conserv* 115:329–341.

Williamson CE, Dodds W, Kratz TK and Palmer MA (2008) Lakes and streams as sentinels of environmental change in terrestrial and atmospheric processes. *Front Ecol Environ* 6:247–254.

Willumsen PA, Nielsen JK and Karlson U (2001) Degradation of phenanthrene-analogue azaarenes by *Mycobacterium gilvum* strain LB307T under aerobic conditions. *Appl Microbiol Biotechnol* 56: 539–544.

Woese CR and Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* 74:5088–5090.

Wolinksy E (1979) Nontuberculous mycobacteria and associated diseases. *Am Rev Respir Dis* 119:107–159.

Wu QL and Hahn MW (2006) High predictability of the seasonal dynamics of a species-like Polynucleobacter population in a freshwater lake. *Environ Microbiol* 8:1660–1666.

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, 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.

Ye L and Zhang T (2013) Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. *Appl Microbiol Biotechnol* 97:2681–2690.

Zwart G, Crump BC, Kamst-van Agterveld MP, Hagen F and Han SK (2002) Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat Microb Ecol* 28:141–155.

GLOSSARY

Allochthonous Refers to organisms or compounds originating or created in a place other than where found. In ecology, it refers to foreign organisms or compounds.

Autochthonous Refers to organisms or compounds originating or created in the place where they were found. In ecology, it refers to indigenous organisms or compounds.

Autotroph Organism that convert inorganic carbon, such as CO_2 into organic compounds.

Beta diversity Difference in community composition among sites.

Biodiversity The genetic, taxonomic, and functional variety of all forms of life on Earth, encompassing the interactions among them and the processes that maintain them.

Biogeography Study of the distribution of biodiversity over space and time.

Class In taxonomy, refers to a set of orders that have similar characteristics.

Community An assemblage of taxa found together in a specific habitat at a certain time, and interacting with each other's.

Dispersal Movement of individuals among local population in another population.

Diversity The degree of variation of living things present in a particular ecosystem.

Ecological drift processes that randomly change the abundance of taxa.

Eutrophic Describing water bodies that are richly supplied with nutrients, promoting the proliferation of algae and aquatic plants.

Family In taxonomy, refers to a set of genus that have similar characteristics.

Functional trait Refers to a feature of an organism that has demonstrable links to the organism's function (ecosystem role) or functioning (performance).

Genus In taxonomy, refers to a set of species that have similar characteristics.

Heterotroph Organism that use organic carbon compounds, such as dissolved organic carbon matter and particulate organic matter as source of carbon. In opposition to autotrophs, these organisms cannot fix carbon.

Lineage Group of organisms that constitutes a monophyletic branch, related by descent from a common ancestor.

Mass effects The quantitative effects of dispersal on local population dynamics. Emigration from a population may have negative effects on its demography, while immigration may have positive (rescue) effects.

Mesotrophic Describing water bodies that are intermediate in characteristics between oligotrophic and eutrophic water bodies.

Metacommunity Set of local communities that are linked by dispersal of multiple potentially interacting species.

Myxotroph Organism that both use the inorganic carbon and the organic carbon as source of carbon. They are autotrophs and heterotrophs.

Neutral processes Variation in community composition determined by stochastic effects of dispersal and demography among taxa with equivalent niches.

Oligotrophic Describing water bodies that have low nutrient concentration.

Order In taxonomy, refers to a set of families that have similar characteristics.

Patch dynamic Describe the ecological systems are mosaic of patches exhibiting non-equilibrium transient dynamics and together determining system-level structure and function.

Phylum Refers to taxonomic categories. Within prokaryotes, it constitutes the highest taxonomic rank.

Plankton A collective term for various drifting organisms of the pelagic zone.

Polymictic Corresponds to lakes having no stable thermal stratification but exhibiting perennial circulation.

Primary producer An organism capable of converting atmospheric carbon dioxide into organic matter. See autotroph.

Relative abundance The quantitative pattern of rarity and commonness among species in a sample or a community richness.

Selection (natural) A difference, on average, between the survival or fecundity of individuals with certain phenotypes compared with individuals of the same species with other phenotypes.

Speciation Process by which new species emerge through evolutionary forces.

Species A monophyletic and genomically coherent cluster of individual organisms that show a high degree of overall similarity in many independent characteristics, and is diagnosable by a discriminative phenotypic property

Species richness The number of species in a sample, in a site, or in a region.

Species sorting Variation in community composition by the optimization of fitness among species across patches.

Taxon (plural **taxa**) Any named group at any taxonomic rank. Within the *Bacteria* kingdom a taxon could also refers to a set of strains with close phenotypic and genotypic characteristics.

Waterborne Refers to organisms that are transported by water or transmitted in water.

Appendix

Appendix 1 Biotic and abiotic parameters analyzed during the 49 lakes campaigns and over the monthly monitoring of Créteil Lake by each laboratory.

LEESU UMR MA 102 University of Paris-Est	**MCAM** UMR 7245 CNRS/MNHN	**ESE** UMR CNRS 8079 University of Paris-Sud	**BIOEMCO** UMR CNRS 7618 ENS
pH*	Depth max	*Densities of:*	Seston*
PAR*	NO_3^-	Cyanobacteria*	Zooplankton*
Temperature*	NH_4^+	Small picoeukaryotes*	
Thermal stratification*	N_{total}*	Large picoeukaryotes *	
Extinction coefficient*	PO_4^{3-}*	Heterotrophic bacteria*	
Conductivity*	P_{total}*	Protists / Large heterotrophic	
O_2 saturation (%)*	Total Chlorophyll *a* *		
NO_2^-	Total phytoplankton		
TSS*	Cyanobacteria		
TOC*			
DOC*			
POC*			
Metals*			
PAH*			
*Escherichia coli**			
Intestinal enterococci*			
NTM*			
Bacterial assemblage*			

Abreviations: PAR, photosynthetically active radiation; TSS, total suspended solid; TOC, total organic carbon; DOC, dissolved organic carbon; POC, particulate organic carbon; PAH, polycyclic aromatic hydrocarbon; NTM, nontuberculous mycobacteria

* Refers to the abiotic parameters also analyzed during the monthly campaigns of Créteil Lake.

** Refers to the parameters only analyzed during the monthly campaigns Créteil Lake.

Appendix 2 (a) 16S rRNA gene sequence variability map (based on 16S rRNA gene sequence of *Escherichia coli*). Red sequences indicate variable regions, underlined red sequences indicate hypervariable regions and blue sequences indicate conserved regions. From Barker and colleagues (2003). (b) Efficiency of 'universal' probes. Efficiency tested *in silico* using the probeBase online resource (http://www.microbial-ecology.net/probebase/search.asp).

a



b

	Conserved region	Probe name	Sequence	Mistmaches number	% bacterial sequences amplified	% chloroplast sequences amplified	Authors
a		27F	AGAGTTTGATCMTGGCTCAG	0 1 2	13 18 20	14 20 21	Hongoh et al. 2003
		27F	GAGTTTGATCMTGGCTCAG	0 1 2	16 19 20	17 21 22	Lee et al. 2010
		27F	AGRGTTYGATYMTGGCTCAG	0 1 2	14 19 20	17 21 22	Lane 1991
		39F	TGGCTCAGRWYGAACGCTRG	0 1 2	18 23 25	20 22 23	Hongoh et al. 2003
		63F	CAGGCCTAACACATGCAAGTC	0 1 2	5 15 27	1 2 21	Hongoh et al. 2003
		64F	BGYCTWANRCATGCAAGTYG	0 1 2	26 57 61	12 42 44	Hongoh et al. 2003
		68F	TNANACATGCAAGTCGRRCG	0 1 2	43 60 63	22 42 45	McAllister et al. 2011
		8F 616V	AGAGTTTGATYMTGGCTCAG	0 1 2	14 19 20	15 21 22	Kim et al. 2009
		8F pA	AGAGTTTGATCCTGGCTCAG	0 1 2	10 16 19	11 18 21	Frank et al. 2007
		9bF	GRGTTTGATCCTGGCTCA	0 1 2	11 18 20	12 20 22	Huber et al. 2002
		GM3F	AGAGTTTGATCMTGGC	0 1 2	13 18 20	15 20 22	Muyzer et al. 1995
b		pBR-V1.ASF	AGTGGCGGACGGGTGAGTAA	0 1 2	11 49 63	41 51 59	Jonasson et al. 2002
		pBR-V1.ASR	TTACTCACCCGTCCGCCACT	0 1 2	11 49 63	41 51 59	Jonasson et al. 2002
c		338 R	TGCTGCCTCCCGTAGGAGT	0 1 2	77 81 84	55 62 64	Fierer et al. 2008
		338F	ACTCCTACGGGNGGCNGCA	0 1 2	79 83 84	55 62 65	Ahmad et al. 2011
		338F	**TCCTACGGGAGGCAGCAGT**	**0** **1** **2**	**76** **82** **84**	**55** **63** **63**	**Nadkarni et al. 2002**
		343F	TACGGRAGGCAGCAG	0 1 2	79 84 85	59 68 69	Nossa et al. 2010
		347F	GGAGGCAGCAGTRRGGAAT	0 1 2	74 80 82	62 68 69	Nossa et al. 2010

c	b341F	CCTACGGGAGGCAGCAG	0	78	56	Juck et al. 2000
			1	78	63	
			2	85	65	
	Bact340F	TCCTACGGGAGGCAGCAGT	0	76	55	Li et al. 2010
			1	81	63	
			2	84	65	
	bact363F	CAATGGRSGVRASYCTGAHS	0	77	58	Armougom and Raoul 2009
			1	82	69	
			2	86	70	
	Bakt_341F	CCTACGGGNGGCWGCAG	0	81	56	Herlemann et al. 2011
			1	84	63	
			2	85	68	
	P338f	ACTCCTACGGGAGGCAGCAG	0	77	55	El Fantroussi et al. 1999
			1	81	62	
			2	83	64	
	Primer 3 R	GCCTACGGGAGGCAGCAG	0	0	0	Muyzer et al. 1993
			1	0	0	
			2	0	0	
	R357	CTGCTGCCTYCCGTA	0	79	59	Armougom and Raoul 2009
			1	84	68	
			2	85	69	
	U341F	CCTACGGGRSGCAGCAG	0	78	56	Baker et al. 2003
			1	83	63	
			2	85	68	
d	517F	GCCAGCAGCCGCGGTAA	0	75	70	Wang and Qian 2009
			1	80	72	
			2	81	73	
	518F	CCAGCAGCCGCGGTAAT	0	69	26	Muyzer et al. 1993
			1	80	71	
			2	81	72	
	518R	ATTACCGCGGCTGCTGG	0	69	26	Muyzer et al. 1993
			1	80	71	
			2	81	72	
	518R	WTTACCGCGGCTGCTGG	0	71	27	Lee et al. 2010
			1	80	71	
			2	81	72	
	520F	AYTGGGYDTAAAGNG	0	77	75	Cleasson et al. 2009
			1	81	78	
			2	82	78	
	530F	GTGCCAGCMGCNGCGG	0	78	71	Dowd et al. 2008
			1	81	72	
			2	85	73	
	533F	TGCCAGCAGCCGCGGTAA	0	75	69	Hongoh et al. 2003
			1	80	70	
			2	81	72	
	533R	TTACCGCGGCTGCTGGCAC	0	74	69	Huse et al. 2008
			1	80	72	
			2	81	72	
	534R	ATTACCGCGGCTGCTGGC	0	69	26	Nossa et al. 2010
			1	79	71	
			2	81	72	
	536R	CAGCMGCCGCGGTAATWC	0	0	0	Liu et al. 1997
			1	0	0	
			2	0	0	
	Bact531R	CTNYGTMTTACCGCGGCTGC	0	63	52	Armougom and Raoul 2009
			1	77	70	
			2	79	72	
	P609D	GGMTTAGATACCCBDGTA	0	71	67	Lucena et al. 2010
			1	76	71	
			2	77	72	
	P609R	TACHVGGGTATCTAAKCC	0	71	67	Lucena et al. 2010
			1	76	71	
			2	77	72	
	U515F	GTGCCAGCMGCCGCGGTAA	0	75	70	Walters et al. 2011
			1	80	72	
			2	81	72	
	U529R	ACCGCGGCKGCTGGC	0	78	70	Baker et al. 2003
			1	81	72	
			2	82	73	
	Uni522R	GWATTACCGCGGCKGCTG	0	69	27	Amann et al. 1995
			1	79	71	
			2	82	72	
e	784F	AGGATTAGATACCCT	0	63	9	Nossa et al. 2010
			1	68	10	
			2	76	70	
	784F	AGGATTAGATACCCTGGTA	0	60	1	Andersson et al. 2008
			1	67	9	
			2	73	13	
	796R	AGGGTATCTAATCCT	0	63	9	Nossa et al. 2010
			1	68	10	
			2	76	70	
	802R	TACNVGGGTATCTAATCC	0	70	67	Cleasson et al. 2009

		1	76	71	
		2	77	72	
803R	CTACCRGGGTATCTAATCC	0	65	3	Nossa et al. 2010
		1	75	64	
		2	76	71	
U786R	CTACCAGGGTTATCTAATC	0	0	0	
		1	61	1	
		2	73	13	
805R	GACTACCAGGGTATCTAAT	0	60	1	Frank et al. 2007
		1	72	7	
		2	76	65	
806R 797R	**GGACTACHVGGGTATCTAAT**	**0 1 2**	**68 75 76**	**58 68 69**	**Walters et al. 2011**
Ab789F	TAGATACCCSSGTAGTCC	0	5	2	Baker et al. 2003
		1	71	58	
		2	75	65	
B785 R	CTACCAGGGTATCTAATCC	0	61	1	Juck et al. 2000
		1	72	9	
		2	76	67	
Bact806R	GGACTACCAGGGTATCTAATCCTG	0	59	1	Li et al. 2010
		1	65	1	
		2	72	5	
Bakt_805R	GACTACHVGGGTATCTAATCC	0	68	61	Herlemann et al. 2011
		1	75	68	
		2	76	72	
E806R	GGACTACCAGGGTATCTAAT	0	60	1	Wang and Qian 2009
		1	71	5	
		2	75	63	
P699F	YAACGAGCGMRACCC	0	59	56	
		1	61	61	
		2	61	61	
P699R	GGGTYKCGCTCGTTR	0	59	56	
		1	61	61	
		2	61	61	
907R	CCGTCAATTCMTTTGAGTTT	0	58	57	Muyzer et al. 1995
		1	65	60	
		2	66	61	
908R	CGTCAATTCMTTTGAGTT	0	59	58	Hongoh et al. 2003
		1	65	60	
		2	66	61	
909F	ACTCAAAKGAATWGACGG	0	60	58	Berry et al. 2011
		1	65	60	
		2	66	61	
909F	AAACTYAAARRAATTGACGG	0	53	56	Keijser et al. 2008
		1	65	60	
		2	66	61	
917F	GAATTGACGGGGRCCC	0	54	51	Nossa et al. 2010
		1	62	54	
		2	64	57	
917F	GAATTGACGGGGRCCCGC	0	54	51	Degnan and Ochman 2012
		1	61	54	
		2	63	54	
917Fw	GAATTGACGGGGRCCCGCA	0	53	51	Keijser et al. 2008
		1	61	54	
		2	63	54	
926R	CCGTCAATTCCTTTRAGTTT	0	51	56	Liu et al. 1997
		1	64	60	
		2	66	61	
967F	CAACGCGAAGAACCTTACC	0	38	31	Sogin et al. 2006
		1	51	53	
		2	54	55	
967F	ATACGCGAGGAACCTTACC	0	7	0	Dethlefsen et al. 2008
		1	8	0	
		2	48	32	
967R	GGTAAGGTTCTTCGCGTTG	0	38	31	Sogin et al. 2006
		1	51	53	
		2	54	55	
968F	AACGCGAAGAACCTTAC	0	38	31	Nuebel et al. 1996
		1	52	54	
		2	61	56	
970F	CGCGAAGAACCTTACC	0	1	1	Dethlefsen et al. 2008
		1	2	1	
		2	2	1	
986F	CNACGCGAAGAACCTTANC	0	39	31	Claesson et al. 2009
		1	53	54	
		2	61	55	
Bac927R	ACCGCTTGTGCGGGCCC	0	26	47	Lane 1991
		1	54	53	
		2	61	54	
Bac967Fc	CAACGCGCAGAACCTTACC	0	4	0	Huber et al. 2007
		1	44	32	

e

f

f	Bac967Fd	ATACGCGARGAACCTTACC	0	7	0	Huber et al. 2007
			1	48	32	
			2	61	54	
	bio-pJBS-	GCAACGCGAAGAACCTTACC	0	38	31	Jonasson et al. 2002
			1	51	53	
			2	53	54	
	bio-pJBS-	GGTAAGGTTCTTCGCGTTGC	0	38	31	Jonasson et al. 2002
			1	51	53	
			2	53	54	
g	1027R	CGACRRCCATGCANCACCT	0	0	0	Claesson et al. 2009
			1	0	0	
			2	0	0	
	1046R	CGACARCCATGCASCACCT	0	56	60	Dethlefsen et al. 2008
			1	61	62	
			2	61	62	
	1046R	CGACAGCCATGCANCACCT	0	30	59	Sogin et al. 2006
			1	60	62	
			2	61	62	
	1050R	ACGACAGCCATGCANC	0	30	60	Degnan and Ochman 2012
			1	60	62	
			2	61	62	
	1061R	CRRCACGAGCTGACGAC	0	59	58	Andersson et al. 2008
			1	61	62	
			2	61	62	
	1061R	CACGRCACGAGCTGACGAC	0	55	34	Degnan and Ochman 2012
			1	61	59	
			2	61	62	
	1061Rv	TCACGRCACGAGCTGACGAC	0	55	34	Keijser et al. 2008
			1	61	59	
			2	61	62	
	1099F	GYAACGAGCGCAACCC	0	52	56	Nossa et al. 2010
			1	60	61	
			2	61	61	
	1100R	GGGTTNCGNTCGTTG	0	49	55	Dowd et al. 2008
			1	60	61	
			2	61	61	
	B1055f	ATGGCTGTCGTCAGCT	0	30	60	Ferris et al. 1996
			1	60	62	
			2	61	62	
	Bac1046Rb	CGACAACCATGCANCACCT	0	27	0	Huber et al. 2007
			1	58	60	
			2	61	62	
	B-V3.ASR	ACGACAGCCATGCAGCACCT	0	23	0	Jonasson et al. 2002
			1	38	60	
			2	60	62	
	U1053F	GCATGGCYGYCGTCAG	0	31	60	Baker et al. 2003
			1	60	62	
			2	61	62	
	GM12R	CGTCATCCMCACCTTCCTC	0	43	3	MacGregor and Amann 2006
			1	55	47	
			2	59	59	
	U1053R	CTGACGRCRGCCATGC	0	31	60	Baker et al. 2003
			1	60	62	
			2	61	62	
h	1391R	GACGGGCGGTGWGTRCA	0	31	34	Walker and Pace 2007
			1	33	37	
			2	34	38	
	1391R	GACGGGCGGTGTGTRCA	0	30	34	Dethlefsen et al. 2008
			1	33	37	
			2	34	38	
	1401R	CGGTGTGTACAAGACCC	0	6	1	Nuebel et al. 1996
			1	29	35	
			2	34	38	
	1378R	CGGTGTGTACAAGGCCCGGGAAC	0	20	29	Heuer et al. 1997
			1	30	34	
			2	33	38	
	1407R	GACGGGCGGTGTGTRC	0	30		Nossa et al. 2010
			1	33	37	
			2	34	38	
	UA1406R	ACGGGCGGTGTGTRCAA	0	30		Lane 1991
			1	33	37	
			2	34	38	
	UA1406R	ACGGGCGGTGWGTRCAA	0	32	34	Baker et al. 2003
			1	33	38	
			2	34	38	
	Uni1390R	GACGGGCGGTGTGTACAA	0	28	33	Zheng et al. 1996
			1	32	37	
			2	33	37	
	Uni1392R	ACGGGCGGTGTGTRC	0	31	34	Lane et al. 1985
			1	33	37	
			2	34	38	

1492R	RGYTACCTTGTTACGACTT	0	9	11	McAllister et al. 2011
		1	10	21	
		2	11	22	
1492R	NTACCTTGTTACGACT	0	10	21	Berry et al. 2011
		1	11	24	
		2	12	24	
GM4R	TACCTTGTTACGACTT	0	11	23	Muyzer et al. 1995
		1	11	24	
		2	12	25	
U1510R	GGTTACCTTGTTACGACTT	0	4	7	Baker et al. 2003
		1	9	17	
		2	10	22	

Appendix 3 R script for the variance analysis partitioning
This script describe an example of analyze performed on the entire bacterial community

```
# XXX.txt = file containing the bacterial communities dataset
# YYY.txt = = file containing the independent variable dataset

### Data importation/preparation
# ************************************************
rm(list=ls(all=FALSE))
setwd('//vmware-host/Shared Folders/adelaideroguet/Desktop/R_Pulse/')
library(vegan)
library(PCNM)
library(randomForest)
library(packfor)

predata.BC<-read.table("XXX.txt", h=T)
data.BC <- sweep(predata.BC, 1, rowSums(predata.BC), '/'); apply(data,1,sum) # check

data.env1 <- read.table("YYY.txt", h=T)
data.env<-data.env1[,-c(2:3)]
data.xy<-data.env1[,2:3]

# ********************************************************************************************
### VARIANCE PARTITIONING ANALYSIS
# ********************************************************************************************

# 1. Data preparation before the variance partitioning analyze
# ************************************************
apply(data.BC,1,sum)
data.BC.h<-decostand(data.BC, method="hellinger")

# 2. Local factors
# ************************************************
# Recode regional qualitative variables into dummy binary variables + quantitative variables were centered
and scaled.
TS.data.BC <- model.matrix(~data.env[,3])[,-1]
BV.data.BC<-model.matrix(~data.env[,5])[,-1]
quant_loc.data.BC<-scale(data.env[,c("pH","Temp","Cond","COD")])
loc.data.BC <- cbind(quant_loc.data.BC, TS.data.BC, BV.data.BC)

# Forward selection of the environmental variables
env.loc.rda.data.BC <- rda(data.BC.h~loc.data.BC)
env.loc.R2a.data.BC <- RsquareAdj(env.loc.rda.data.BC)$adj.r.squared
(env.loc.fwd.data.BC <- forward.sel(data.BC.h, loc.data.BC, adjR2thresh=env.loc.R2a.data.BC, nperm=999))
select.loc.data.BC<-cbind(# selected environmental variables)

# 3. Spatial and dispersal
# ************************************************
# Spatial factors: creation/selection of the PCNM axes
xy.pcnm.pre.data.BC<-PCNM(dist(data.xy))
select.data.BC<-which(xy.pcnm.pre.data.BC$Moran_I$Positive=="TRUE" &
xy.pcnm.pre.data.BC$Moran_I$p.value<=0.05)
xy.pcnm.data.BC <- as.data.frame(xy.pcnm.pre.data.BC$vectors)[,select.data.BC]
#xy.pcnm<-read.table("xy.pcnm.txt", h=T)
test.data.BC<-rda(data.BC.h~., xy.pcnm.data.BC)
anova.cca(test.data.BC)
```

```
RsquareAdj(test.data.BC)$adj.r.squared
xy.pcnm_select.data.BC <- forward.sel(data.BC.h, xy.pcnm.data.BC,
adjR2thresh=RsquareAdj(test.data.BC)$adj.r.squared)
xy.pcnm_select.Data.BC_1<-xy.pcnm.data.BC[,c(xy.pcnm_select.data.BC$order)]
test1.Data.BC<-rda(data.BC.h~ xy.pcnm_select.data.BC_1)
#test1<-rda(data.h~., xy.pcnm_select_1)
anova.cca(test1.data.BC)
RsquareAdj(test1.data.BC)$adj.r.squared # Need to check if the PCNM selection do not surfit the model

# 4. Dispersal factors
# ***********************************************
Type_RH.data.BC <- model.matrix(~data.env[,11])[,-1]
RUTP.data.BC <- model.matrix(~data.env[,13])[,-1]
disp.data.BC<-cbind(Type_RH.data.BC, RUTP.data.BC)

# Forward selection of the dispersal variables
disp.rda.data.BC <- rda(data.BC.h~disp.data.BC)
disp.R2a.data.BC <- RsquareAdj(disp.rda.data.BC)$adj.r.squared
(disp.fwd.data.BC <- forward.sel(data.BC.h, disp.data.BC, adjR2thresh=disp.R2a.data.BC, nperm=999))
select.spa_disper.data.BC<-cbind(#selected dispersal variables, xy.pcnm_select.data.BC_1)

# 5. Variance partitioning analysis
# ***********************************************
(data.varpart <- varpart(data.BC.h, select.loc.data.BC, select.spa_disper.data.BC))
plot(data.varpart, digits=2)

# Tests of the unique fractions [a] and [b]
# ***********************************************
# Fraction [a], pure local
anova.cca(rda(data.BC.h, select.loc.data.BC, cbind(select.spa_disper.data.BC)))
# Fraction [b], pure spatial/dispersal
anova.cca(rda(data.BC.h, select.spa_disper.data.BC, cbind(select.loc.data.BC)))
```

Appendix 4 R script for the spatial neutral community model (Sloan *et al.* 2006)
This script describe an example of analyze performed on the entire bacterial community

```
# XXX.txt = file containing the bacterial communities dataset

### Data importation/preparation
# ********************************************
rm(list=ls(all=FALSE))
setwd('Users/adelaideroguet/Desktop/R_Pulse/')
library(MASS)

predata.BC<-read.table("XXX.txt", h=T)
data.BC <- sweep(predata.BC, 1, rowSums(predata.BC), '/'); apply(data,1,sum) # check

# ********************************************************************************
# Determination of the frequencies and the relative abundances of the bacterial T-RFs/OTUs
# ********************************************************************************

# Mean Relative Abundance (mra)
# ********************************************
premra<-as.data.frame(colSums(data.BC, na.rm = FALSE, dims = 1))
names(premra)[1]<-"mra"
mra<-(premra[1])/nrow(data)
head(mra)
sum(mra) # sum = 1?

# Frequencies (fre)
# ********************************************
occurrence<-as.data.frame(BinaryTransformation(data, 0)) # convert abundance dataset into binary
dataset
prefre<-as.data.frame(colSums(occurrence, na.rm = FALSE, dims = 1))
names(prefre)[1]<-"fre"
fre<-(prefre[1])/nrow(data)
head(fre)
sum(fre)
plot(mra$mra, fre$fre, xlab="Mean relative abundance", ylab="Occurrence frequency", pch=20)

# ********************************************************************************
# FIT TO THE SPATIAL NEUTRAL COMMUNITY MODEL
# ********************************************************************************
data.fit<-as.data.frame(cbind(fre$fre, mra$mra))
colnames(data.fit)<-c("fre.fit","mra.fit"); head(data.fit)

scale<- # for T-RFLP analyses, m = 5e-3; for Illumina MiSeq analyses, m = 1e-5
f <- function(x, a) {1-(pbeta(scale, a*x, a*(1-x)))}
fit <- nls(fre.fit ~ f(mra.fit, a), data=as.data.frame(data.fit), start=list(a=20), weights)
summary(fit)
with(data.fit, 1 - sum(resid(fit)^2)/sum((fre.fit - mean(fre.fit))^2))
# Function above determines the following equation: 1 - RSS/AdjSS.
```

```
# RSS refers to the residual sum of squares from the fitted model, and AdjSS refers to the adjusted
# sum of squares or the sum of squares of the deviations of the observed responses from their
# mean
```

Appendix 5 R script for the temporal neutral community model (Ofiţeru *et al.* 2010). This script describe an example of analyze performed on the entire bacterial community

```
# XXX.txt = file containing the bacterial communities dataset

### Data importation/preparation
# ***********************************************
rm(list=ls(all=FALSE))
setwd('Users/adelaideroguet/Desktop/R_Pulse/')

predata.BC<-read.table("XXX.txt", h=T)
data.BC <- sweep(predata.BC, 1, rowSums(predata.BC), '/'); apply(data,1,sum) # check

Date1<-read.table("Date.txt", header=TRUE, dec=',')
Date<-as.data.frame(Date1)

# Loop to determine which T-RFs fit the neutral model
# ***********************************************
dim_data<-ncol(data)
r2dXt=0

for(j in 1:dim_data){
p.value<-0
TRF<-as.data.frame(data[,c(j)])
dim<-nrow(TRF)-1

# Variable dX creation
dX=0
for (i in 1:dim) {
dX[i]<-(TRF[c(i+1),]-TRF[c(i),])
}

# Variable dt creation
dt=0
for (i in 1:dim) {
dt[i]<-(Date[c(i+1),]-Date[c(i),])
}

# Variable X creation
X=0
for (i in 1:dim) {
X[i]<-TRF[c(i+1),]
}

dXt<-dX/dt
Xweighted<-X; Xweighted[Xweighted==0]<-0.000000001
r2dXt[j]<-summary(lm(dXt~X, weights=(Xweighted*(1-Xweighted))^(-1)))$adj.r.squared
}
```

```
# Export the R-squared to observed which T-RFs in the dataset are well shaped by the neutral
model
write.table(r2dXt, '/Users/adelaideroguet/Desktop/r2dXt.txt')


# For the T-RFs that fit the neutral model
# ********************************************
para<-lm(dXt~X, weights=(Xweighted_161*(1-Xweighted_161))^(-1))

# Check residuals normality
shapiro.test(residuals(para))
hist(residuals(para))

intercept<-coefficients(para)[1] # m0
slope<-coefficients(para)[2] # m1
std_residuals_error<-sqrt(deviance(para)/df.residual(para)) # √(2/a)
a<-2/(std_residuals_error^2))
Ntm<-(-slope*a))
p<-(intercept*a)/Ntm)
```

PUBLICATIONS AND COMMUNICATIONS

Publications

Roguet A, Laigle GS, Therial C, Bressy A, Soulignac F, Catherine A, Lacroix G, Jardillier L, Bonhomme C, Lerch TZ and Lucas FS. Neutral community model explains the bacterial community assembly in freshwater lakes. *FEMS Microbiology Ecology* 91(11), fiv125.

Roguet A, Therial C, Saad M, Boudahmane L, Moulin L and Lucas FS. High mycobacterial diversity in recreational lakes. *Antonie van Leeuwenhoek* (in press).

Roguet A, Therial C, Bressy A, Catherine A, Lacroix G, Jardillier L and Lucas FS. Spatial distribution processes differ between bacterial taxonomic groups in freshwater lakes (in preparation).

Roguet A, Therial C, Bressy A, Tran V, Lemaire B, Dubois P, Catherine A, Lacroix G, Jardillier L, Moulin L and Lucas FS. Mycobacteria is everywhere, but, only the environment selects? (in preparation).

Radomski N, Roguet A, Lucas FS, Veyrier F, Cambau E, Accrombessi H, Moilleron R, Behr MA and Moulin L (2013) *atpE* gene as a new useful specific molecular target to quantify *Mycobacterium* in environmental samples. *BMC Microbiology*, 13:277.

Kaisermann A, Roguet A, Nunan N, Maron PA, Ostle N and Lata JC (2013) Agricultural management affects the response of soil bacterial community structure and respiration to water-stress. *Soil Biology and Biochemistry*, 66: 69-77.

Oral communications

Roguet A, Soulignac F, Scarati Martins R, Alves de Sousa Neto JN, Mendes dos Santos F, Lemaire BJ, Lucas FS and Vinçon-Leite B (2015) Modelling the impact of a stormwater inflow on *Escherichia coli* densities in an urban shallow lake. SEFS9 (Geneva, Swiss).

Roguet A, Thérial C, Maloufi S, Bressy A, Catherine A, Vinçon-Leite B, Jardillier L, Moulin L and Lucas FS (2014) Caractérisation des facteurs favorisant la présence de pathogènes humains dans les lacs urbains. JRI (Narbonne, France).

Roguet A, Thérial C, Boudahmane L, Saad M, Moulin L and Lucas FS (2013). Caractérisation des habitats favorables aux mycobactéries non tuberculeuses dans les lacs urbains. AFEM (Parent, France).

Posters

Roguet A, Thérial C, Boudahmane L, Saad M, Moulin L and Lucas FS (2013). Identification of non-tuberculous mycobacteria reservoirs in urban lakes facing global change pressures. SAME13 (Stresa, Italy).

Roguet A, Maloufi S, Bressy A, Catherine A, Vinçon-Leite B, Lacroix G, Jardillier L and Lucas FS (2012). Land use impact on *Escherichia coli* densities in urban and peri-urban lakes. JILO (Clermont-Ferrand, France).

Caractérisation des pressions anthropiques et environnementales influençant le compartiment bactérien dans les lacs peu profond

Bien que présentes dans tous les écosystèmes lacustres, la composition ainsi que l'abondance des communautés bactériennes peut varier à l'échelle régionale mais également à l'échelle locale. Une meilleure compréhension des facteurs responsables de ces patrons biogéographiques permettrait d'améliorer la connaissance des fonctionnements de ces systèmes aquatiques et donc de leur réponse potentielle face aux pressions anthropiques.
Dans ce contexte, cette thèse a visé à étudier la biogéographie du compartiment bactérien dans un ensemble de lacs peu profonds. Les objectifs principaux de cette étude ont été d'évaluer aux échelles régionale et locale les facteurs responsables des patrons biogéographiques pour (i) l'ensemble de la communauté bactérienne et (ii) pour un groupe bactérien spécifique, i.e. les mycobactéries non-tuberculeuses. Pour atteindre ces objectifs deux approches complémentaires ont été entreprises. Tout d'abord, la variabilité spatiale à l'échelle régionale a été évaluée par l'échantillonnage en Ile-de-France de 49 lacs pendant trois étés. Couplé à cette approche, une étude plus fine a été entreprise pour caractériser la dynamique spatiotemporelle du compartiment bactérien par le suivi mensuel pendant deux ans et de six évènements pluvieux importants au sein du lac de Créteil (Val de Marne).
À l'échelle régionale, la variabilité spatiale de la structure de la communauté bactérienne pour les trois années de suivi (caractérisé par T-RFLP) a été prédite à hauteur de 76% (r carré moyen) par les processus stochastiques et moins de 14% par les facteurs déterministes incluant les paramètres environnementaux (statut trophique) et les processus de dispersion (connexion du lac à une rivière et les axes PCNM). L'analyse de la composition de la communauté bactérienne par séquençage à haut débit (MiSeq Illumina) a mis en évidence des résultats similaires à ceux acquis par T-RFLP. Cependant, cette analyse a révélé que l'importance des processus impliqués dans les patrons biogéographiques pouvait évoluer en fonction des phylums ou des classes bactériens considérés. La variabilité spatiale des densités de mycobactéries (PCR en temps réel), était quant à elle expliquée à hauteur de 50% par des facteurs déterministes (pH de l'eau, concentration en fer labile et connexion à une rivière).
À l'échelle locale, le suivi du lac de Créteil n'a révélé aucune variation spatiale significative (le long des transects horizontal et vertical) de la structure de la communauté bactérienne ainsi que des densités de mycobactéries. Par contre une étude spécifique sur deux lacs a révélé des variations significatives de densité et la diversité des mycobactéries au sein de différents compartiments des lacs. À l'inverse, d'importantes variations temporelles de la structure des communautés bactériennes ont été observées au cours des deux années de suivi, principalement associées aux variations de température de l'eau. Par ailleurs, bien que relativement stable au cours des deux années de suivi, les variations de densités des mycobactéries ont seulement été prédites par les processus stochastiques à hauteur de 35%.

Mots-clés : biogéographie, communauté bactérienne, mycobactérie non-tuberculeuse, lac, pressions anthropiques, biologie moléculaire

Characterization of anthropogenic and environmental pressures influencing the bacterial compartment in shallow lakes

Although bacteria are widespread in lacustrine environments, their composition and abundance vary at the regional and also at the local scale. A better understanding of the factors responsible for these biogeographic patterns would improve our knowledge of these aquatic systems and thus their potential response to anthropogenic pressures.
In this context, this thesis studied the biogeography of the bacterial compartment in a set of shallow lakes located in the Paris area. The main objectives of this study were to assess at the regional and local scale the factors responsible for the biogeographical patterns on (i) the entire bacterial community, and (ii) a specific bacterial group, i.e. the nontuberculous mycobacteria. To achieve these objectives two complementary approaches were undertaken. First, at the regional scale, the spatial variability was assessed by sampling 49 lakes during three consecutive summers. A finer study was also performed to characterize the spatio-temporal dynamics of the bacterial compartment over a two-year monthly monitoring and during six important rain events within the Créteil Lake (Val de Marne).
At the regional scale, the spatial variability of the bacterial community structure for the three summers (assessed by T-RFLP) was predicted for 76% (mean r-squared) by stochastic processes and less than 14% by deterministic factors including environmental parameters (trophic status) and dispersal-related process (connection to a river and PCNM axes). The analysis of the bacterial composition by high-throughput sequencing (Illumina MiSeq) showed similar tendencies to those acquired by T-RFLP. However, this analysis revealed that the importance of the processes involved in biogeographical patterns could vary according to the bacterial phyla or classes considered. Spatial variability of mycobacterial densities (real-time PCR) was explained up to 50% by deterministic factors (water pH, amount of labile iron and connection to a river).
At the local scale, the monitoring of Créteil Lake revealed no significant spatial variation (along the horizontal and vertical transect) on the structure of the bacterial community and mycobacterial densities. However, a specific study of two lakes showed that mycobacterial density and diversity significantly varied among the different compartments of the lakes. Inversely, significant temporal variations on the bacterial community structure were observed over the two-year of monitoring, mainly related to water temperature changes. Although mycobacterial densities were relatively stable over the Créteil Lake monitoring, their variations were only predicted by stochastic processes up to 35%.

Key-words: biogeography, bacterial community, nontuberculous mycobacteria, lake, anthropogenic pressures, molecular biology