



**HAL**  
open science

# Spatio-temporal structuring of microbial communities in freshwater ecosystems

Gwendoline David

► **To cite this version:**

Gwendoline David. Spatio-temporal structuring of microbial communities in freshwater ecosystems. Ecosystems. Université Paris-Saclay, 2020. English. NNT : 2020UPASL049 . tel-03134024

**HAL Id: tel-03134024**

**<https://theses.hal.science/tel-03134024v1>**

Submitted on 8 Feb 2021

**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.

# Spatio-temporal structuring of microbial communities in freshwater ecosystems

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n° 577, Structure et dynamique des  
systèmes vivants (SDSV)

Spécialité de doctorat : Sciences de la vie et de la santé

Unité de recherche : Université Paris-Saclay, CNRS, AgroParisTech,  
Ecologie Systématique et Evolution, 91405, Orsay, France.

Référent : Faculté des sciences d'Orsay

**Thèse présentée et soutenue à Orsay, le 9 décembre 2020, par**

**Gwendoline DAVID**

## Composition du jury :

<b>Stéphane Bazot</b> Professeur, Université Paris-Saclay	Président
<b>Urania Christaki</b> Professeure, Université du Littoral Côte d'Opale	Rapporteuse et examinatrice
<b>Jean-François Humbert</b> Directeur de recherche, INRAE	Rapporteur et examinateur
<b>Karin Rengefors</b> Professeure, Lund University	Examinatrice
<b>Ludwig Jardillier</b> Professeur, Université Paris-Saclay	Co-Directeur de thèse
<b>Purificación López-García</b> Directrice de recherche, Université Paris-Saclay	Co-Directrice de thèse
<b>Elisa Thébaud</b> Chargée de Recherche, Sorbonne Université	Co-Encadrante de thèse
<b>Gwendal Restoux</b> Chargé de Recherche, INRAE	Co-Encadrant de thèse



*A ylb,*



## Remerciements

Ça y est, c'est la fin d'une belle aventure. Les microorganismes y auront joué un rôle prépondérant, mais il faut bien le dire, sans des macroorganismes de qualité autour de moi, ces quelques années n'auraient pas été les mêmes...

First, I would like to thank my jury members, who kindly accepted to evaluate my work. Many thanks to my reviewers Urania Christaki and Jean-François Humbert, and to my examiners, Karin Rengefors and Stéphane Bazot.

Merci à ma grande équipe d'encadrants. Un grand merci tout d'abord à Ludwig pour m'avoir donné l'opportunité de travailler sur ce sujet passionnant. Je te remercie pour m'avoir accompagnée dans mes premiers pas de chercheuse, pour m'avoir encouragée et conseillée pendant ces quatre années. Un grand merci également à Puri, pour ta détermination, ta rigueur scientifique, tes encouragements et pour nous pousser toujours un peu plus loin. Et enfin, un grand merci à Elisa et Gwendal pour avoir été disponibles dès que j'en avais besoin. J'ai beaucoup apprécié les journées de travail à vos côtés à Jussieu et à Jouy. Merci pour tous vos conseils et votre aide. Ce travail n'aurait pas été le même sans vous.

Merci aux membres de mon comité de thèse, Julie Leloup, Sébastien Barot et Didier Casane, pour vos conseils, votre intérêt et votre vision extérieure lors de nos réunions annuelles.

Merci à Pierre Capy et Sandrine Le Bihan pour le suivi du côté de l'École Doctorale, pour leur patience et leur aide dans la résolution de tous les soucis techniques rencontrés pendant la thèse.

Un immense merci à Pao, pour avoir toujours été là. Je me rappelle avoir dit à ma copine Laure en arrivant au labo, "Pao, je pense que c'est la personne qui remonte le moral quand ça ne va pas". Et au final, tu as été bien plus que ça. Merci pour les sorties de terrain, les très nombreuses filtrations, ta pédagogie, mais surtout ton amitié, les mille tours de la retenue de Bures le midi, les conseils littérature, théâtre, cinéma et j'en passe. Les pauses dans ton bureau me manquent déjà (mais surtout toi) !

Merci à mes copains de promo de thèse. Luis, thank you for being the best desk neighbour, for

your crazyness, for your curiosity for all living beings, and for being such a good friend. Guillaume, pas seulement mon co-bureau mais aussi mon coloc, apparemment, se voir tous les jours au labo ne suffisait pas... Merci pour ta gentillesse, pour ta bonne humeur, pour le poivre de sapin et bien sûr pour tes plats trop pimentés. Merci à Jodie, pour toutes tes folles aventures qui auront bien rythmé nos journées, mais surtout pour ton enthousiasme et ton soutien.

Merci à Maria, pour le chocolat, pour le réconfort, pour les repas du midi dans le jardin. Cette fin de thèse n'aurait vraiment pas été la même sans ta présence et tes attentions quotidiennes. Merci d'avoir pris soin de moi.

Merci à tout le reste de la d(r)eam team ! Thomas, merci pour les jeux de mots, l'inventivité, les gadgets improbables et les qualités de producteur/réalisateur. Tout ça va bien me manquer. Fabian, pour les "C'est la vie". Guifré, merci pour nous avoir accueillis dans ton bureau au début de la thèse et pour ta présence tout du long. Rafael (Rrrafouu), pour ton aide à nos débuts avant que tu ne nous abandonnes et pour ta gentillesse ! Miguel, merci pour ton enthousiasme de tous les jours. Julien, merci pour toutes les discussions, scientifiques ou non et pour le partage de tes réflexions. Merci à Laura, Philippe, Naoji, Ana, Jolien. Merci à Brittany, Jazmin et Ferial, bon courage pour vos thèses, girls !

Merci à tous les membres du laboratoire Ecologie, Systématique et Evolution. Ce fut enrichissant et agréable de passer ces quelques années avec vous. Et merci à Emmanuelle pour ta bonne humeur, ta patience et ton aide inestimable face aux démarches administratives !

Merci à Judith Legrand, Nicolas Delpierre et Stéphane Bazot de m'avoir fait confiance pour prendre en charge leurs TPs et TDs. Cette expérience de monitorat aura été très formatrice pour moi. Et un grand merci à Marianne pour sa bonne humeur et sa gentillesse, et pour avoir fait en sorte que tous mes enseignements se passent toujours au mieux.

Merci à Antoschka, Frrred et Lefe et à toute l'équipe Kickass Якутя pour m'avoir emmenée au fin fond de la Sibérie pour une mission incroyable et dépaysante. Merci mille fois de m'avoir donné une telle opportunité, je n'aurais jamais pensé dans ma vie avoir l'occasion de couper des m<sup>2</sup> d'herbe aux ciseaux assise sur des polygones.

Merci à Jean-Nicolas, Laurent et Cybill pour m'avoir encadrée pendant mon stage de master 2. Je n'aurais pu espérer mieux comme stage de recherche : du terrain, des manips, de l'analyse de données et de la motivation pour rédiger mon premier article. Cependant, pas du tout merci à Laurent pour la déformation, maintenant je vois des cymbalaria partout et tout le temps.

Vielen Dank Bernd! You welcomed me during 3 months in Magdeburg, and it was a great experience, professionally and humanely.

Mille mercis à Maman pour croire en moi depuis toute petite, me motiver toujours un peu plus et bien sûr pour toutes les relectures attentives, à Pap's pour tes idées farfelues, à Leslie (le yang de mon yin) pour la vie parisienne et les sorties culturelles entre poule poule, à ma Mamie pour les confitures et la couture, à ma Grand-mère pour ses "c'est comme ça", à mon Grand-père, à Marie pour les hérissons et les conseils lectures, à Béné pour sponsoriser ma vaisselle, à Simon (la colocation fut brève, mais intense), Didi, Mat, Mélou d'être des germains si chouettes, à Vincent pour le bateau. A Saïf pour le tofu aux olives et le sticky rice (promis, pas que pour ça). A Laurent et Elisabeth pour les repas à Gif, les coings et votre présence. A Bruno pour sa merveilleuse statuette de Dodo. Merci à Joëlle et Marc pour les expos et les séjours sur Ré. Merci à Francis pour toutes les vacances et à Izia pour les dessins.

Merci à tous les copains sans qui la vie serait quand même drôlement moins chouette ! A ma Nenni si précieuse qui me supporte depuis 15 ans. A ma Astrid (ou Astroud pour les intimes), la meilleure des dentistes, pour les moments rennais passés et à venir. A mes acolytes princesses et à nos week-ends entre copines : Clémence, pour ta bonne humeur h24, le vélo/vin, les soirées-douche à Moulins, les couronnes de fleurs et les plus belles vacances (liste clairement non exhaustive); Laure, mon double, pour les post-it, les calendriers de l'avent, la passion pour Pio, le soutien indéfectible et tant d'autres choses. A ma Manon, merci pour ta présence quotidienne (coucou copine !), pour le serpent dans la botte, pour Margaux et pour toutes les aventures qui arrivent prochainement... A Lisa, pour la collocation à Massy, les licornes, toutes les soirées jeux et pour ta folie, surtout. A Baptou et Laure, les deux Mours, pour l'accueil toujours trop bien, les appels pour dire des blagues et le parrainage de Bobby. A la pitchou et associés, Roro (et Denis), Lolo (les pousse-Doline me manquent presque), Aline (trop hâte de venir dans ton château bb) et Coup-coup (la meilleure créatrice de sauces). A Lucas et ses cuillères, Constant (dommage qu'on puisse pas imiter Camille à l'écrit), Jéroums et tous les copains/copines de l'ENGEES. Au super her'eau de l'ombre pour sa subtilité, à Thomas et Emilie, à Céline, à Théo et son bob, à Pierre. A Justine, parce que je sais que tu n'es jamais bien loin de moi. Thanks to Siri, and I hope we will see each other in Berlin! Aux copains du lycée, Charles et ses leçons d'architecture de nuit, Gaëlle, Arielle, Laurent, et soon to be Dr. Foni. Merci à Christoph, François, Philippe et Prunelle pour m'avoir permis de reprendre la musique de la meilleure façon qui soit.

Pour terminer, un très grand merci à Maxime. Merci d'être toujours là à mes côtés pour me motiver et croire en moi. Merci de ton aide inestimable sur R, pour avoir toujours plein d'idées à proposer. Merci d'accepter toutes mes lubies sans me juger. Merci pour les macarons et la cuisine en général, les plantes, les vehiculo longo, les montées et descentes d'escaliers, les soirées jeux, ta gentillesse, ta curiosité, ton humour et j'en passe. En résumé... merci pour la vie ensemble !

Et bien évidemment, je remercie chaleureusement Par Jupiter, les colorriages et les dinosaures.





# Contents

Table of contents . . . . .	i
List of figures . . . . .	v
List of tables . . . . .	xi
<b>1 General Introduction</b>	<b>1</b>
1.1 Overview of aquatic ecosystems dynamics . . . . .	5
1.1.1 Characteristic of aquatic ecosystems . . . . .	5
1.1.2 Spatial changes . . . . .	6
1.1.3 Temporal changes . . . . .	8
1.1.4 Functional specificities of small freshwater ecosystems . . . . .	9
1.2 Microorganisms are key components of aquatic ecosystems . . . . .	11
1.2.1 Wide microbial diversity in freshwater . . . . .	11
1.2.2 Implications of microorganisms in ecosystem functioning . . . . .	16
1.3 How to elucidate ecosystem functioning? . . . . .	19
1.3.1 Abiotic factors influence microorganisms community composition . . . . .	19
1.3.2 Interactions between microorganisms shape communities . . . . .	22
1.3.3 Microbial communities are structured in space and time . . . . .	25
1.4 Aims of the thesis . . . . .	32
<b>2 Materials and methods</b>	<b>35</b>
2.1 Presentation of the ecosystems . . . . .	37
2.1.1 Lake Baikal . . . . .	37
2.1.2 Small freshwater ecosystems in the Parc Naturel Régional de la Haute- Vallée de Chevreuse . . . . .	38
2.2 Data collection . . . . .	40
2.2.1 Spatial survey - Lake Baikal . . . . .	40
2.2.2 Temporal survey - Small freshwater ecosystems . . . . .	40
2.3 Samples preparation and processing . . . . .	42

2.3.1	Extraction and amplification of DNA . . . . .	42
2.3.2	Sequencing . . . . .	43
2.3.3	Processing of the sequences . . . . .	43
2.4	Data analysis . . . . .	45
2.4.1	Multivariate statistical analysis . . . . .	45
2.4.2	Co-occurrence networks . . . . .	50
<b>3</b>	<b>Environmental drivers of plankton protist communities along latitudinal and vertical gradients in the oldest and deepest freshwater lake</b>	<b>53</b>
3.1	Context and objective . . . . .	55
3.2	Manuscript . . . . .	56
3.2.1	Abstract . . . . .	56
3.2.2	Introduction . . . . .	57
3.2.3	Material and methods . . . . .	58
3.2.4	Results and discussion . . . . .	61
3.2.5	Concluding remarks . . . . .	71
	<b>Supplementary Material</b>	<b>74</b>
3.A	Supplementary Tables . . . . .	74
3.B	Supplementary Figures . . . . .	83
<b>4</b>	<b>Small freshwater ecosystems with dissimilar microbial communities exhibit similar temporal patterns</b>	<b>93</b>
4.1	Context and objective . . . . .	95
4.2	Manuscript . . . . .	96
4.2.1	Abstract . . . . .	96
4.2.2	Introduction . . . . .	97
4.2.3	Material and methods . . . . .	98
4.2.4	Results . . . . .	102
4.2.5	Discussion . . . . .	110
	<b>Supplementary Material</b>	<b>117</b>
4.A	Supplementary Tables . . . . .	117
4.B	Supplementary Figures . . . . .	124

<b>5</b>	<b>Dissimilar protist and prokaryote communities display similar intra- and inter-annual temporal patterns in small freshwater ecosystems</b>	<b>133</b>
5.1	Context and objective . . . . .	135
5.2	Manuscript . . . . .	136
5.2.1	Abstract . . . . .	136
5.2.2	Introduction . . . . .	136
5.2.3	Material and Methods . . . . .	138
5.2.4	Results . . . . .	143
5.2.5	Discussion . . . . .	153
	<b>Supplementary Material</b>	<b>160</b>
5.A	Supplementary Tables . . . . .	160
5.B	Supplementary Figures . . . . .	171
<b>6</b>	<b>Topology of seasonal co-occurrence networks reveals similar community structuration in different small freshwater ecosystems</b>	<b>183</b>
6.1	Context and objective . . . . .	185
6.2	Manuscript . . . . .	186
6.2.1	Abstract . . . . .	186
6.2.2	Introduction . . . . .	186
6.2.3	Material and methods . . . . .	189
6.2.4	Results . . . . .	192
6.2.5	Discussion . . . . .	200
6.2.6	Conclusion . . . . .	203
	<b>Supplementary Material</b>	<b>205</b>
6.A	Supplementary Tables . . . . .	205
6.B	Supplementary Figures . . . . .	215
<b>7</b>	<b>General Discussion</b>	<b>225</b>
7.1	Microbial community composition in freshwater ecosystems . . . . .	227
7.1.1	Global composition of microbial communities . . . . .	228
7.2	Spatial changes in microbial community composition . . . . .	230
7.2.1	Effect of depth on microbial communities . . . . .	230
7.2.2	Variation of microbial communities depending of the sampling station . .	230
7.3	Temporal evolution of microbial communities . . . . .	231

7.4	Exploration of the interactions in the communities . . . . .	234
7.4.1	Spatio-temporal structure of co-occurrence networks . . . . .	234
7.4.2	Distribution of the functional groups in the ecosystems . . . . .	237
<b>8</b>	<b>Perspectives and Conclusions</b>	<b>239</b>
8.1	Perspectives . . . . .	241
8.1.1	Explore the missing pieces in ecosystem functioning . . . . .	241
8.1.2	Determine the proper sampling frequency . . . . .	244
8.1.3	Focus on Candidate Phyla Radiation . . . . .	245
8.2	Conclusions . . . . .	247
	<b>Bibliographie</b>	<b>251</b>
	<b>Résumé en français</b>	<b>289</b>
	<b>Curriculum Vitae</b>	<b>299</b>

# List of Figures

1.1	Extract of a book written by L. Joblot and published in 1754 (posthumously)	3
1.2	Zonation of marine systems and freshwater lakes	7
1.3	View of the phylogenetic tree of life	12
1.4	Rank-abundance curve representing the abundance of individuals ( <i>e.g.</i> OTUs) ranked according to their abundance	13
1.5	Representation of the marine food webs showing the role of microbial communities.	17
1.6	Possible biological interactions and their effects between microorganisms.	23
1.7	Illustration of the scales used to investigate microbial community structuration	26
1.8	Variation of the environmental parameters in the water column in a subtropical ocean	27
1.9	Example of a spatial survey along both vertical and horizontal gradients	28
1.10	Example of analysis based on long temporal surveys	30
2.1	Location of Lake Baikal, in Siberia (Russia)	37
2.2	Location of the five studied small freshwater ecosystems	39
2.3	Possible interpretations of the linear model applied on the time-lag analysis	49
3.1	Sampling sites and overall planktonic protist community composition in Lake Baikal	63
3.2	Marine signature taxa detected in Lake Baikal plankton	66
3.3	Non-metric multidimensional scaling (NMDS) analysis of Lake Baikal plankton samples as a function of protist community similarities	69
3.4	Box plots showing the distribution of relative abundances of major functional categories of protists in Lake Baikal plankton	70
3.5	Co-occurrence networks of planktonic protists 700 in the Lake Baikal water column	72
3.B.1	Box plots of physico-chemical parameters measured in Lake Baikal water samples	84
3.B.2	Accumulation curves for OTUs identified in Lake Baikal plankton samples before and after rarefaction	85

3.B.3	Box plots showing diversity (richness) and evenness values in Lake Baikal water samples as a function of basin and depth . . . . .	86
3.B.4	Rank-abundance curves of protist OTUs in Lake Baikal plankton samples . . . .	87
3.B.5	KRONA representation of the global diversity of Holomycota in Lake Baikal plankton . . . . .	88
3.B.6	Non-metric multidimensional scaling (NMDS) analysis of Lake Baikal plankton samples based on differences in protist community composition . . . . .	89
3.B.7	Non-metric multidimensional scaling (NMDS) analysis of Lake Baikal plankton samples as a function of protist community similarities based on SWARM-determined OTUs . . . . .	90
3.B.8	Block-model representation of the networks of the planktonic protists in Lake Baikal for the three major depth categories . . . . .	91
4.2.1	(a) Cumulative curves showing the total number of prokaryotic OTUs detected in each ecosystem from April 2011 until April 2013 during the temporal survey in the five small freshwater ecosystems; (b) Richness and diversity estimators computed for all the samples; (c) Histograms showing the relative abundance of 16S rDNA amplicon reads assigned to high-rank taxa averaged over the two-year survey . . . . .	103
4.2.2	Venn diagram showing the number of OTUs shared by several ecosystems or specific to an ecosystem . . . . .	105
4.2.3	Canonical correspondence analyses conducted on all OTUs affiliated either to microbial eukaryotes or prokaryotes retrieved in the five ecosystems over the two-year survey . . . . .	107
4.2.4	Time-lag analyses (TLA) were computed for each ecosystem between sampling dates over increasing time-lags over the two-year survey (from one-month to 23-months) . . . . .	109
4.2.5	Wisconsin standardized pairwise Bray-Curtis dissimilarities were calculated between samples separated from 1 to 12 months for abundant and rare OTUs belonging to microbial eukaryotes, prokaryotes and whole microbial community detected in the five ecosystems over the two-year survey . . . . .	111
4.B.1	Rank-abundance histograms of all OTUs detected in the five ecosystems (A), OTUs with a mean relative abundance of at least 0.5% (B) and rank-abundance histograms of taxonomic groups in each small freshwater ecosystem . . . . .	125

4.B.2	Histograms of monthly relative proportions of 16S rDNA amplicon reads assigned to high-rank taxa . . . . .	126
4.B.3	Histograms of average relative proportions of 16S rDNA amplicon reads assigned to archaeal high-rank taxa, from the five ecosystems over the 2-years survey . .	127
4.B.4	Non-metric multidimensional scaling (NMDS) plot, built on Wisconsin standardized Bray–Curtis dissimilarities between combined microbial eukaryote and prokaryote OTUs of all samples collected in the five ecosystems over the 2-years survey . . . . .	128
4.B.5	Hierarchical cluster analysis build using Wisconsin standardized Bray-Curtis dissimilarities between every samples collected over the 2-years survey for the five small freshwater ecosystems . . . . .	129
4.B.6	Canonical correspondance analysis (CCA) conducted on abundant microbial eukaryote (A) and prokaryote (B) Wisconsin-standardized OTUs retrieved in the five systems over the 2-years survey . . . . .	130
4.B.7	Non-metric MultiDimensional Scaling (NMDS) plots, based on Bray-Curtis distances computed after Wisconsin standardization of all OTU frequencies in samples from the five ecosystems . . . . .	131
4.B.8	Box plots of turnover rates computed for rare and abundant OTUs affiliated to microbial eukaryotes and prokaryotes in the five small freshwater ecosystems over the 2-years survey. . . . .	132
5.2.1	Time-lag analyses (TLA) were computed for each ecosystem between sampling dates over increasing time lags over the 8-year survey (from one-month until 29-month) . . . . .	144
5.2.2	Histograms showing the relative abundance of 16S and 18S rDNA amplicon reads assigned to high-rank taxa average over the 8-year survey in the five ecosystems. . . . .	146
5.2.3	Venn diagram showing the number of rarefied OTUs over the 8-year survey; (B) Non-metric multidimensional scaling plot constructed with Wisconsin-standardized Bray-Curtis dissimilarities applied on rarefied data between all samples . . . . .	148
5.2.4	Co-occurrence network inferred by the PLN approach and representing the global structure of microorganisms over the five ecosystems . . . . .	159
5.B.1	Temporal dynamics of the physico-chemical parameters and the chlorophyll a concentration measured over the 8-year survey . . . . .	172
5.B.2	Principal Component Analysis (PCA) on all physico-chemical parameters measured in the five ecosystems during the 8-year survey . . . . .	173



5.B.3(A) Richness and evenness were computed on the rarefied microbial eukaryotic and prokaryotic communities in the samples collected during the 8-year survey in the five ecosystems; (B) Cumulative curves showing the total number of eukaryotic and prokaryotic rarefied new OTUs detected in each ecosystems . . . .	174
5.B.4 Rank-abundance histograms of all rarefied OTUs in the entire dataset for microbial eukaryotes and prokaryotes . . . . .	175
5.B.5 Box plots of turnover rates computed for rare and abundant rarefied OTUs, affiliated to microbial eukaryotes and prokaryotes in the five ecosystems between samples regrouped by season . . . . .	176
5.B.6(A) Successive turnover rates computed for rare and abundant rarefied OTUs, affiliated to microbial eukaryotes and prokaryotes in the five ecosystems between samples regrouped by sampling years; (B) Mean value of the turnover rate for each ecosystem . . . . .	177
5.B.7 Pie-chart summarizing the results of Canonical correspondence analyses (CCA) conducted on all rarefied OTUs affiliated either to microbial eukaryotes or prokaryotes in all samples collected during the 8-year survey . . . . .	178
5.B.8 Non-metric Multi-Dimensional Scaling (NMDS) plots, based on pair-wise Bray-Curtis dissimilarities computed after Wisconsin standardization of rarefied OTU in samples for each of the five ecosystems . . . . .	179
5.B.9 Histograms of the recurrent OTUs per ecosystem in terms of diversity and abundance . . . . .	180
5.B.10 Block-model representation of the co-occurrence network inferred by the PLN approach and representing the global structure of microorganisms over the five ecosystems . . . . .	181
5.B.11 Module representation of the co-occurrence network inferred by the PLN approach and representing the global structure of microorganisms over the five ecosystems . . . . .	182
6.2.1 Seasonal co-occurrence networks over the eight-year survey in the four studied ecosystems . . . . .	193
6.2.2 Principal Component Analysis (PCA) of the topological properties of the seasonal network . . . . .	195
6.B.1 Block-model representation of the seasonal networks over the eight-year survey in the four ecosystems . . . . .	216

6.B.2	Barplots presenting the OTU composition of the block model groups (BMG) for the season networks of the Etang des Vallées . . . . .	217
6.B.3	Barplots presenting the OTU composition of the block model groups (BMG) for the season networks of La Claye . . . . .	218
6.B.4	Barplots presenting the OTU composition of the block model groups (BMG) for the season networks of Mare Gabard . . . . .	219
6.B.5	Barplots presenting the OTU composition of the block model groups (BMG) for the season networks of Ru Sainte Anne . . . . .	220
6.B.6	Representation of the links between network topology and diversity indexes . . .	221
6.B.7	Histogram showing the composition of the eukaryotic nodes in the networks . . .	222
6.B.8	Histogram showing the composition of the prokaryotic nodes in the networks . .	223
7.4.1	Co-occurrence network of the microbial communities in Mare Gabard in spring using three different methods . . . . .	236



# List of Tables

2.1	General characteristics of the studied ecosystem . . . . .	38
2.2	Example of a reduced data-frame used for the analyses, exposing the read numbers for each samples . . . . .	45
3.1	Permutational multivariate analyse of variance (PERMANOVA) of Lake Baikal plankton samples across basin, sampling site and depth . . . . .	67
3.A.1	Sampling sites in Lake Baikal water column and associated physico-chemical, sequence and diversity data . . . . .	75
3.A.2	Identification, phylogenetic affinity and abundance of eukaryotic OTUs identified in Lake Baikal plankton . . . . .	77
3.A.3	Diversity, abundance and distribution of eukaryotic lineages previously thought to be exclusively marine identified in Lake Baikal plankton . . . . .	78
3.A.4	ANOSIM analyses between pairs of Lake Baikal plankton sample groups defined as a function of depth, basin of origin and coastal vs. pelagic location . . . . .	79
3.A.5	Tentative classification of phylogenetic lineages within broad functional categories	80
3.A.6	Functional and phylogenetic classification of OTUs used for network construction	81
3.A.7	Properties of networks built upon Lake Baikal plankton OTUs for the three depth categories of the water column . . . . .	82
4.2.1	Permutational multivariate analyses of variance (PERMANOVA) partitioning of the whole microbial community, prokaryotes and microbial eukaryotes, based on Bray-Curtis dissimilarities calculated using Wisconsin standardized OTU relative abundances among hierarchical-defined clusters 1 and 2 . . . . .	108
4.A.1	Names, numbers of reads, OTUs and diversity estimates obtained for the different samples . . . . .	118
4.A.2	Number of abundant and all OTUs affiliated to prokaryotes and microbial eukaryotes in the five ecosystems and in the five ecosystems considered together . .	121

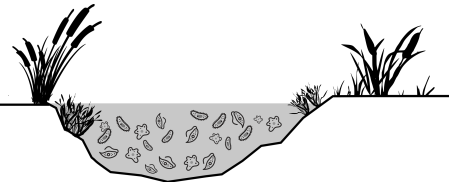
4.A.3	Linear and polynomial models were computed based on time-lag analyses (TLA) between sampling dates over increasing time-lags over the 2-years survey (from one-month to 23-months) for the five ecosystems . . . . .	122
4.A.4	RV index were calculated based on co-inertia analyses to assess co-dynamic between prokaryotes and microbial eukaryotes in the five ecosystems and in all ecosystems considered together . . . . .	123
5.2.1	Repartition of all the rarefied OTUs in 4 categories, with the percentage of OTUs and their total relative abundance within the categories for microbial eukaryotes and prokaryotes . . . . .	151
5.A.1	Summary of the physico-chemical composition of the five ecosystems . . . . .	161
5.A.2	Description of the samples used in the analyses. Names, number of rarefied microbial eukaryotic and prokaryotic OTUs, evenness for microbial eukaryotes and prokaryote and number of reads before rarefaction for microbial eukaryotes and prokaryotes . . . . .	162
5.A.3	Total number of OTUs, mean number of OTUs and number of abundant OTUs per ecosystem for microbial eukaryotes and prokaryotes . . . . .	166
5.A.4	ANOSIM analysis between each pair of ecosystem based on Bray-Curtis dissimilarities computed after Wisconsin standardization of rarefied OTUs (9999 permutations) . . . . .	166
5.A.5	Permutational multivariate analyses of variance (PERMANOVA) calculated using Wisconsin standardization on rarefied OTUs for microbial eukaryotes and prokaryotes . . . . .	167
5.A.6	Linear models were computed based on the time-lag (TLA) calculated between sampling dates over increasing time-lags over the 8-year survey (from one-month to 30-month) for the five ecosystems . . . . .	168
5.A.7	Canonical correspondence analyses (CCA) conducted on all rarefied OTUs affiliated either to microbial eukaryotes or prokaryotes and on physico-chemical parameters in the samples collected during the 8-year survey in each ecosystem .	169
5.A.8	ANOSIM analysis conducted for each ecosystems between each pair of season, based on Bray-Curtis dissimilarities computed after Wisconsin standardization of rarefied OTUs (9999 permutations) . . . . .	170
6.2.1	Summary of the seasonal co-occurrence networks properties in the four ecosystems	194
6.A.1	Linear regressions conducted between network properties and diversity indexes .	206

6.A.2	Summary of the seasonal co-occurrence networks properties in the four ecosystems	207
6.A.3	Summary of the five most connected nodes in the seasonal co-occurrence networks for each ecosystem . . . . .	208
6.A.4	Output of the “degree” model, quantify the effect of both the node’s lifestyle and the season . . . . .	209
6.A.5	Details of the pairwise comparison of the “degree” model between the season within the different ecosystems . . . . .	210
6.A.6	Summary of percentage of the type of interactions developed in the seasonal co-occurrence networks for each ecosystem . . . . .	211
6.A.7	Output of the “parasite” model, quantifying the effect of the season on the percentage of edges established with putative parasitic OTUs in networks . . . . .	214



# Chapter 1

## General Introduction



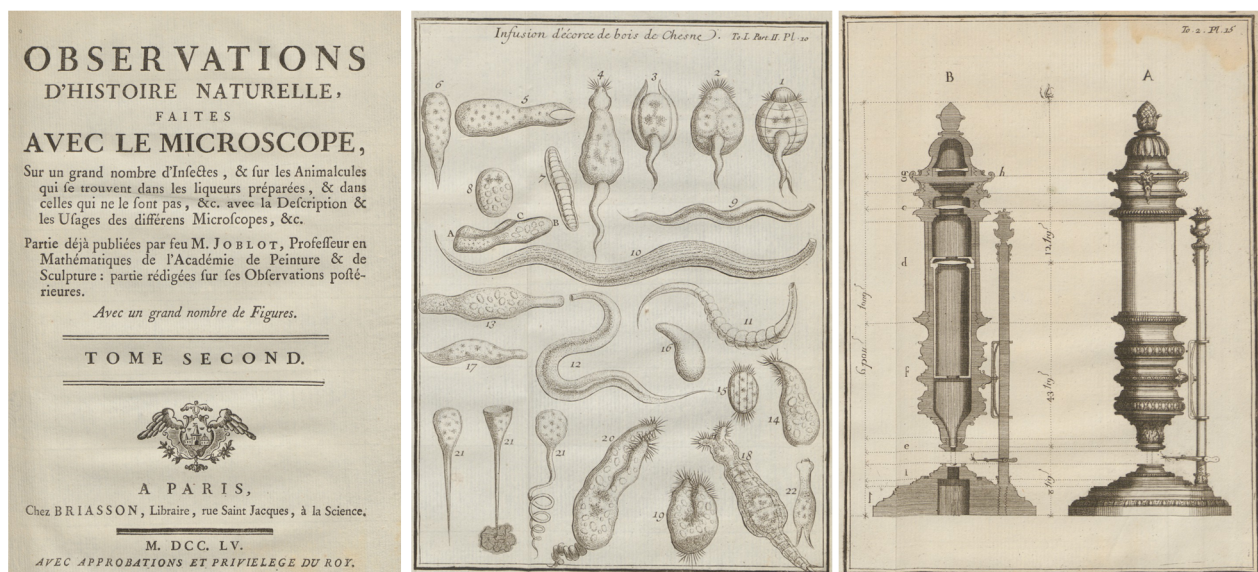
“Despite perhaps being regarded as the cinderellas of aquatic science, temporary waters represent significant components of the global landscape.”

*D. Williams*





Since the mid-seventeenth, microbial communities have been explored in a large range of environments. The first observations of microorganisms in the nature were only descriptive and resulted from serendipity and were conducted by two scientists. In 1674, Antoni van Leeuwenhoek (1632-1723), a draper from Delft, built a hand-made microscope to investigate the quality of its fabrics (Corliss, 1975). Fascinated by his observations, he investigated various environments, in particular freshwater ponds, and was the first to describe protists, that is microbial eukaryotes. At that time, they were named “*animalculum*”, from the Latin “tiny animal”. In France, Louis Joblot (1645-1723) a mathematician, built microscopes and published drawings of his observations in 1718 and in several publications later (see Figure 1.1, Lechevalier, 1976). The microorganisms observed at that time were among the largest ones, and nowadays we know that they are covering an important size fraction, ranging from 0.2 to 100  $\mu\text{m}$  (Azam and Malfatti, 2007). From this period, the fascination for microorganisms, also called the “unseen majority” did not stop growing. Indeed, to decipher the diversity of microbial communities, it is essential to investigate all types of environments. Therefore, even if aquatic systems have been among the first ones investigated because water is crucial for life, various types of ecosystems have been and are investigated nowadays (*e.g.* soil, digestive track, extreme environments).



**Figure 1.1** *Left panel*: Front page of the book written by L. Joblot and published in 1754 (posthumously), with the mention of “animalcules”; *Middle Panel*: Illustrations of L. Joblot, describing his microscopic observations of an infusion of oak bark. *Right panel*: Technical drawing of one of the Joblot’s microscopes (Source of the images: BNF, Bibliothèque Nationale de France, <https://gallica.bnf.fr>)

Aquatic ecosystems are still widely studied, because of their major importance at different levels. At the global scale, aquatic ecosystems are important reservoirs of liquid water. They are known to regulate climate, and to have a role in biogeochemical cycles (Worden et al., 2015;

Cavicchioli et al., 2019). For instance, oceanic investigations revealed that 50% of the photosynthesis on Earth is conducted in marine waters (Seymour et al., 2017). Aquatic ecosystems offer various types of habitat, and harbour a wide diversity of life forms, from mammals to bacteria. At a local and anthropogenic scale, they provide diverse ecosystem services. Among others, they are source of drinking water, of food through the fishes, and allow the irrigation of crops. Water-bodies also constitute commercial paths, as well as recreational areas. However, the study of aquatic environments is challenging, as they are constituted of a wide variety of systems, including for instance marine coastal areas, wide lakes, deep ocean, rivers, or small ponds.

To understand the functioning of ecosystems, it is important to describe the kind of organisms that occur, to decipher the spatial distribution and temporal dynamics of organisms, and to apprehend their interactions with other organisms (biotic interactions) and their environment (abiotic interactions). This is what I addressed in my thesis, by focusing on microorganisms of the three domains of life in Lake Baikal and in small freshwater ecosystems. In this introduction, I first depict the diversity of aquatic ecosystems on Earth. Then, I expose the importance of microorganisms in these environments, in terms of diversity and function. Finally, I present the different factors influencing microbial communities, as well as the importance of considering aquatic ecosystems as dynamic environments, that present both spatial and temporal variations.

## 1.1 Overview of aquatic ecosystems dynamics

Three quarters of the surface of Earth, also called Blue Planet, are covered by water. Marine systems contain 97% of water, while the remaining 3% is terrestrial and defined as freshwater. Water on Earth is constantly renewed through the hydrological cycles and circulates from one reservoir to another. The study of marine systems is the domain of oceanography, from the Greek word “*okeanos*” (Pinet, 2019). Ocean exploration started with the Greeks by 4000 BC, but first expeditions to investigate marine systems and more particularly marine plankton were conducted by Charles Darwin (1809-1882). Indeed, the expeditions on the *Beagle* and the *Challenger* (1835, 1872-1876) allowed him to discover radiolarians and diatoms. Oceanography covers the investigation of the whole marine volume, from surface water to deep sea and hydrothermal vents. The study of inland freshwater systems is called limnology, from the Greek word “*limne*” meaning the lake. This term was proposed by F.A. Forel in 1904 following his work on Lake Geneva (Forel, 1904). It covers the investigation of all freshwater systems, both lentic (*i.e.* stagnant) and lotic (*i.e.* running).

### 1.1.1 Characteristic of aquatic ecosystems

Two of the major differences between marine and inland freshwater systems are the salinity and the nutrient state. Marine environments cover up to 70% of Earth surface, and they are characterized by their salinity, which is relatively stable around 35‰ on average (Madsen, 2015). They have a weak nutrient load, and are thus mainly oligotrophic except in up-welling zones where the ecosystem productivity is high (Madsen, 2015). Freshwater systems are defined by the USGS (United States Geological Survey) as a water system containing less than 1000 mg/L of dissolved solids. In comparison to marine systems, freshwater systems are more diversified and they present a wide range of nutrient states, from oligotrophic to highly eutrophic waters.

Freshwaters can be divided into lentic and lotic ecosystems. In addition, they can be divided into size-categories (*e.g.* “small” or “large”), but are not accurate because there is an environmental continuum of freshwater ecosystems sizes (Wellborn et al., 1996). Freshwater ecosystems are also classified based on their lifespan. For instance, lentic ecosystems range from small ephemeral ponds generated after a rain event to large and profound lakes formed millions years ago (Wellborn et al., 1996). Traditionally, lakes are opposed to ponds, even if their definitions remain unclear. Indeed, some authors limit the ponds to water-bodies with a surface  $\leq 0.001 \text{ km}^2$  (Downing, 2010), while the European Pond Conservation Network (EPCN) goes up to  $0.02 \text{ km}^2$  width and a maximal depth of 8 m (Indermuehle et al., 2008).

Therefore, the definition of pond covers a large variety of ecosystems, from the small puddle created after the rain to larger permanent water-bodies (Biggs et al., 2017). In the case of lotic systems, small ecosystems correspond to small tributaries and brooks, whereas large ecosystems correspond to rivers presenting a wide floodplain. As for ponds, there is no clear definition of small streams. They can be defined as the small order streams, according to the Strahler's classification (Strahler, 1957; Moss, 2010). For instance, a stream with no tributaries, that is of order 1, is a small stream. Their width is also often used to classify them. The small lotic systems are generally defined as having a maximal width of 1.5 m (Biggs et al., 2017).

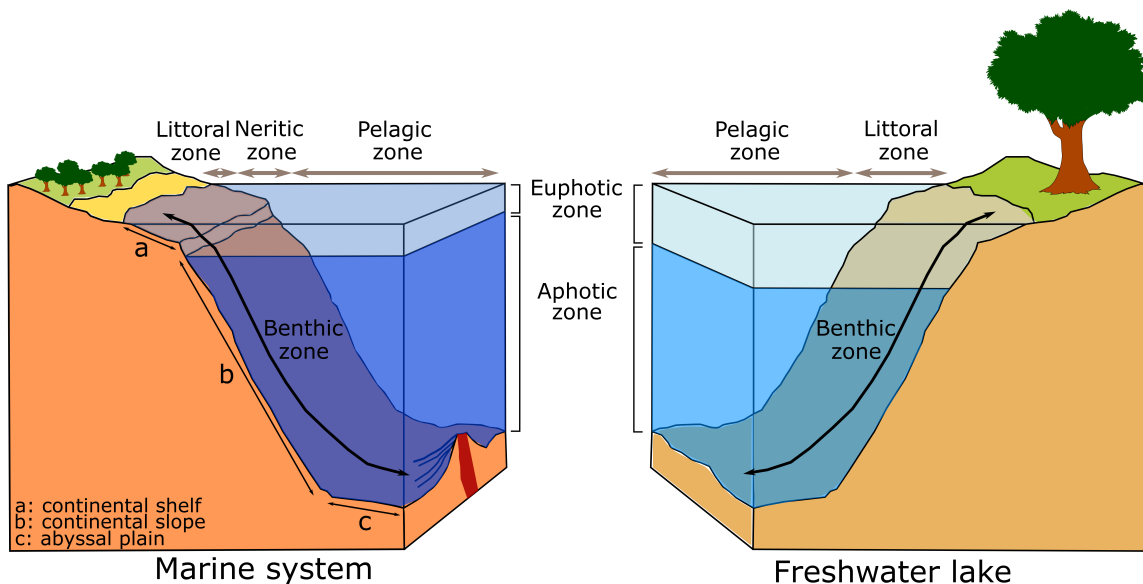
### 1.1.2 Spatial changes

Within aquatic ecosystems, multiples types of habitats are present due to structural spatial variations. Indeed, there are differences in horizontal and vertical dimensions, leading for instance to changes in exposure to light, temperature (Lewis, 2009; Madsen, 2015; Brönmark and Hansson, 2017). In lotic systems, habitat variations are observed from upstream to downstream. The environmental conditions evolve from the source to the sea, as described by the River Continuum Concept (Vannote et al., 1980). Along the water course, there is a progressive shift of abiotic conditions, communities and therefore of functioning of the ecosystem (Vannote et al., 1980). For instance, the quotient between production and respiration evolves with the stream order. Also, the water quality tends to decline along the water course, leading to a spatial eutrophication of the system (Vannote et al., 1980). These variations result in changes in the composition of microbial communities (Zancarini et al., 2017).

Similarly, for marine and lentic systems, there are horizontal spatial differences, from littoral to pelagic areas. These ecosystems can be divided into two to three horizontal zones (see *e.g.* Figure 1.2 adapted from Madsen, 2015). The littoral zone corresponds to the shallowest area. It is influenced by the shore and the light goes through the whole water column. This zone is often characterized by a wide biodiversity and hosts rooted aquatic plants (Brönmark and Hansson, 2017). In marine systems, an intermediate zone exists between the littoral and pelagic zones, called the neritic zone which covers the continental shelf. Finally, the pelagic zone corresponds to the offshore areas, which is also called limnetic zone for freshwater systems. Compared to the pelagic zone, the littoral area is more subject to perturbations. Therefore, microbial communities vary along this horizontal gradient (Obertegger et al., 2018).

Finally, ecosystems also change from the surface to the bottom. On this vertical axis, the light is the main factor delimiting the different zones (see Figure 1.2), since it is determinant in the photosynthetic process (Sarmiento and Gruber, 2006). The light penetration segregates

two main areas, which are called the euphotic and the aphotic zones. The euphotic zone is the area where photosynthesis can occur (Lewis, 2009). The aphotic area is the depth under which the light is  $\leq 1\%$  of the surface irradiance (Lewis, 2009; Pinet, 2019). Therefore, its depth varies greatly depending on water masses. In addition to the light stratification, four depth-levels can be identified in aquatic systems (Lewis, 2009; Moss, 2010; Kurt, 2019). First, the neuston layer that corresponds to the air-water interface is an area of nutrients accumulation (Kurt, 2019). Then, the water column, that corresponds to the global volume of water filling the marine system or lake, from the surface to the benthic zone. Finally, at the bottom of the ecosystems are found the benthic zone and the sediments. The benthic zone is the interface between the water column and the lacustrine sediments. The functioning of small freshwater systems is however quite different from what is described above, because of their small water volume and shallower depth. In these systems, the light can often reach the sediments leading to the absence of an aphotic area, as observed by Forel in 1904 (Oertli et al., 2005). Hence, primary production can occur in the whole water column, which is not happening in larger lakes. However, light penetration can be reduced by resuspension of sediments due to the mixing induced by winds (Blottière et al., 2017).



**Figure 1.2** Comparison of the zonation of marine systems (left panel) and freshwater lakes (right panel). Inspired from Madsen (2015) and Brönmark and Hansson (2017).

### 1.1.3 Temporal changes

Aquatic ecosystems can be characterized by their temporal dynamics, at different scales. Short-time changes in ecosystems can be observed yearly, following seasonal changes. In lakes, over the course of a year, the water column often experiences several phases of mixing and periods of stratification (Lewis Jr, 1983). Some lakes show a seasonal vertical zonation (Lewis, 2009; Brönmark and Hansson, 2017). In temperate regions, the dimictic lakes are the most common. They alternate between periods of stratification and water column mixing twice a year. This inter-annual variability affects the communities inhabiting the ecosystems, with for instance the mixing of microorganisms from the different depth layers and the resuspension of cells from the sediments (Shade et al., 2008; Linz et al., 2017). In small freshwater ecosystems, the small buffering capacity results in a fast response to any environmental perturbation or allochthonous input (Hurst, 2019). For instance, inputs from surrounding terrestrial ecosystems can induce disruptions which will quickly modify water chemistry and communities. Therefore, it leads to continuous temporal changes of the conditions. So, small freshwater ecosystems follow local and regional weather conditions (Blottière et al., 2017). Moreover, small freshwater ecosystems are characterized by a very mobile water column. During the year, they experience strong variability in their hydrological regime, resulting in significant variations of their depth, extreme variations corresponding to a temporarily dryness of the ecosystem. If the dryness state is recurrent, the ecosystem is then classified as “temporary”. Temporary streams represent on average more than 60% of the entire waterways (McDonough et al., 2011). In small freshwater ecosystems, the water depth remains low throughout the year, which results in an absence of thermal stratification, contrary to what is observed in lakes. However, recent studies revealed that daytime stratifications can occur, combined to nocturnal mixing (Andersen et al., 2017; Martinsen et al., 2019). This impacts the distribution of oxygen in the ecosystem on very short scales, that may lead to oxygen depletion at the bottom of small freshwater ecosystems (Briée et al., 2007; Andersen et al., 2017). Further investigations should be conducted in the various small ecosystems to better decipher their mixing functioning, in order to avoid misinterpretations (Biggs et al., 2017).

Long-time changes are also observed in aquatic ecosystems, especially because freshwater ecosystems have a given life time. Indeed, lakes are created by different natural processes (*e.g.* tectonic, withdrawal of ice, volcanic; Brönmark and Hansson, 2017). During their life time, lentic freshwater ecosystems are shrinking, going through ecological succession (Carpenter, 1981). Similarly to the spatial eutrophication in lotic systems, a temporal eutrophication is observed in lentic systems (Anneville et al., 2002). This phenomenon is often enhanced by human activities (Moss, 2010). The successional stages of the freshwater ecosystems and their

filling by sediments can be observed easily in small freshwater ecosystems, because they occur much faster than in larger systems. Indeed, if ponds are not properly managed, they can fill up and quickly become terrestrial ecosystems (Angélibert et al., 2004).

Hence, aquatic ecosystems are dynamic at different time scales. To assess their functioning and understand the dynamics of the microbial communities inhabiting them, it is therefore important to take into account multiple time scales. In this thesis, I will focus on small ecosystems that have a lifespan long enough to allow a survey of several years, typically 10 to 20 years (see the ecosystem presentation in Figure 2.2 and Table 2.1).

#### 1.1.4 Functional specificities of small freshwater ecosystems

For a long time, small freshwater ecosystems have been left apart in the scientific literature. Indeed, ponds were considered to be of much less interest than lakes, and small streams appeared to be of negligible interest with respect to downstream reaches (De Meester et al., 2005; Moore and Richardson, 2003; Larned et al., 2010). Indeed, they were considered as less important in the ecosystems because of their small size, or considered as small versions of bigger ecosystems (Céréghino et al., 2007). Also, it has been hypothesized that the ecosystem size was positively correlated to biological richness, which made small ecosystems less interesting to study than larger ones (Reche et al., 2005). However, the link between microbial diversity and ecosystem size is controverted, and the few studies dedicated to microbial communities in small freshwater ecosystems revealed a high diversity despite their small ecosystem size (Lindström et al., 2007; Simon et al., 2015; Teittinen and Soininen, 2015). Therefore, there is nowadays a regain of interest for this type of ecosystem in the limnological field, because of their high frequency, diversity, and major ecological importance (see *e.g.* Gomi et al., 2002; Søndergaard et al., 2005; Downing, 2010; Biggs et al., 2017; Mullins and Doyle, 2019; Pätzig et al., 2020). Downing (2010) estimated that ponds with a surface  $\leq 0.001 \text{ km}^2$  represent an amount of  $3.2 \times 10^9$  waterbodies for a total surface of 0.8 billion of  $\text{km}^2$ . Most of the microbial studies on small freshwater ecosystems have been conducted on aquaculture ponds, because of the economic importance for fish production. For instance, the impact of cyanobacterial blooms is studied with the aim to reduce fish mortality; and microbial communities are investigated to improve water quality (Sevrin-Reyssac and Pletikosic, 1990; Liu et al., 2020). There is therefore a lack of studies in small freshwater ecosystems in natural environments.

Small freshwater ecosystems have a clear ecological importance. As they are numerous, they harbour a wide diversity of environmental conditions and localization (*e.g.* rural, forest or urbanized areas). Most of them are still weakly impacted by humans and therefore constitute



preservation areas for various species (Cantonati et al., 2012; Biggs et al., 2017). A wide range of species inhabits small freshwater ecosystems, either lentic or lotic. Even if fishes are often absent, ponds and streams harbour diverse communities of invertebrates, amphibians, or microorganisms, among others (Céréghino et al., 2007; Downing, 2010; Simon et al., 2015; Teittinen and Soininen, 2015; Semlitsch et al., 2015). According to Biggs et al. (2017), ponds are of major importance for endangered species preservation. Moreover, they play a significant role not only for water species but also for terrestrial species. For instance, they act as drinking areas for mammals, and they have a positive impact on pollinator populations (Stewart et al., 2017). Moreover, small freshwater ecosystems play an important role in the global cycling of carbon and methane (Downing, 2010). It has been shown that the carbon dioxide and the methane concentration in water are decreasing with increasing lake size (Holgerson and Raymond, 2016). However, despite their small size, ponds have been identified as systems emitting but also sequestering carbon dioxide. They are also emitting methane, with an amount corresponding to 40% of the aquatic continental methane emission (Abnizova et al., 2012; Holgerson and Raymond, 2016; Yvon-Durocher et al., 2017). Therefore, better understanding their functioning is of major importance in the context of global climate change.

Finally, small freshwater ecosystems are also an important actor of ecosystem services for humans (Biggs et al., 2017). Ponds can be considered as potential water supply sources. Many man-made ponds were constructed with the objective of storing water, but we can also see them as preserved areas for drinking water. They can act as reservoirs of water and as retainers of the running water after rains, contributing to the global limitation of water-flow in the catchment area. Their ability to reduce pollutant loads has also been investigated, as well as their role of nutrient retention areas (Mallin et al., 2002; Gomi et al., 2002; Cheng and Basu, 2017). Small streams also play a major role in the global freshwater cycle as they are flowing in the direction of higher order streams, rivers, but also lentic systems (Biggs et al., 2017). Finally, they can have a strong cultural impact. Man-created ponds can relate a part of the region history, in particular of the agricultural development. Ponds can be created for recreational purposes, in particular next to or within urbanized area. They are also used as material to teach to students or to vulgarize freshwater biodiversity and ecology (Mullins and Doyle, 2019). Nowadays, and for all the above reasons, conservation programs encourage preservation of small freshwater ecosystems (Angélibert et al., 2004; Oertli et al., 2009; Oertli, 2018; Hill et al., 2018). This concerns also the preservation of micro-diversity, with the particular case of the Krauthügel pond in Austria. In 2012, this temporary pond and its buffer area have been declared as “Natural Monument for Single-celled Organisms”, to protect its microbial diversity (Cotterill et al., 2013).

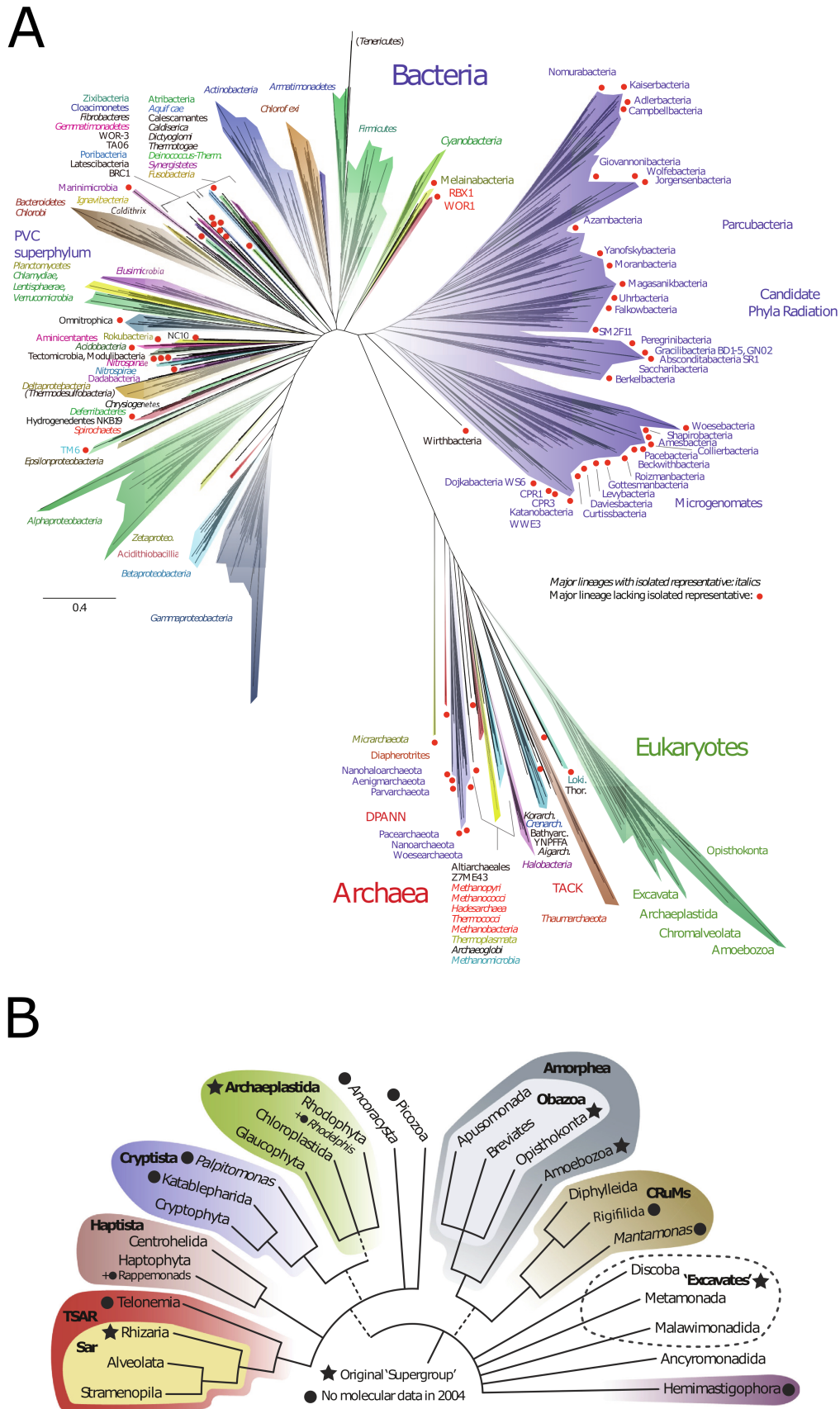
## 1.2 Microorganisms are key components of aquatic ecosystems

Microorganisms are ubiquitous in a large amount of ecosystems, where they are taking part in multiple processes within the systems (Worden et al., 2015). Since Woese et al. work in 1990 on the phylogenetic marker of life, they are distributed in a phylogenetic tree of life composed of three branches, corresponding to the following domains: bacteria, archaea, and eukaryota (see Figure 1.3). Between these three domains, microorganisms share common ecological features (*e.g.* small cell size, wide and diverse populations), but they also present strong differences, in physiology, metabolic capacities, or trophic behaviour for instance (review by Massana and Logares, 2013). Therefore, it is important to focus on all the microbial diversity to investigate the functioning of aquatic ecosystems. In this thesis, I will focus on the microbial communities of the three domains of life living in freshwater ecosystems.

### 1.2.1 Wide microbial diversity in freshwater

Since the beginning of microbiology, microbial diversity has been investigated in various types of environments using a wide range of methods, including observation, culture, sequencing and metagenomic. Today, massive-sequencing technologies (also referred as next-generation sequencing, NGS) such as metabarcoding allow us to explore a large proportion of the microbial world, including the detection of new taxa (Hug et al., 2016). These sequencing techniques are generating a large number of sequences, which are then regrouped into Operational Taxonomic Units (OTUs), or Amplicon Sequence Variants (ASVs), both being a proxy for species. However, among the entire microbial biodiversity detected by molecular approaches, a large proportion of the microorganisms remains uncultivated, which results in a lack of knowledge about their implication in the ecosystems.

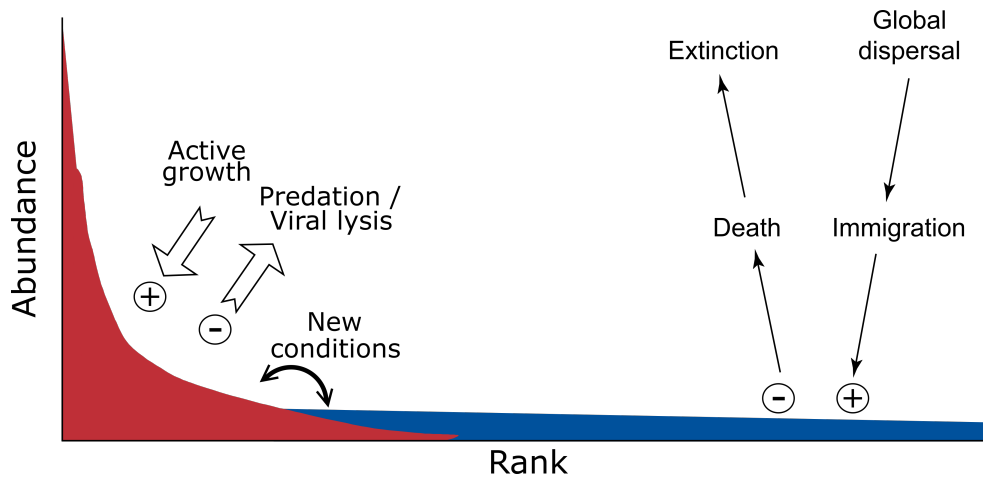
A large diversity of microorganisms has been detected in lakes and river, for both prokaryotic and eukaryotic communities (Zwart et al., 2002; Newton et al., 2011; Pernthaler et al., 2013; Debroas et al., 2017). By comparing the bacterial communities at the surface of marine and freshwater systems, Barberán and Casamayor (2010) have found that inland waters harbour more diverse communities. However, microbial eukaryotes, also called protists, are less studied in the literature (Caron et al., 2009; Debroas et al., 2017). Moreover, most of the studies have been conducted in lakes, and we are still missing a lot of knowledge on the diversity in smaller ecosystems (Simon et al., 2015; Biggs et al., 2017).



**Figure 1.3** Upper panel (A): View of the phylogenetic tree of life from Hug et al. (2016) build on sequenced genomes. The red dots indicate lineages lacking an isolated representative. Lower panel (B): Details on the phylogenetic tree of eukaryotes based on a consensus of recent phylogenomic studies. Figure taken from Burki et al. (2020).

### 1.2.1.1 Structure of microbial communities

Microbial communities are typically separated into two categories defined by their abundance (see Figure 1.4 from Pedrós-Alió, 2006). A small part of the taxa is dominating the community. They are well adapted to the conditions in the ecosystem and are therefore in an optimal growth situation. On the other side, as they are numerous, they are more likely prone to predation, parasitism, and viral lysis. Indeed, abundant microorganisms are more likely to be controlled by top-down processes, following the “Killing the winner hypothesis” (Thingstad, 2000; Pernthaler, 2005). The “rare biosphere”, defined by Sogin et al. (2006), is constituted of microorganisms occurring at very low frequency and rarely detected. Among those rare taxa, some can become punctually highly abundant and are therefore defined as “conditionally rare taxa” (Shade et al., 2014). The other taxa constitute a seed bank and can be inactive or dormant (Stoeck et al., 2007; Lennon and Jones, 2011; Simon et al., 2016). Rare taxa have also an important role in the ecosystem functioning, as in the biogeochemical cycles of nitrogen and sulphur for instance (Pester et al., 2010).



**Figure 1.4** Rank-abundance curve representing the abundance of individuals (*e.g.* OTUs) ranked according to their abundance. The red area corresponds to the abundant community. The blue area corresponds to the rare community. Figure adapted from Pedrós-Alió (2006).

The community assembly and the maintenance of a high diversity in ecosystems are still under investigation (Langenheder and Lindström, 2019). However, four main processes have been defined as the drivers shaping microbial communities, *i.e.* selection, dispersal, ecological drift, and speciation (Vellend, 2010). Selection, also called “species sorting”, is the variation of reproductive success across individuals and species due to biotic and abiotic pressures (Stegen et al., 2013; Langenheder and Lindström, 2019). Therefore, selection has long been considered as one of the most important processes structuring community composition. Dispersal corresponds to the movement of microorganisms across space, and it is a passive process (Nemergut et al.,

2013). The size of the cells can induce a limitation of the potential for dispersion, as well as the ability to transit into a cyst formation (Logares et al., 2020). Ecological drift corresponds to the random changes in the relative abundance of microorganisms (Logares et al., 2018). It is a stochastic process. Finally, speciation is the generation of new species in an environment (Vellend, 2010; Logares et al., 2018).

The origin of the immigration and the dispersal in freshwater ecosystems is still poorly known. Some studies indicate that inlet inflow of lakes plays an important role, whereas atmospheric deposition, sediment resuspension, and precipitation are not effective sources of immigration (Jones and McMahon, 2009; Comte et al., 2017). However, these conclusions could vary depending on the ecosystem studied. Indeed, the success of an immigration process depends both on the ability the invader has to invade and on the ability the local community has to resist to invasion. For instance, the immigration of new species can be easier during disturbance events. Further studies are still needed to better identify the dispersal mechanisms (Lindström and Langenheder, 2012; Kinnunen et al., 2018; Langenheder and Lindström, 2019).

### 1.2.1.2 Overview of eukaryotic and prokaryotic phyla in freshwaters

In ecosystems, the most abundant microorganisms are often affiliated to a restricted group of phyla. These phyla are often the taxa always found in the considered type of ecosystems. Here, I will present several prokaryotic and eukaryotic phyla that are classically detected in freshwater environments, and give insight into their ecology.

Five major bacterial phyla are often detected at high abundances in freshwater systems, the proteobacteria, the actinobacteria, the bacteroidetes, the cyanobacteria, and the verrucomicrobia (Zwart et al., 2002; Newton et al., 2011; Salcher, 2014; Madigan et al., 2014; Kurt, 2019). In general, archaea have a low abundance in freshwater environment (Kirchman, 2018). Among them, members of the thaumarcheota are found to be as the most abundant (Salcher, 2014). Proteobacteria is the most diverse phylum and is composed of gram-negative bacteria. It is divided into six classes, *i.e.* the alpha-, beta-, delta-, epsilon-, gamma-, and zeta-proteobacteria; the most abundant in freshwater being the alpha-, beta-, and gammaproteobacteria (Zwart et al., 2002). All classes of proteobacteria differ in terms of metabolism. Some members of the proteobacteria are phototrophs and therefore use the light as energy source. This concerns the orders rhizobiales, rhodobacterales, rhodospirillales (alphaproteobacteria), rhodocyclales (betaproteobacteria), and chromatiales (gammaproteobacteria) (Imhoff et al., 2018). Alpha- and betaproteobacteria are usually abundant and ubiquitous in freshwater systems (Salcher, 2014). Alphaproteobacteria are obligate or facultative aerobes, and they have optimal growth

rates in oligotrophic environments (Madigan et al., 2014). In majority, they are obligate or facultative aerobes. The microorganisms associated to betaproteobacteria are in general fast growing, and are in majority chemoorganotrophs and lithoautotrophs (Newton et al., 2011; Madigan et al., 2014). Gammaproteobacteria represent the largest and most diverse class of proteobacteria. At the same time, they present a large diversity of metabolisms and ecology, with members that can be phototrophic, chemoorganotrophic, or chemolithotrophic. Because of these ecological differences between the members of the proteobacteria, I will consider in this thesis proteobacteria at the class level rather than at the phyla level. Actinobacteria are gram-positive bacteria that can contribute up to 50% of the epilimnion communities in lakes (Newton et al., 2011). They are small size bacteria, most of them being chemoorganotrophic. Bacteroidetes are gram-negative bacteria, which feature a wide metabolic diversity (Madigan et al., 2014). Most of them are saccharolytic, with aerobic or fermentative metabolisms. They are known to be efficient in the degradation of humic material. Cyanobacteria are a group of oxygenic phototrophic bacteria, using CO<sub>2</sub> as carbon source. They are highly diverse in terms of morphology and ecology. Some of the members of cyanobacteria can produce cyanotoxins, which can be harmful for the environment (Wood et al., 2020). Verrucomicrobia are all aerobic or facultatively aerobic bacteria which can ferment sugars (Madigan et al., 2014). They are known to be symbiotic with microbial eukaryotes.

In eukaryotic communities, the superphyla alveolata, cryptophyta, and the phylum stramenopiles are the protists classically dominating freshwater communities (Debroas et al., 2017). The alveolates are characterized by the presence of cytoplasmic sacs located under the cell membrane, called “alveoli” (Madigan et al., 2014). Among the alveolata, various lifestyles are represented, such as predation, phototrophy, and parasitism (Adl et al., 2012). Two phyla are particularly abundant in freshwater, the ciliates and the dinoflagellates. The major characteristic of ciliates is the cilia at the surface of their membrane (Adl et al., 2012; Madigan et al., 2014). They are major predators in freshwater systems and their cilia are involved in food capture (Pernthaler, 2005). They feed in majority on algae, detritus, and protozoans. The dinoflagellates have classically two long flagellates. They are photosynthetic, mixotrophs, or heterotrophs and present a wide diversity in terms of size and shape (Kirchman, 2018). They are also characterized by the ability to form resting cysts (Bravo and Figueroa, 2014). The group of the stramenopiles constitute one of the most diversified eukaryotic group. It is composed of a majority of phototrophs and heterotrophs. In the phototrophic stramenopiles, the diatoms (Bacillariophyta) are among the most abundant algae in aquatic systems. Finally, cryptophytes are in majority phototrophs, but some of them can also be parasites or endosymbionts.

Among the non-abundant phyla in freshwater ecosystems, some have nevertheless an important role in ecosystem functioning (see section 1.2.1.1). However, we are still missing the

role of a large amount of taxa in freshwater ecosystems. For instance, Marine Stramenopiles (MAST) represent an eukaryotic phylum, initially detected in marine systems and composed of 19 classes (Massana et al., 2004, 2014). These organisms were first thought to be only found in marine systems, but then several lineages have also been detected in freshwaters (Simon et al., 2014). This type of marine-freshwater transitions is rare, which makes their study interesting (Cabello-Yeves and Rodriguez-Valera, 2019). Moreover, they are known to be mostly heterotrophic and to act as major predators in marine systems (Massana et al., 2014). However, we do not know if they have the same role and ecological importance in freshwater systems as in marine environments. Similarly, we are missing information of the role and function of microorganisms affiliated to the Candidate Phyla Radiation (CPR). The CPR is a recently discovered prokaryotic clade that includes 74 potentially phylum-level lineages and represents up to 25% of the bacterial diversity (Hug et al., 2016; Castelle et al., 2018; Castelle and Banfield, 2018). For now, no members of the CPR have a cultivate representative, implying that their ecology is poorly known (Hug et al., 2016). CPR members have been discovered in a wide range of ecosystems and are characterized by a small genome, which limits their biosynthetic capacities (Brown et al., 2015). This underlines that CPR members are probably symbionts of other microbes (Nelson and Stegen, 2015). Therefore, they might represent a significant part of the parasitic load in the ecosystems and some of them have already been identified as parasites, *e.g.* Saccharibacteria or Dojkabacteria (Castelle and Banfield, 2018). In terms of ecosystem functioning, the Parcubacteria super-phyla has been identified as involved in the nitrogen cycle by conducting denitrification (Danczak et al., 2017). Hence, exploring the diversity and the temporal variation of microbial eukaryotes and prokaryotes in small freshwater ecosystem can help to elucidate new aspects of the ecology of poorly known taxa.

## 1.2.2 Implications of microorganisms in ecosystem functioning

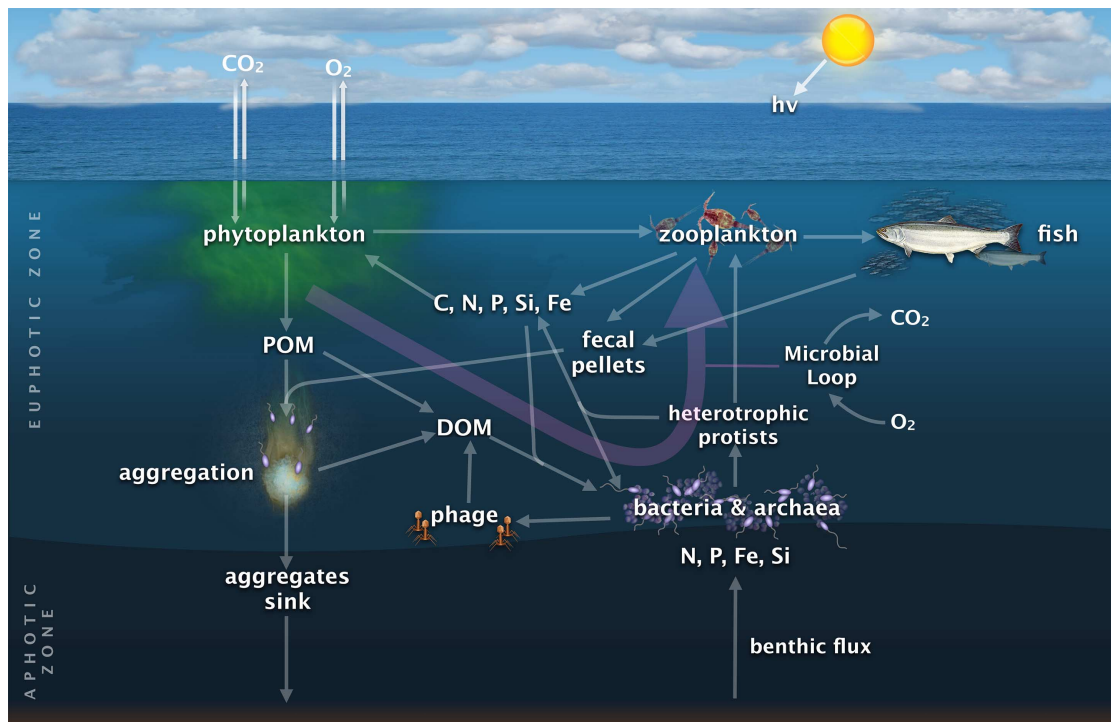
Microorganisms are key players in the ecosystems. They are involved in the aquatic food web through the microbial loop and are also major actors of all biogeochemical cycles (Worden et al., 2015).

### 1.2.2.1 Microbial loop

The microbial loop describes the microbial path of the aquatic food web and is included in the cycle of the dissolved organic matter (DOM) (see Figure 1.5, Azam et al., 1983; Pomeroy et al., 2007; Worden et al., 2015). These processes have been first described in marine environments,

but they are also occurring in freshwater ecosystems (Sherr and Sherr, 1988).

There are four major steps in the microbial loop (see Figure 1.5, Worden et al., 2015). Briefly, the phytoplankton, *i.e.* the primary producers, releases great amounts of DOM. Then, heterotrophic bacteria use a large proportion of this DOM as food source. These bacteria are mainly predated by heterotrophic flagellates and finally, these latter are grazed by ciliates and are preys of zooplankton. This predation allows to recycle nutrients such as carbon, nitrogen, and phosphorus. The microbial loop allows to incorporate DOM to higher trophic levels through heterotrophic bacterial activity (Azam et al., 1983). Indeed, a large part of the DOM produced by the phytoplankton is only accessible to heterotrophic bacteria and archaea (Azam and Malfatti, 2007). Viruses are an integral part of the microbial loop and have a major effect on prokaryotic cells (Fenchel, 2008; Worden et al., 2015). Indeed, cell viral lysis also contributes to the pool of DOM available in the aquatic environment. Moreover, viruses regulate microbial biomass and composition.



**Figure 1.5** Representation of the marine food webs showing the role of microbial communities. The microbial loop is highlighted with the purple arrow. POM corresponds to the Particulate Organic Matter and DOM to the Dissolved Organic Matter. Figure taken from Worden et al. (2015).



### 1.2.2.2 Biogeochemical cycles

Biogeochemical cycles correspond to the circulation of substance (*e.g.* nutrients) in the biosphere (Bolle, 2013). For all living organisms, six major elements (*i.e.* H, C, N, O, S, and P) are necessary to run the cellular machinery and to produce carbohydrates, proteins, nucleic acids (Falkowski et al., 2008). Microorganisms are actively involved in the cycles of these elements, and act in the transformation and mineralization of the organic matter (Del Giorgio and Cole, 1998). They are essential for the functioning of ecosystems, in particular for macro-organisms (Falkowski et al., 2008). Indeed, microorganisms are able to transform nutrients, which make them available for other organisms. For instance, the only way to make nitrogen available for organisms such as plants is through the reduction of atmospheric  $N_2$  to ammonia (nitrogen fixation, Van Der Heijden et al., 2008; Falkowski et al., 2008).

#### Examples of the implication of microorganisms in three cycles

**Carbon and oxygen** Carbon and oxygen cycles are tightly linked: oxygen is stored in diverse organic compounds as well as in  $CO_2$ . Carbon is present in reduced forms, such as methane ( $CH_4$ ) and organic carbon. It is also present in oxidized forms, such as carbon dioxide and inorganic carbon (limestone). In the carbon cycle, the principal way to synthesize organic carbon is through photosynthesis and chemosynthesis. Therefore, photosynthetic organisms constitute the basis of the carbon cycle. For instance, it has been shown that picophytoplankton (*i.e.* microorganisms smaller than 3  $\mu m$  performing photosynthesis) is highly involved in the carbon dioxide fixation in oceans (Jardillier et al., 2010). Globally, the phytoplankton represents up to 50% of the global photosynthesis on Earth, that is of the carbon fixation (Seymour et al., 2017).

**Nitrogen** Microorganisms are major players in the four processes of the nitrogen cycle (Falkowski et al., 2008; Fowler et al., 2013). First, they are involved in nitrogen fixation by converting the atmospheric  $N_2$  into ammonia (Madigan et al., 2014). Then, they are able to assimilate ammonia and transform it into organic compounds. Finally, aerobic organisms conduct the nitrification by oxidizing ammonia into nitrates, whereas anaerobic bacteria conduct denitrification by transforming nitrate into  $N_2$  (Kuenen, 2008).

## 1.3 How to elucidate ecosystem functioning?

One of the major objectives of microbial ecology is to understand how ecosystems are functioning. Several steps are necessary to progress in the resolution of this global question:

1. Describe community composition and structuration.

To understand an ecosystem, we first have to figure out which or what type of organisms inhabits it. This step can also allow the detection of previously undiscovered taxa.

2. Identify the abiotic factors influencing community structuration.

Microbial communities can be influenced by a wide range of physico-chemical parameters. Exploring the links between communities and their environments will help us to understand how communities are structured. Indeed, these abiotic parameters constitute the bottom-up controls of community composition.

3. Decipher the interactions between the members of the communities.

Environmental conditions cannot completely explain community composition because organisms are in constant interactions with each other. Understanding the principal relationships between organisms will therefore help us to identify the role of the different members of the community. Microbial interactions consist in top-down (*e.g.* predation) and sideways (*e.g.* competition) controls driving community composition.

Moreover, these three steps have to be integrated in a global context. Indeed, studying an ecosystem at one time point or one spatial location is not enough to decipher microbial communities diversity and complexity. For instance, temporal surveys can help to identify bottom-up and top-down factors shaping the community composition (Muylaert et al., 2002). Therefore, to study an ecosystem and to be able to identify the ecological processes within, it is important to focus on various spatio-temporal scales.

### 1.3.1 Abiotic factors influence microorganisms community composition

The role played by a large set of abiotic parameters has been investigated over the years and proves to impact the microbial community composition. A wide range of factors (*e.g.* temperature, nutrients availability) can govern the communities but it is difficult, time consuming, and

expensive to measure them all. Classically, we focus on the environmental conditions as well as on resources and nutrient availability (Madigan et al., 2014). The environmental conditions are often resumed to the temperature, the pH, the dissolved oxygen, the light, and the osmotic conditions. The nutrient availability refers to the concentrations of the diverse chemical forms of carbon, nitrogen, phosphorus, and other macronutrients. These various forms reflect the different phases of biogeochemical cycles, which allow to have a global picture of the processes. It can also include the concentration of micronutrients, such as copper or zinc. The characterisation and study of these parameters help to define environmental gradients that can provide a more suitable environment for certain organisms than to others. In general, the trophic state index of the ecosystems, *i.e.* oligotrophic, mesotrophic, or eutrophic one, can play a role in the composition of microbial communities (Smith, 2003; Yannarell et al., 2003; Monchamp et al., 2018).

Temperature is often depicted as the most important factor influencing aquatic community composition and distribution, for both macro and micro-organisms (Crump and Hobbie, 2005; Gilbert et al., 2012; Brönmark and Hansson, 2017). Temperature influences microbial richness (Fuhrman et al., 2008; Kirchman, 2018). Its effect on communities can vary depending on the ecosystems, as well as on the organisms considered. For instance, in a surface marine study by Logares et al. (2020), temperature was found to be the most important measured parameter impacting the selection of prokaryotes, but it had a smaller effect on picoeukaryotes. Furthermore, temperature is also a structuring factor at the global scale. Indeed, it has been shown that in oceans, temperature is partly responsible of the decrease of bacterial diversity at the poles compared to lower latitudes (Sunagawa et al., 2015). In freshwater systems, temperatures have an indirect effect on communities, through the water mixing processes (Brönmark and Hansson, 2017). Indeed, variations in water temperature induce stratification-mixing dynamics in the water-column. Therefore, it leads to disturbances in the ecosystems, resulting in habitat heterogeneity and changes in the microbial communities (Shade et al., 2008).

In temperate freshwater aquatic habitat, the pH value usually varies between 6 and 9 (Kirchman, 2018). In freshwater systems, pH often reflects the functioning of the catchment area (Brönmark and Hansson, 2017). Indeed, pH depends on the hydrology and geology of the surrounding environment. It can change after the input of acidifying substances (*e.g.* acidic rain, Schindler, 1988) and depends on ecosystem productivity. The pH gives information on the equilibrium relationship between carbon dioxide and bicarbonate in the ecosystem (Brönmark and Hansson, 2017).

The oxygen level depends on the diffusion from the atmosphere as well as on the release of photosynthetic organisms (Moss, 2010). Furthermore, an increase of temperature alters

the concentration of dissolved oxygen as oxygen is more soluble in cold water (Brönmark and Hansson, 2017). The oxygen level can impact biological processes, like for instance the denitrification which requires anaerobic conditions (Lam and Kuypers, 2011). In ecosystems, oxygen is required for aerobic respiration which concerns almost all eukaryotic microorganisms, as well as macroorganisms. Therefore, variations of the level of dissolved oxygen can impact the development of microorganisms. For instance, microorganism that are adapted to oxygen-depleted environments such as CPR will benefit from a decrease in dissolved oxygen concentration (Castelle et al., 2018). In eutrophic systems, both high and low levels of oxygen are observed. High levels of oxygen can be observed during the day, due to the high photosynthetic activity of the phytoplankton (Paerl et al., 2018). However, eutrophication induces a global reduction of oxygen in freshwater ecosystems, especially in the hypolimnion (Foley et al., 2012). This oxygen depletion phenomenon is caused by the respiration and the decomposition of the algal blooms (Graneli et al., 2008). Therefore, eutrophication and the changes in oxygen concentration can induce perturbations in microbial community composition (Filstrup et al., 2014).

Light availability (in terms of quality and quantity) can be of major importance for microbial communities. First, the diurnal cycles of dark and light phases impact microbial communities for the primary production, as light is crucial for photosynthetic organisms (Trench-Fiol and Fink, 2020). It also impacts indirectly other communities, such as bacteria using the dissolved organic matter released in ecosystems by phytoplankton (Worden et al., 2015). Therefore, microbial communities are changing on a daily basis, which highlights the need for selecting the proper temporal scale depending on the research question. Intensity of light also affects microbial communities, with for instance Chrysophyceae that seems to be associated with higher light levels in marine systems (Kirkham et al., 2011). Seasonal variations of light intensity can therefore change community composition. Indeed, light intensity has been depicted as a factor regulating the germination of cyanobacteria and diatoms and therefore their recruitment from the sediment to the freshwater communities (Rengefors et al., 2004).

The conductivity of a medium reflects its ability to conduct electricity. It is tightly linked to salinity, that is the concentration of dissolved salts in water. It is also an indicator of the osmotic conditions in the aquatic environment. An increase of conductivity can benefit to salinity tolerant species. For instance, the conductivity has been shown to be a driver of diatoms community composition (Aarnio et al., 2019). Finally, salinity has been identified as a major factor of community dissimilarities in a wide range of environments (Lozupone and Knight, 2007).

The concentration of nutrients in freshwater ecosystems is partly linked to the catchment

area, to the bedrock type, as well as to the plants inhabiting the ecosystem and its banks. Microbial communities in lakes are tightly linked to the concentration of nutrients, as proven for instance by studies of bacteria in oligotrophic lakes (Logue et al., 2012). Among the nutrients, phosphorous is often depicted as the most limiting factors in lakes, mainly because of its low solubility in water (Tiessen, 2008). Therefore, it is an important environmental driver of phytoplankton composition, because it generates competition between microorganisms (Vadstein, 2000; Salcher et al., 2007).

### 1.3.2 Interactions between microorganisms shape communities

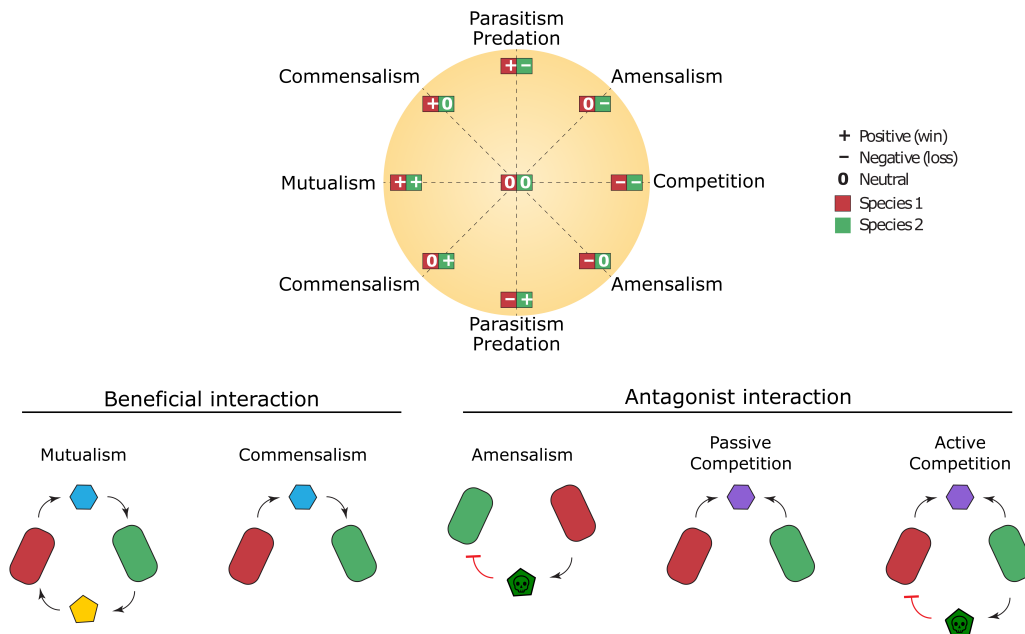
Abiotic factors are drivers of microbial community composition. However, they are not sufficient to explain the entire community variability, even with a large panel of parameters took into account (Muylaert et al., 2002; Lima-Mendez et al., 2015). Therefore, it is also necessary to focus on biotic interactions within the organisms in ecosystems.

Microbial interactions are very diverse and happen within and between the three domains of life. However, they are difficult to study and researchers face a wide range of challenges. Interactions between microorganisms are difficult to reproduce *in vitro*. It is due to the fact that less than 1% of the microbial diversity is cultivated, the rest constituting the “uncultured microbial majority” (Rappé and Giovannoni, 2003; Hug, 2018). Microbial interactions are also complex to investigate through direct observations because of their small size and the abundance of microbial cells. Indeed, a tremendous number of interactions can take place at the same time between the members of a single community, making experimental identifications difficult (Deutschmann et al., 2020). Finally, we can consider that, even if we could observe interactions *in vitro*, there would be no guaranty that they would be the same in environment. Despite these difficulties, interactions between microorganisms have been studied for a long time and, nowadays, co-occurrence studies help to better decipher the complexity of the relationship within communities.

#### 1.3.2.1 A variety of interactions between microorganisms

There is a large range of interactions types between microorganisms, which can benefit or impair the involved organisms (see Figure 1.6; Faust and Raes, 2012; Karimi et al., 2017). They can be defined as antagonist or beneficial, they can have different strength, and they can involve specialist or generalist organisms. Interactions result in positive, negative, or neutral effects for

the protagonists (Lidicker, 1979). Moreover, the interactions between two organisms can vary over time, depending on the environmental conditions (Hoek et al., 2016).



**Figure 1.6** Possible biological interactions and their effects between microorganisms. The interaction can result in a positive (+), negative (-), or neutral (0) effect for each protagonist (Faust and Raes, 2012; Lidicker, 1979). Illustration of the interactions in the case of metabolic exchange are provided. Dark green pentagons are for toxic compounds (Cavaliere et al., 2017; Karimi et al., 2017).

**Beneficial interactions** There are two types of beneficial interactions, commensalism and mutualism (Karimi et al., 2017). Commensalism interactions benefit only to one of the involved organisms. The other organism is not affected by the interaction. In the case of an exchange of metabolites, commensalism is also called cross-feeding (Faust and Raes, 2012). For instance, commensalism corresponds to the release of dissolved organic matter by phytoplankton which benefits to bacteria (Seymour et al., 2017). Mutualism interactions benefit to both organisms. For instance, they can consist in the exchange of metabolisms by both organisms.

**Antagonist interactions** Antagonist interactions imply that one of the organisms is negatively affected by the interaction. Amensalism interactions result in a negative effect on one organism with no benefit for the other. This type of interactions can occur when one organism modifies its environment or releases metabolites that happen to be toxic or inhibiting for other organisms. Competition interactions are negative for both protagonists. They can be passive or active (Cavaliere et al., 2017). Passive competition occurs when two organisms are interested by the same limiting resource or by the need of space and light. This interaction becomes

active when one of the organisms intentionally releases toxic compounds in order to inhibit the other competitors. Organisms that release biochemical substances influencing others are called allelopathic organisms (Fuhrman et al., 2015). Predatory and parasitic interactions are called win-lose interactions. Predation happens when one organism, either heterotrophic or mixotrophic, feeds on another organism. For instance, protists are known to be major predators of bacterial communities (Jardillier et al., 2005). Predation was thought to be restricted to protists but it has also been detected recently in a planctomycete bacteria (Shiratori et al., 2019). Parasitism consists in an interaction where an organism is entirely dependant to another, and physically associated to it. Organisms that parasite can be fed, hosted, and protected by their host, which results in a beneficial interaction for the parasite and a harmful one for the host. Some parasites can exploit several hosts during their life cycle.

A good example of a hub of interaction is the phycosphere surrounding the phytoplankton (Seymour et al., 2017). This area is a theatre of diverse types of interactions generated by the release of metabolites and dissolved organic matter by the phytoplankton.

### 1.3.2.2 Visualisation of the interactions

As previously mentioned, interactions between microorganisms can be studied using various methods. Laboratory approaches such as fluorescence in situ hybridization (FISH), EpicPCR, single cell genomics, or NanoSIMS allow to investigate the associations between microorganisms (Yoon et al., 2011; Musat et al., 2016; Spencer et al., 2016; Grujic et al., 2018). For instance, the FISH technic allows to target microorganisms using specific probes. Combined with single-cell techniques, it can allow to detect associated organisms in a predation or parasitic interaction. However, these techniques are time consuming and they need to design specific probes for the targeted microorganisms since the diversity of probes is restricted to a few taxa. Inferring interactions using co-occurrence analyses allows targeting interactions of potential interest among the large diversity of possible of interactions. After the identification of interactions, laboratory approaches are complementary to be able to validate or not the predicted relationships.

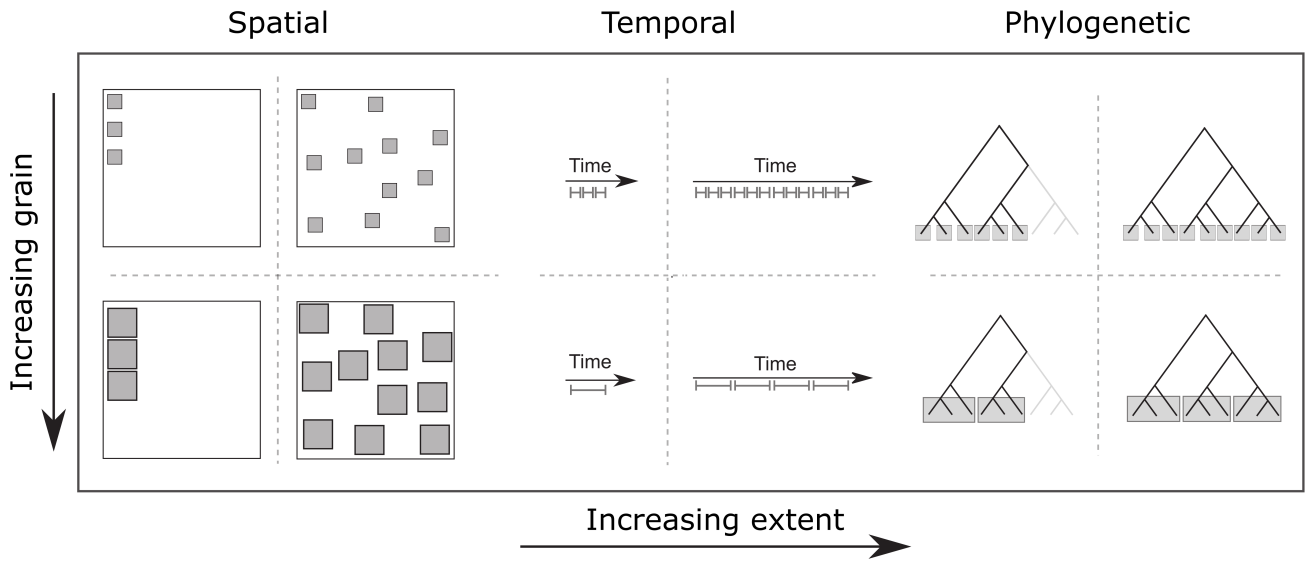
Microbial network inference is based on statistical co-occurrences of sequence reads generated using high-throughput sequencing technologies (Faust and Raes, 2012). Predictions of microbial associations are based on community matrix, *i.e.* a table containing the number of reads obtained for each OTU and site. The networks are based on the significant and non-random co-occurrences between taxa. Briefly, if two taxa present the same trends, such as rise or decrease of their respective abundances at the same time, then they can be considered as

interacting positively (Faust and Raes, 2012). A large variety of methods has been elaborated to infer association networks between microorganisms, several of them being reviewed in Röttgers and Faust (2018). They differ in several ways and try to overcome the issues inherent to NGS datasets. They can be based on various measurements of correlations (*e.g.* Pearson correlations, matrix of variance-covariance). They can take into account the sparseness of the dataset (*i.e.* the number of 0 in the matrix, representing either an absence of the specie or a non-detection). The methods have also different ways to treat the compositionality of the datasets (Gloor et al., 2017). Data are compositional because they are composed of samples with unequal sequencing depth and are not true abundance counts. Finally, the methods try to avoid spurious correlations, for instance due to the effect of environmental parameters or because two species are positively correlated to a third one and therefore are correlated together (Kurtz et al., 2015).

### 1.3.3 Microbial communities are structured in space and time

In all types of ecosystems, microbial communities are not static and they change in space and time. In order to get a good overview of the microbial communities in an ecosystem and to study their dynamics, it is not suitable to focus on a single sampling point. Large spatial and/or temporal surveys are needed to capture the complexity of the microbial communities. The spatial scale can be for instance the survey over several localisations on the surface of a pond (Lear et al., 2014). The temporal scale can be the sampling of one ecosystem every month over six years (Gilbert et al., 2012). Generally, the larger the number of samples is, the more organisms are detected. Indeed, variation over space and time generates habitat heterogeneity and, therefore, potential variations in the microbial community composition. Moreover, considering more samples increases the probability to detect rare taxa and allows to study the rare biosphere (Hugoni et al., 2013; Shade et al., 2014). The taxonomic scale represents the scope at which the microbial community is considered. For instance, we can focus on the three domains of life, or only on a specific phylum within one domain. Finally, the surveys can be realized at different levels (Ladau and Eløe-Fadrosch, 2019). The grain of the analysis corresponds to the breadth of each sample. For instance, it can be the volume of water collected for the analysis (*e.g.* 100 ml or 3 l) or the frequency of the temporal survey (*e.g.* daily or monthly). The extent corresponds to the breadth of the all survey (*e.g.* entire surface of one large ecosystem or only littoral zone) To study the microbial communities in an ecosystem, it is possible to focus on three scales (spatial, temporal and taxonomic) and at two levels (grain and extent) (see Figure 1.7, Ladau and Eløe-Fadrosch, 2019).





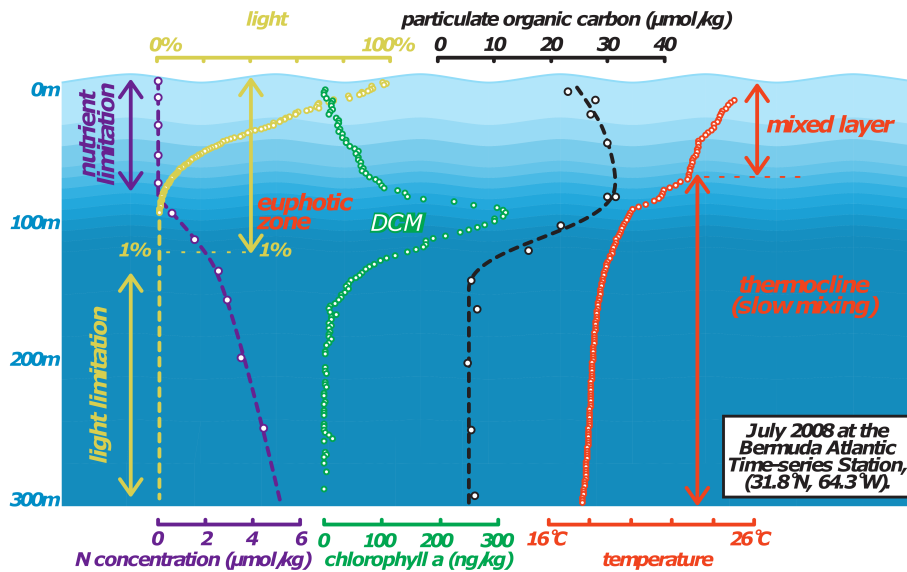
**Figure 1.7** Illustration of the scales used to investigate microbial community structuration. Ecological scale has three axes: space, time, and phylogeny. Each of these has two components: grain, that is the scope of the unit of observations, and extent, that is the scope of the entire set of observations (adapted from Ladau and Eløe-Fadros, 2019).

Incorporating different spatio-temporal dimensions in ecology can help to answer the hypothesis of Baas-Becking, “everything is everywhere, but the environment selects” (Baas-Becking, 1934). Indeed, the variations in space and time of environmental conditions and microbial community composition can help to decipher community structuration and the role of abiotic and biotic factors on the communities. Large dataset of microbial communities can be constituted thanks to the development of new sequencing techniques that are faster and cheaper. However, they still require a high amount of time and money to be conducted, which makes them rather scarce (Faust et al., 2015). Moreover, most of the spatial and/or temporal survey focus on bacterial community composition while microbial eukaryote communities are often overlooked. Besides, literature also lacks of studies considering simultaneously the three domains of life (Caron et al., 2009; Debroas et al., 2017).

### 1.3.3.1 Spatial variability of microbial communities

The properties of aquatic ecosystems are highly dependent of their spatial characteristics. Depending on the size of the ecosystem, on its mixing, and on its proximity to the littoral area or to an input water zone for instance, the environmental conditions and the habitats will change. With spatial surveys, the biogeography of microbial communities can be better understood, while either hotspots or in the contrary deserts of biodiversity can be discovered (Gonzalez

et al., 2012). Spatial variation of community composition can happen at different scales, ranging from the microscale to kilometres. The microscale corresponds to the cell-to-cell distances, which are close interactions between microorganisms (Cordero and Datta, 2016). In this thesis, I will not consider the microscale changes of microorganisms, because I aim at focusing on community composition on larger scale.

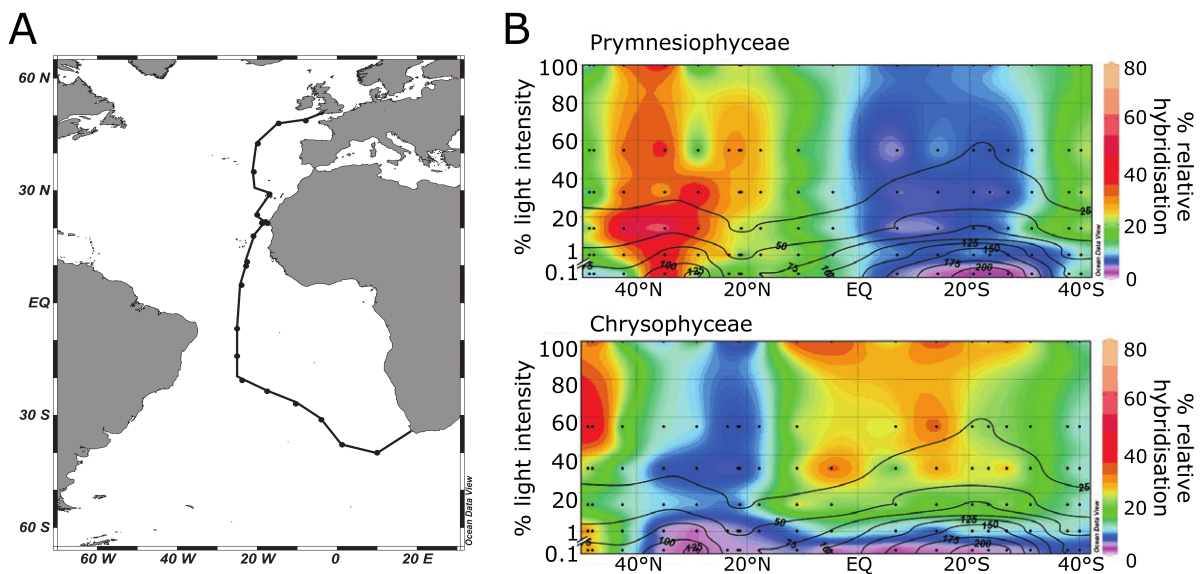


**Figure 1.8** Variation of the environmental parameters in the water column in a subtropical ocean (taken from Sigman and Hain, 2012 and generated with data collected at the Bermuda Atlantic Time-series Station (BATS) in July 2008). DCM corresponds to the Deep Chlorophyll Maximum which is a water layer enriched in chlorophyll.

Spatial variations of communities can be considered following a vertical gradient, that is the distribution of microorganism along the water column. According to lake and marine systems functioning, environmental conditions along the water column vary strongly (see section 1.1 and Figure 1.8). These variations correspond for instance to gradients of salinity, pressure, temperature, or dissolved oxygen (Shade et al., 2008; Tobias-Hünefeldt et al., 2019). Concomitantly, depth investigation of microbial communities in marine systems reveal a vertical zonation of taxa (DeLong et al., 2006). In lakes, there are few studies that explore depth distribution of microbial communities (Kurt, 2019). However, they allowed to detect structuring factors of communities (*e.g.* in the distribution of microbial eukaryotes in the southern basin of Lake Baikal or in temperate lakes in France; Yi et al., 2017; Keshri et al., 2018). Vertical surveys give the opportunity to make assumptions on the functional role of taxa. For instance, Lepere et al. (2010) investigated the structure of microbial eukaryotes in three lakes with three different trophic statuses. On the basis of the distribution of taxa and knowing the mixing regime of the different lakes it is possible to discuss the potential importance of mixotrophy in the functioning of the lakes.

The horizontal gradient corresponds to the distribution of communities at different locations (latitude/longitude). The sampling can be conducted within one ecosystem at several points of the surface, or in several distinct ecosystems distributed over a given study area. The horizontal gradient has also its importance in lotic water-bodies, from upstream to downstream. Indeed, changes in the composition of microbial communities were detected along the water course, indicating changes of water chemistry and possible immigration from the bank and sediments (Crump et al., 2007). In freshwater ecosystems, several studies have compared different ecosystems over the same geographical area in order to investigate the biogeographic patterns. For instance, Eiler and Bertilsson (2004) sampled bacterial communities in four lakes and showed that despite geographic proximity and under similar cyanobacterial bloom events, communities were significantly different.

Combination of horizontal and vertical gradients can help resolve global distribution of taxa. This approach is more common in marine systems than in freshwater systems. For instance, Kirkham et al. (2011) studied the distribution of two marine picophytoplanktonic groups, the Pymnesiophyceae and the Chrysophyceae, using a spatial survey conducted in the Atlantic. The two classes revealed strongly complementary distributions along the transects, with differences in the N:P requirements and on the latitudinal repartition (Figure 1.9).



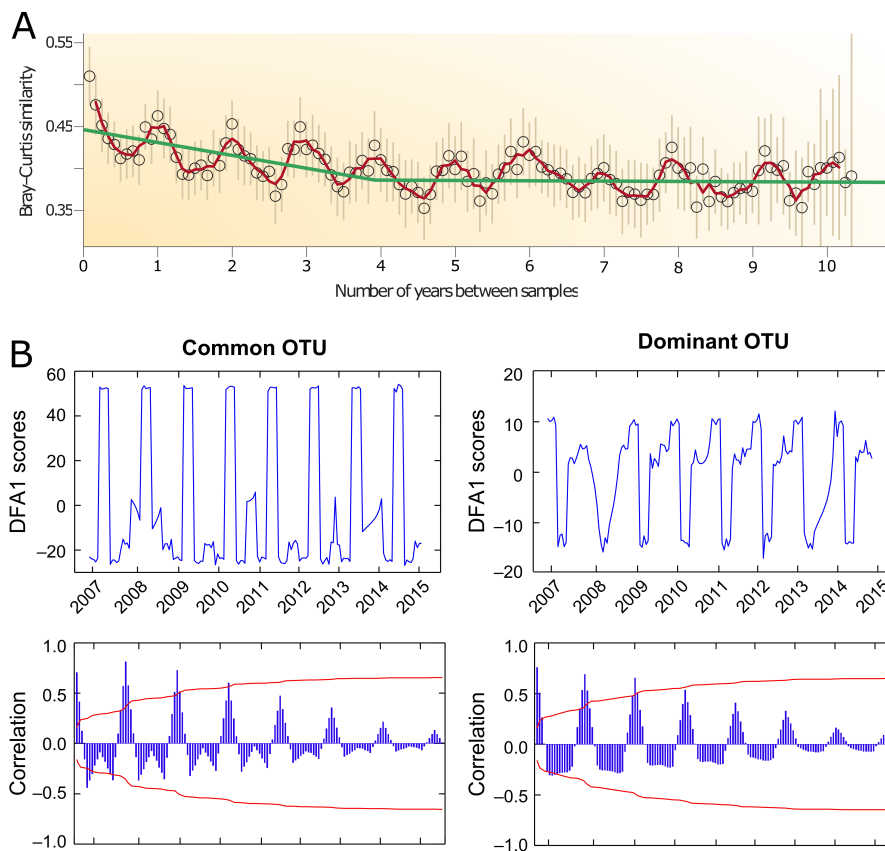
**Figure 1.9** Example of a spatial survey along both vertical and horizontal gradients (Kirkham et al., 2011). (A) Location of the sampling stations of the AMT15 cruise in the Atlantic Ocean. The stations are indicated by black dots. Each station was sampled at six depths, defined by a percentage of surface light intensity penetrating the water column. (B) Distribution patterns of two groups of photosynthetic picoplankton based on dot blot hybridization technique. Black lines indicate the depth in meters. Black dots mark the sampling points.

In conclusion, spatial surveys over the vertical gradient can help understand better the distribution of organisms along the water column. Over the horizontal gradient, they help unravel the variations resulting from the biogeography and the influence of ecosystem structure (*i.e.* coastal versus pelagic areas). However, it is important to be careful when considering time in spatial studies. The observed biogeographic patterns can be biased if time is not taken into account during the sampling (*e.g.* several weeks between two samples). Moreover, spatial patterns are conducted using discrete time points, which does not allow to depict the entire communities inhabiting the systems.

### 1.3.3.2 Temporal evolution of microbial communities

Temporal evolution of community composition is of major importance to understand the functioning of an ecosystem, its ecology and the structuration of its communities (Moreira and López-García, 2019). Temporal surveys are a way to determine if microbial communities remain stable over time, to detect periodic and non-periodic patterns, and to evaluate the effect of perturbations of the communities, as for instance drought events (Simon et al., 2016). They can be conducted at different time scales (*i.e.* from hours to years) as microbial communities are dynamic over multiple time scales (Fuhrman et al., 2015). First, community changes can be investigated at an hour scale. This time scale corresponds for instance to day and night variations (Trench-Fiol and Fink, 2020). It can also be useful to investigate the immediate effect of a perturbation, for instance of nutrients amendment (Newton and McMahon, 2011). It can help to capture variations in the physiological response of communities. Indeed, the generation time of organisms can be of one day for the fastest, like copiotrophic organisms (Mangot et al., 2013; Fuhrman et al., 2015). For instance, a day versus night study allowed to show that coral reef microbial communities collected at the same period are more similar than the ones sampled at day and night within one site (Kelly et al., 2019). Then, the communities can change over daily to weekly time ranges. This frequency of sampling allows to study the effect of environmental perturbations, such as the effect of algal blooms. It is also a good scale to infer interactions between microorganisms with co-occurrence networks. In the case of lakes, this time range allowed to study the impact of the mixing regime of lakes on microbial communities (Shade et al., 2008). Finally, time surveys can be conducted at monthly to seasonal range. Monthly surveys are the most common in literature. These surveys can help identifying global effect of climatic variation, regional changes, or changes in the trophic status of the ecosystems. For instance, using community dissimilarity estimations or autocorrelation of time-series, long surveys allowed to identify strong seasonal patterns in marine and freshwater environments (see Figure 1.10, Gilbert et al., 2012; Fuhrman et al., 2015; Yan et al., 2017; Giner et al., 2018). Therefore,

temporal surveys at different frequencies are needed to investigate the various dynamics and the ecology of microorganisms inhabiting freshwater ecosystems. Every year, there are some regular variations of the environmental conditions in ecosystems (*e.g.* succession of seasons). In addition, there are unusual perturbations (*e.g.* a warmer winter than usually) that disturb the regular variations. Therefore, if the duration of the survey is too short, these perturbations will not be detected like such. Hence, to obtain reliable conclusions on the patterns detected, long surveys are needed.



**Figure 1.10** Example of analysis on long temporal surveys. *Upper panel (A)*: Bray-Curtis similarity plot of samples collected on a monthly basis at the San Pedro Ocean Time-series over ten years. Each data point is the average of pairwise measurements of samples collected times the number of months apart. The sinusoidal curve indicates that community composition follows a seasonal pattern (Figure taken from Fuhrman et al., 2015). *Lower panel (B)*: Seasonal changes over a nine-year survey in Lake Donghu (China). Temporal patterns were revealed using discriminant function (to resume the dynamics of a pool of species) combined to an autocorrelation analysis (Figure taken from Yan et al., 2017).

Several long time-series are being conducted in marine and freshwater systems and follow the same environments over several years. Long term marine surveys are reviewed in Buttigieg et al. (2018). Long term freshwater surveys are conducted in various ecosystems, such as peri-alpine

lakes with the “Observatoire des Lacs”<sup>1</sup> or north-temperate lakes with the US Long-Term Ecological Research Network<sup>2</sup>. These long time surveys investigate for instance the link between microbial communities and abiotic parameters, the temporal changes in community composition, the interactions between microbial communities or even the impact of climate change Anneville et al. (2002); Karl and Church (2014); Herren and McMahon (2018); Obertegger et al. (2019). The grain of these studies is often based on a monthly sampling, but rarely on smaller frequencies. For instance, the Linnaeus Microbial Observatory<sup>3</sup>, consists in a weekly sampling since 2011, and focuses on the three domains of life (bacteria, phytoplankton, zooplankton) as well as on viruses. Among other results, the data collected with the LMO allow to decipher at a finer scale the variability of microbial communities and their influence on biogeochemical cycles (Bunse et al., 2019).

Microbial ecology studies benefit from long time surveys on the understanding of microbial communities’ structuration processes. However, these surveys remain scarce in freshwater systems. In those ecosystems, surveys focus mostly on bacterioplankton composition, with time scales ranging from five to ten years and they covered different environments such as freshwater bog lakes and Lake Mendota (Wisconsin, USA), a shallow lake in Estonia, Lake Donghu (China), Lake Tiefwaren (Germany) (Shade et al., 2007; Rösel et al., 2012; Kara et al., 2013; Tammert et al., 2015; Linz et al., 2017; Yan et al., 2017). Studies on phytoplankton are also conducted, using microscopy most of the time. For instance, a study in Lake Mondsee (Austria) of 6 years compared the pattern detected between seasonal and a monthly sampling of the phytoplanktonic communities (Bergkemper and Weisse, 2018). The longest temporal survey focusing on both eukaryotic and prokaryotic communities that we found in literature was conducted in Lake Tovel (Italy), with a monthly survey over four years (Obertegger et al., 2019). We also found a study covering 20 years of survey in Lake Gjersjøen (Norway) (Xiao et al., 2014). The sampling was conducted weekly, but on discontinuous periods. However, even if they used high throughput sequencing of both eukaryotic and prokaryotic communities, they did not investigate the temporal variations of community composition. In the case of small freshwater ecosystems, we were not able to find long term surveys on microbial. However, a six-year survey of the water quality and the macrophytes communities revealed that increasing the time of sampling improves the understanding of the ecosystems because it allows to capture more hydrological variability (Pätzig et al., 2020). This underlines the importance to conduct long time surveys to study the functioning of small freshwater ecosystems.

---

<sup>1</sup><https://www6.inrae.fr/soere-ola/Sites-d-etude>

<sup>2</sup><https://lter.limnology.wisc.edu>

<sup>3</sup><https://lnu.se/en/research/searchresearch/linnaeus-microbial-observatory-lmo/>

## 1.4 Aims of the thesis

As explained in section 1.1, freshwater ecosystems are diverse and can be divided between lentic and lotic ecosystems, as well as between large and small ecosystems. This results into a wide variability in their functioning. Large freshwater ecosystems represent a large volume of water, but are rather rare in the landscape (Downing, 2010). These ecosystems are expected to be stabler in time, because their large volume of water induces a strong buffering effect. At the same time, they are more subject to spatial variations within the ecosystem, through horizontal and vertical gradients. On the contrary, small freshwater ecosystems are common and numerous. Therefore, to have a good overview of the latter, it is better to sample several of them in order to capture variability in terms of abiotic composition, shape, or size for instance. Moreover, their small size makes them highly sensitive to global environmental variations and to punctual perturbations. Hence, it is important essential to focus on their variability over time at different time scales.

In this thesis, I aim at investigating the functioning of freshwater ecosystems, through the lens of microorganisms communities. The study of any kind of ecosystems requires multiple steps: explore the diversity and the autoecology of the communities inhabiting the ecosystems; determine the influence of top-down and bottom-up factors on the communities; conduct analyses with consistent spatiotemporal scales to have a broad picture of the different mechanisms. To do so, I focused on the biggest lake in the world, that is Lake Baikal, and five common and small ecosystems, that are four ponds and one brook from the Ile-de-France region (France). I used two approaches, consisting in a spatial survey in Lake Baikal, and a temporal survey of the five distinct small freshwater ecosystems. These two dimensions are important to capture the maximum of the community complexity. Finally, I focused on the microbial diversity and on the impact of abiotic and biotic factors on the communities using various tools, that is multivariate statistics and co-occurrence networks inference. I summarized all the methodological aspects in a chapter dedicated to the material and methods (Chapter 2). A more detailed version is presented in a specific section of each chapter.

### **Describe the eukaryotic and prokaryotic diversity in two types of freshwater ecosystems**

Before trying to understand the mechanisms taking place in an ecosystem, it is mandatory to explore and to determine the communities that inhabit it.

Lake Baikal is a highly studied ecosystem because of its importance on Earth, *e.g.* it is the largest freshwater volume. However, despite several studies on its metazoan communities, its microbial communities are still missing descriptions. Especially, there is a small number of studies focusing on the global scale, that is taking into account the entire lake surface and depth range. Therefore, I conducted a description of Lake Baikal microbial eukaryotic communities over the horizontal and vertical dimensions in summer (Chapter 3).

Small freshwater ecosystems are poorly studied ecosystems. The current literature indicates that they host highly diverse microbial communities, but we lack information on both their microbial eukaryotes and prokaryotes communities. In complement to a previous work on the microbial eukaryotic communities over a monthly survey of two years (from April 2011 to April 2013), I investigated the prokaryotic communities over the same period (Chapter 4). In order to get a more integrative view of the communities, we conducted in the same ecosystems a seasonal survey of eight years (October 2013 to January 2019) using a more recent sequencing method (Chapter 5). For the two-year monthly survey, we focus only on the small microbial communities (size in the range between 0.22 and 5  $\mu\text{m}$ ) because they are less described and known. For the seasonal eight-year survey, we considered a larger part of the communities, in order to capture potential predators and attached organisms (size in the range between 0.22 and 30  $\mu\text{m}$ ).

## **Describe the spatio-temporal dynamics of microbial communities from the three domains of life**

The spatio-temporal dimension is crucial for an integrative view of the functioning of aquatic ecosystems. The dimensions of Lake Baikal allowed us to conduct a large spatial survey, in the horizontal gradient, that is at different points at the surface of the lake, and in the vertical gradient, that is at different depths. This large survey allowed us to identify the factors that drive community composition in summer in the lake (Chapter 3). However, the temporal dimension is also very important. Some temporal variability exists in Lake Baikal, but the Lake's response to any environmental change is slow because of its large size. On the contrary, small freshwater ecosystems are very sensitive to environmental variations because of their small volume of water that induces a weak buffering capacity. Moreover, small ecosystems are easier to sample on a regular basis due to their small size and proximity. Hence, I aimed at describing the temporal dynamics of the global microbial communities over time in five small freshwater ecosystems. A monthly survey offered a detailed view of the dynamics (Chapter 4). A seasonal survey allowed us to check if the observed temporal patterns can be confirmed on a longer time



---

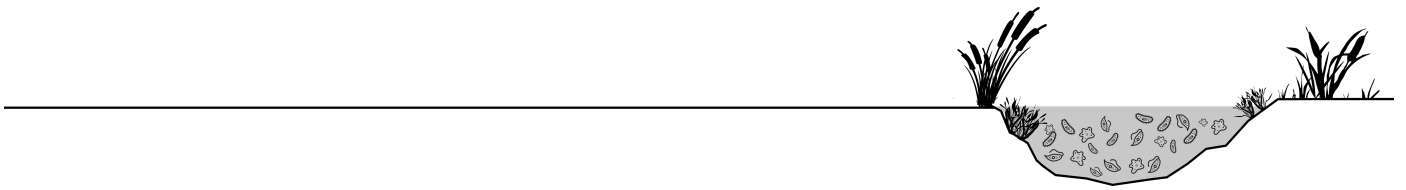
scale, or only result from a particular event that happened during the short survey. Also, as the seasonal survey was longer, it gave us the opportunity to investigate temporal patterns at the OTU scale (Chapter 5).

### **Explore the microbial interactions through co-occurrence networks**

Microbial communities are structured by abiotic and biotic factors. In order to explore the interactions between the microorganisms detected in our ecosystems, I conducted co-occurrence analysis. The inference of interaction networks at the various depth levels in Lake Baikal allowed us to determine if the depth induced different structures of communities (Chapter 3). In the small freshwater ecosystems, I explored the global community structure by building an interaction network combining the data of the five ecosystems (Chapter 5). Then, I focused on temporal patterns, in order to determine whether the structure of the communities was different between seasons, and/or if it was specific of each season (Chapter 6).

# Chapter 2

## Materials and methods

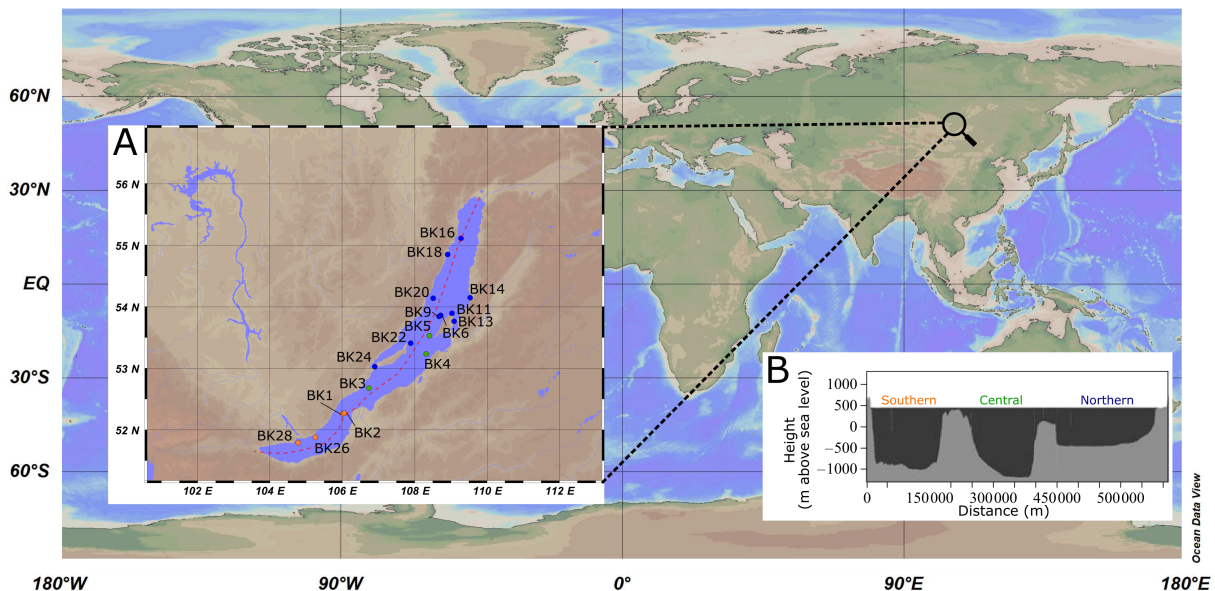




## 2.1 Presentation of the ecosystems

### 2.1.1 Lake Baikal

Lake Baikal is located in Russia, in Southern Siberia (see Figure 2.1). It is the deepest lake in the world, with a maximal depth of 1642 m, and represents up to 20% of the world's freshwater resources. It covers a wide surface and contains an approximate volume of 23600 km<sup>3</sup> of water (Table 2.1). The water composition is very specific, with high oligotrophic level, water transparency, and dissolved oxygen level, but also low salt content and dissolved organic level (Pavlichenko et al., 2015; Kurilkina et al., 2016; Khodzher et al., 2017; Cabello-Yeves and Rodriguez-Valera, 2019). Lake extends over a large region, from latitude 55°51'40" N to latitude 51°28'15" N and has a length of 636 km and a width of 79 km. It is commonly divided into three basins, distinguished by their bathymetry (see Figure 2.1, panel B). Contrary to the South basin, the North basin is well preserved from anthropic disturbances. For this study, the sampling was conducted over the entire Lake area, from North to South and at multiple depths and from the surface to the deepest locations. This work was conducted in collaboration with the Limnological Institute (Irkutsk, Russia).



**Figure 2.1** Location of Lake Baikal, in Siberia (Russia). *Panel A* represent Lake Baikal with the position of the 17 sampling sites. The red dot line corresponds to the longitudinal profile of the lake, which is represented in the *panel B*. The three basins are delimited by depth variations. Longitudinal bathymetric profile was taken from Sherstyankin et al. (2006). Maps were generated using the software Ocean Data View.

## 2.1.2 Small freshwater ecosystems in the Parc Naturel Régional de la Haute-Vallée de Chevreuse

### 2.1.2.1 Location of the five studied ecosystems

The “Parc Naturel Regional de la Haute Vallée de Chevreuse” (PNR-HVC) is located in the South West of Paris. This park was created in 1985 with the objective of preservation of the environment and the patrimony in the area. Its surface covers up to 630 km<sup>2</sup>. The PNR-HVC geological context is typical of the region, with successions of silt, millstone, sand and clays. Because of the permeability of the upper substrate and the impermeability of the clays, the region favours the formation of wetlands (Lefebvre, 2015). Around 1300 standing small fresh ecosystems have been counted in the PNR-HVC. In 2011, five small freshwater ecosystems of the PNR-HVC were selected with the help of the PNR-HVC managers (see Figure 2.2). All ecosystems are located less than 10 km apart but present specific environmental characteristics, despite their geographic proximity. Furthermore, they are located at about 30 min drive from our laboratory which facilitates their regular sampling during the year (De Meester et al., 2005).

**Table 2.1** General characteristics of the studied ecosystems

	Lake Baikal	Etang des Vallées	La Claye	Mare Gabard	Saint Robert	Ru Sainte Anne
Approximate perimeter (m)	2125.10 <sup>3</sup>	430	100	55	70	
Approximate width (cm)						30-100
Approximate surface (m <sup>2</sup> )	31722.10 <sup>6</sup>	11000	480	200	350	
Approximate maximal depth (cm)	1640.10 <sup>2</sup>	150	60	50	70	25

### 2.1.2.2 Description of the five ecosystems

The five studied ecosystems are called Etang des Vallées (EV), La Claye (LC), Mare Gabard (MG), Saint Robert (SR) and Ru Sainte Anne (RSA) (Figure 2.2). Among them, four are lentic ecosystems, *i.e.* still water, and one is a lotic ecosystem, *i.e.* running water. The ecosystems are representative of the diversity of small freshwater ecosystems in the Ile-de-France region.

They present different areas, providing us the benefits of the specificities of multiple size ranges (see Table 2.1, Oertli et al., 2002).

- Lentic ecosystems: Etang des Vallées is the largest and the deepest of the five ecosystems (GPS coordinates: 48°41'23.0" N; 1°54'59.2" E). It is located on the side of a national road, and rarely disturbed by human activity because fishing is restricted. La Claye is a pond located in the middle of a forest, on a peat substratum (GPS coordinates: 48°36'31.72" N; 1°56'17.33" E). It belongs to a complex of several ponds, and has undergone several drought events during the sampling period. Mare Gabard is also a pond located in the forest (GPS coordinates: 48°39'15.83" N; 1°55'20.26" E). The latter is supplied by a small brook, and presents strong sulphur odours, especially when the sediments are disturbed. Saint Robert is a pond located in an urban area, within a small hamlet (GPS coordinates: 48°39'54.82" N; 1°56'45.28" E). It is often covered with duckweed (likely *Lemna minor*). Until July 2016 Saint Robert harboured a large population of ducks that left the pond after a drought event.
- Lotic ecosystem: Ru Sainte Anne is a brook with highly variable flow and water level (GPS coordinates: 48°36'45.91" N; 1°58'16.61" E). The sampling point for the brook is located in a forest, close to a house, and to a national road. Many animals (*e.g.* wild boars) come there to drink.



**Figure 2.2** Location of the five studied small freshwater ecosystems

## 2.2 Data collection

### 2.2.1 Spatial survey - Lake Baikal

The sampling campaign in Lake Baikal was conducted in July 2017. A total of 65 samples were collected in 17 sites from a research vessel (see Figure 2.1, Panel A). One to eight samples were collected over the water column at each site, depending on the bottom depth. At the bottom, particular precautions were taken to avoid sediment disturbances. Water was collected using a 5 litres Niskin bottle for samples from 5 to 500 m depth, and a 10 litres bottle for samples from 500 to 1450 m. The water was processed immediately onboard, using a serial filtration procedure (pre-filtration through a 30  $\mu\text{m}$  pore-size Millipore, followed by a filtration onto a 0.22  $\mu\text{m}$  membranes (Nucleopore, Whatman membranes). Samples were then preserved in absolute ethanol at  $-20^{\circ}\text{C}$ . Three samples were treated differently on sites BK16 at 800 m and BK28 at 1000 and 1300m, using CellTraps instead of classical membranes.

For each sample, several physico-chemical parameters were measured: water temperature, pH, salinity, conductivity, redox potential and percentage of dissolved oxygen. All the measurements were conducted in unfiltered water using a Multi 350i (WTW, Weilheim, Germany) multiparameter probe.

### 2.2.2 Temporal survey - Small freshwater ecosystems

#### 2.2.2.1 Field sampling

The sampling of the five small freshwater ecosystems took place between April 2011 and January 2019. Two different time surveys were conducted, that are referred to as the monthly survey and the seasonal survey. The monthly survey was conducted every month from April 2011 to April 2013. The seasonal survey was conducted every three months (once per season, usually April, July, October and January) from October 2011 to January 2019. I participated to the sampling from October 2016 to January 2019.

We applied the same protocol to all the samples. Each time, the sampling of the five ecosystems was conducted within a restricted time frame (in max. 3 days). This allowed to have similar meteorological conditions between ecosystems, which is crucial to compare their communities. We collected the water at the same spot (deepest point) at surface in the five

ecosystems, using sterile 500 ml containers. For each ecosystem, 15 litres of water were collected in sterilised plastic carboys. Several physico-chemical parameters (water temperature, conductivity, dissolved oxygen concentration and pH) were measured *in-situ*, using a multiparameter probe WTW-Multi 350i: water temperature, conductivity, dissolved oxygen and pH.

Back to the laboratory, we immediately performed serial filtrations of the water samples through 100 and 30  $\mu\text{m}$  nylon mesh (Millipore), then through 5 and 0.22  $\mu\text{m}$  Nucleopore membranes (Whatman). The total volume of water filtered varied between 40 and 2490 ml. Membranes were then stored at  $-80^\circ\text{C}$  in 2 ml Eppendorf tubes, until DNA extraction.

### 2.2.2.2 Measure of physico-chemical parameters in the laboratory

The same day of sampling, we measured the concentrations of dissolved nutrients using colorimetric tests (Hach-Lange, Düsseldorf, Germany) once back to the laboratory. For all the samples, we measured the concentration of nitrites ( $\text{NO}_2^-$ ), nitrates ( $\text{NO}_3^-$ ), ammonia ( $\text{NH}_3$ ) and orthophosphates ( $\text{PO}_4^{3-}$ ) from the 30  $\mu\text{m}$  filtration. For future analyses of organic carbon concentration (DOC), water filtered through 0.22  $\mu\text{m}$  membrane was frozen at  $-20^\circ\text{C}$ . Finally, we collected plankton biomass smaller than 100  $\mu\text{m}$  on two GF/F filters that were conserved at  $-20^\circ\text{C}$  until chlorophyll *a* extraction.

We measured DOC concentration using a TOC VCSH analyzer (Shimadzu, Tokyo, Japan) following the protocol of Rochelle-Newall et al. (2014). We measured chlorophyll *a* concentrations using a protocol of ethanol pigment extraction, adapted by Claude Yéprémian from Ritchie (2006). Briefly, frozen filters were dried for one hour at  $37^\circ\text{C}$ , then grounded in 7 ml absolute ethanol and heated at  $70^\circ\text{C}$  for 20 min. Samples were centrifuged at 4000 rpm and 1 ml of the supernatant was collected in order to measure optical densities at 665 and 750 nm with a spectrophotometer (DR5000, Hach-Lange). We estimated the chlorophyll *a* concentration with the formula:

$$[\text{Chla}] = 11.9 \times \frac{\text{OD}_{665} - \text{OD}_{750}}{w} \times \frac{V_e}{V_f} \quad (2.1)$$

where  $[\text{Chla}]$  is the chlorophyll *a* concentration ( $\mu\text{g.l}^{-1}$ ), “ $\text{OD}_x$ ” the optical density where  $x$  is the wave length (nm) of optical density measurement, “ $w$ ” the width of the spectroscopic cuvette, “ $V_e$ ” the volume of ethanol (ml) used for the pigment extraction and “ $V_f$ ” the volume of water initially filtered (in litres) through the membranes. The value was arbitrarily validated only when differences between the two replicates were lower than 10 percent.



## 2.3 Samples preparation and processing

The acquisition of the sequence followed the same steps for the small freshwater ecosystems and for Lake Baikal. In the small freshwater ecosystems, we targeted both eukaryotic and prokaryotic communities, whereas for Lake Baikal, we only focused on the eukaryotic communities.

### 2.3.1 Extraction and amplification of DNA

I extracted DNA using the “PowerSoil DNA extraction” kit (Quiagen), following the manufacturer instructions. I first cut the filters using sterile scalpel or razor blades. For the monthly survey in the ponds, we only used the filters of 0.22 pore-size, covering the 0.22-5  $\mu\text{m}$  size-fraction. For the seasonal survey in the ponds and the Baikal spatial survey, we considered the size fraction ranging from 0.22 to 30  $\mu\text{m}$ , using the filters of 0.22 and 5  $\mu\text{m}$  pore-size. This size fractionation allows to filter higher volumes but especially to obtain samples enriched in diverse functional groups, *i.e.* picoplankton (0.2-2  $\mu\text{m}$ ), nanoplankton (2-20  $\mu\text{m}$ ) and microplankton (20-30  $\mu\text{m}$ ).

Portions of the targeted genes were amplified using a polymerase chain reaction (PCR). For the eukaryotic community, the targeted region of the 18S ARNr gene was located on the V4 region and was of about 550 pb long. To do so, I used 565F/1134R-UnonMet primers, the latter being biased against metazoa (Bower et al., 2004). For the study of the prokaryotes, I also targeted a portion of the hyper-variable region V4 of the 16S ARNr gene. To do so, I used the universal U151F/U806R primers. The targeted region had a length of about 290 pb. All primers were tagged using molecular identifiers (MIDs) of 10 pb, in order to track the sample of origin for each sequence, *i.e.* MID49 for January 2017 in Etang des Vallées. As I was sequencing together several samples, MIDs determine from which sample each sequence came from during the cleaning step of the sequences (see 2.3.3).

I conducted the amplifications in a 25  $\mu\text{l}$  mix containing 1.5 mM of  $\text{MgCl}_2$ , 0.2 mM of dNTPs, 0.3 mM of each primer, 0.5 U of Taq Polymerase (Taq Platinum, Invitrogen, Carlsbad, CA, USA) and between 0.3 and 3  $\mu\text{l}$  of the sample DNA. PCR cycles were composed of three steps: (i) a denaturation step (2 min at 94°C) (ii) followed by 35 amplification cycles (15 sec at 94°C, then 30 sec at 55°C for the eukaryotes or 30 sec at 58°C for the prokaryotes, and finally 90 sec at 72°C), (iii) and a final extension step (10 min at 72°C). I pooled five PCR products for each sample. Finally, I purified each pool using a purification QIAquick kit (Quiagen, Hilden, Germany). All the sampling and the bench work was performed by Paola Bertolino, Marianne

Simon and myself.

### 2.3.2 Sequencing

I measured the DNA quantity in each purified pool and I sent a tube with a mix containing an equivalent amount of amplified DNA of all pools to sequencing.

Sequences of the monthly survey were produced by pyrosequencing 454 GS FLX Titanium technology (Roche) by Beckman Coulter Genomics (Danvers, USA). Their length was of about 550 bp. The sequences of the seasonal and Baikal samples were obtained with Illumina MiSeq (Eurofins Genomics, Ebersberg, Germany). Their length was of about 300 bp.

### 2.3.3 Processing of the sequences

The sequences obtained through both technologies had to be cleaned before any analysis, to eliminate artefacts and sequencing mistakes.

#### 2.3.3.1 Monthly survey - Pyrosequencing 454 FLX Titanium

The treatment of the pyrosequencing sequences was conducted through a local pipeline, by Drs. Marianne Simon and Philippe Deschamps (DEEM Team, ESE), following several steps:

1. Cleaning of the sequences: First, reads were cleaned using the AmpliconNoise software to eliminate *(i)* errors in the primers, *(ii)* too short reads and *(iii)* errors that were produced during PCR amplification or pyrosequencing (Quince et al., 2011).
2. OTUs building: Clean reads were grouped into Operational taxonomic units (OTU). Eukaryotic and prokaryotic sequences were respectively gathered at 98% and 97% of identity. The grouping was conducted with the average linkage clustering method, using AmpliconNoise (Quince et al., 2011). The most abundant sequence in each OTU was selected as the reference sequence. All OTUs represented by only one sequence (*i.e.* singleton) were deleted.
3. OTUs affiliation: The representative sequences were blasted against two databases, SILVA and PR2, in order to determine their taxonomic affiliation.

4. Final modifications: All sequences were attributed to their sample of origin using the MIDs. To limit the presence of potential chimeras, OTUs detected in only one ecosystem were checked using KEYDNATool (<http://KeyDNATools.com>) and then manually blasted. OTUs affiliated to mitochondria and chloroplasts were deleted.

### 2.3.3.2 Seasonal and spatial survey - Illumina MiSeq

I conducted the treatment of the Illumina MiSeq sequences in collaboration with Guillaume Reboul and Philippe Deschamps (DEEM Team, ESE). The treatment steps are detailed below. Step 1 to 3 were performed using a local pipeline.

1. Cleaning of the sequences: Paired-end sequences were merged using the “flash” tool, generating Clean Merged Reads (CMRs) (Magoč and Salzberg, 2011). CMRs containing errors either in the primers or in the MIDs were removed. The remaining correct primers and MIDs were removed from the CMRs using the cutadapt tool (Martin, 2011). Thereafter, the CMRs were dereplicated into unique sequences using the “vsearch” tool (Rognes et al., 2016). Furthermore, chimeras were removed using a *de novo* process. For this, the sequences dataset and an algorithm are used. The sequences are pairwise compared in order to detect sequences (i.e. the chimeras) that are composed of portions of other sequences.
2. OTUs construction: Sequences were gathered into OTUs at 95% of identity, for both eukaryotes and prokaryotes, using the cd-hit tool (Li and Godzik, 2006; Fu et al., 2012). This taxonomic threshold grouped microorganisms at the genus level (Yarza et al., 2014). The cd-hit tool performs a classic *de novo* clustering of the OTUs (i.e. an iterative process group amplicons together one by one depending on their pairwise sequence similarity). Each OTU was represented by its most abundant sequence (i.e. reference sequence). Singletons were eliminated from the dataset.
3. OTUs affiliation: The reference sequence of each OTU was compared to a local database using the “vsearch” tool (Rognes et al., 2016). The local database was based on cultivated and environmental sequences obtained from SILVAv132 and PR2v4.5 (Quast et al., 2012; Guillou et al., 2012). OTUs affiliated to mitochondria and chloroplast were removed.
4. Taxonomic verification: Sequences with a percentage of identity with an environmental best-hit lower than 80% were manually checked. Their affiliations were compared to reference sequences with a phylogenetic placement. When the position in the phylogenetic tree

was incorrect, I conducted a blast in NCBI (Sayers et al., 2020). The correct affiliation was then modified by hand in the dataset.

## 2.4 Data analysis

### 2.4.1 Multivariate statistical analysis

Subsequent to the data acquisition and preparation steps, data-frames were created, which are two dimensional arrays, with samples in row and OTUs in column (see example in Table 2.2). In complement to the data-frames, the taxonomic affiliations of the OTUs were summarized in another table. To answer the ecological questions on the studied ecosystems, I computed several multivariate statistical analysis and indicators estimations on the data. Statistical analyses were performed using R software (R Core Team, 2019).

**Table 2.2** Example of a reduced data-frame used for the analyses, exposing the read numbers for each samples. The samples are in row, and the OTUs are in column.

	X118553	X118622	X118871	X119360
EV21	35	1	420	0
EV27	16	0	107	3
EV33	6	0	63	1

Prior to any analysis, I conducted a normalization of the data. Metabarcoding data are compositional, which means that they are not equivalent to real counting of OTUs abundance (Tsilimigras and Fodor, 2016; Gloor et al., 2017). It is mainly due because the total number of reads obtained depends on data collection, on the PCR amplification and on the sequencing technology. Therefore, it can be irrelevant to compare two samples without a preliminary normalization. For the pyrosequencing data, I applied two types of normalization:

- For simple community comparison: a total-sum normalization, where the number of reads for each OTU was transformed into a frequency, depending on the total number of reads per sample.
- For statistical analysis based on Bray-Curtis dissimilarities (see 2.4.1.1): a Wisconsin double standardization, where the reads of each OTU were first divided by their column maximum and then divided by the total number of reads in each sample. This method reduces the impact of abundant OTUs versus rare OTUs, which can be amplified with kind of dissimilarity calculation.

I conducted a rarefaction of the data for the Illumina MiSeq datasets. To do so, I selected the sample with the smallest number of reads. Then, for each sample, I selected a subsample of OTUs and reads to obtain a total number of reads equal the smallest number of reads selected previously.

### 2.4.1.1 Community composition in and within the ecosystems

The first step for analysing the microbial communities in a specific ecosystem consisted in the description of the community composition. In this work, I applied several indexes and analyses:

#### 1. Dissimilarities between samples

Several of the following analyses are based on the calculation of differences in the composition of pairwise samples. I calculated these dissimilarities between a sample X and a sample Y using the Bray-Curtis dissimilarities function  $d$  (Equation 2.2, Bray and Curtis, 1957).

$$d = \frac{\sum_{i=1} (|x_i - y_i|)}{\sum_{i=1} (x_i + y_i)} \quad (2.2)$$

where  $x_i$  and  $y_i$  are the frequencies of the  $i$  OTU in the X and Y samples. If  $d$  is equal to 0, then the two samples are identical. If  $d$  is equal to 1, then the two samples share no common OTUs.

#### 2. Diversity indexes

Three indexes were used in this study to describe community composition in samples. They were all computed using the `vegan` R package (Oksanen et al., 2007).

- *Richness*

The richness represents the total number of OTUs detected in one sample.

- *Evenness*

The evenness indicates whether the sample is composed of OTUs with relatively similar abundances, or composed of a few very abundant OTUs belonging to a community mostly composed of rare OTUs. The evenness  $e$  was calculated following Pielou (1966) (Equation 2.3).

$$e = \frac{\sum_i f_i \times \ln(f_i)}{\ln(S)} \quad (2.3)$$

where  $f_i$  is the OTU <sub>$i$</sub>  frequency and S the number of OTUs observed in the sample.

- *Simpson index*

The Simpson index  $D$  represents the probability that two random sequences belong to two different OTUs. It indicates the level of diversity found in the ecosystem. If  $D$  is equal to 1, the sample presents no diversity.  $D$  was calculated following Simpson (1949) (Equation 2.4)

$$D = 1 - \sum_i f_i^2 \quad (2.4)$$

where  $f_i$  is the OTU <sub>$i$</sub>  frequency.

### 3. Community composition comparison

As I was handling samples collected at various locations, depth and time periods, using the differences in terms of community composition was the best way to compare the data.

- *Venn diagrams*

In order to compare the OTU richness, Venn diagrams were generated using the `gplots` R package. They determine the amount of shared or exclusive OTUs within the sampling sites.

- *Non-metric multidimensional scaling (nMDS)*

This analysis presents in two dimensions the similarities between samples, based on Bray-Curtis dissimilarities. Several nMDS analyses were computed using the `vegan` R package. If the samples clustered into groups, a complementary analysis of similarities (ANOSIM) between groups was performed. It was useful to quantify and to statistically test for the differences between observed groups.

- *Hierarchical clustering analysis*

This clustering method combines similar samples into clusters in a global dendrogram. Bray-Curtis dissimilarities were used, combined to the Ward's method (Murtagh and Legendre, 2014). The analysis was conducted using the `stats` R package.

#### 2.4.1.2 Impact of the abiotic factors on community variation

The variations of the community composition can be explained by various factors, including abiotic factors such as physico-chemical parameters sampling period, or sampling depth. Therefore, I conducted two analyses to evaluate the impact of abiotic factors on community composition. They were conducted using the `vegan` R package.

- *PERMANOVA*

Permutational Analyses of Variance assesses globally the impact of various factors on

community composition (Anderson, 2014). For instance, in the temporal analysis, I used the role of the habitat (pond or stream), the ecosystem, the season, and the year of sampling as explanatory parameters.

- *Canonical Correspondence Analysis*

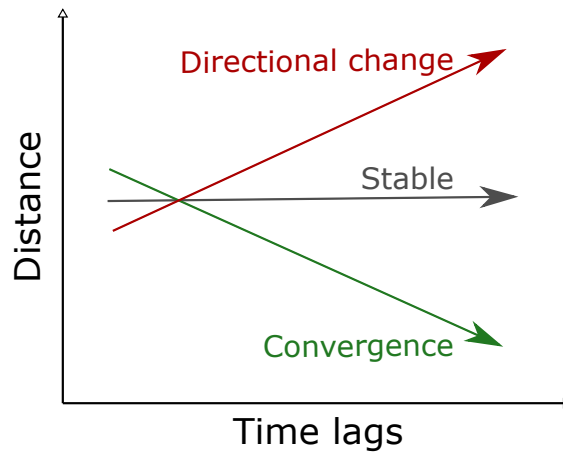
I used Canonical Correspondence Analysis (CCA) to estimate the contribution of each physico-chemical parameter to the community variance. In this analysis, the community tables were constrained by the physico-chemical table. At the same time, the putative multicollinearity between physico-chemical parameters was checked with the variance inflation factor (VIF). Physico-chemical parameters with a strong multicollinearity were removed to avoid bias in the analysis (under the condition that  $VIF > 10$ ). Finally, a forward selection was conducted to identify the most significant physico-chemical parameters in the model (by means of Monte Carlo permutation tests, on the basis of 999 permutations). I also used this analysis to estimate the influence of eukaryotic phyla on prokaryotic composition and inversely.

### 2.4.1.3 Temporal evolution of the communities

I studied the temporal evolution of communities by carrying out four different analyses. The aim was to identify patterns in the dynamic of the global community or of individual OTUs. It was thus possible to evaluate the rate of change of the community composition over time and to compare temporal changes in eukaryote and prokaryote community composition.

- *Time-lag analysis*

The composition of communities and abiotic factors are changing over time. Time-lag analysis (TLA) is a way to quantify the rate of change in the ecosystems (Collins et al., 2000). This approach is commonly conducted to decipher multiple temporal patterns (Fuhrman et al., 2015). For the communities, I compared the Bray-Curtis dissimilarities between all the possible sampling dates over increasing time lags. I used Euclidean distances for physico-chemical parameters. Linear regressions were computed to estimate the global evolution of community dissimilarities (or of environmental parameters) over time. The slope and the significance of the regression line indicated the degree and direction of the changes, *i.e.* convergence, divergence (directional change) or stability (Figure 2.3). To test for the accuracy of the observation, a randomization procedure was conducted on the data, with 5000 permutations, using the `lmPerm` R package (Wheeler et al., 2016). Polynomial regressions were computed to identify cycling patterns (*e.g.* seasonal). The best fitted polynomial model was selected using the Akaike Information Criterion (AIC) (Akaike, 1973).



**Figure 2.3** Possible interpretations of the linear model applied on the time-lag analysis. Adapted from Collins et al. (2000)

- *Co-inertia analysis*

I investigated the synchrony of community composition using co-inertia analysis (CIA) (Bock et al., 2018). CIA were computed using the `ade4` R package (Dray et al., 2007). With this analysis, I identified similar structure between two datasets within the same samples (*i.e.* the eukaryotic and the prokaryotic datasets). The correlation of the multivariate vectors, summarized as the RV index, assesses the strength of the synchrony between the two communities. It was tested using 999 permutations.

- *Turnover of OTUs*

In complement to the evolution of community comparison, I calculated the turnovers which indicates the type of changes occurring over time. For instance, if the turnover was high, the dissimilarities between samples could be explained by drastic changes in community composition. Turnovers were calculated between pairwise seasons (*e.g.* summer versus winter) or between samples collected at the same season but for different years. Turnovers were computed using the `codyn` R package, following the formula:

$$\text{Turnover} = \frac{S_{\text{gained}} + S_{\text{lost}}}{S_{\text{total}}} \quad (2.5)$$

where  $S_{\text{gained}}$  and  $S_{\text{lost}}$  are the numbers of species gained and lost between the first and the second time points respectively and  $S_{\text{total}}$  is the total number of species observed in the time points.

- *Identification of recurrent OTUs*

Recurrent OTUs were detected using a method developed by Giner et al. (2018). The



method is based on the calculation of a recurrence index (RI). I used the RI to identify the OTUs presenting a temporal pattern, *i.e.* occurring at the same period several times during the survey. The index calculation is based on the comparison between an autocorrelation function of the OTU (that-is a correlation of the time series with itself and a null distribution generated with a randomization of OTU abundances (1000 occurrences). RI was calculated for all OTUs in each ecosystem using the `ecolutils` R package (Salazar, 2018). However, OTUs occurring less than 4 times during the survey were excluded from the analysis. Indeed, these OTUs were not detected enough times to present a recurrence pattern.

## 2.4.2 Co-occurrence networks

### 2.4.2.1 PLN models

A large amount of models has been developed to generate microbial co-occurrence networks. Here, I applied a recent method based on Poisson Log Normal models (PLN), called PLNmodel (Chiquet et al., 2019). With this method, co-occurrence networks are inferred using inverted matrix of variance-covariance, called precision matrix. Using this matrix avoid the detection of associations resulting from indirect covariations (between two OTUs due to their direct association to a third OTU for instance) (Kurtz et al., 2015). PLNmodel is based on the counts of the OTUs reads that are considered as Poisson distributions. The model selection is conducted by choosing one optimal penalty parameter. This parameter controls the number of edges in order to obtain the most stable network. The optimal penalty was selected using the Bayesian Information Criterion (BIC), among a 50-size grid of penalties (Chen and Chen, 2008). Finally, the model takes into account several explicative parameters, such as the sequencing depth (*i.e.* total number of reads per sample) and various covariables, either quantitative (*e.g.* abiotic factors) or qualitative (*e.g.* season). These covariables can improve the model by taking into account environmental variability, but more importantly, by avoiding spurious interactions (Chiquet et al., 2019; Röttjers and Faust, 2018).

During this thesis, I constructed seasonal networks for the small freshwater ecosystem dataset. I built a network for each season and each ecosystem, using OTUs detected in at least 6 seasons over the 8 years of survey (*e.g.* 6 over the 8 autumn periods sampled) and having a mean abundance higher than 0.05%. I used three abiotic parameters (conductivity, dissolved oxygen, pH) as covariables in the model. I also constructed networks with Lake Baikal dataset. I have built one network for each depth class, on the OTUs present in 20% of

the samples, and with a mean abundance higher than 0.01%. I considered the sampling basin (north, central or south) as covariable in order to avoid spurious associations

### 2.4.2.2 Visualization of the networks

Based on the precision matrix, I built the networks using the `ggnet2` R package. The OTUs are represented by nodes, and the direct covariations between them by the edges. I used graphical option to colour nodes according to their taxonomy or lifestyle, and to differentiate positive from negative nodes.

### 2.4.2.3 Topology of the networks

For each network, I calculated classical networks properties, using the `igraph` R package (Csardi et al., 2006). The first characteristics are the total number of nodes (*i.e.* the number of OTUs implemented in the model) and the number of connected nodes (*i.e.* the number of nodes involved in the final network). Then, I counted the number of edges, with a distinction between the edges representing a positive or a negative relationship. For each node, I estimated the node degree, *i.e.* the average number of edges that connected the nodes. I also calculated the mean degree of all the nodes for each network. Finally, I calculated three indicators called average path length, clustering coefficient and connectance. The average path length represents the mean of all the shortest path between two nodes in the network (Berry and Widder, 2014). The clustering coefficient was used to estimate the tendency of nodes to cluster together. It represents the probability that two nodes having a similar neighbour are connected to each other (Delmas et al., 2019). The connectance quantifies the proportion of links that are actually established in the network compared to the number of possible potential links. It is equal to one when all possible links are established. I calculated the connectance  $c$  using a formula adapted to undirected networks (Equation 2.6, Barrat et al., 2008).

$$c = \frac{2E}{N \times (N - 1)} \quad (2.6)$$

where  $E$  is the number of edges and  $N$  the number of nodes in the network.

Globally, co-occurrence networks presented complex structures. To simplify them, I broke down networks into groups. For each network, I applied a stochastic block model approach based on the precision matrix using the `blockmodels` R package (Leger, 2016). This method

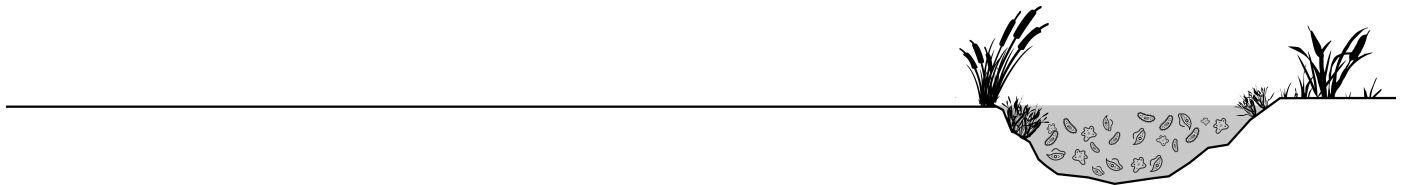
synthetizes the overall network structure by gathering nodes into a blockmodel group (BMG) and characterized by similar modes of interactions. For instance, the nodes can be highly connected to the rest of the network or poorly connected to the network but featuring high interconnections between the BMG members.

To summarize the information provided by the different properties of the networks, I conducted principal component analysis (PCA). The dataset was constituted of the following variables, centred and scaled: percentage of connected nodes, percentage of positive links, clustering coefficient, connectance, mean degree, average path length.

I used generalized linear models to estimate the variation of properties depending on different effects, *e.g.* ecosystem, season, or lifestyle. They were based on a quasi-binomial distribution. The first model, the model “degree”, was used to estimate the impact of different effects on the node degree. The lifestyle of the OTU and the interaction with the variables season and ecosystem were used as co-variables. An ANOVA conducted on the results of the model allowed us to estimate the importance and the significance of the chosen co-variables. The second model, the model “parasite”, was used to estimate the effect of season on the number of edges established with putative parasitic OTUs. A supplementary covariable was added to the model, in order to take into account the initial proportion of putative parasites in the networks. I tested the different covariable combination of both models using the `emmeans` R package (Lenth et al., 2018).

## Chapter 3

# Environmental drivers of plankton protist communities along latitudinal and vertical gradients in the oldest and deepest freshwater lake





## 3.1 Context and objective

Lake Baikal is located in Siberia (Russia) and is a “one-of-a-kind” lake. Especially, it is the oldest lake on Earth, it contains the largest freshwater volume and it is the deepest lake. However, even so it has been intensively investigated, it remains poorly investigated in terms of microbial communities. A survey was conducted by our team in July 2017 to the structure of the microbial communities inhabiting the lake over horizontal and vertical profiles. Water and sediment samples were collected, in order to have a detailed insight into the microbial composition of the lake over the summer period. As Lake Baikal reaches a maximal depth of  $\sim 1640\text{m}$ , it is the unique freshwater ecosystem where it is possible to explore the variation of community composition along so deep vertical profiles. Moreover, Lake Baikal can be viewed as “freshwater sea”, which differentiate it from other freshwater lakes and can lead to specific communities and functioning. Metabarcoding data obtained using Illumina MiSeq on 18S rRNA genes, allowed me to:

1. Describe and compare microbial eukaryotic communities in Lake Baikal over horizontal (latitude) and vertical (depth) gradients.
2. Investigate for the potential presence of typical marine lineage in the freshwater sea that is Lake Baikal.

## 3.2 Manuscript

Gwendoline M. David<sup>1</sup>, David Moreira<sup>1</sup>, Guillaume Reboul<sup>1</sup>, Nataliia V. Annenkova<sup>2</sup>, Luis J. Galindo<sup>1</sup>, Paola Bertolino<sup>1</sup>, Ana I. López-Archilla<sup>3</sup>, Ludwig Jardillier<sup>1</sup> and Purificación López García<sup>1</sup>

<sup>1</sup>Université Paris-Saclay, CNRS, AgroParisTech, ESE, 91405, Orsay, France

<sup>2</sup>Limnological Institute, Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

<sup>3</sup>Departamento de Ecología, Universidad Autónoma de Madrid, Madrid, Spain

Submitted in *Environmental Microbiology*, currently under review. Available in preprint on bioRxiv (<https://www.biorxiv.org/content/10.1101/2020.09.26.308536v1>).

### 3.2.1 Abstract

Identifying which abiotic and biotic factors determine microbial community assembly is crucial to understand ecological processes and predict how communities will respond to environmental change. While global surveys aim at addressing this question in the world's oceans, equivalent studies in large freshwater systems are virtually lacking. Being the oldest, deepest and most voluminous freshwater lake on Earth, Lake Baikal offers a unique opportunity to test the effect of horizontal versus vertical gradients in community structure. Here, we characterized the structure of planktonic microbial eukaryotic communities (0.2-30  $\mu\text{m}$  cell size) along a North-South latitudinal gradient ( $\sim 600$  km) from samples collected in coastal and pelagic waters and from surface to the deepest zones (5-1400 m) using an 18S rRNA gene metabarcoding approach. Our results show complex and diverse protist communities dominated by alveolates (ciliates and dinoflagellates), ochrophytes and holomycotan lineages, with cryptophytes, haptophytes, katablepharids and telonemids in moderate abundance and many low frequency lineages, including several typical marine members, such as diplomonads, syndinians and radiolarians. Depth had a strong significant effect on protist community stratification. By contrast, the effect of the latitudinal gradient was marginal and no significant difference was observed between coastal and surface open water communities. Co-occurrence network analyses showed that epipelagic communities are much more interconnected than meso- and bathypelagic communities and suggest specific biotic interactions between autotrophic, heterotrophic and parasitic lineages that influence protist community structure. Since climate change is rapidly affecting Siberia and Lake Baikal, our comprehensive protist survey constitutes a useful reference to monitor ongoing community shifts.

### 3.2.2 Introduction

Of all ecosystems, freshwater reservoirs are the most dynamic and concentrate a high biodiversity (Rolls et al., 2018). Freshwater ecosystems are particularly vulnerable to climate change owing to a higher exposure and sensitivity to increasing temperature and other altered conditions, limited dispersal across these fragmented habitats and little-known, but likely modest, resilience potential (Woodward et al., 2010; Markovic et al., 2017). Since microorganisms are crucial in biogeochemical cycles, the impact of climate change will strongly depend on how they will respond to environmental challenge (Cavicchioli et al., 2019). Permafrost-covered areas in the Arctic region (Schuur et al., 2015) and forest-steppe ecotones in Siberia are among the most heavily impacted regions by global warming (Mackay et al., 2017). This includes Lake Baikal, in southern Siberia, which is the oldest (ca. 30 Myr), deepest, and most capacious freshwater lake on Earth (Müller et al., 2001). Lake Baikal is rapidly changing, as can be told from trends in hydrological and hydrochemical processes (Moore et al., 2009; Shimaraev and Domysheva, 2013). The lake sediments represent a continuous record of past climate for over 12 million years (Kashiwaya et al., 2001; Prokopenko et al., 2002) such that Lake Baikal is a unique model to understand and predict microbial community change and how this is linked to carbon cycling and hydrological processes. A mandatory prerequisite for such a task is to have comprehensive information about the existing microbial community structure. However, if the broad biodiversity of Lake Baikal metazoans, including many endemisms (1455 out of 2595 species described), has been amply documented in the past two centuries, that of microbial life is highly fragmentary. One of the reasons relates to the large dimensions of the lake, which is around 640 km long, attains a depth of ca. 1650 m and contains around 20% of the Earth's unfrozen freshwater (Sherstyankin et al., 2006; UNDP-GEF, 2015). This, together with its geographical location and its association to a rifting zone make Lake Baikal unique and listed as UNESCO World Heritage Site (UNDP-GEF, 2015). The lake is divided in three major basins (Northern, Central, Southern) by, respectively, the Academician Ridge and the Selenga river delta (Mats and Perepelova, 2011). Its surface freezes in winter for several months, favoring coastal downwelling and deep-water oxygenation (Schmid et al., 2008; Moore et al., 2009). As a result, Lake Baikal ultraoligotrophic waters are globally cold ( $\sim 4^{\circ}\text{C}$ ) and oxygen-rich down to the bottom (Schmid et al., 2008; Moore et al., 2009; Shimaraev and Domysheva, 2013; Troitskaya et al., 2015). Baikal also uniquely hosts methane hydrates, which are stabilized by the low temperatures and high pressures (De Batist et al., 2002; Granin et al., 2019). All these features make Lake Baikal akin a freshwater sea. Microbial diversity in Lake Baikal plankton was first studied by classical observation and cultural approaches (Maksimova and Maksimov, 1972; Maksimov et al., 2002; Belkova et al., 2003) before molecular tools started to be applied at the beginning of the century (Glöckner et al., 2000) and expanded more recently with the



generalization of high-throughput sequencing. Several 16S rRNA gene-based metabarcoding studies have targeted pelagic bacteria diversity (Kurilkina et al., 2016; Zakharenko et al., 2019; Belikov et al., 2019; Wilburn et al., 2019) and, more recently, metagenomic analyses have been used to characterize planktonic prokaryotic communities from sub-ice (Cabello-Yeves et al., 2018) and deep waters (Cabello-Yeves et al., 2020), virus-bacteria assemblages in coastal waters (Butina et al., 2019) or viruses from the pelagic zone (Potapov et al., 2019). Microbial eukaryotes have only been partially studied by 18S rRNA gene metabarcoding. Several of these studies focused on phytoplankton, either on specific groups, such as diatoms (Zakharova et al., 2013) or dinoflagellates (Annenkova et al., 2011), or on whole communities, from winter sub-ice waters (Bashenkhaeva et al., 2015) to spring blooms (Mikhailov et al., 2015, 2019b,a). Remarkably few studies have aimed at charactering the diversity of all microbial eukaryotes, especially in a comparative manner. Yi et al. (2017) analyzed protist diversity by 454 sequencing of 18S rRNA gene V9-region amplicons along the Southern basin water column (52-1450m). More recently, Annenkova et al. (2020a) determined the community structure of small protists (0.45-8  $\mu\text{m}$  cell-size fraction) from surface waters (1-15-50 m) across the lake via 18S rRNA gene V4-region metabarcoding and suggested that some clades within known protist groups might be endemic. Nonetheless, we still lack a comprehensive view about how microbial eukaryotes distribute in the lake plankton, across basins and throughout the complete water column and, crucially, which are the most influential parameters determining community structure. In this work, we carry out a wide-ranging comparative study of Lake Baikal planktonic protist communities in the 0.2-30  $\mu\text{m}$  cell-size range using a 18S rRNA gene metabarcoding approach to study distribution patterns and to test whether depth, latitude or the coastal versus pelagic location determine community structure. With this aim, we analyze 65 samples from 17 sites across a  $\sim$ 600 km latitudinal North-South transect along the three lake basins and from littoral shallow areas to deep water columns covering the epi-, meso- and bathypelagic region. Our results show complex and diverse protist communities that are mostly structured by depth and that include several typical marine lineages in low abundance. Network analyses show that epipelagic communities are much more interconnected than meso- and bathypelagic communities, suggesting potential specific biotic interactions between autotrophs, heterotrophs and parasites.

### 3.2.3 Material and methods

Lake Baikal water samples were collected at different depths from seventeen sites distributed along a North-South transect during a French-Russian research cruise in the summer of 2017. Sites were chosen to cover littoral (Moore et al., 2009) and open water (Shimaraev and Domyshева, 2013) samples, including the deepest zones in the three major basins of the lake. In

total, 65 water samples were collected from depths ranging from 5 to 1400 m; deep samples were collected far from the bottom to avoid sediment disturbance (Supplementary Table 3.A.1). Samples were collected with Niskin bottles (5 l for epipelagic waters, 10 l for meso- and bathypelagic waters). The physicochemical parameters of lake waters were measured with a multiparameter probe Multi 350i (WTW, Weilheim, Germany). The water was sequentially filtered onboard immediately after collection through 30  $\mu\text{m}$  and 0.22  $\mu\text{m}$  pore-size Nucleopore filters (Whatman, Maidstone, UK) and 0.2  $\mu\text{m}$  pore-size Cell-Trap units (MEM-TEQ Ventures Ltd, Wigan, UK). Volumes of water samples filtered through Cell-Traps were smaller (samples indicated with an asterisk in Figure 3.1). The recovered biomass and biomass-containing filters were fixed in absolute ethanol and stored at  $-20\text{ }^{\circ}\text{C}$  until processed. DNA purification, 18S rRNA gene-fragment amplification and sequencing DNA was purified using the Power Soil<sup>TM</sup> DNA purification kit (Qiagen, Hilden, Germany). 18S rRNA gene fragments ( $\sim 530$  bp) encompassing the V4 region were PCR-amplified using EK-565F-NGS (5'-GCAGTTAAAAGCTCGTAGT-3') and UNonMet (5'-TTTAAGTTTCAGCCTTGCG-3'), the latter biased against metazoans (Bower et al., 2004). Primers were tagged with specific 10 bp molecular identifiers (MIDs) for multiplexed sequencing. To minimize PCR-associated biases, five PCR reaction products per sample were pooled. PCR reactions were conducted in 25  $\mu\text{l}$  reaction mixtures containing 0.5-3  $\mu\text{l}$  of eluted DNA, 1.5 mM  $\text{MgCl}_2$ , 0.2 mM dNTPs, 0.3  $\mu\text{M}$  primers and 0.5 U Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA) for 35 cycles (94 $^{\circ}\text{C}$  for 30 s, 55-58 $^{\circ}\text{C}$  for 30-45 s, 72 $^{\circ}\text{C}$  for 90 s) preceded by 2 min denaturation at 94 $^{\circ}\text{C}$  and followed by 5 min extension at 72 $^{\circ}\text{C}$ . Pooled amplicons were purified using QIAquick PCR purification kit (Qiagen, Hilden, Germany). Amplicons were sequenced using paired-end (2x300 $\times$  bp) Illumina MiSeq (Eurofins Genomics, Ebersberg, Germany). Sequences have been deposited in GenBank under the BioProject number PRJNA657482 (BioSamples SAMN15830589 to SAMN15830657).

### 3.2.3.1 Sequence and phylogenetic analyses

We used an in-house bioinformatic pipeline to process raw sequences. Paired-end reads were merged with FLASH (Magoč and Salzberg, 2011) under strict criteria and assigned to specific samples based on their MIDs. MID and primer sequences were trimmed using CUTADAPT (Martin, 2011). Cleaned merged reads were next dereplicated to unique sequences using VSEARCH (Rognes et al., 2016), which was also used to detect and eliminate potential chimeras. Non-chimeric sequences from all samples were pooled together to define operational taxonomic units (OTUs) at a conservative threshold of 95% identity for 18S rRNA genes using CD-HITEST (Fu et al., 2012) and SWARM (Mahé et al., 2015). Singletons were excluded from subsequent analyses. OTUs were assigned to taxa based on their similarity with a local

18S rRNA database build from SILVA v128 (Quast et al., 2012) and PR2v4.5 (Guillou et al., 2012). OTUs less than 80% identical to their best environmental hit were blasted against the GenBank nr database (<https://ncbi.nlm.nih.gov/>) and assigned manually by phylogenetic placement analyses. Briefly, the closest hits to our OTUs in SILVA and PR2 were aligned with full 18S rDNA reference sequences covering the eukaryotic diversity using MAFFT (Katoh and Standley, 2013). After removal of uninformative sites with trimAl (Capella-Gutiérrez et al., 2009), we built a tree with full reference sequences with IQ-tree (Nguyen et al., 2015) under a GTR+G+I sequence evolution model. OTU sequences were aligned to the reference alignment and then placed in the reference phylogenetic tree 150 using EPA-ng (Barbera et al., 2019). OTUs with no reliable affiliation were maintained as “Unclassified”. Maximum likelihood phylogenetic trees of diplomonad and radiolarian OTUs were reconstructed from specific MAFFT alignments including their closest blast hits and reference sequences with PhyML (Guindon and Gascuel, 2003) applying a GTR+G+I (4 categories) model of sequence evolution. Bootstrap values were obtained from 100 replicates.

### 3.2.3.2 Statistical analyses

We generated a table of eukaryotic OTU read abundance in the different samples of Lake Baikal for diversity and statistical analyses (Supplementary Tables 3.A.1 and 3.A.2). To avoid biases due to differences in absolute numbers of reads per sample, we rarefied our sequences to the second smaller number of reads (9771 in BK16.500m). BK28.100m was excluded from this process due to its lower number of reads. Statistical analyses were conducted on these data with R (R Core Team, 2019). Richness and diversity indices were calculated using the ‘vegan’ package (Oksanen et al., 2007). Evenness was calculated according to Pielou (Pielou, 1966). To see if these indices were significantly different between sampling depths and basins, we performed Wilcoxon tests between the groups distributions using R. Likewise, to test the effect of sampling point, basin and depth class on protist community composition across samples, we conducted permutational multivariate analyses of variance (PERMANOVA) based on Wisconsin-standardized Bray-Curtis dissimilarities, using the `adonis` function of the `vegan` package. Across-sample community composition differences were visualized using non-metric multidimensional scaling (NMDS) analysis, also on Wisconsin-standardized Bray-Curtis dissimilarities. To connect communities according to specific origin we drew ellipses with the ‘`ade4`’ package (Dray et al., 2007). To test the significance of groups revealed by NMDS, we applied analysis of similarity (ANOSIM) tests with 999 permutations. Principal component analysis (PCA) of abiotic parameters based on centered and scaled data was performed with `FactoMineR` (Lê et al., 2008).

### 3.2.3.3 Network analysis

We built co-occurrence networks for each depth category (epipelagic, mesopelagic, bathypelagic) using a multivariation Poisson lognormal model with the ‘PLNmodels’ R-package (Chiquet et al., 2019) in order to account for depth-class differences between samples and potential additional covariables (specifically the sampling basin). We retained for the analysis OTUs present in more than 20% of samples and abundances higher than 0.01%. For model selection, we used Bayesian information criteria with a 50 size grid of penalties. Networks were visualized with the ‘ggnet’ R-package. To further analyze network structure, we carried out a block model analysis using a stochastic block model approach on the binary co-occurrence network using the ‘blockmodel’ R-package (Leger, 2016), which synthesizes the overall network structure by gathering nodes in groups with similar modes of interactions. Network properties were calculated using the igraph R-package (Csardi et al., 2006). Properties included the number of positive and negative edges, the total number of nodes and number of connected nodes. Network mean degrees correspond to the average number of established edges. The average path length indicates the mean number of edges necessary to link a given node randomly to another. Network complexity was estimated using two indicators: connectance and clustering coefficient. The connectance was calculated as  $c = \frac{2E}{N \times (N-1)}$ , where E is the number of edges and N the number of nodes (Barrat et al., 2008). Connectance is 1 when all possible links are established. The clustering coefficient is the probability that two nodes having a similar neighbor are connected to each other (Delmas et al., 2019). It varies between 0 and 1; low values indicated poor connectivity.

## 3.2.4 Results and discussion

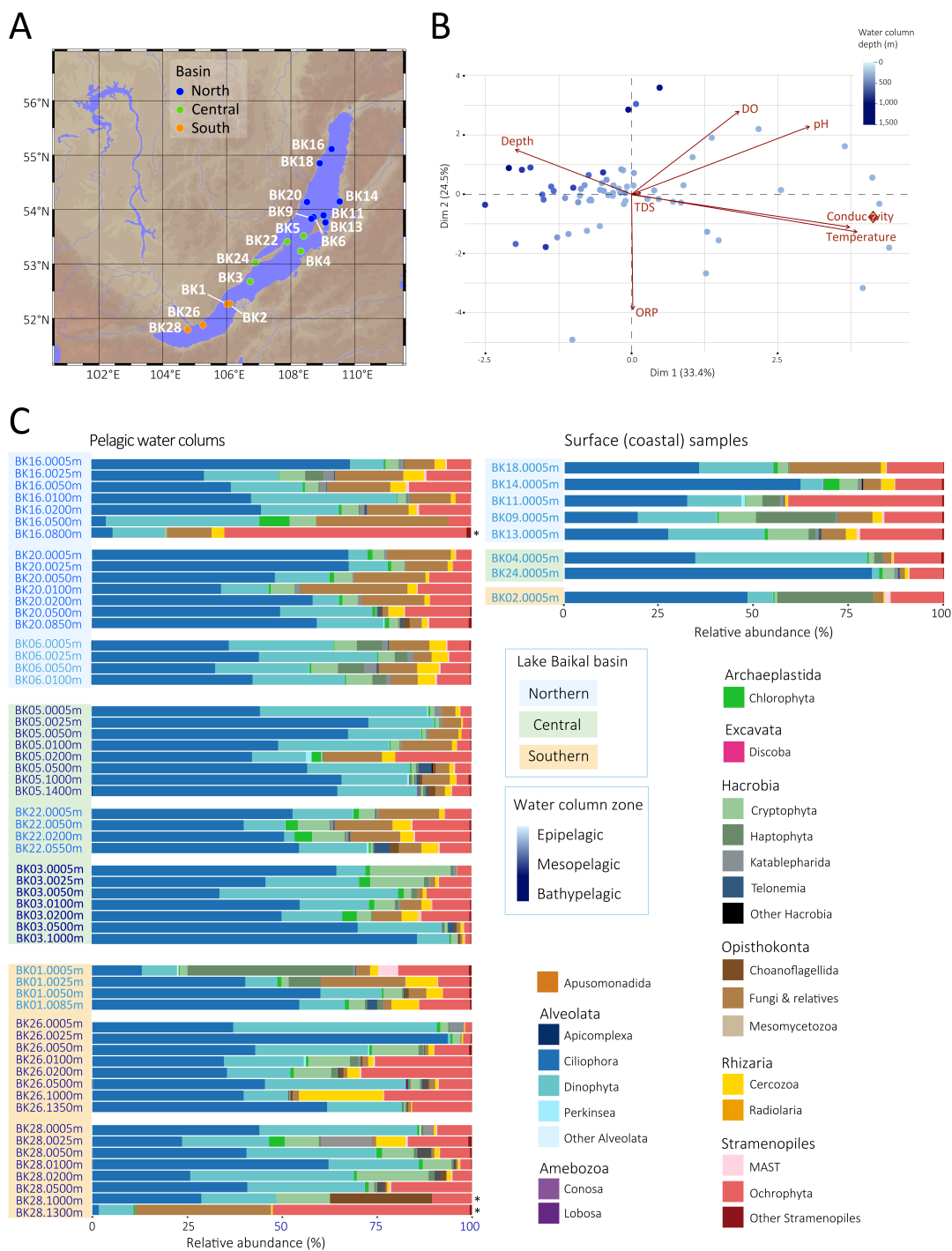
### 3.2.4.1 Abiotic variables across sampling sites

We collected Lake Baikal water samples along the Northern, Central and Southern basins from the same established depths in the water column (except for the deepest sample, which was collected close to the bottom but at sufficient distance –minimum 45 m– to avoid sediment influence) (Figure 3.1A; Supplementary Table 3.A.1). Samples from coastal areas were always collected at 5 m depth in the water column. The measured physicochemical parameters were remarkably stable across sites and depths. Temperature ranged from 3.6 to 15.3°C, but was globally low (average 5.7°C; only five surface samples exceeded 9.5°C), and significantly higher in epipelagic samples (Supplementary Figure 3.B.1). pH ranged from 7.45 to 8.47. Salinity was extremely low (always 0.0 PSU) as, accordingly, conductivity and total dissolved solids (TDS).

Dissolved oxygen was high (mean 79.5%). Like temperature, pH, conductivity and dissolved oxygen in mesopelagic waters were significantly lower than in epipelagic samples. Bathypelagic parameters were similar to those of the mesopelagic zone but more variable. In terms of basins, temperature, pH, conductivity and dissolved oxygen were higher in the Southern basin, which is also more impacted by human activities and pollution, notably aromatic hydrocarbons and mercury brought by the Selenga river (Adams et al., 2018; Roberts et al., 2020), although only oxygen and, marginally, conductivity were significantly different (Supplementary Figure 3.B.1). The two main axes of a PCA considering these abiotic parameters explained 58% of the variance (Figure 3.1B). Surface samples correlated with higher temperature, conductivity and, to a lower extent, pH and dissolved oxygen. These observations suggest that depth, as a proxy for light accessibility but also temperature and other abiotic parameters, might be a strong environmental driver for community structure.

#### 3.2.4.2 Composition of planktonic protist communities

To study the diversity and relative abundance of microbial eukaryotes in Lake Baikal plankton, we concentrated cells in the 0.2-30  $\mu\text{m}$  diameter fraction by successive filtration steps. This fraction thus integrated pico- (0.2-2  $\mu\text{m}$ ), nano- (2-20  $\mu\text{m}$ ) and small microplankton (20-30  $\mu\text{m}$ ), covering a wider protistan spectrum than some previous comparative studies (Annenkova et al., 2020a). We purified DNA and massively sequenced (MiSeq Illumina, 2x300 bp) multiplexed 18S rRNA gene V4 region amplicons. After discarding low-quality reads, we generated 6 405 343 high-quality merged paired-end sequences that we clustered in operational taxonomic units (OTUs) at different thresholds. We determined 27 504 OTUs and 9 700 OTUs at, respectively, 98% and 95% sequence identity (CD-HIT). SWARM yielded 11 590 OTUs (Supplementary Table 3.A.1), only slightly higher than the number of OTUs defined at the latter cut-off. For our subsequent comparative analyses, we deliberately retained OTUs defined at 95% sequence identity threshold. Many diversity studies focus on exact sequence variants after sequence error correction (Callahan et al., 2016) that can inform about individual strain variation. However, for the purpose of this comparative study, we chose to use conservatively defined OTUs that, on average (this varies across phylogenetic groups), correspond to the genus or species-genus level (Caron et al., 2009). This taxonomy cutoff level is relevant for broad comparative ecological studies (members of the same genus are likely to have similar general functions, despite inter-strain or species-specific niche differences), while operationally diminishing the number of handled OTUs. Based on sequence MIDs, the abundance of the different OTUs was determined for each sample (Supplementary Table 3.A.1). To avoid potential biases in diversity and relative abundance estimates linked to differences in the total number of reads, we rarefied



**Figure 3.1** Sampling sites and overall planktonic protist community composition in Lake Baikal. **A**, map of Lake Baikal showing the different sampling sites across the three major lake basins (indicated by colors). **B**, Principal component analysis (PCA) of samples according to their associated physicochemical parameters. The number near the points correspond to the sampling site, and the color of the points indicates their sampling depth. TDS, total dissolved solids; DO, dissolved oxygen; ORP, oxidation-reduction potential. Blue tones indicate the sampling depth in the water column, as indicated. **C**, Relative abundance of different high-rank eukaryotic taxa in Baikal plankton based on read counts for the defined OTUs. The asterisk indicates samples retrieved from Cell-Traps (Methods). Color codes for sample basin and depth origin as well as for the different taxa are indicated.

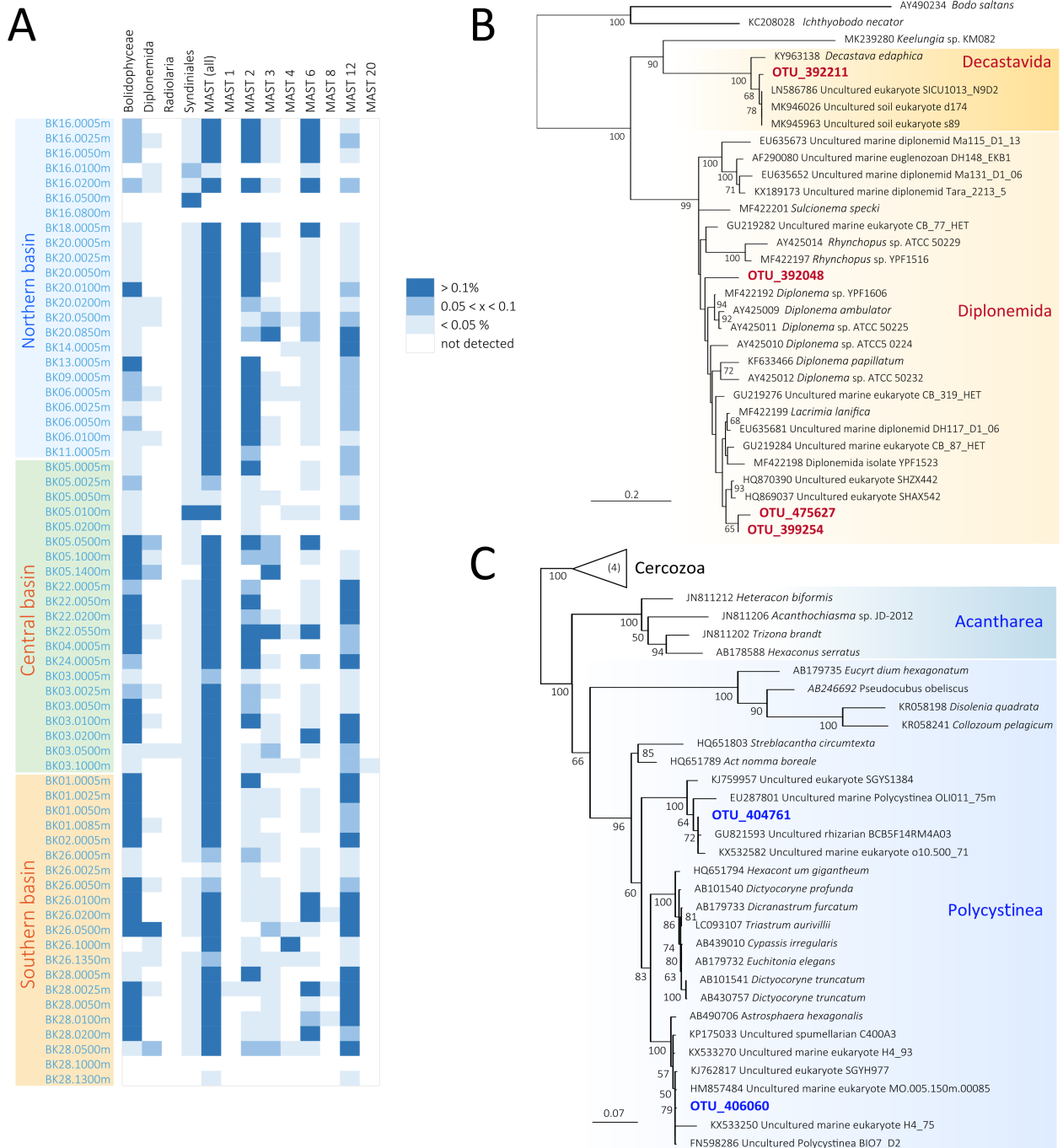
sequences to the same number across samples, which resulted in a global number of 4 570 genus-level OTUs. Nonetheless, accumulation curves showed that the diversity of planktonic protists was far from reaching saturation, even at the conservative genus level (Supplementary Figure 3.B.2). Richness significantly decreased in deep as compared to surface waters; so did evenness (Supplementary Figure 3.B.3). A lower evenness may be partly explained by the lower cell abundance in deeper waters, as the counts for each OTU become more aleatory. We did not observe richness differences across lake basins, but evenness appeared significantly higher in the Northern basin. From a phylogenetic perspective, our defined OTUs affiliated to at least 27 eukaryotic phyla belonging to several major eukaryotic supergroups (Figure 3.1C; Supplementary Table 3.A.2): the SAR clade (Stramenopiles, Alveolata, Rhizaria), Amoebozoa, Archaeplastida, Excavata, Opisthokonta and Hacrobia. Although we considered Hacrobia as originally described (Okamoto et al., 2009), they should be possibly split in two or more groups as the eukaryotic phylogeny progressively resolves (Burki et al., 2020). Ciliates and dinoflagellates (Alveolata), Ochrophyta (Stramenopiles) and Holomycota (Fungi and related lineages within the Opisthokonta) dominated plankton samples representing, respectively 48.4%, 21.5%, 12.6% and 8% relative sequence abundance. Cryptophyta, Haptophyta, Kathablepharida and Tenomemida displayed moderate abundances (0.5 to 5% reads) and were followed by a long tail of lower-frequency taxa in rank:abundance curves (Supplementary Figure 3.B.4). The major dominant groups were similar in all depths, with small variations in the deepest waters. Ciliates were by far the most abundant in terms of sequence reads. However, this observation is to be pondered by the fact that, in ciliate somatic macronuclei, rRNA genes are amplified several thousand times (e.g.  $\sim 9000$  copies in *Tetrahymena thermophila* (Ward et al., 1997)), such that their relative abundance in term of cells is certainly much lower. Although diatoms (Bacillariophyta, Ochrophyta), several of them considered endemic, are well known in Lake Baikal plankton (Moore et al., 2009; Zakharova et al., 2013; Bashenkhaeva et al., 2015; Mikhailov et al., 2019b; Roberts et al., 2018), they represented only 6.1% ochrophyte reads distributed in 64 OTUs. Optical microscopy on board showed that diatoms were numerous, but their long frustules prevented most of them from being retained in the analyzed plankton fraction. Members of the Holomycota were very diverse. Classical fungi represented ca. 60% holomycotan sequences, most of them corresponding to chytrids, although the Dicarya (Ascomycota, Basidiomycota) were relatively abundant too (Supplementary Figure 3.B.5). Most Dicarya belonged to typical terrestrial fungi entering the lake waters with river in-flow or from the surrounding land. However, chytrids (flagellated fungi) are more likely to be truly planktonic organisms. Interestingly, members of Rozellida (Cryptomycota) and Aphelida, were also relatively abundant, making up to almost 40% of the holomycotan sequences. Rozellids and aphelids, together with their microsporidian relatives are parasites (Karpov et al., 2014; Bass et al., 2018). Although rozellids (cryptomycotes) are often included within fungi, they are phagotrophic organ-

isms, unlike fungi (which are osmotrophs), and they branch more deeply than apheids in the Holomycota tree (Torruella et al., 2018). Our data suggest that the majority of actual fungal-like planktoners in Lake Baikal are parasites. Overall, despite methodological differences, our identified plankton protist communities were consistent with previous studies in surface waters or in a water column previously sampled in the Southern basin, with ciliates, dinoflagellates and ochrophytes being highly represented (Yi et al., 2017; Annenkova et al., 2020a).

### 3.2.4.3 Marine signature taxa

Although marine-freshwater transitions are thought to be rare (Mukherjee et al., 2019) and salinity, a major driver of microbial community composition (Lozupone and Knight, 2007), high-throughput environmental studies are revealing an increasing number of typically marine eukaryotic lineages in freshwater systems. Among those are members of the parasitic perkinsids (Bråte et al., 2010), haptophytes (Simon et al., 2013), Bolidophyceae (Richards and Bass, 2005; Annenkova et al., 2020a) and several Marine Stramenopiles (MAST) clades (Massana et al., 2004, 2006), such as MAST-2, MAST-12, MAST-3 and possibly MAST-6 (Simon et al., 2014). Recently, diplomonads, a cosmopolitan group of oceanic excavates particularly abundant and diverse in the deep ocean (Lara et al., 2009; De Vargas et al., 2015) were identified in deep freshwater lakes (Yi et al., 2017; Mukherjee et al., 2019). Likewise, Syndiniales, a clade of parasitic alveolates (often parasitizing their dinoflagellate relatives) widely distributed in oceans (López-García et al., 2001; Guillou et al., 2008), were recently identified in Baikal surface plankton (Annenkova et al., 2020a). We identified members of all these lineages in our large Lake Baikal plankton dataset, albeit mostly in low proportions (Figure 3.2A; Supplementary Table 3.A.3). Bolidophytes and, collectively, MAST clades were nonetheless relatively abundant in the lake. However, MAST clades are not monophyletic and they exhibited different abundance patterns. Clades previously detected in freshwater systems, MAST-2, MAST-6, MAST-12 and to a lesser extent MAST-3, were relatively abundant. But MAST clades not previously observed in other freshwater systems, including MAST-1, MAST-4, MAST-8 and MAST-20 occurred in very low proportions in a few samples. In addition to the rare diplomonads, which were widely but sporadically present across Lake Baikal samples (Figure 3.2A-B), we identified OTUs belonging to the emblematic Radiolaria, to our knowledge never before identified in freshwater plankton. These OTUs were members of the Polycystinea (Figure 3.2C) and exhibited extremely low frequencies. The low abundance of some of these typically marine lineages partly explains the fact that they failed to be detected in previous studies of freshwater systems, suggesting that these ecological transitions have been so far underestimated (Paver et al., 2018). However, an additional explanation might be found in the particular features of the Lake Baikal, includ-





**Figure 3.2** Marine signature taxa detected in Lake Baikal plankton. **A**, heat map showing the relative abundance of different typically marine taxa across Baikal plankton samples. The frequency of the different phylogenetic groups is indicated by different shades of blue. **B**, Maximum Likelihood (ML) phylogenetic tree of OTUs belonging to diplomonemids and a related group of euglenozoan excavates (594 unambiguously aligned positions). **C**, ML tree of radiolarian OTUs (534 unambiguously aligned positions).

ing its considerable depth, marked oligotrophy and even the presence of deep-venting (Müller et al., 2001; Sherstyankin et al., 2006; UNDP-GEF, 2015), which make it qualify in all points but salinity as a freshwater sea.

**Table 3.1** Permutational multivariate analyse of variance (PERMANOVA) of Lake Baikal plankton samples across basin, sampling site and depth. PERMANOVA was calculated using Wisconsin standardization on rarefied OTUs belonging to the 65 studied plankton samples. Df, degrees of freedom

Effect	Df	F.Model	R2	P value
Basin	2	1.8369	5.3%	***
Sampling site	15	1.1857	23.7%	***
Depth Sampling site	9	1.2861	16.3%	***

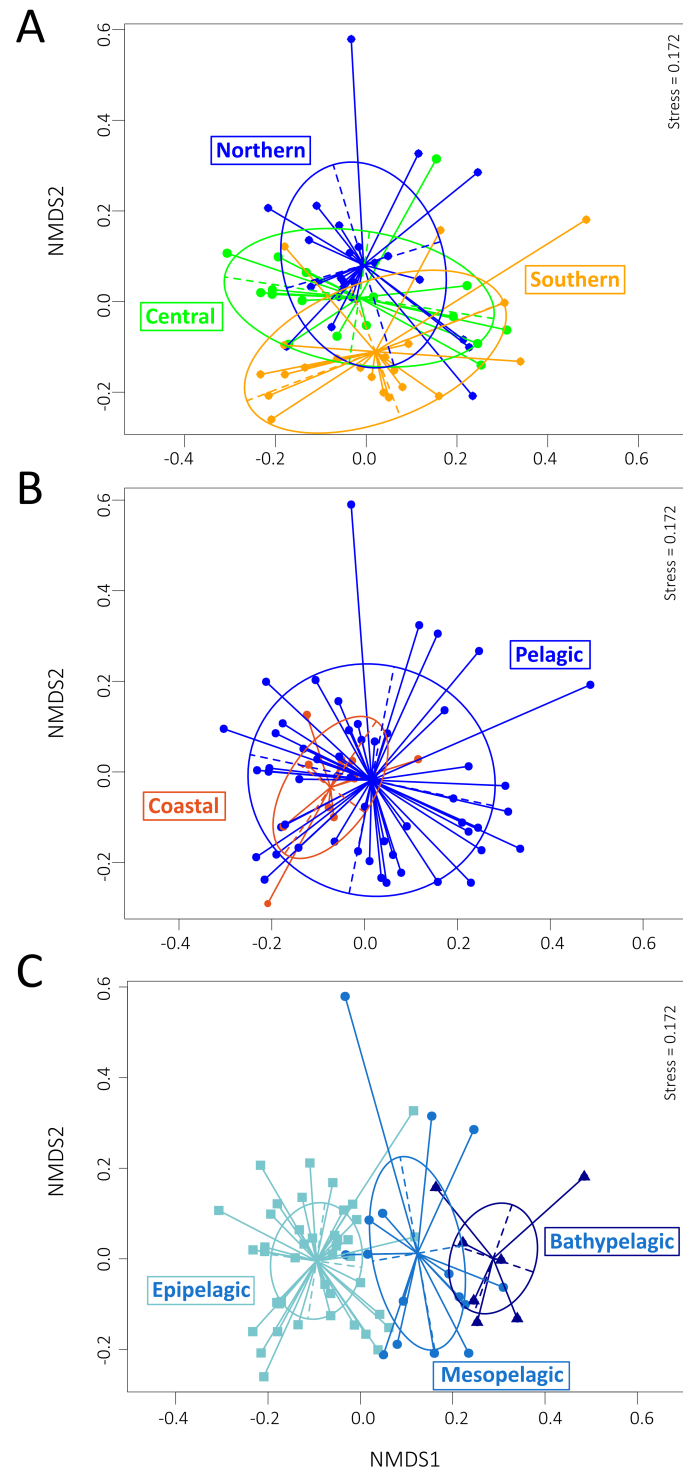
#### 3.2.4.4 Environmental drivers of protist community structure

To test whether planktonic protist communities were influenced by abiotic factors (clearly correlated to sample spatial origin; Figure 3.1), we carried out permutational multivariate analysis of variance (PERMANOVA) of Wisconsin-standardized Bray-Curtis distances between communities as a function of sample spatial origin. PERMANOVA tests revealed significant differences in microbial eukaryotic communities as a function of basin (latitudinal region), sampling site (coordinates) and depth within sampling sites (Table 3.1). However, the most influential effects were those of the water column location, 23.7%, which combine latitudinal and vertical determinants, and depth within each single water column, i.e. vertical variation alone (16.3%). The effect of the sampling basin was significant but small (5.3%). To better visualize differences between communities, we carried out an NMDS analysis on the global Bray-Curtis distance matrix. Points from most water columns did not show a marked differentiation, as most water columns overlapped to some extent (Supplementary Figure 3.B.6). Likewise, samples from different basins did not show a clear differentiation, although samples from the Southern basin tended to segregate from the two other basins (Figure 3.3A). Samples from coastal versus open waters did not segregate at all (Figure 3.3B). However, planktonic communities clearly segregated as a function of the water column zonation, with epipelagic, mesopelagic and bathypelagic communities well separated in the NMDS plot (Figure 3.3C). NMDS analyses based on SWARM-defined OTUs yielded very similar results (Supplementary Figure 3.B.7). These observations were statistically supported by ANOSIM tests, which showed significant and marked differences among communities according to depth, significant but weak differences according to basin origin, and no correlation at all between coastal and pelagic samples (Supplementary Table 3.A.3). These results suggest that depth is the major environmental factor structuring Lake Baikal protist communities. Depth is in turn a proxy for a variety of abiotic parameters,

notably light, but also, despite their limited variation, temperature, dissolved oxygen, conductivity and pH (Figure 3.1). These environmental variables and others, such as the nature of dissolved organic matter (TDS amount does not vary significantly; Figure 3.1B), are likely to influence prokaryotic communities as well (Kurilkina et al., 2016). Consequently, the nature of prey available for bacterivorous protists is possibly different. This may, in turn, select for protists with particular preying affinities, such that biotic interactions with other planktonic members may be also important determinant factors of community structure and function.

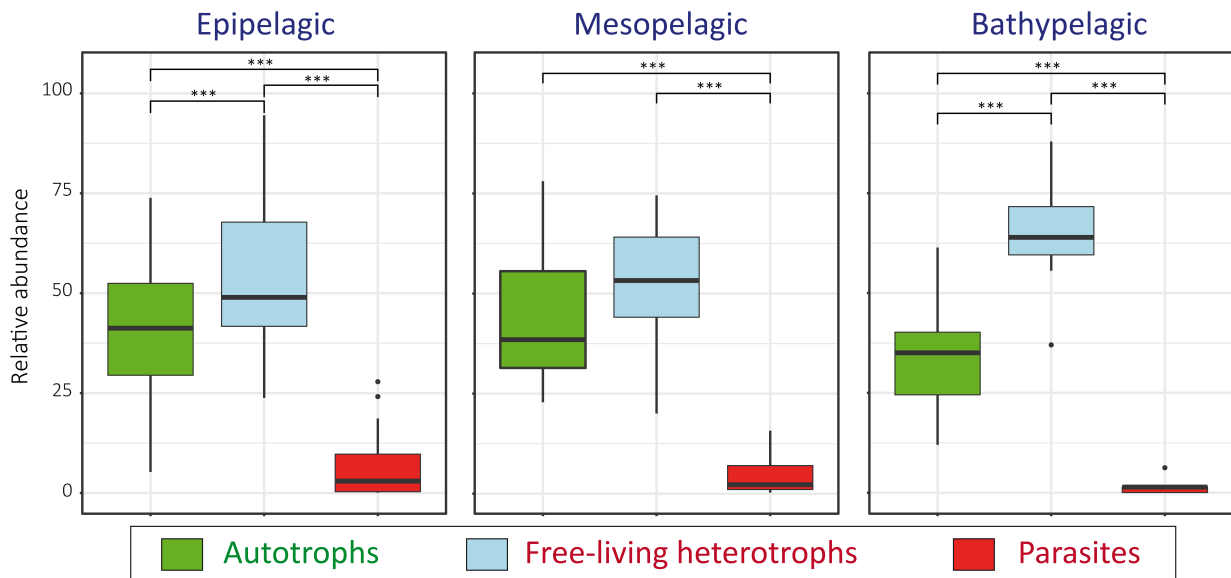
#### 3.2.4.5 Functional groups and biotic interactions

To look for potential ecological interactions between members of protist communities, we first explored the distribution of major functional classes with depth. We attributed protists to three major categories based on knowledge about the lifestyle and ecological function of the corresponding phylogenetic lineages: autotrophs, free-living heterotrophs and parasites (Supplementary Table 3.A.5). We acknowledge that these are very broad categories and that many photosynthetic organisms can be mixotrophs (Massana, 2011; Mitra et al., 2016). However, information about mixotrophy is still scarce and it is difficult to predict this ability from sequence data only. Therefore, our category “autotrophs” included also photosynthetic organisms that can additionally use heterotrophic feeding modes. Free-living heterotrophs include predatory protists but also osmotrophic organisms feeding on organic matter, such as fungi or some Stramenopiles. The relative abundance of the three functional categories in Lake Baikal significantly followed the same trend in the three water column zones, with autotrophs being less abundant than heterotrophs and parasites being in much lower proportion (Figure 3.4). Low proportions of parasitic protists are consistent with affordable parasite loads for an ecosystem, as was previously observed (Simon et al., 2015). Nonetheless, the relative amount of parasitic lineages diminished with depth, potentially suggesting that a relatively important proportion of protists identified in deep waters might be inactive. This is indeed likely the case for most photosynthetic organisms that were identified below the epipelagic region. Although the proportion of autotrophs diminished with depth, they still made up to 30% of the total in bathypelagic waters. As mentioned, some of these protists may be mixotrophic and prey on bacteria or other protists in the dark water column. However, the majority of photosynthetic lineages may simply be inactive, dormant or on their way to decay, serving as food for the heterotrophic component of microbial communities. The presence of relatively abundant photosynthetic protists in the Baikal dark water column and sediments is well documented (Zakharova et al., 2013; Yi et al., 2017), low temperatures possibly helping their preservation during sedimentation. Finally, free-living heterotrophs were the most abundant functional category throughout



**Figure 3.3** Non-metric multidimensional scaling (NMDS) analysis of Lake Baikal plankton samples as a function of protist community similarities. The NMDS plot was constructed with Wisconsin standardized Bray-Curtis dissimilarities between all samples. **A**, plankton samples highlighted by basin origin. **B**, plankton samples from coastal, shallow sites versus open water sites. **C**, samples grouped according to their depth origin in the water column; epipelagic (<200 m), mesopelagic (200-500 m), bathypelagic (>500 m).

the water column. This might seem at odds with a pyramidal food-web structure whereby primary producers should be more abundant than consumers. However, several factors might explain this. First, the presence of ciliates likely introduces a positive bias in this functional category. Second, many autotrophs might be, on average, larger than heterotrophic protists and their biomass exceed that of consumers. Finally, many heterotrophic protists might depend on bacteria or on larger organisms (e.g. fungi degrading decaying plant material).



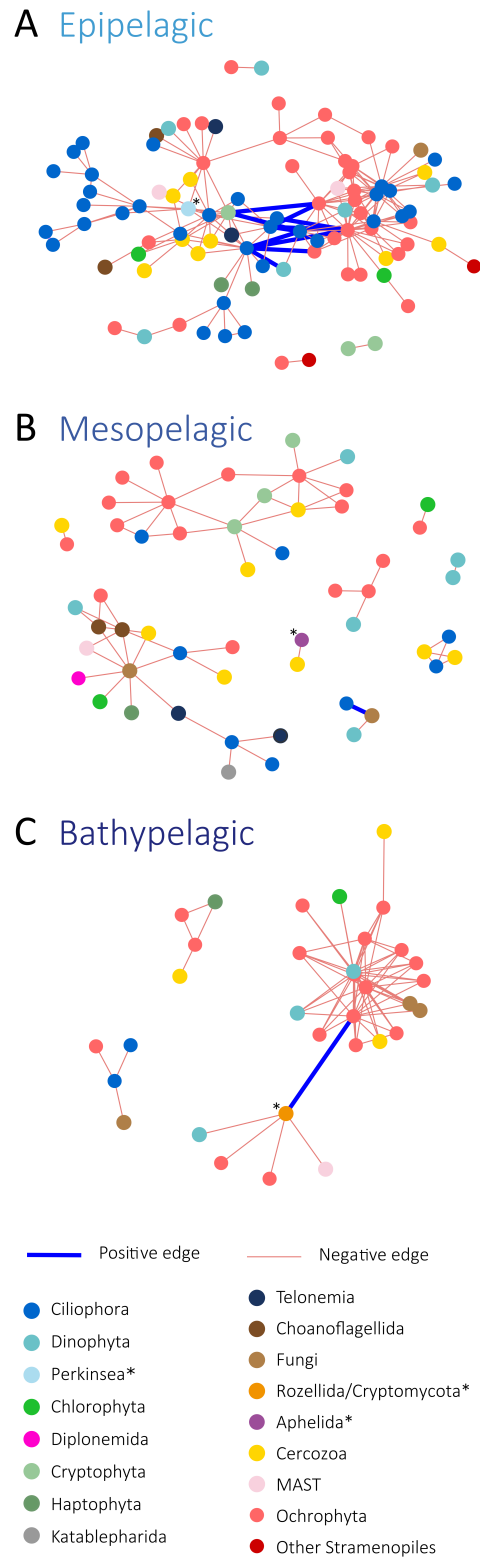
**Figure 3.4** Box plots showing the distribution of relative abundances of major functional categories of protists in Lake Baikal plankton. The three plots show the relative abundance of sequences affiliated to autotrophic, heterotrophic and parasitic protists for each sampling depth class. The thickest line inside each box represents the median on the distribution, bottom and top borders of boxes correspond to the first and the third quartiles and whiskers extend to minimal and maximal distances. Significant differences between distributions are indicated with stars ( $p$ -values  $<0.05$ ,  $<0.005$  and  $<0.0005$  are respectively indicated by one, two and three stars). For the assignment of taxa to functional categories, see Supplementary Table 3.A.5.

To further explore biotic interactions, and given that protist community differences were essentially seen throughout the water column, we reconstructed co-occurrence networks of OTUs found in epipelagic, mesopelagic and bathypelagic zones. To build the networks, we retained OTUs present in more than 20% of samples at relative abundances higher than 0.01% (Supplementary Table 3.A.6). The structure of the three networks was markedly different (Figure 3.5). The epipelagic network was denser, having more interconnected OTUs, more positive interactions and several hub-type OTUs that interact with many OTUs. Mean node degrees were also higher in the epipelagic network (Supplementary Table 3.A.7). Meso- and bathypelagic networks had less connected nodes and most correlations were negative. Only one positive interaction was observed in mesopelagic waters (ciliate-fungus) and in bathypelagic

waters (rozellid-ochrophyte). The latter might suggest a specific parasitic interaction. Although bathypelagic waters exhibited the least connected nodes, both the connectance and the clustering coefficient of the network were the highest. A block-model representation of the three networks indicated the occurrence of pairs of OTU sets sharing similar properties that were highly interconnected with each other and only loosely to other sets (Supplementary Figure 3.B.8). Collectively, our network data suggest that epipelagic communities are more active and have more positive and negative interactions, whereas in very deep waters, communities are more stable with a restricted but strongly connected core.

### 3.2.5 Concluding remarks

Lake Baikal in Southern Siberia is a unique freshwater system by its volume, maximum depth (1 642 m) and topographical features that include rifting associated with hydrothermalism. With its highly oligotrophic waters, it amounts to an inner freshwater sea in all points but an extremely low salinity. Freshwater ecosystems are particularly threatened by climate change and, being located in Southern Siberia, one of the most rapidly changing zones, Lake Baikal is being severely impacted (Mackay et al., 2017). Yet, despite the importance of the lake and its uniqueness, its microbial planktonic communities have been only partially studied and we lack reference comprehensive comparative community data to assess ongoing and future change and infer how it may affect microbial functions and the ecology of the lake. In this study, we have analyzed the composition of microbial eukaryotic communities in plankton collected from different water columns along a transect of  $\sim 600$  km across the three lake basins, from surface (5 m) to high depth (1 400 m) and from littoral to open waters. Our study shows widely diverse communities covering all eukaryotic supergroups, with ciliates, dinoflagellates, chrysophytes and flagellated fungi plus related lineages (rozellids, aphelids) being the most relatively abundant, together with cryptophytes, haptophytes, katablepharids, telonemids and several MAST lineages. Interestingly, confirming previous observations in Lake Baikal, we observed members of typically marine lineages, including bolidophytes, syndineans, diplomemids and, for the first time, radiolarians. These observations suggest that the salinity barrier is relatively easy to cross and that the “marine” determinants might be more related with the oligotrophic nature of the system and the occurrence of a deep water column than with salinity itself. Despite the relatively homogeneous values of several physicochemical parameters, planktonic protist communities were highly and significantly stratified in Lake Baikal, suggesting that depth, as a proxy for light but also temperature, pH, oxygen and conductivity, is a major determinant of community structure. By contrast, the effect of latitude (basins) was minor, if not negligible. Consistent with vertical stratification, the relative proportion of



**Figure 3.5** Co-occurrence networks of planktonic protists 700 in the Lake Baikal water column. **A**, network obtained from epipelagic (<200 m) samples across the lake. **B**, network obtained from mesopelagic (200-500 m) samples. **C**, network obtained from bathypelagic (>500 m samples). Networks were built on OTUs present in more than 20% samples and having a relative abundance higher than 0.01%. OTUs are represented by nodes and direct covariations between them, by edges. Nodes and taxa labeled with an asterisk correspond to putative parasites.

---

autotrophs, free-living heterotrophs and parasites is altered with depth, where photosynthetic lineages are still present but, like parasites, in lower proportions. Biotic factors are also important in structuring Lake Baikal communities. Co-occurrence network analyses showed highly interconnected communities in surface waters, with positive and negative interactions. By contrast, deep, bathypelagic communities exhibit much less connected OTUs, although they are strongly, and mostly negatively, correlated. This might be suggestive of much more diluted and potentially inactive populations, but with a conserved core of highly interconnected OTUs. Our results pave the way for future comparative analyses of protist communities through time, notably in the context of rapid climate change that is affecting Siberia and Lake Baikal.



# Supplementary Material

## 3.A Supplementary Tables

**Supplementary Table 3.A.1** Sampling sites in Lake Baikal water column and associated physico-chemical, sequence and diversity data. Operational taxonomic units (OTUs) were defined at 95% sequence identity, except when indicated (SWARM). Singletons are excluded from the OTU counts.

Basin	Sampling site	Sample name	Sampl. depth (m)	Sampling Date	Latitude	Longitude	Bottom depth (m)	Temperature (°C)	pH	Cond. (µS/cm)	TDS (mg/L)	ORP (mV)	DO (%)	Raw reads	Clean merged reads	# OTUs	# SWARM OTUs	# OTUs after raref.	Simpson	Biveness				
North	BK16	BK16.5m	5	02.07.2017	55°06'259" N	109°16'104" E	846	4.7	7.88	75.7	73.5	106.7	76	176014	132046	954	1216	248	0.87	0.51				
		BK16.25m	25					4.6	7.67	75.7	69.6	79.5	69.5	129314	103211	1156	1128	357	0.94	0.6				
		BK16.50m	50					4.4	7.82	74.3	73	106.1	80.5	109882	91554	854	1052	276	0.94	0.6				
		BK16.100m	100					4.3	8.16	74.7	70.9	106.1	71.8	70538	68416	197	167	104	0.85	0.55				
		BK16.200m	200					4.2	7.81	74.4	74.1	105.5	73	150492	137629	980	1309	286	0.91	0.56				
		BK16.500m	500					4.2	7.68	74.2	77.4	101.8	68	10085	9771	41	26	41	0.89	0.67				
		BK16.800m,CT	800					4.2	7.68	73.8	76.1	67.6	71.8	56767	51846	261	54	135	0.82	0.52				
		BK18	BK18.5m					5	03.07.2017	54°51'372" N	108°54'14" E	36	5.6	8.07	77.5	76.1	100.3	94	136920	96373	1178	1039	322	0.91
	BK20	BK20.5m	5	03.07.2017	54°08'361" N	108°30'613" E	896	5.6	7.92	73.9	74.8	136.1	136.1	75.8	130943	102135	518	712	147	0.7	0.4			
		BK20.25m	25					4.9	7.9	73.6	76.1	134.8	78	149603	102765	698	927	203	0.73	0.42				
		BK20.50m	50					4.8	7.87	73.8	76.1	132.6	78.6	161996	105383	806	1168	229	0.91	0.53				
		BK20.100m	100					4.5	7.86	73.3	77.4	130.4	80.6	102618	89102	752	868	237	0.9	0.55				
		BK20.200m	200					4	7.74	72.9	74.8	131.8	77	101431	77741	727	727	229	0.83	0.48				
		BK20.500m	500					4.8	7.62	73.3	76.7	117.6	80.3	340083	275886	1261	1515	357	0.87	0.5				
		BK20.850m	850					6.5	7.71	75.9	74.1	88.5	68	92074	86915	621	674	366	0.87	0.56				
		BK14	BK14.5m					5	01.07.2017	54.09.926 N	109.31.465 E	129	5.5	7.63	78.1	68.3	118.9	84	174174	171428	663	274	268	0.93
	BK13	BK13.5m	5	01.07.2017	53°53'53.10"N	109°1'25.98"E	13	6.2	7.9	84	66	80.9	65	94831	85391	851	1237	307	0.95	0.62				
	BK09	BK09.5m	5	30.06.2017	53°51'346" N	108°42'911" E	13	5.9	7.92	80.4	74.8	124.1	77	146124	126201	1079	1264	323	0.92	0.58				
BK06	BK06.5m	5	30.06.2017	53°50'658" N	108°40'195" E	156	4.8	7.92	76.1	78	113.9	80	88294	78071	1217	991	378	0.9	0.53					
	BK06.25m	25					4.4	7.91	76.3	75.4	118	80	49540	46572	664	659	282	0.92	0.57					
	BK06.50m	50					5.1	7.82	78.9	77.8	115	82	233612	207530	1919	1477	414	0.94	0.6					
	BK06.100m	100					4.6	7.85	77.6	73.1	109.1	75	146124	102945	1000	1276	277	0.92	0.57					
	BK11	BK11.5m					5	30.06.2017	53°46'12.03"N	109°5'9.56"E	10	15.3	8.33	82.2	76.7	136.3	83.1	50396	35832	567	585	332	0.93	0.58
	BK05	BK05.5m					5	29.06.2017	53°31'096" N	108°224'583" E	1512	5.5	7.9	77.1	78.7	134.3	80	115229	107228	811	1033	236	0.85	0.47
BK05.25m	25	4.1	7.81	74.1	77.4	134.8	77					48005	44526	543	478	228	0.73	0.41						
BK05.50m	50	4.2	7.81	73.8	78	134.6	76.6					103557	88299	1067	1001	309	0.84	0.46						
BK05.100m	100	4.2	7.81	74.3	78	134.7	77					96709	80346	1222	1240	389	0.89	0.5						
BK05.200m	200	4	7.84	74.2	76.7	133.3	70.5					129517	74958	236	103	114	0.94	0.7						
BK05.500m	500	3.7	7.7	74.4	75.4	127.3	71.5					160563	156621	456	305	151	0.82	0.51						
Central	BK05.1000m	1000	1400	3.7	7.55	73.6	73.5	122.7	63.5	39227	36443	362	306	229	0.81	0.46								
	BK05.1400m	1400															3.6	7.7	74.5	80	114	77	66984	62868

Continued on next page

Supplementary Table 3.A.1 – Continued from previous page

Basin	Sampling site	Sample name	Sampl. depth (m)	Sampling Date	Latitude	Longitude	Bottom depth (m)	Temperature (°C)	pH	Cond. ( $\mu\text{S}/\text{cm}$ )	TDS (mg/L)	ORP (mV)	DO (%)	Raw reads	Clean merged reads	# OTUs	# SWARM OTUs	# OTUs after raref.	Simpson	Evenness
Central	BK22	BK22.5m	5	04.07.2017	53°23'52.4" N	107°53'09.4" E	592	4.9	7.86	74.7	76.5	86	83.8	170009	135771	1050	1159	281	0.82	0.49
		BK22.50m	50					5.2	7.85	75.8	74.5	80.6	70.8	87940	85032	879	813	337	0.94	0.63
		BK22.200m	200					7.8	7.77	75.2	72.2	72.8	68.3	143354	133777	825	701	264	0.93	0.62
		BK22.550m	550					4.6	7.76	77	76.7	29.6	114	102066	97459	649	634	271	0.85	0.53
	BK04	BK04.5m	5	29.06.2017	53°14'59.6" N	108°24'58.3" E	NA	9.5	8.47	87.3	72.8	44.4	80.6	62438	55549	726	838	311	0.89	0.53
	BK24	BK24.5m	5	05.07.2017	53°00'52.3" N	106°53'65.9" E	35	7.8	8.11	81.3	76.7	31.8	98	131337	114477	743	685	248	0.85	0.48
		BK03.5m	5					8.6	7.74	84	78.7	176.4	83	120296	107489	544	511	149	0.57	0.31
		BK03.25m	25					4.8	7.67	75.5	77.4	179.9	82.6	132170	109849	915	879	279	0.87	0.5
		BK03.50m	50					4.7	7.63	75.2	78.7	181.5	80.3	106227	97251	798	791	250	0.9	0.52
	BK03	BK03.100m	100	28.06.2017	52°41'40.1" N	106°44'20.8" E	1083	4.3	7.6	74.7	76.1	183.2	77.6	105172	96013	1065	831	320	0.89	0.55
		BK03.200m	200					4.3	7.6	74.6	74.8	184	73.3	59072	57656	344	107	172	0.91	0.63
		BK03.500m	500					4	7.45	74.5	76.7	189.6	71.3	72297	69907	363	383	188	0.73	0.39
	BK03.1000m	1000					4.8	7.48	79.1	79.3	195	68	144322	140470	436	301	156	0.7	0.38	
	BK01.5m	5					12.2	7.98	95.2	80.6	204.4	94.5	68506	59443	812	655	336	0.8	0.51	
BK01	BK01.25m	25					8.1	7.56	86.8	83.2	224	83.5	124684	114488	1198	565	347	0.8	0.43	
	BK01.50m	50	28.06.2017	52°15'70" N	106°02'90" E	105	8	7.54	78.9	74.5	234.3	26	89182	76475	1043	743	340	0.87	0.5	
	BK01.85m	85					6.5	7.88	83.5	83.2	205.6	87.7	59454	44623	685	640	306	0.91	0.57	
	BK02.5m	5	28.06.2017	52°15'70" N	106°02'90" E	10	13.7	7.62	97.5	80.6	202	83.3	77234	60742	647	533	296	0.84	0.48	
South	BK26	BK26.5m	5					8.7	8.18	83.6	78.7	134	87.1	129657	103961	727	864	218	0.77	0.39
		BK26.25m	25					6.5	8.06	76.1	76.7	138.2	86.1	142012	128621	567	514	157	0.47	0.28
		BK26.50m	50					4.7	7.87	74.4	78	131	80.2	188707	150315	1204	954	312	0.88	0.53
		BK26.100m	100	05.07.2017	51°52'63.8" N	105°15'29.4" E	1408	4.6	7.84	73.9	78	130.1	81.4	144537	114062	1102	977	343	0.91	0.54
	BK26.200m	200					4.1	7.8	73.3	79.3	124.8	82.6	228022	188147	1301	828	326	0.91	0.56	
	BK26.500m	500					4	7.71	73.2	78.7	119.9	78.5	72429	55468	693	691	306	0.86	0.51	
	BK26.1000m	1000					3.9	7.64	73	78.7	109.5	82.8	36746	31456	324	299	204	0.87	0.5	
		BK26.1350m	1350					3.9	8.13	79	85.2	80.2	97.2	114918	109195	378	459	157	0.82	0.45
		BK28.5m	5					11.7	8.28	89	76.7	90.7	88	197087	166833	830	1102	177	0.79	0.43
		BK28.25m	25					5.6	8.14	77.1	71.5	95.3	104	321169	235066	2973	1993	698	0.94	0.59
BK28	BK28.50m	50					7	7.9	78	70	98.3	70	70786	53923	965	755	361	0.89	0.53	
	BK28.100m	100	06.07.2017	51°47'8.17" N	104°46'44.9" E	1381	4.6	7.87	75.2	77.4	94.2	85.4	111378	84474	1042	866	349	0.84	0.46	
	BK28.200m	200					4.6	7.8	75.4	77.3	107.8	83.7	81308	64784	993	1075	353	0.89	0.51	
	BK28.500m	500					4.9	7.69	76.1	78	96.9	85	133522	118083	943	1257	297	0.92	0.54	
	BK28.1000m.CT	1000					6.4	7.77	78.7	76.7	87.3	77	3234	3022	9	10	9	0.82	0.8	
	BK28.1300m.CT	1300					4.8	8.02	80.7	81.3	45.6	112	111545	107430	220	101	94	0.7	0.38	

**Supplementary Table 3.A.2** Identification, phylogenetic affinity and abundance of eukaryotic OTUs identified in Lake Baikal plankton.

(too large to be displayed in this document, available in excel format)

**Supplementary Table 3.A.3** Diversity, abundance and distribution of eukaryotic lineages previously thought to be exclusively marine identified in Lake Baikal plankton.

Phylogenetic group	Number of OTUs	Number of reads	No. samples where they occur
Bolidophyceae	85	9442	58
Diplonema	4	505	21
Radiolaria	2	22	1
Syndiniales	73	1098	58
All MAST	95	23722	61
Specific MAST clades	MAST1	2	1
	MAST2	8250	54
	MAST3	1296	41
	MAST4	219	10
	MAST6	3453	50
	MAST8	4	3
	MAST12	10487	59
	MAST20	11	1

**Supplementary Table 3.A.4** ANOSIM analyses between pairs of Lake Baikal plankton sample groups defined as a function of depth, basin of origin and coastal vs. pelagic location. ANOSIM were calculated upon 999 permutations between pairs of sample groups.

	R	p-value
Epi-Meso	0.3821	0.0001
Epi-Bathy	0.7262	0.0001
Meso-Bathy	0.7262	0.0001
Southern-Central	0.09237	0.0001
Southern-North	0.2201	0.0001
Central-North	0.1442	0.0001
Coastal-Pelagic	0.05955	0.2339

**Supplementary Table 3.A.5** Tentative classification of phylogenetic lineages within broad functional categories.

<b>Phylum</b>	<b>Super_Group</b>	<b>Lifestyle</b>
Chlorophyta	Archaeplastida	Autotrophs
Cryptophyta	Hacrobia	Autotrophs
Dinophyta	Alveolata	Autotrophs
Haptophyta	Hacrobia	Autotrophs
Ochrophyta	Stramenopiles	Autotrophs
Rhodophyta	Archaeplastida	Autotrophs
Charophyta	Archaeplastida	Autotrophs
Apusozoa	Apusozoa	Free-living heterotrophs
Bicoecea	Stramenopiles	Free-living heterotrophs
Breviatea	Amoebozoa	Free-living heterotrophs
Centroheliozoa	Hacrobia	Free-living heterotrophs
Cercozoa	Rhizaria	Free-living heterotrophs
Choanoflagellida	Opisthokonta	Free-living heterotrophs
Ciliophora	Alveolata	Free-living heterotrophs
Conosa	Amoebozoa	Free-living heterotrophs
Discoba	Excavata	Free-living heterotrophs
Fungi	Opisthokonta	Free-living heterotrophs
Katablepharidophyta	Hacrobia	Free-living heterotrophs
Labyrinthulea	Stramenopiles	Free-living heterotrophs
Lobosa	Amoebozoa	Free-living heterotrophs
MAST	Stramenopiles	Free-living heterotrophs
Metamonada	Excavata	Free-living heterotrophs
Radiolaria	Rhizaria	Free-living heterotrophs
Telonemia	Hacrobia	Free-living heterotrophs
Picozoa	Uncertain	Free-living heterotrophs
Mesomycetozoa	Opisthokonta	Free-living heterotrophs
Pirsonia	Stramenopiles	Free-living heterotrophs
Syndiniales	Alveolata	Putative parasites
Cryptomycota	Opisthokonta	Putative parasites
Hyphochytriomyceta	Stramenopiles	Putative parasites
Oomycota	Stramenopiles	Putative parasites
Apicomplexa	Alveolata	Putative parasites
Opalinata	Stramenopiles	Putative parasites
Perkinsea	Alveolata	Putative parasites
Aphelida	Opisthokonta	Putative parasites
Uncertain_Alveolata	Alveolata	Uncertain
Uncertain_Amoebzoa	Amoebozoa	Uncertain
Uncertain_Opisthokonta	Opisthokonta	Uncertain
Uncertain_Stramenopiles	Stramenopiles	Uncertain

**Supplementary Table 3.A.6** Functional and phylogenetic classification of OTUs used for network construction. Only OTUs that were present in more than 20% of samples with a relative abundance higher than 0.01% were considered.

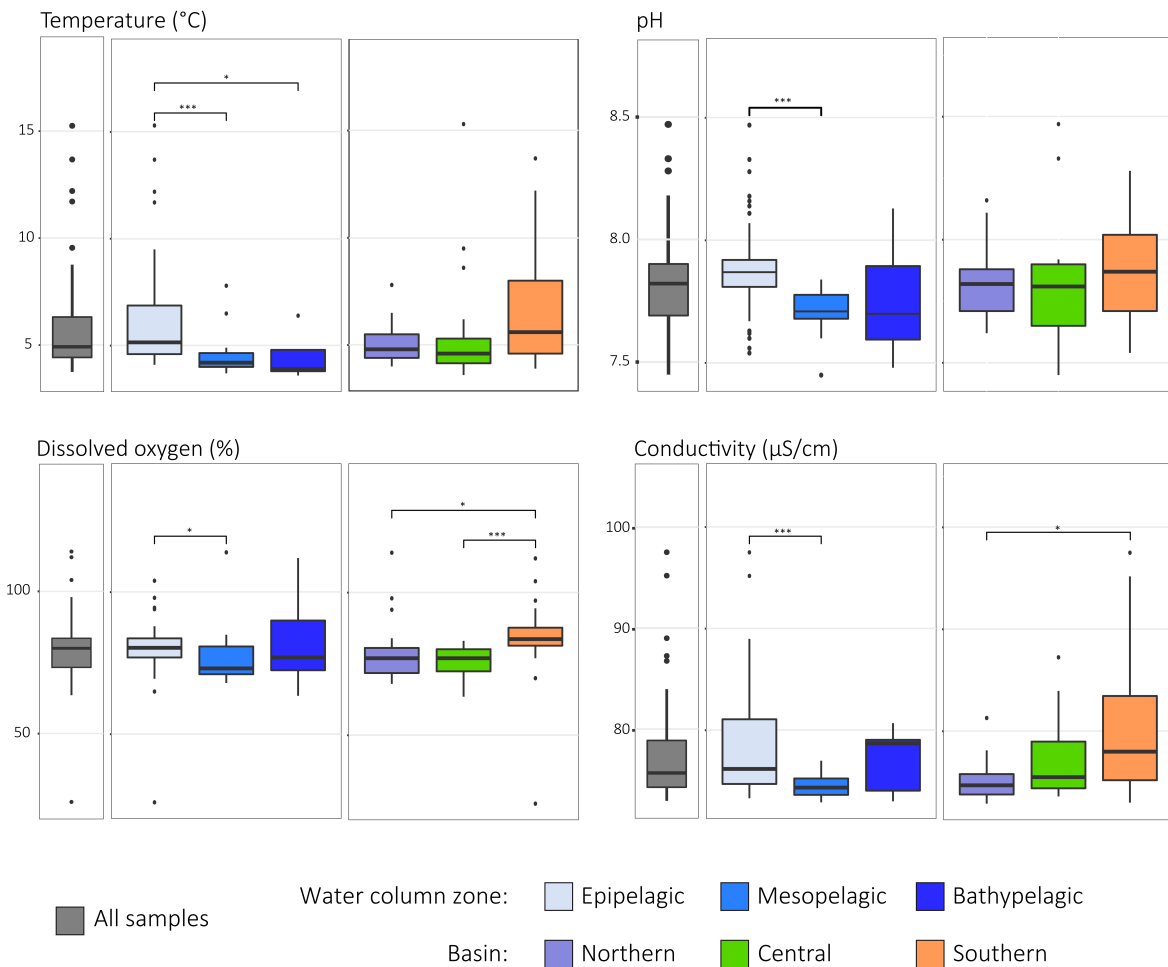
	Epipelagic		Mesopelagic		Bathypelagic	
	Not connected	Connected	Not connected	Connected	Not connected	Connected
<b>Functional classification</b>						
Autotrophs	69	45	104	29	59	23
Heterotrophs	35	51	59	25	74	9
Parasites	5	1	9	1	7	1
<b>Phylogenetic affiliation</b>						
Apicomplexa	1	0	1	0	1	0
Apusozoa	0	0	0	0	0	0
Bicoecea	2	1	1	0	3	0
Centroheliiozoa	0	0	0	0	0	0
Cercozoa	13	9	12	8	13	3
Chlorophyta	9	2	13	2	7	1
Choanoflagellida	0	2	2	2	4	0
Ciliophora	11	33	23	8	33	2
Conosa	0	0	1	0	1	0
Cryptomycota	3	0	2	0	3	1
Cryptophyta	11	3	8	3	13	0
Dinophyta	13	6	15	6	8	3
Discoba	0	0	0	1	1	0
Fungi (+Aphelida)	5	1	10	3	11	3
Haptophyta	5	2	3	1	3	1
Hyphochytriomyceta	0	0	0	0	0	0
Katablepharidophyta	1	0	0	1	1	0
Labyrinthulea	0	0	0	0	0	0
Lobosa	0	0	0	0	0	0
MAST	3	2	8	1	4	1
Mesomycetozoa	0	0	1	0	1	0
Ochrophyta	31	32	65	17	28	18
Oomycota	1	0	3	0	2	0
Opalinata	0	0	0	0	0	0
Perkinsea	0	1	3	0	1	0
Picozoa	0	0	0	0	0	0
Pirsonia	0	1	0	0	0	0
Radiolaria	0	0	0	0	0	0
Syndiniales	0	0	0	0	0	0
Telonemia	0	2	1	2	2	0
Uncertain_Alveolata	0	0	0	0	0	0
Uncertain_Opisthokonta	0	0	0	0	0	0
Uncertain_Stramenopiles	0	0	0	0	0	0



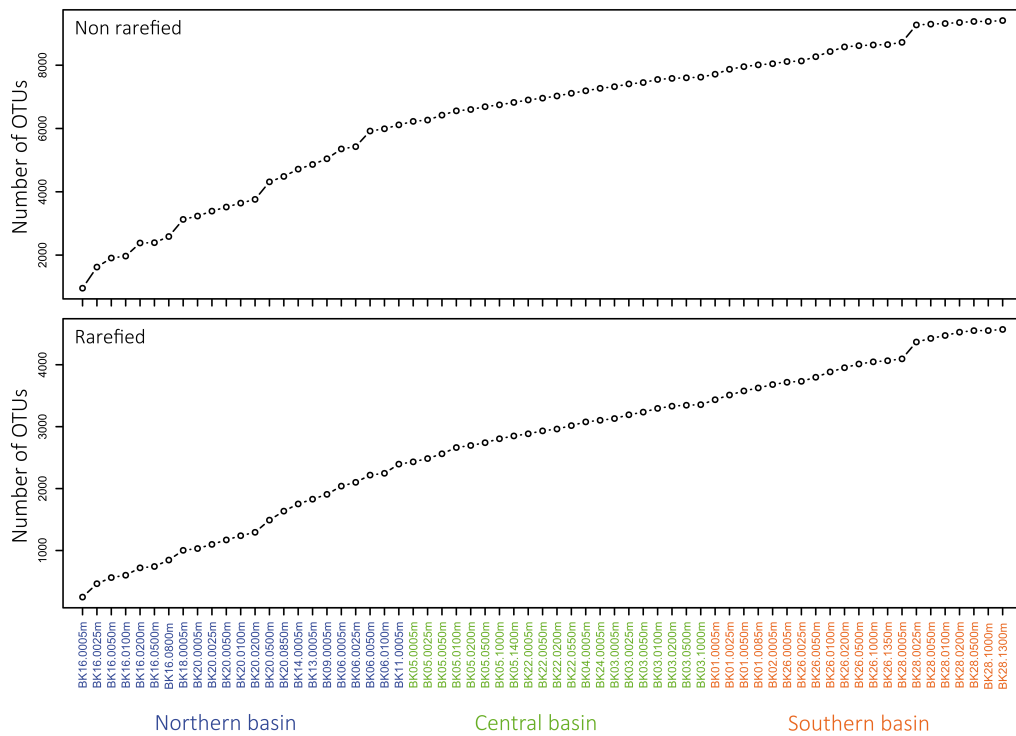
**Supplementary Table 3.A.7** Properties of networks built upon Lake Baikal plankton OTUs for the three depth categories of the water column. Only OTUs that were present in more than 20% of samples with a relative abundance higher than 0.01% were considered.

	Epipelagic		Mesopelagic		Bathypelagic	
	Basin	%	Basin	%	Basin	%
Covariable	Basin		Basin		Basin	
Positive edges	10	4.80	1	1.60	1	1.30
Negative edges	198	95.2	61	98.4	78	98.7
Total number of edges	208		62		79	
Total number of nodes	206		227		173	
Connected nodes	97	47.1	55	24.2	33	19.1
Connectance	0.04		0.041		0.14	
Clustering coefficient	0.31		0.25		0.43	
Most connected node	X406443 (27)		X403371 (9)		X401791, X403173 (18)	
Average path length	3.51		2.5		2.11	
Mean node degree	2.02		0.55		0.91	

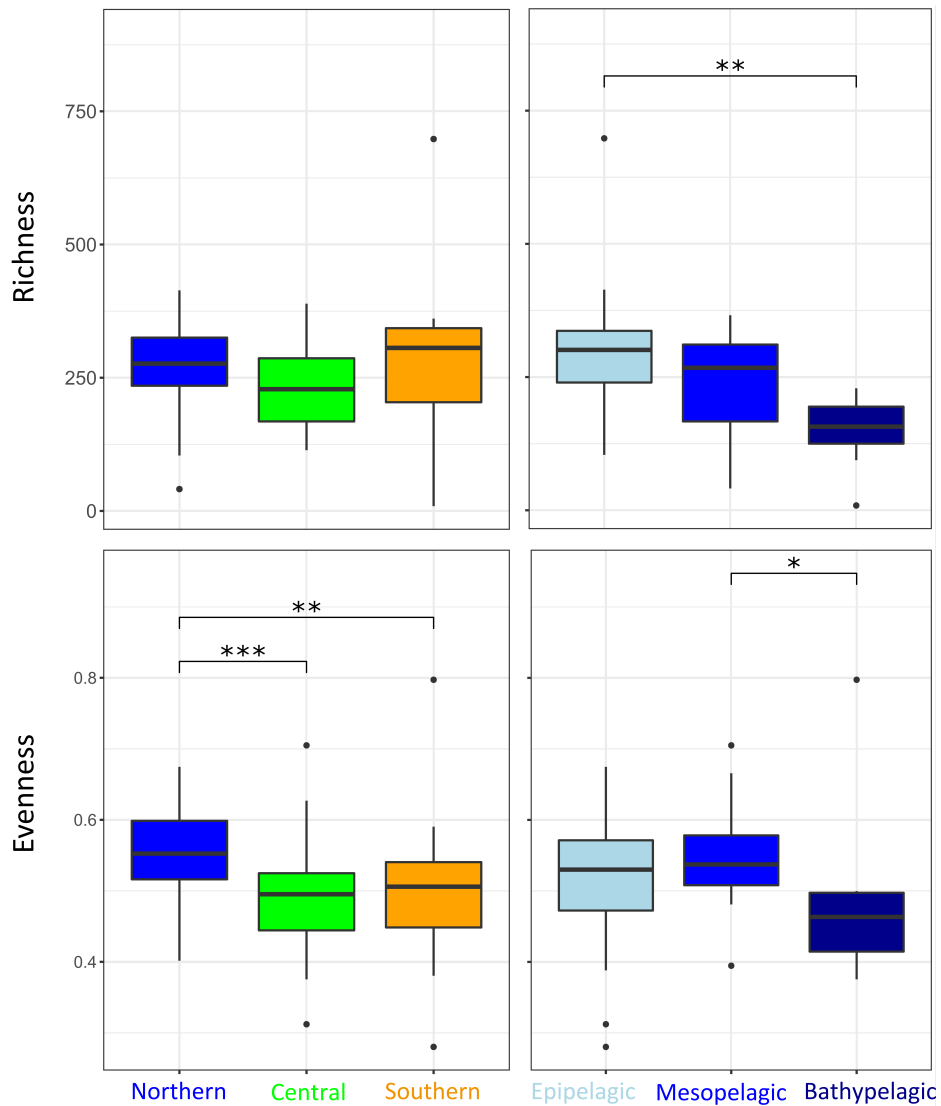
## 3.B Supplementary Figures



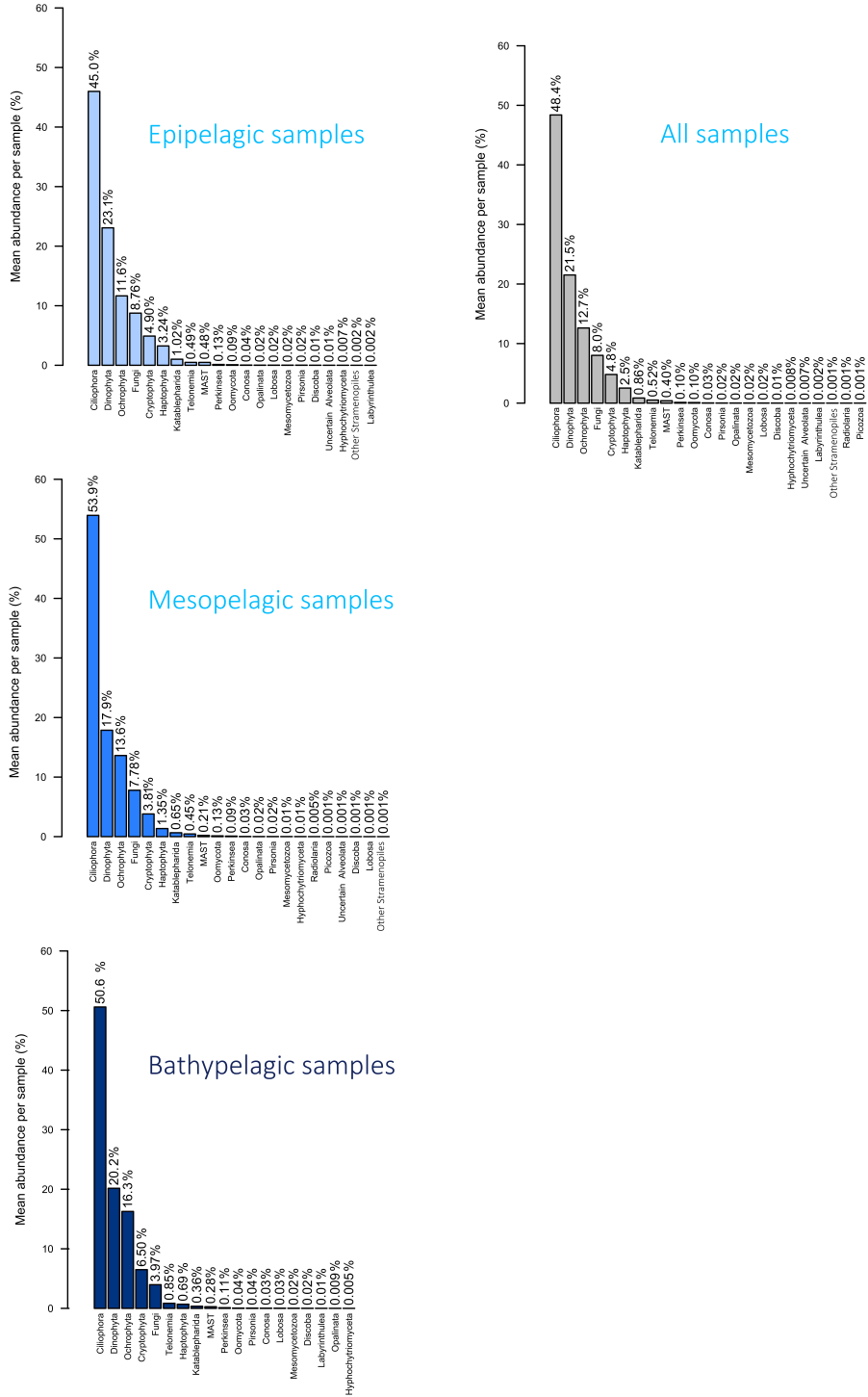
**Supplementary Figure 3.B.1** Box plots of physico-chemical parameters measured in Lake Baikal water samples. Average values and variation are shown for all the Baikal samples and as a function of basin and depth. Significant differences between pairs of samples are indicated (p-value  $\leq 0.05^*$ ; p-value  $< 0.001^{***}$ ).



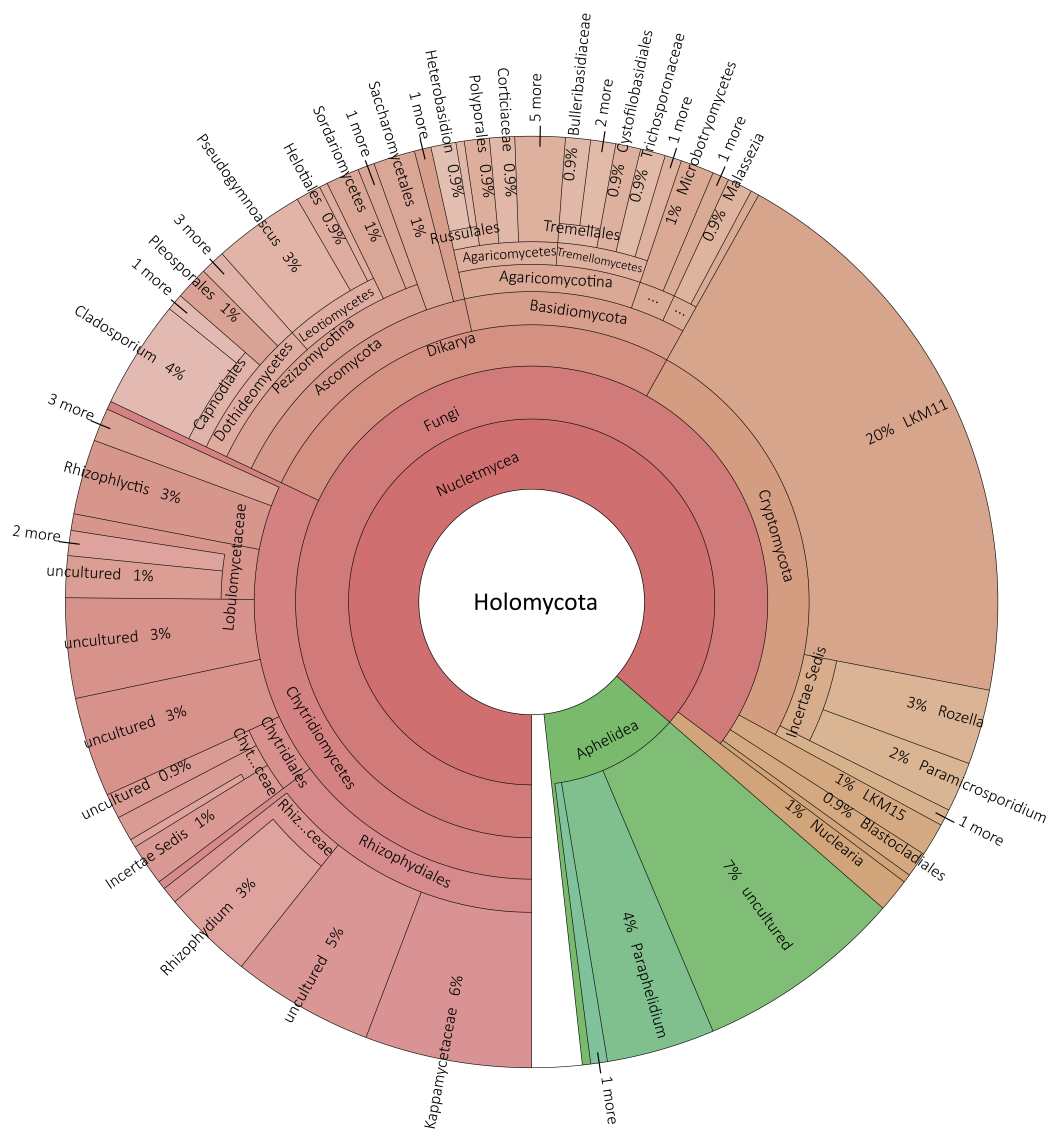
**Supplementary Figure 3.B.2** Accumulation curves for OTUs identified in Lake Baikal plankton samples before and after rarefaction.



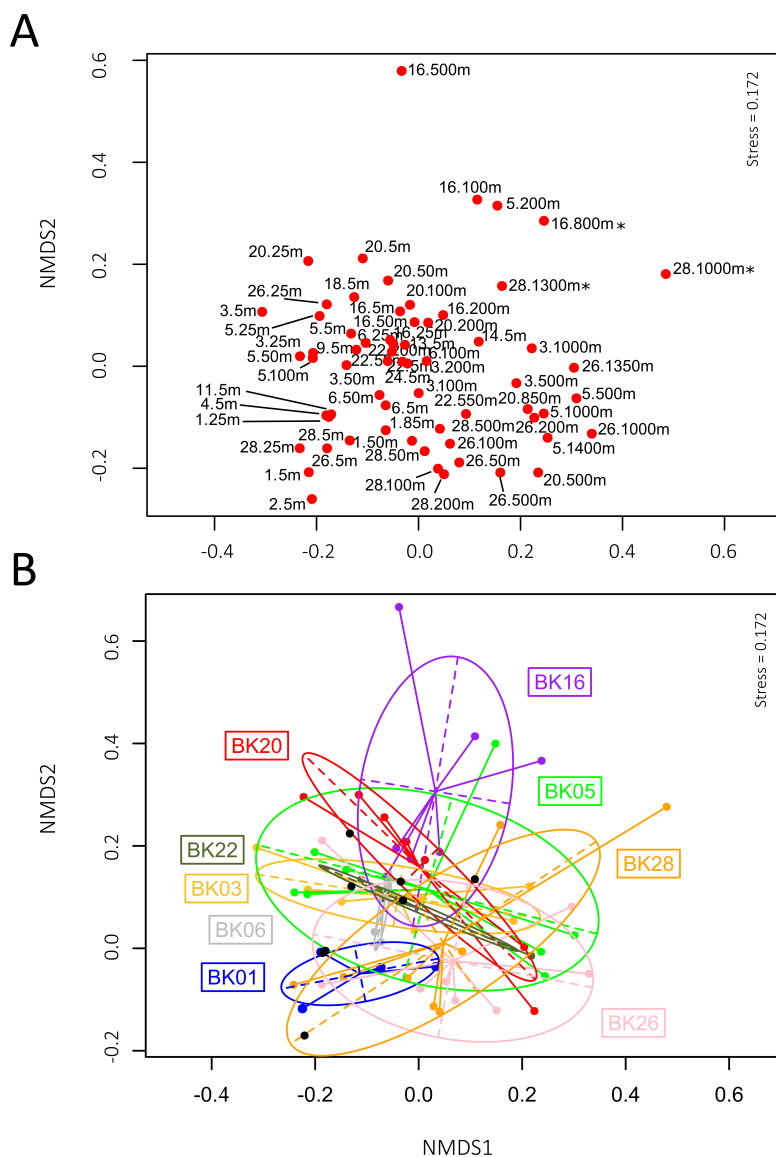
**Supplementary Figure 3.B.3** Box plots showing diversity (richness) and evenness values in Lake Baikal water samples as a function of basin and depth. Richness and evenness were calculated on OTUs defined at 95% 18S rRNA gene sequence identity (~genus level). Significant differences between pairs of samples are indicated ( $p$ -value  $\leq 0.05^*$ ;  $p$ -value  $< 0.001^{***}$ ).



**Supplementary Figure 3.B.4** Rank:abundance curves of protist OTUs in Lake Baikal plankton samples. Rank:abundance curves are presented globally for the lake and by lake depth category.

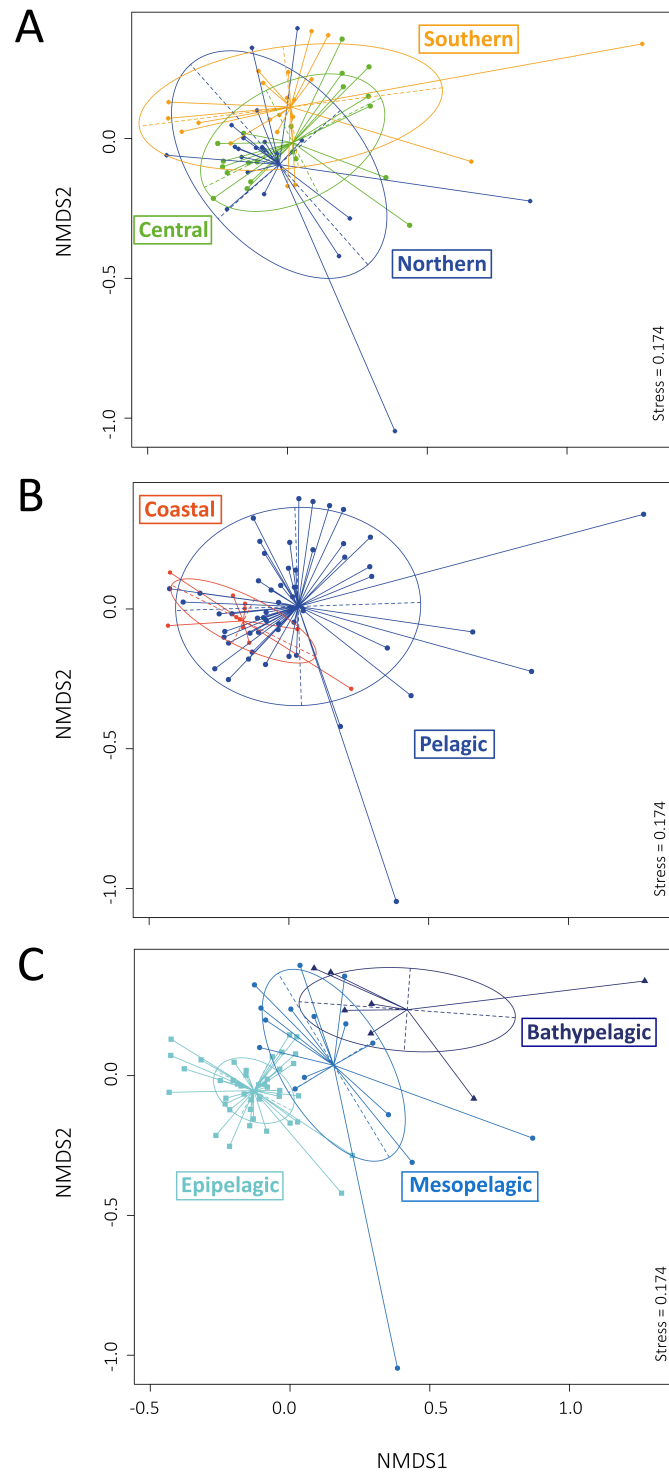


**Supplementary Figure 3.B.5** KRONA representation of the global diversity of Holomycota in Lake Baikal plankton. Holomycota is one of the two branches of Opisthokonta, including Fungi and related lineages (Cryptomycota, Aphelida, Nuclearia). The graph was produced using the SILVA classifier (<https://www.arb-silva.de>).



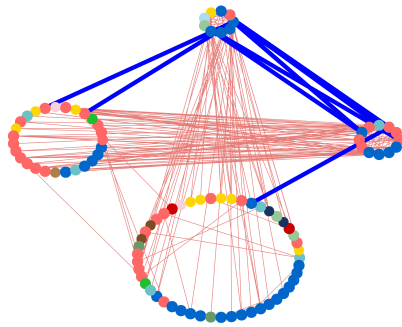
**Supplementary Figure 3.B.6** Non-metric multidimensional scaling (NMDS) analysis of Lake Baikal plankton samples based on differences in protist community composition. Community structure differences are based on Bray-Curtis distances established upon the distribution and relative frequency of OTUs. **A**, NMDS showing all sampling points. **B**, NMDS on the same points taking into account different water column samples. Black dots correspond to surface (5 m deep), coastal samples.



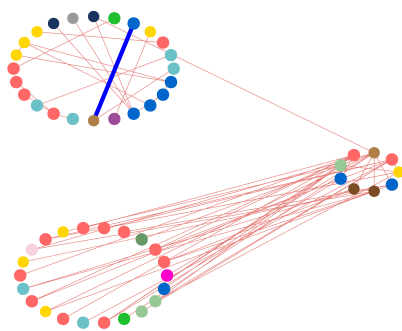


**Supplementary Figure 3.B.7** Non-metric multidimensional scaling (NMDS) analysis of Lake Baikal plankton samples as a function of protist community similarities based on SWARM-determined OTUs. The NMDS plot was constructed with Wisconsin-standardized Bray-Curtis dissimilarities between all samples. **A**, plankton samples highlighted by basin origin. **B**, plankton samples from coastal, shallow sites versus open water sites. **C**, samples grouped according to their depth origin in the water column; epipelagic (<200 m), mesopelagic (200-500 m), bathypelagic (>500 m).

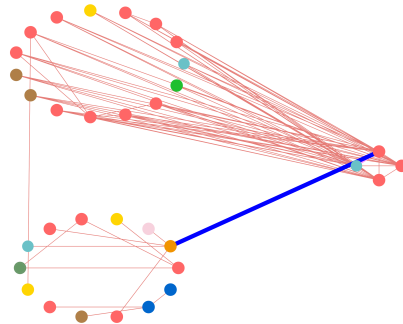
## A Epipelagic



## B Mesopelagic



## C Bathypelagic



— Positive edge  
— Negative edge

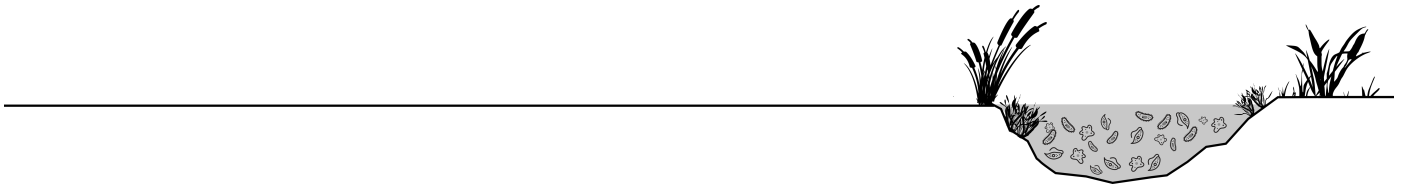
- Ciliophora
- Dinophyta
- Perkinsea
- Chlorophyta
- Diplonemida
- Cryptophyta
- Haptophyta
- Katablepharidophyta
- Telonemia
- Choanoflagellida
- Fungi
- Rozellida/Cryptomycota
- Aphelida
- Cercozoa
- MAST
- Ochrophyta
- Other Stramenopiles

**Supplementary Figure 3.B.8** Block-model representation of the networks of the planktonic protists in Lake Baikal for the three major depth categories. Networks were built on OTUs present in more than 20% samples and having a relative abundance higher than 0.01%. OTUs are represented by nodes and direct covariations between them, by edges.



## Chapter 4

Small freshwater ecosystems with  
dissimilar microbial communities exhibit  
similar temporal patterns





## 4.1 Context and objective

In the previous chapter, I investigated the microbial eukaryote communities in Lake Baikal, the biggest freshwater ecosystem on Earth. In this chapter, I focused on fairly common but also mostly unexplored ecosystems. Indeed, small freshwater ecosystems are overlooked in the literature, even if they present a clear ecological importance. Therefore, the knowledge of the diversity and community composition of microbial communities inhabiting these ecosystems is still scarce. Furthermore, their temporal evolution is also poorly understood. A previous study conducted in the laboratory by M. Simon, a former PhD student, described the composition of small microbial eukaryotes in five small freshwater ecosystems located in the south-west of Paris (France) (Simon et al., 2015). The sampling was conducted monthly during two years and the data were obtained by metabarcoding, using pyrosequencing 454 FLX of 18S rRNA genes. This study revealed that these small freshwater ecosystems harboured very diverse microbial eukaryotic communities. Moreover, the microbial eukaryotic communities were found to be highly dynamic, and presented strong seasonal patterns. Here, I completed the dataset by analyzing the prokaryotic communities. The prokaryotes dataset was obtained similarly to the eukaryote dataset. Therefore, I used the same samples but I enlarged the study to the whole microbial community (bacteria, archaea and microbial eukaryotes) to understand its structuring over time. My objectives were:

1. To decipher the bacterial and archaeal community composition in these small freshwater ecosystems. Indeed, the diversity of these communities is mostly unknown in small freshwater ecosystems. To do so, we used amplicon sequences obtained by pyrosequencing 454 FLX of the 16S rRNA genes.
2. To identify a potential seasonality of prokaryotic communities. As the microbial eukaryotes revealed strong seasonal patterns, I determined whether prokaryotes shared the same temporal dynamics, or if their temporal patterns were different.
3. To describe the temporal variations of whole microbial communities in the size fraction 0.22 to 5  $\mu\text{m}$ . With an approach based on time-lag analysis and turnover, I described the global dynamics of community composition.
4. To identify the abiotic parameters involved in microbial community composition and to determine their relative importance.

## 4.2 Manuscript

Gwendoline M. David<sup>1</sup>, Purificación López-García<sup>1</sup>, David Moreira<sup>1</sup>, Benjamin Alric<sup>2</sup>, Philippe Deschamps<sup>1</sup>, Paola Bertolino<sup>1</sup>, Gwendal Restoux<sup>3</sup>, Emma Rochelle-Newall<sup>4</sup>, Elisa Thébault<sup>4</sup>, Marianne Simon<sup>1</sup> and Ludwig Jardillier<sup>1</sup>

<sup>1</sup>Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, 91405, Orsay, France

<sup>2</sup>Irstea, UR RiverLy, Laboratoire d'écotoxicologie, centre de Lyon-Villeurbanne, 5 rue de la Doua CS 20244, F-69625, Villeurbanne, France

<sup>3</sup>GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

<sup>4</sup>Sorbonne Université, UPEC, CNRS, IRD, INRAE, Institute d'Ecologie de des Sciences de l'Environnement de Paris, iEES-Paris, F-75005 Paris, France

Submitted in *Molecular Ecology*, currently under review.

### 4.2.1 Abstract

Despite small freshwater ecosystems are biodiversity reservoirs and significantly contribute to greenhouse fluxes, their microbial communities remain largely understudied. Yet, microorganisms intervene in biogeochemical cycling and impact water quality. Because of their small size, these ecosystems are in principle more sensitive to disturbances, seasonal variation and pluri-annual climate change. However, how microbial community composition varies over space and time, and whether prokaryotic and eukaryotic communities behave similarly, remain fully open questions. Here, we aim at unravelling the composition and intra/inter-annual temporal dynamic patterns for both prokaryotes and microbial eukaryotes in a set of small freshwater ecosystems. We monitored prokaryote composition during 24 consecutive months in four ponds and one brook from North-Western France by 16S rDNA amplicon sequencing (microbial eukaryotes were previously investigated for the same systems). Unexpectedly for oxic environments, bacterial Candidate Phyla Radiation (CPR) were highly diverse and locally abundant. Our results suggest that microbial community structure is mainly driven by environmental conditions acting over space (ecosystems) and time (seasons). A low proportion of OTUs (<1%) was shared by the five ecosystems despite their geographical proximity, making microbial communities almost unique in each ecosystem and highlighting the strong selective influence of local environmental conditions. Marked and similar seasonality patterns were observed for prokaryotes and microbial eukaryotes in all ecosystems despite strong turnovers of rare OTUs.

Over the two-year survey, microbial community composition varied despite relatively stable environmental parameters. This suggests that biotic associations play an important role in inter-annual community assembly.

### 4.2.2 Introduction

Small freshwater ecosystems such as ponds or shallow lakes remain largely unexplored, despite being widespread and numerically very important ( $3.2 \times 10^9$ ) and, covering altogether a large area (0.8 billion km<sup>2</sup>) (Downing, 2010). Collectively, they contribute up to 15% of CO<sub>2</sub> and, especially, 40% of CH<sub>4</sub> emissions (Holgerson and Raymond, 2016; Yvon-Durocher et al., 2017). They exhibit highly diverse physicochemical conditions and are reservoirs of an important biodiversity in terms of small invertebrates and microorganisms as identified by classical observation (Dudgeon et al., 2006) and a few molecular diversity studies (e.g. Simon et al., 2015, 2014; Qin et al., 2016; Liu et al., 2020). Because of their small volume, small aquatic ecosystems have lower buffering capacity toward environmental variation at short i.e. seasons and longer, multiple-year time scales. Consequently, their communities and functioning are in principle more reactive to environmental changes than those of larger ecosystems (McKee et al., 2003; Downing et al., 2006; Downing, 2010). Understanding the temporal dynamics of organisms inhabiting these overlooked ecosystems is one of a first step to understand and predict their functioning in the long run. Indeed, because their fundamental roles in ecosystem functioning (Azam et al., 1983; Bar-On et al., 2018; Cho and Azam, 1988), changes in microbial community composition can have profound impacts on ecosystem processes due to changes in their functional capacities (Bell et al., 2005). However, most studies of microbial communities in freshwater systems have focused on their composition and spatial distribution (Debroas et al., 2017; Grossmann et al., 2016; Richards and Bass, 2005; Richards et al., 2015; Shade et al., 2013; Zhong et al., 2016; Zwirgmaier et al., 2008). Temporal surveys in aquatic systems have been only occasionally carried out, mostly at relatively short scales (days to months) (Bunse and Pinhassi, 2017; Fuhrman et al., 2015; Scherer et al., 2017). However, longer-term studies eventually reflecting seasonal patterns are still very rare (Gilbert et al., 2012; Karl and Church, 2014; Linz et al., 2017). Furthermore, most microbial diversity studies, including temporal surveys, focus on either prokaryotic or eukaryotic communities, but an integrated view of how the two change with time, eventually revealing underlying biotic interactions, are extremely scarce and restricted to large environments such as oceans and lakes (Biller et al., 2018; Obertegger et al., 2019; Steele et al., 2011). Yet, differences in overall metabolic capabilities, physiology, trophic features, generation time and dispersal might reflect in distinct temporal patterns of prokaryotes (bacteria, archaea) versus microbial eukaryotes. For instance, microbial eukaryotes have



much more restricted metabolic capabilities than prokaryotes widely diverse energy and carbon metabolisms. In contrast, many microbial eukaryotes are predators, including some microalgae (Bjorbækmo et al., 2020; Sherr and Sherr, 1994; Zubkov and Tarran, 2008), while predation is exceptional in prokaryotes (Shiratori et al., 2019). Here, we investigate the bacterial and archaeal community composition during 24 consecutive months in four shallow ponds and one brook located in the natural regional park of the Haute Vallée de Chevreuse (France) using 16S rDNA amplicon sequencing. We subsequently compare the temporal patterns of prokaryotic communities to those of microbial eukaryotes previously analyzed for the same ecosystems (Simon et al., 2015), analyzing them as a function of several environmental parameters using diverse metrics and multivariate statistical analyses. Although the five ecosystems harbour specific microbial communities, their temporal patterns were similar. Interestingly, both the prokaryotic and eukaryotic communities also show similar inter-annual temporal dynamics.

### 4.2.3 Material and methods

#### 4.2.3.1 Sampling and measurement of physicochemical parameters

We sampled monthly at the deepest site surface waters from five representative small and shallow freshwater ecosystems (the four ponds Etang des Vallées (EV), La Claye (LC), Mare Gabard (MG), Saint Robert (SR), and the brook Ru Sainte Anne (RSA)), previously described (Simon et al., 2014), between April 2011 and April 2013 (24 sampling dates) using sterilized plastic carboys. A total of 106 samples were collected (LC and RSA dried out occasionally). Single samples were analysed at each time point of this study since our protocol ensured very similar replicates (Simon et al., 2014). The five ecosystems are located in the Natural Regional Park of the Chevreuse Valley (Ile-de-France, France), less than 10 km apart. During our two-year survey, the Ru Sainte Anne and La Claye experienced several dry periods, from August to September 2012 for Ru Sainte Anne, and from July to December 2011 and in September 2012 for La Claye. These five ecosystems differ in their environmental conditions (Simon et al., 2015). Water temperature (T), dissolved oxygen concentration ( $O_2$ ), pH and conductivity (Cond) were measured in situ using a multiparameter probe Multi 350i (WTW, Weilheim, Germany). Concentrations of dissolved nutrients (total phosphorus ( $P_{tot}$ ), orthophosphate ( $PO_4^{3-}$ ), total nitrogen ( $N_{tot}$ ), nitrite ( $NO_2^-$ ), nitrate ( $NO_3^-$ ) and ammonia ( $NH_3$ )) were measured on the sampling day using colorimetric tests following manufacturer instructions (Hach-Lange, Düsseldorf, Germany). Samples filtered onto 0.2 $\mu$ m pore-size Nuclepore membranes (Whatman) were immediately frozen at -20°C until measurement (Rochelle-Newall et al., 2014) of dissolved organic carbon (DOC) using a TOC VCSH analyzer (Shimadzu, Tokyo, Japan). Water samples

were transported back to the laboratory and immediately serially filtered through 100, 30, 5 and 0.2  $\mu\text{m}$  pore-size Nuclepore filters (Whatman). Filters were immediately stored frozen at  $-80^{\circ}\text{C}$  until DNA extraction.

#### **4.2.3.2 DNA extraction, amplification and sequencing of 16S and 18S rDNA fragments**

DNA was extracted from cells collected in the smallest size-fraction (0.2—5  $\mu\text{m}$ ) using the PowerSoil DNA extraction kit (Qiagen, Hilden, Germany) as previously described (Simon et al., 2014). 16S rDNA fragments of ca. 290 bp, encompassing the V4 hypervariable region, were PCR-amplified using the universal primers U515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACVSGGGTATCTAAT-3') targeting both bacteria and archaea (Baker et al., 2003; Takai and Horikoshi, 2000). These primers were tagged with different 10 bp Molecular IDentifiers (MIDs) to allow differentiation of PCR products from 20 distinct samples pooled together for multiplex sequencing. PCR amplification was conducted in a total volume of 25  $\mu\text{l}$  using 1.5 mM  $\text{MgCl}_2$ , 0.2 mM dNTPs, 0.3  $\mu\text{M}$  primers, 0.3-3  $\mu\text{l}$  of DNA sample and 0.5 U HotStart Taq polymerase (Taq Platinum, Invitrogen, Carlsbad, CA, USA). Twenty-five amplification cycles (94°C for 30 s, 58°C for 45 s and 72°C for 90 s) were preceded by a 3-min denaturation step at 94°C and ended by a 10 min for final extension at 72°C. The products of five PCR reactions were pooled together for each sample. These pools were then purified with the QIAquick PCR purification kit (Qiagen, Hilden, Germany). Equivalent amounts of purified amplicons of 20 different samples were pooled and pyrosequenced using the 454 GS-FLX Titanium technology from Roche (Beckman Coulter Genomics, Danvers, MA, USA). In the case of eukaryotes, 18S rDNA fragments of ca. 550 bp, encompassing the V4 hypervariable region, were PCR amplified (Simon et al., 2014) with primers EK-565F and 18S-EUK-1134-R—UnonMet biased against Metazoa (Bower et al., 2004).

#### **4.2.3.3 Analysis of 16S and 18S rDNA 454-pyrosequences**

A total of 1 705 835 16S rDNA amplicon reads was obtained. Reads were filtered through an in-house pipeline to remove spurious sequences. Briefly, pyrosequences without the two flanking primer sequences or containing errors in primer/MIDs sequences or positions with undetermined bases were discarded. To further eliminate PCR and pyrosequencing errors, pairwise flowgram alignments were clustered and treated with the PyroNoise program of Amplicon-Noise (Quince et al., 2011). Filtered reads were then clustered into Operational Taxonomic Units (OTUs) with a 97% similarity cut-off (Stackebrandt and Goebel, 1994) using CD-HIT

(Fu et al., 2012). Singletons (OTUs composed of only one single sequence) were excluded. Each OTU was represented by its most abundant sequence, used as reference. Reference sequences were assigned to taxonomic groups based on similarity with sequences of the SILVA 128 database (<http://www.arb-silva.de>). Reads affiliated to mitochondria and chloroplast were excluded from the dataset. At the end, 1 131 900 high-quality reads were retained and assigned to 16 573 OTUs. Sequences from all OTUs were then attributed to samples according to their MIDs. OTU sequences of 0.2-5  $\mu\text{m}$  microbial eukaryotes were obtained following a similar procedure to that followed for 18S rDNA amplicons (Simon et al., 2015). Sequencing data for the last four months of the temporal series in Mare Gabard could unfortunately not be generated. Prokaryotic sequences have been deposited at GenBank under Bioproject number PRJNA539931.

#### 4.2.3.4 Statistical analyses

All statistical analyses were conducted with the R software (<http://cran.r-project.org>) (R Core Team, 2019). Missing physico-chemical values (13% of the physico-chemical dataset) were replaced by the overall average value of the parameter and nutrient concentrations below kit detection limits (19% of the physico-chemical dataset) were set as the limit value minus a small random number (normal distribution  $\mu = 0$ ;  $\sigma = 10^{-6}$ ) for all statistical analyses. To take into account the measured environmental factors altogether in the diverse statistical analyses, their values were normalized ( $\mu = 0$ ;  $\sigma = 1$ ). Relative OTU abundances were obtained by dividing the number of reads for each OTU by the total number of reads retrieved in the sample. OTUs were divided into two groups based on their relative abundance and temporal frequencies as in Simon et al. (2015) : abundant OTUs, with a mean relative abundance superior to 0.1% and detected in at least 25% of the samples within one ecosystem, and rare OTUs, represented by all the remaining OTUs. All subsequent statistical analyses based on Bray-Curtis dissimilarities (Bray and Curtis, 1957) were conducted on Wisconsin-standardized OTU relative frequencies to balance the weight between abundant OTUs and rare OTUs. For global microbial community analyses, i.e. considering eukaryotic and prokaryotic datasets together, tables of Wisconsin-standardized OTU frequencies were pooled into a single table without changing the initial relative OTU frequencies. In each ecosystem, richness, evenness and diversity index were calculated. Richness was evaluated by rarefaction analyses as the estimated number of OTUs in a random subsample of each sequence library, which is of the same size as the smallest one (Hurlbert, 1971). Evenness,  $e$ , was calculated according to Pielou (1966):  $e = \frac{\sum_i f_i \times \ln(f_i)}{\ln(S)}$  with  $f_i$  the frequency of OTU $_i$  and  $S$  the observed number of OTUs in the sample. Diversity was estimated according to the Gini-Simpson index (Simpson, 1949; Lee et al., 2013) calculated

as follows using the diversity function of the ‘vegan’ R-package (Oksanen et al., 2007):  $D = 1 - \sum_i f_i^2$  where  $f_i$  is the OTU<sub>*i*</sub> frequency. Microbial community structuring was investigated in the different ecosystems. We drawn Venn diagrams to show shared OTUs between ecosystems, using the ‘gplots’ R-package (Warnes et al., 2016). Hierarchical cluster analysis was build using Bray-Curtis dissimilarities and the Ward2 algorithm (Murtagh and Legendre, 2014) in the hclust function of the ‘stats’ R-package to determine clusters of microbial communities based on the distances between every samples of the five ecosystems. Patterns in the temporal dynamics of prokaryotes and eukaryotes were explored at intra- (seasons) and inter-annual (between years) scales. Pairwise Bray-Curtis dissimilarities were computed between samples collected the first year and samples from the same system the second year over increasing time-lags (from one-month to 12-months) to characterize the temporal dynamics of microbial communities at the year-scale. Time-lag analyses (TLA, Collins et al., 2000) between sampling dates over increasing time-lags (from one-month to 23-months) were computed using Bray-Curtis dissimilarities for microbial communities and Euclidean distances for environmental parameters to determine their variability over consecutive years. We computed polynomial regressions (fitting better than sinusoidal regressions) of Bray-Curtis dissimilarities over time to estimate the frequency of any putative temporal cyclic pattern and linear regressions over time to estimate global shifts in the microbial community compositions and environmental parameters over the two-year survey using the lm function of the ‘stats’ R-package. The best fitted polynomial model was determined based on the Akaike Information Criterion (AIC, Akaike, 1973). In TLA, the slope of the linear regression indicates the rate and direction of temporal changes. To test for the significance of slopes for each ecosystem, a randomization procedure was applied with 5 000 permutations using the lperm function of the ‘lmPerm’ R-package (Wheeler et al., 2016). A null distribution was estimated from each slope that was calculated on randomized data. The observed slopes were then compared to the distribution of F-values obtained under the null hypothesis. OTU turnovers were calculated between seasons or between years of the two-year survey, using the ‘codyn’ R-package based on the following formula (Diamond, 1969; Hallett et al., 2016; MacArthur and Wilson, 1963):  $\text{Turnover} = \frac{S_{\text{gained}} + S_{\text{lost}}}{S_{\text{total}}}$ . Canonical correspondence analyses were conducted to determine the contribution of each environmental factor to the variation of microbial community composition. Microbial community assembly was investigated considering the potential interplay between habitats (i.e. ecosystems) and time. To assess how close are microbial community compositions between sampling sites or seasons, non-metric multidimensional scaling (NMDS) analyses were performed based on Bray-Curtis dissimilarities using the ‘vegan’ R-package. To quantify the role played by the ecosystem clusters (i.e. groups of ecosystems defined in the hierarchical cluster analysis), the habitat (i.e. ecosystems), the time (i.e. year of sampling or the sampling season) in the structuring of the microbial communities, we conducted permutational multivariate analyses of variance

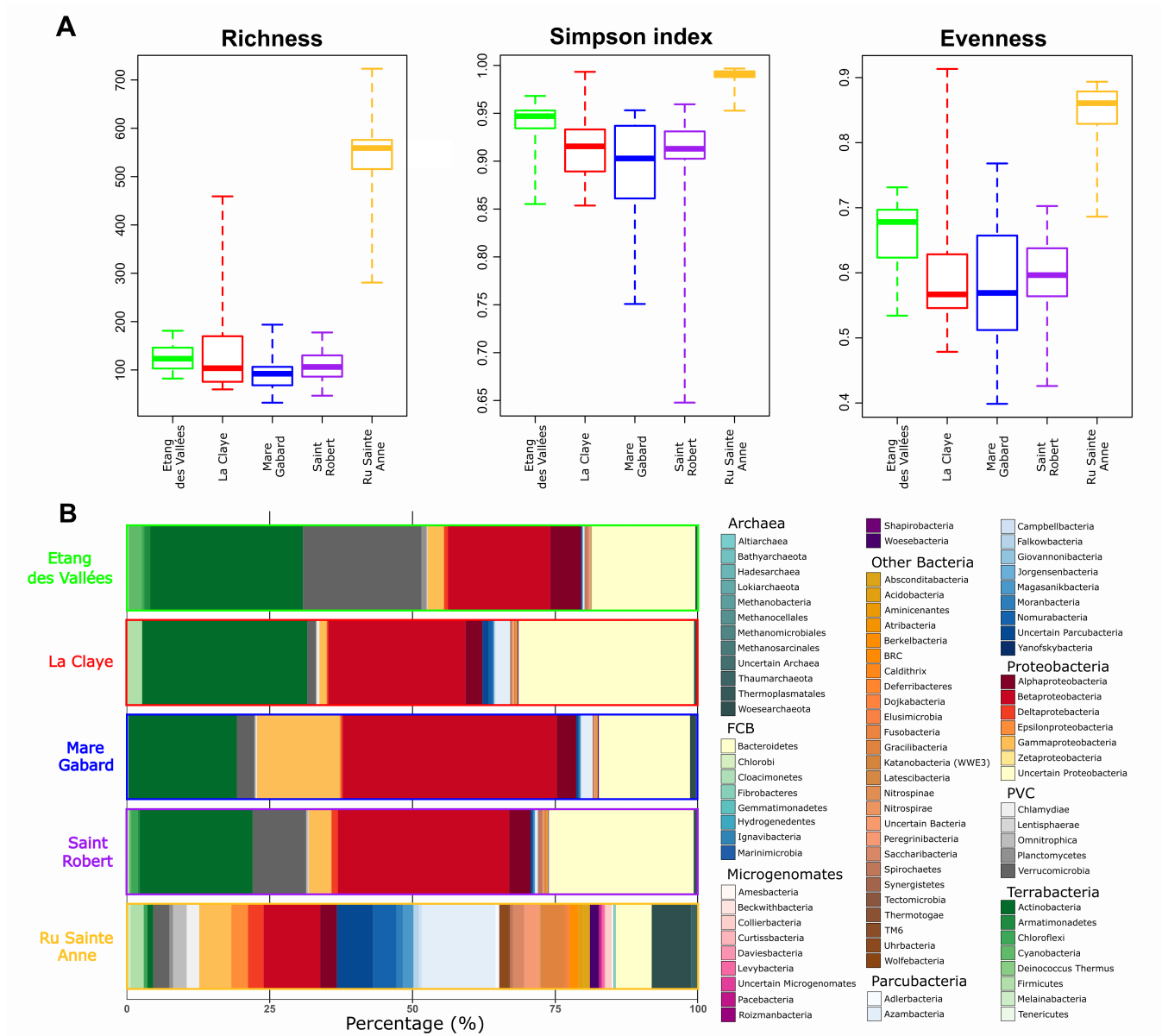
(PERMANOVA) using Bray-Curtis dissimilarities calculated on OTU relative frequencies using the `adonis` function of the ‘vegan’ R-package. We explored biotic interactions by inferring how prokaryotes and microbial eukaryotes may influence each other assembly over time in each ecosystem. To do so, putative multicollinearity between environmental parameters was checked by computing the variance inflation factors (VIF) with the function ‘`vif.cca`’ of the ‘vegan’ R-package. A VIF value above 10 indicated a strong multicollinearity. Environmental factors were excluded from the analysis if they were characterized by multicollinearity. Forward selection was then used to sort environmental parameters explaining significantly community composition variation for CCA model (Monte Carlo permutation tests, 999 permutations). Correlation of multivariate vectors (RV index) between prokaryotes and microbial eukaryotes was assessed based on co-inertia analyses in order to assess synchrony of their dynamics and was tested by permutations (999 permutations) provides the variability explained between the two communities (Dolédec and Chessel, 1994). CCA combined to forward selection were conducted to determine how prokaryotes explain the variance of microbial eukaryotes and inversely.

#### 4.2.4 Results

We generated prokaryotic 16S rDNA amplicon data (Supplementary Table 4.A.1) for plankton samples (106 samples; 0.2-5  $\mu\text{m}$  size-fraction) collected monthly for two years from five shallow freshwater ecosystems for which protist communities had been previously studied (Simon et al., 2015). These spatially close ecosystems varied in their physico-chemical conditions. The main differences were lower pH (6.2 on average) and nitrite concentrations ( $0.0086 \text{ mg.l}^{-1}$  on average) in Mare Gabard, lower ammonia concentrations ( $0.07 \text{ mg.l}^{-1}$  on average) in La Claye, variable ammonia inputs ( $<0.1$  to  $2.65 \text{ mg.l}^{-1}$ ) in Saint Robert, higher oxygen (79.2% on average) and lower orthophosphate ( $0.05 \text{ mg.l}^{-1}$ ) concentrations in Etang des Vallées, and higher conductivity ( $657.3 \mu\text{S.cm}^{-1}$ ) in the stream Ru Sainte Anne. All the ecosystems were characterized by strong temporal variations in water temperature (thermal amplitude of  $19.4^\circ\text{C} \pm 2.3$  on average, for the five ecosystems) whereas the other environmental parameters had no apparent temporal variability (Simon et al., 2015).

##### 4.2.4.1 Prokaryotic community composition

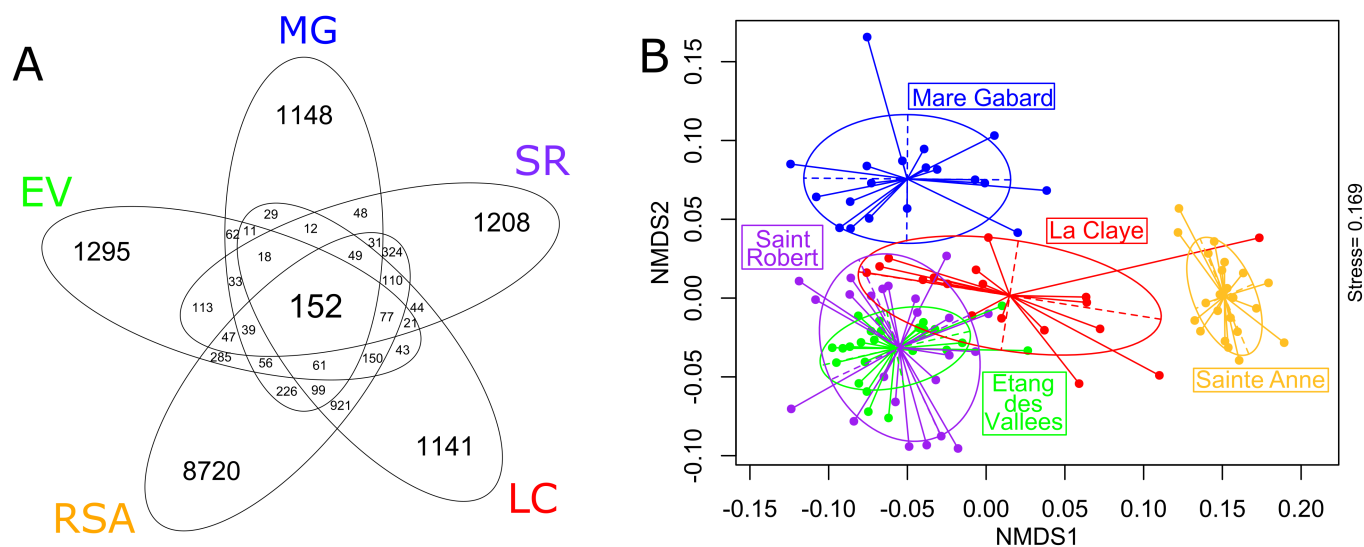
From the total number of retained high-quality 16S rRNA gene reads (1 131 900), we determined a total of 16 573 prokaryotic OTUs (14 701 bacterial and 1 872 archaeal). The total number of OTUs in the Ru Sainte Anne brook was higher than in the four ponds, where similar amounts of



**Figure 4.2.1** (a) Cumulative curves showing the total number of prokaryotic OTUs detected in each ecosystem from April 2011 until April 2013 during the temporal survey in the five ecosystems (Etang des Vallées, La Claye, Mare Gabard, Saint Robert and Ru Sainte Anne). Dashed lines indicate missing data due to the drought: Ru Sainte Anne was dried in August and September 2012, and La Claye from the end of July to December 2011 and in September 2012. (b) Richness and diversity estimators computed for all the samples. Richness corresponds to the expected number of OTUs in random subsamples of the size of the smallest sequence library (1309 reads in Mare Gabard in May 2011). The thickest line inside each box represents the median on the distribution, bottom and top borders of boxes correspond to the first and third quartiles and whiskers extend to minimal and maximal distances. (c) Histograms showing the relative abundance of 16S rDNA amplicon reads assigned to high-rank taxa averaged over the two-year survey, from the five ecosystems (CPR: Candidate Phyla Radiation, FCB: Fibrobacteres-Chlorobi-Bacteroidetes, PVC: Planctomycetes-Verrucomicrobia-Chlamydiae). One star (\*) indicates Microgenomates and two stars (\*\*) shows Parcubacteria.

OTUs were detected (Figure 4.2.1A), an observation consistent with that of eukaryotic diversity (Simon et al., 2015). Consequently, richness, evenness and diversity (Simpson index) were similar in the four ponds and reached higher values (mean richness of 546, Simpson index of 0.99 and evenness of 0.85) in the Ru Sainte Anne (Figure 4.2.1B). The spread of the quartiles of the box plots suggested that these indices varied strongly over time and differently among ecosystems (Figure 4.2.1B). Overall, only 271 OTUs (considering only one occurrence for OTUs detected in several ecosystems; 1.6% of all OTUs) were abundant (i.e. detected in each ecosystem in at least 25% of samples and with  $\geq 0.1\%$  relative abundance per sample) but represented 84.0% of the total number of reads (Supplementary Table 4.A.2). The vast majority of the prokaryotic diversity was thus composed of rare OTUs (Supplementary Figure 4.B.1A and 4.B.1B). Prokaryotic OTUs affiliated to 78 different phyla, which we classified into 8 super-groups of different taxonomic levels: Archaea (not further subdivided in taxa owing to their lower relative diversity and abundance in our samples), Fibrobacteres-Chlorobi-Bacteroidetes (FCB), Microgenomates, Parcubacteria, Proteobacteria, Planctomycetes-Verrucomicrobia-Chlamydiae (PVC), Terrabacteria and Other Bacteria (Figure 4.2.1C). The highest number of phyla (71 against 64 on average for the other ecosystems) was found in Ru Sainte Anne. Bacteria clearly dominated (98.0% of all reads) the prokaryotic community in the five ecosystems. Archaea were globally rare at the exception of Ru Sainte Anne, where they reached their higher abundances (8.1% of all reads). Woesearchaeota prevailed within Archaea (1.7% of all reads), whereas Bacteria were dominated in the five ecosystems by four phyla: Proteobacteria (on average, 34.1% of reads; including, by decreasing frequency order, Beta-, Gamma-, Alpha-, Delta- and Epsilon-proteobacteria), Actinobacteria (20.0%), Bacteroidetes (19.3%), and Verrucomicrobia (9.0%). Although Proteobacteria was the most abundant phylum in all ecosystems, the other phyla contributed differently to the structure of the prokaryotic communities in the five ecosystems (Supplementary Figure 4.B.1C). The bacterial Candidate Phyla Radiation (CPR) was represented in our samples by 27 phyla that accounted for 9.9% of all reads. They corresponded mainly to the super-phyla Parcubacteria (67.2% of CPR-affiliated reads), Microgenomates (8.5%) and to the phyla Gracilibacteria (8.1%). In contrast with its relatively small frequency in the ponds, Parcubacteria was the most abundant group in Ru Sainte Anne (28.4% of all reads).

The five ecosystems harboured very diverse prokaryotic communities at the OTU level. Only 152 out of the 16 573 OTUs were shared by all the ecosystems (Figure 4.2.2A). Although they represented less than 1% of all OTUs, they contributed to 61.0% of all reads in the five ecosystems. Shared OTUs were affiliated to 23 bacterial phyla and only to 2 archaeal phyla (Woesearchaeota and Methanobacteria). Abundant OTUs represented 59.9% of the shared OTUs and belonged to the most frequent phyla (i.e. Proteobacteria, Bacteroidetes, Actinobacteria, Verrucomicrobia). Among the rest of shared OTUs, i.e. the rare ones, Archaea were prominent along with Parcubacteria, Proteobacteria and Bacteroidetes. Prokaryotic commu-



**Figure 4.2.2** Venn diagram showing the number of OTUs shared by several ecosystems or specific to an ecosystem. All samples from the two-year survey are pooled for each ecosystem. (b) Non-metric multidimensional scaling plot, built on Wisconsin-standardized Bray–Curtis dissimilarities between all samples. Stress value is displayed on the right border of the plot.

nities were mainly composed of shared OTUs in La Claye (5.2% OTUs and 80.8% reads) and non-shared ones in Ru Sainte Anne (98.7% OTUs and 68.4% reads). We used NMDS analysis based on OTU frequencies to highlight the dissimilarities between the prokaryotic communities of the five ecosystems (Figure 4.2.2B). These communities seemed to be more similar in Etang des Vallées and Saint Robert, where Proteobacteria, Actinobacteria, Bacteroidetes and Verrucomicrobia were the most abundant phyla (Supplementary Figure 4.B.2). Mare Gabard harboured the highest abundance of Proteobacteria, which represented more than half of the prokaryotic community in this ecosystem (56% of all reads). Ru Sainte Anne had a more homogenous prokaryotic community structure at the phylum level and hosted the highest proportions of Azambacteria and Woesearchaeota. The temporal dynamics of the prokaryotic phyla was characterized by small changes at the month scale in the five ecosystems (Supplementary Figure 4.B.2). The highest abundances of Archaea were recorded in Ru Saint Anne in November 2011 (29.1% of all reads), July 2012 (25.2%) and, to a lower extent, in November 2012 (18.6%) and December 2012 (20.9%), and were mainly driven by Woesearchaeota (Supplementary Figure 4.B.3). Proteobacteria, Actinobacteria and Bacteroidetes showed very different temporal dynamics between ecosystems. Proteobacteria reached their higher abundances in June 2012 in Mare Gabard (81.8% of all reads) and Saint Robert (60.4%) and in October 2011 in Ru Sainte Anne (49.8%). Actinobacteria peaked in Etang des Vallées in March 2013 (50.3%) and at similar levels in June and November 2012 in La Claye (46.7% and 45.0%, respectively). The highest abundances of Bacteroidetes were recorded in Saint Robert in March 2013 (63.7%)

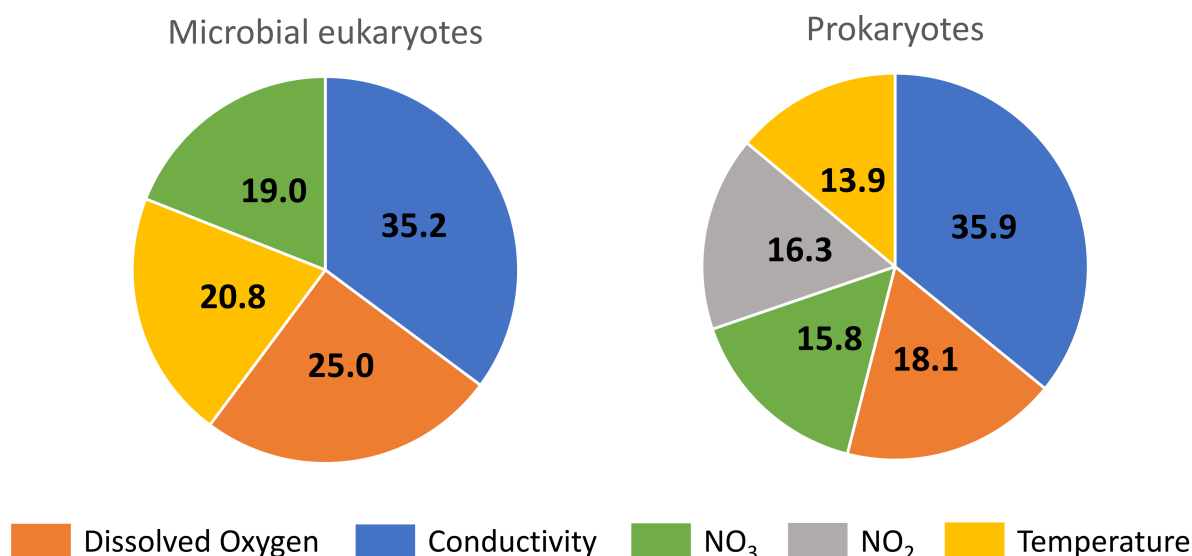


and in January 2012 in Saint Robert and La Claye (54.4% and 46.6%, respectively). Bacteria belonging to the CPR varied also greatly over time and between ecosystems. They dominated in June 2011 and June 2012 (62.9% and 63.5%, respectively) in Ru Sainte Anne and represented up to 25.2% of reads in October 2012 in La Claye. Parcubacteria and Microgenomates exhibited similar temporal patterns. By contrast, Cyanobacteria were low abundant in the five ecosystems all along our survey, even in summer, representing on average 0.7% of all reads.

#### 4.2.4.2 Comparison of eukaryotic and prokaryotic community structures

After analyzing prokaryotic diversity alone, we incorporated data of the small (0.2-5  $\mu\text{m}$ ) eukaryotes present in the same samples (Simon et al., 2015). NMDS analysis of pooled prokaryotic and eukaryotic OTU frequencies revealed marked differences in microbial community composition among the five ecosystems (Supplementary Figure 4.B.4). Hierarchical clustering analysis based on Bray-Curtis dissimilarities of OTU frequencies supported these differences between the ecosystems. As in the case of the prokaryotic community, the Ru Sainte Anne global microbial community was set apart of the cluster composed of the four ponds (Supplementary Figure 4.B.5). Interestingly, the cluster of Ru Sainte Anne samples also contained a sample collected in La Claye just after a drought event (LC37, October 2012). To determine the effect of the environment on the microbial community composition, we conducted CCA on either prokaryotic or eukaryotic OTUs and the significant physico-chemical parameters sorted out by a forward analysis. Conductivity, water temperature, concentrations of nitrite and dissolved oxygen were the only parameters that had a significant effect on the microbial communities (10.0 and 6.9% of variance for prokaryotes and eukaryotes, respectively) (Figure 4.2.3). In addition, the influence of nitrate concentration appeared to be significant only for prokaryotes. Conductivity, a proxy of the total ion concentration, explained a third of the variance of prokaryotes and microbial eukaryotes on average. The remaining four significant abiotic factors contributed similarly to explain the community structure. To further explore how the environment controls the most abundant microbial taxa, we carried out CCA between various physicochemical parameters and the most abundant OTUs (Supplementary Figure 4.B.6). Prokaryotes and microbial eukaryotes in Etang des Vallées were highly related to pH, concentrations of dissolved oxygen, nitrite and nitrate. In Ru Sainte Anne, microbial communities appeared to be only linked to conductivity. There was no apparent correlation between the measured environmental parameters and the abundant prokaryotes and microbial eukaryotes in the other ecosystems.

To assess putative biotic interactions between prokaryotes and microbial eukaryotes, we computed the RV index. Considering all communities in the five ecosystems together, co-inertia analysis shown that the dynamics of prokaryote and protist communities were highly



**Figure 4.2.3** Canonical correspondence analyses conducted on all OTUs affiliated either to microbial eukaryotes or prokaryotes retrieved in the five ecosystems over the two-year survey. OTU proportion were Wisconsin standardized. Among all measured physico-chemical parameters, only conductivity, temperature, concentrations of dissolved oxygen nitrate and nitrite were explained significantly the variance of the OTUs ( $P < 0.05$ ). Considered together, these abiotic parameters explained 6.95 and 9.98% of the global variance of the OTUs belonging to microbial eukaryotes and prokaryotes, respectively. Values within each section of the pie charts indicate the proportion of variance explained by each parameter of the global variance of the OTUs.

and significantly similar with  $RV = 0.745$  ( $p = 0.001$ ), suggesting co-variations as a result of similar environmental triggers or biotic interactions between the two communities. At the ecosystem level, a significant temporal co-variation of microbial communities was observed only in Etang des Vallées and Saint Robert (Supplementary Table 4.A.4). CCA analysis based on OTUs grouped by phyla showed that eukaryotes explained as much as 38.6% of the variance of the prokaryotic community, mainly by four eukaryotic phyla (Cryptomycota, Oomyceta, Marine Stramenopiles, Chrysophyta) for which the co-inertia was significant. Inversely, five prokaryotic phyla explained 21.2% of the eukaryotic variance (Proteobacteria, Actinobacteria, Bacteroidetes, Amesbacteria and Woesebacteria). A PERMANOVA analysis was used to quantify the role played by habitat (i.e. ecosystem) and time (season and year) in the microbial community structuring at the OTU level. Habitat and time had similar influences on the composition of the global microbial, prokaryotic and eukaryotic communities. Differences in the global microbial communities between the Ru Sainte Anne and the other ecosystems were greater (F.model value of 4.739) than between these four remaining ecosystems (F.model value of 2.9607) (Table 4.2.1). To a lower extent, microbial communities shifted over time at seasonal

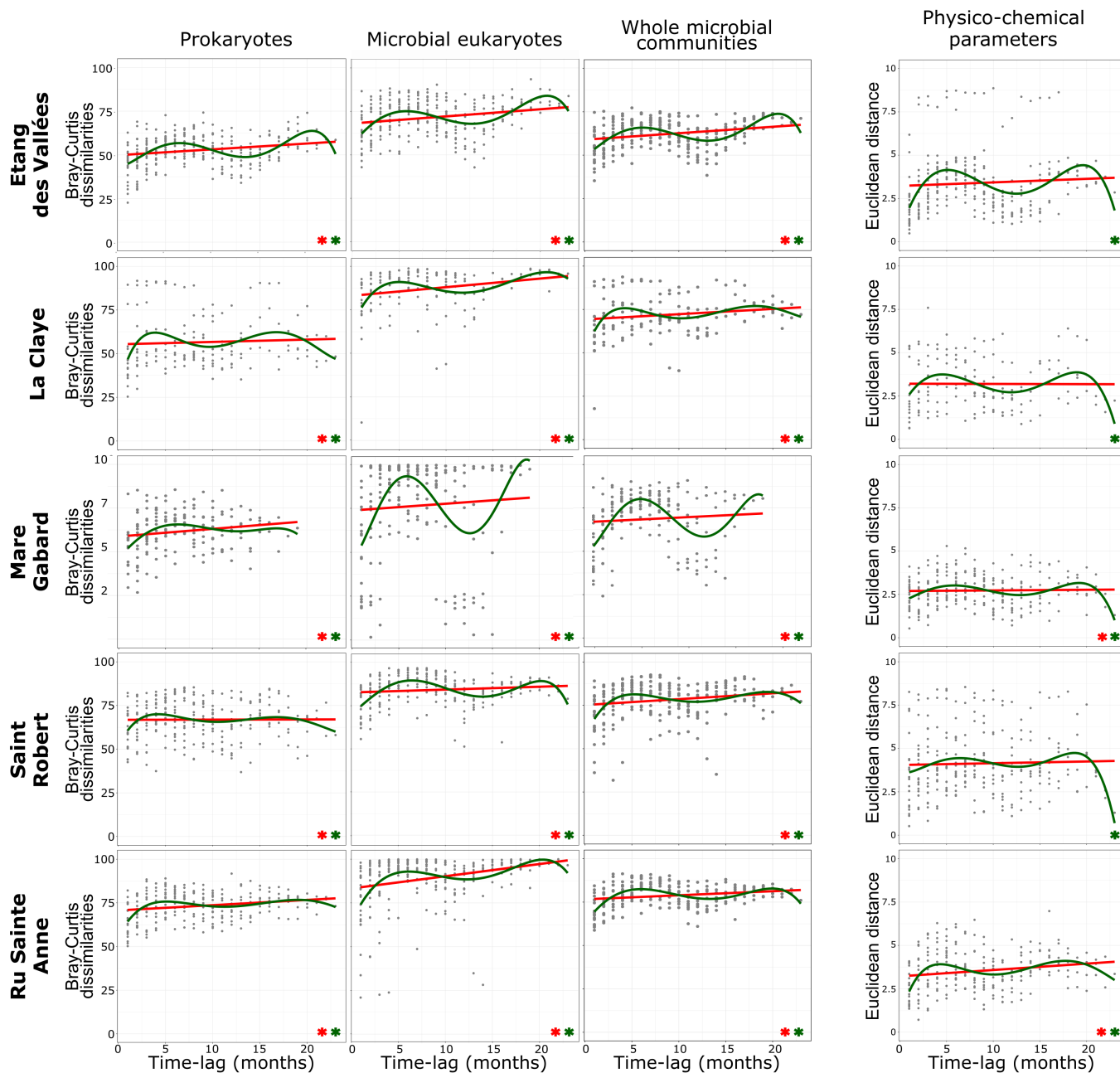
and year scales (F.model values of 1.4079 and 1.2424, respectively).

**Table 4.2.1** Permutational multivariate analyses of variance (PERMANOVA) partitioning of the whole microbial community, prokaryotes and microbial eukaryotes, based on Bray-Curtis dissimilarities calculated using Wisconsin standardized OTU relative abundances among hierarchical-defined clusters 1 and 2 (see Supplementary Figure 4.B.5), location (ecosystem), season and year.

	Cluster				Location   Cluster				Season				Year			
	Df	F.model	r <sup>2</sup>	Pvalue	Df	F.model	r <sup>2</sup>	Pvalue	Df	F.model	r <sup>2</sup>	Pvalue	Df	F.model	r <sup>2</sup>	Pvalue
Whole microbial community	1	4.74	4.1·10 <sup>-2</sup>	5.0·10 <sup>-4</sup>	3	2.96	7.7·10 <sup>-2</sup>	5.0·10 <sup>-4</sup>	3	1.41	3.6·10 <sup>-2</sup>	5.0·10 <sup>-4</sup>	1	1.24	1.1·10 <sup>-2</sup>	1.5·10 <sup>-2</sup>
Prokaryotes	1	4.51	3.9·10 <sup>-2</sup>	5.0·10 <sup>-4</sup>	3	2.65	6.9·10 <sup>-2</sup>	5.0·10 <sup>-4</sup>	3	1.31	3.4·10 <sup>-2</sup>	5.0·10 <sup>-4</sup>	1	1.26	1.1·10 <sup>-2</sup>	9.0·10 <sup>-3</sup>
Microbial eukaryotes	1	4.95	4.1·10 <sup>-2</sup>	5.0·10 <sup>-4</sup>	3	3.30	8.3·10 <sup>-2</sup>	5.0·10 <sup>-4</sup>	3	1.48	3.7·10 <sup>-2</sup>	5.0·10 <sup>-4</sup>	1	1.25	1.0·10 <sup>-2</sup>	1.5·10 <sup>-2</sup>

#### 4.2.4.3 Intra- and inter-annual dynamics of abiotic parameters and microbial community composition

We conducted time-lag analyses based on linear and polynomial regressions of Bray-Curtis distances over time on environmental factors and microbial community composition to assess the temporal patterns at intra- and inter-annual scales. The slopes of the linear regressions of abiotic parameters were either not significant (Etang des Vallées, La Claye and Saint Robert) or only slightly marked (Mare Gabard and Ru Sainte Anne, linear regression slopes of 0.025 and 0.036, respectively) (Figure 4.2.4 and Supplementary Table 4.A.3). By contrast, those for prokaryotes, microbial eukaryotes and all microbes considered together were characterized by significant and positive linear regressions. All the slopes were stronger for the microbial communities than for the abiotic parameters (from 0.091 for La Claye to 0.192 for Mare Gabard). These results suggest no or weak temporal evolution of abiotic parameters when microbial communities diverged increasingly over time. These effects were visible and significant even when considering similar numbers of pairwise Bray-Curtis dissimilarity comparisons for all time-lag periods (data not shown). The model that best fitted the distribution of the time-lag analyses for prokaryotes, eukaryotes and physico-chemical parameters was a fifth order polynomial regression that revealed marked changes in dissimilarities at a 6-month frequency. This analysis revealed significant seasonal patterns for microbial communities and the environmental conditions in the five ecosystems. The dispersion of the data was better explained for the biotic than for the abiotic factors. Pairwise Bray-Curtis dissimilarities calculated on rare and abundant OTU frequencies between the two years of sampling revealed seasonality for prokaryotes, microbial eukaryotes and both considered together. Globally, the temporal dynamics of the abundant OTUs was characterized by seasonality, with higher dissimilarities for samples separated by 5 to 6 months (i.e. opposite seasons, 0.79 average Bray-Curtis dissimilarity) than for temporally



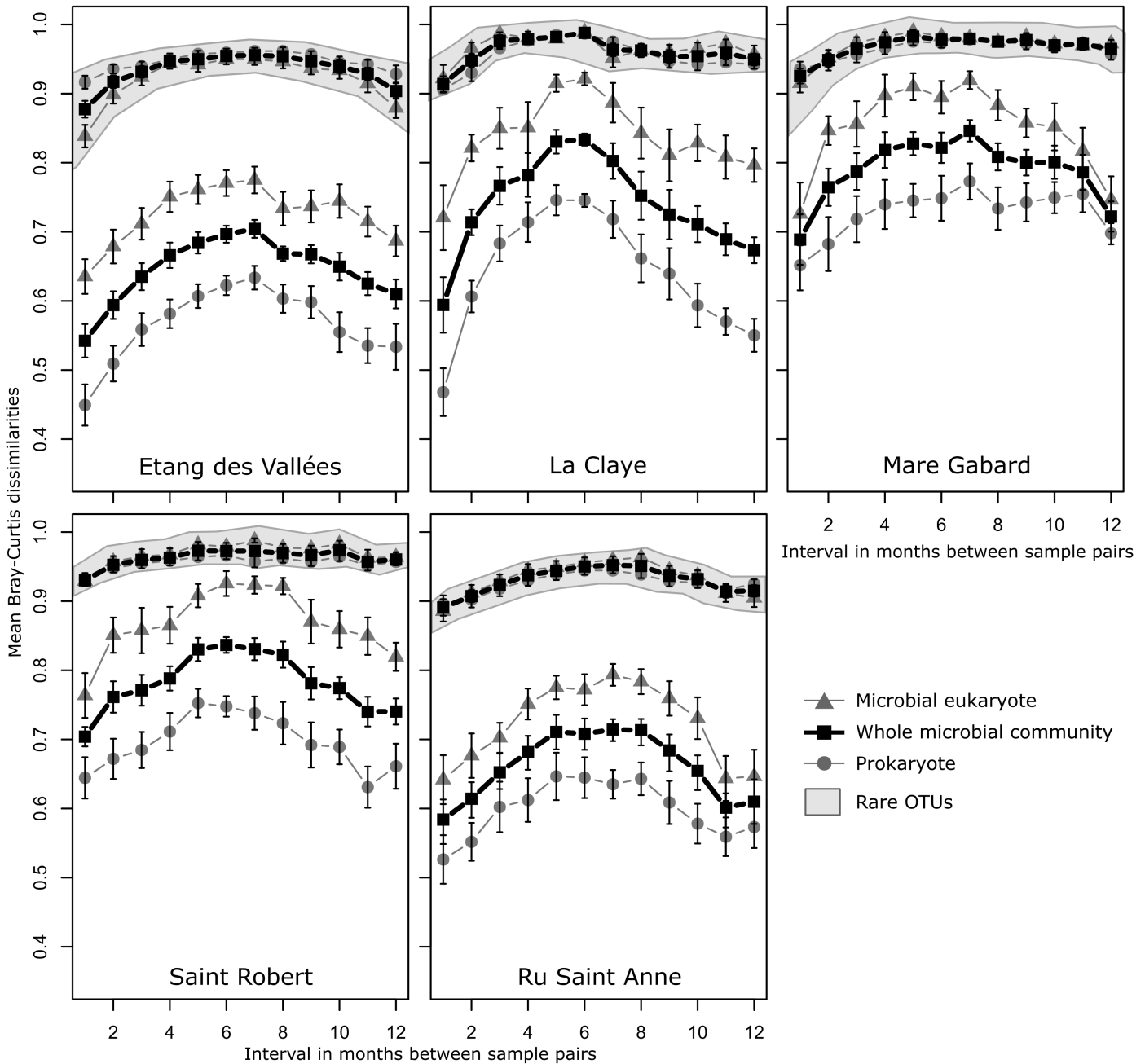
**Figure 4.2.4** Time-lag analyses (TLA) were computed for each ecosystem between sampling dates over increasing time-lags over the two-year survey (from one-month to 23-months). Wisconsin standardized Bray-Curtis dissimilarities were used with all OTUs of prokaryotes, microbial eukaryotes and whole microbial communities while Euclidean distances were applied with environmental parameters. Linear regressions, in red, and polynomial regressions of the 5th order, in green were calculated for each TLA dataset. Significant regressions ( $P < 0.05$ ) are indicated with a green star (linear regressions) or a red star (polynomial regressions).

contiguous samples (consecutive months, 0.62 average Bray-Curtis dissimilarity) (Figure 4.2.5). Rare OTUs showed, to a lower extent, similar patterns (average Bray-Curtis dissimilarity of 0.94 and 0.86 for opposite seasons and consecutive months, respectively). The seasonality of abundant OTUs was most pronounced in La Claye whereas there was no difference in the seasonality of rare OTUs among the five ecosystems. NMDS analysis based on OTU frequencies highlighted that the strongest shifts in community composition occurred between winter and summer (Supplementary Figure 4.B.7), in agreement with previous similar observation for microbial eukaryotes (Simon et al., 2015). Turnover values for the abundant prokaryotic and eukaryotic OTUs were more marked between years (19.6 and 31.4% on average, respectively) than between seasons (10.9 and 22.7% on average, respectively). Turnover indices calculated for each couple of seasons were higher between summer and winter, at the exception of La Claye. Rare OTUs, both prokaryotic and eukaryotic ones, were characterized by higher OTU turnover between seasons and between years than abundant OTUs (Supplementary Figure 4.B.8).

## 4.2.5 Discussion

### 4.2.5.1 Local environmental selection drives different community compositions in small freshwater ecosystems

In order to study the diversity and temporal dynamics of planktonic microorganisms (0.2-5  $\mu\text{m}$  size-fraction) of the three domains of life in small freshwater systems, we generated 16S rDNA amplicon sequences for samples collected over 24 months from 4 ponds and one brook for which we previously generated 18S rDNA datasets from the same original samples (Simon et al., 2015). These systems were characterized by differences in their physico-chemical conditions (Simon et al., 2015). Ru Sainte Anne and La Claye experienced drought events that lasted up to five months. We retrieved a total number of 16 573 OTUs affiliated to prokaryotes (14 701 bacterial OTUs and 1 872 archaeal OTUs) for a total of 3 583 OTUs affiliated to microbial eukaryotes. These numbers are higher than those usually retrieved in other freshwater ecosystems (Bock et al., 2018; Llíros et al., 2014) even after our very stringent analysis of the sequence data (see Methods; Supplementary Tables 4.A.1 and 4.A.2). The higher diversity actually reflects the cumulative effect of observed OTUs along time series and highlights the importance of integrating temporal scales to better depict global microbial biodiversity (Simon et al., 2015). Not surprisingly, microbial communities were mainly composed of rare OTUs (Supplementary Table 2), for both prokaryotes and microbial eukaryotes. Rare OTUs were affiliated to all the 78 detected prokaryotic phyla whereas abundant OTUs ( $\geq 0.1\%$ ) were restricted to 32 phyla. Most prokaryotic phyla were widespread in the five ecosystems; only a few low-abundant ones



**Figure 4.2.5** Wisconsin standardized pairwise Bray-Curtis dissimilarities were calculated between samples separated from 1 to 12 months for abundant (detected in each ecosystem in at least 25% of samples, and with  $>0.1\%$  mean relative abundance per sample) and rare OTUs (detected in each ecosystem in less than 25% of samples, and with  $\leq 0.1\%$  mean relative abundance per sample; curves highlighted with a grey shade) belonging to microbial eukaryotes (triangles), prokaryotes (circles) and whole microbial community (squares) detected in the five ecosystems over the two-year survey. Each point represents mean value of pairwise Bray-Curtis dissimilarities and is plotted with standard errors.

were not detected in all of them. On average, the prokaryotic OTU richness was about five times higher than that of eukaryotes at each sampling date (Figure 4.2.1B), which is in agreement with the global higher diversity of prokaryotes (e.g. Hug et al., 2016). Proteobacteria, Bacteroidetes, Actinobacteria and Verrucomicrobia dominated (Supplementary Figure 4.B.1), as often observed in freshwater lentic ecosystems (Newton et al., 2011; Riemann et al., 2008; Zwart et al., 2002). Cyanobacteria were detected at low abundances, likely because we focused on the small size-fraction (0.2-5  $\mu\text{m}$ ) whereas freshwater Cyanobacteria are usually of larger size or filamentous. CPR bacteria and the Diapherotrites-Parvarchaeota-Aenigmarchaeota-Nanoarchaeota-Nanohaloarchaeota (DPANN) supergroups, which have few cultured representatives and are likely composed of symbionts (Castelle et al., 2018), were observed in the five ecosystems. In particular, Woese archaeota was the most abundant archaeal phylum (Liu et al., 2018) (Figure 4.2.1C, Supplementary Figure 4.B.3). Despite these overall similarities and despite high OTU turnover rates at intra- and inter-annual scales (Supplementary Figure 4.B.8), each ecosystem harboured specific prokaryotic communities mainly composed of non-shared OTUs (Figure 4.2.2 and Supplementary Figure 4.B.5) as observed in other freshwater environments, including lakes (Bock et al., 2018; Linz et al., 2017). Each of the five ecosystems appears to represent a specific habitat for diverse prokaryotic communities because of particular physico-chemical conditions, nutrient composition, availability as previously described for these ecosystems (Simon et al., 2015), and for other freshwater environments (Fisher et al., 2000; Peura et al., 2012). In clear contrast with the four ponds analyzed, the prokaryotic community in the Ru Sainte Anne brook was very different. This lotic ecosystem harboured a prokaryotic community that was characterized by the highest richness and evenness as well as by the highest abundances of Archaea and CPR bacteria (Figure 4.2.1A-C and 4.2.2B, Supplementary Figure 4.B.5). This higher richness might be related to a higher input on allochthonous microbes (e.g. from soil or upstream waters) into the system (e.g. Crump et al., 2012; Teachey et al., 2019). In addition to this major difference in community composition between the brook and the four ponds, we also observed differences among all the five small freshwater ecosystems (Figure 4.2.2, Supplementary Figures 4.B.4 and 4.B.5) despite their spatial proximity. These differences may be the result of a combination of diverse ecological processes (Goldford et al., 2018) that can explain the high OTU turnover rates observed (Supplementary Figure 4.B.8). Many microbes exhibit a high dispersion capacity promoted by natural vectors such as wind, rain, and animals (birds and mammals). However, the capacity to colonize a new environment faces diverse biotic and abiotic filters (e.g. competition, chemistry of the ecosystem, hydrology). Among these factors, competition between autochthonous and dispersing microbes is known to be crucial to limit successful immigration (Comte et al., 2017; Logue and Lindström, 2010). In this regard, we detected 62.7 and 39.6% of prokaryotic and eukaryotic OTUs, respectively, that occurred only once in all ecosystems over the two-year survey. In addition to their very

sporadic occurrence, the abundance of these OTUs was very low, with less than 15 reads on average. These OTUs likely represent dispersing cells that did not succeed to colonize our ecosystems and were no longer detected at subsequent dates. However, this was not the case for all rare OTUs, since some of them reached high abundances punctually. For instance, a rare OTU affiliated to Firmicutes accounted for 2 403 (30.2%) reads in La Claye in July 2012. We can hypothesize that these rare OTUs come from a local seed bank and may bloom when very particular environmental conditions occur. The fact that two of our ecosystems (La Claye and Ru Sainte Anne) experienced drought events that lasted 1 to 5 months and that, each time, microbial communities recovered their original structure within a month after water refill (Supplementary Figure 4.B.2) also strongly advocates for the existence of seed banks in these ecosystems for both the abundant and part of the rare OTUs (Shade et al., 2012; Simon et al., 2016).

#### 4.2.5.2 Prokaryotes and microbial eukaryotes share similar temporal patterns

Our multivariate statistical analyses showed that the microbial community composition in the five freshwater ecosystems was not only driven by the particular environmental conditions of each system but also by their temporal variations at different time scales (Table 4.2.1). Microbial communities can exhibit non-seasonal dynamics due to marked and fast shifts in environmental conditions. These shifts can easily take place in small freshwater ecosystems because their shallow depths and small volumes make them sensitive to local and short-time scale climate variations (Bock et al., 2018; Downing, 2010; Simon et al., 2015). In fact, even if rapid changes can occur in larger ecosystems such as lakes (Mangot et al., 2013), climate has usually a lower impact at short-time scales on lakes and oceans, where microbial seasonal successions are well documented (Fuhrman et al., 2015; Tammert et al., 2015). Nevertheless, despite punctual marked shifts in environmental conditions, we observed global seasonal patterns of both the environmental parameters (Figure 4.2.4), and the prokaryotic and eukaryotic microbial communities (Table 4.2.1 and Supplementary Table 4.A.3, Figures 4.2.4-4.2.5 and Supplementary Figure 4.B.7). These communities were more differentiated at summer and winter, possibly because of higher OTU turnover rates occurring at these two opposed seasons. Concomitantly, communities also experienced composition changes at a monthly scale, more pronounced for rare taxa than for abundant taxa (Supplementary Figure 4.B.8). Interestingly, at a much longer time scale, we also observed constant directional shift (increasing dissimilarity over time) of both the prokaryotic and eukaryotic communities, and the whole microbial community structures over the two-year survey (Figure 4.2.4, Supplementary Table 4.A.3), a phenomenon that has already been observed in larger lakes (Liu et al., 2015). As mentioned



above, diverse simultaneous processes affecting community assembly can explain this marked microbial dynamics, including immigration, the presence of a seed bank, and genetic or ecological drift (Chase, 2007; Evans et al., 2017; Logares et al., 2018). The temporal community composition changes observed in our ecosystems occurred in environmental conditions that were relatively stable at the inter-annual scale (Figure 4.2.4). This may explain partly why seasonal changes of both prokaryotes and microbial eukaryotes were slightly driven (<10% of variance explained) by a restricted subset of measured physico-chemical parameters (temperature, total concentration of ions (conductivity), dissolved oxygen and nitrate concentrations) (Figure 4.2.3 and Supplementary Figure 4.B.6). In addition to possible abiotic factors not measured during this study (e.g. light quantity and quality, wind, rainfall, specific hydrochemistry), biotic interactions were most likely also involved in determining the microbial community assembly as is assumed for other ecosystems (Lindström and Langenheder, 2012; Yannarell and Triplett, 2005). Many microbial eukaryotes prey on prokaryotes (Boenigk and Arndt, 2002; Pernthaler, 2005). Likewise, diverse symbioses (from mutualistic to parasitic ones) are established between microorganisms of the three domains of life (González-Olalla et al., 2018; Zehr et al., 2016) and influencing community composition. In return, the composition of prey and host communities may affect that of their respective predators and symbionts (Šimek et al., 2013). Time series with adapted time scale and sampling frequency are crucial to resolve these biotic interactions between microorganisms (Kelly et al., 2019; Moreira and López-García, 2019). Although our monthly sampling was likely too sparse to accurately assess potential synchrony between prokaryotes and microbial eukaryotes (that can often experience rapid changes on daily or weekly basis), we were able to identify an overall marked correlation between the changes in prokaryote community composition and those of microbial eukaryotes (Supplementary Table 4.A.4). This observation supports, beyond physicochemical change, the occurrence of top-down processes between heterotrophic eukaryotes and prokaryotes (Massana and Logares, 2013).

#### **4.2.5.3 Symbiosis may be a major process in small freshwater microbial ecosystems**

If predation is one of the main recognized biotic interactions at the base of trophic webs, symbioses (including interactions going from parasitism to mutualism) are widespread among microbial communities and can also play key roles (Harding et al., 2018; Hudson et al., 2006). We observed a large proportion of probable symbiotic taxa among the microbial eukaryotes and prokaryotes identified in our ecosystems. For example, based on the very limited biosynthetic capacities encoded in their genomes, CPR bacteria are inferred to live in symbiosis with other prokaryotes (Castelle and Banfield, 2018). Overall, in the five ecosystems, the most abun-

dant CPR phyla were Azambacteria (3.14%), Nomurabacteria (0.92%), Gracilibacteria (0.80%), Saccharibacteria (0.52%) and Peregrinibacteria (0.43%). Interestingly, CPR bacteria have been found mostly in oxygen-depleted, sulfide-rich environments (Castelle et al., 2018). Our sampled ecosystems are somewhat oxygen-depleted but display oxic conditions on the surface. We recorded the highest proportions (44.9% of reads, 46.3% of OTUs) in Ru Sainte Anne (47.2% [O<sub>2</sub>] on average) and low proportions (2.3% of reads, 24.3% of OTUS) in Mare Saint Robert (38.6% [O<sub>2</sub>] on average). We cannot rule out the hypothesis that CPR bacteria were more abundant in the sediments or in the suboxic deeper waters of our ecosystems, and that those that we detected were cells suspended again from the bottom to the surface since these ecosystems are characterized by frequent water column mixing. Nevertheless, we did retrieve in our samples very low numbers of sequences from anaerobic prokaryotes typical of sediments, such as methanogenic archaea (0.14% of all reads in all ecosystems) or sulfate-reducing Desulfobacteraceae (0.068%). This suggests a moderate input of cells from the sediment, so that the much larger amounts of CPR bacteria that we detected most likely correspond to bona fide members of the planktonic community and not cells suspended again from the sediment. Nonetheless, methanogenesis and sulfate-reduction are active local processes associated to the degradation of abundant organic matter (bubbling and strong H<sub>2</sub>S smell during sampling are observed in some systems, e.g. Mare Gabard, especially in warmer seasons). Therefore, CPR and their hosts might be more linked to sulfide availability than to the absence of dissolved oxygen per se. In addition, CPR bacteria were consistently retrieved from the five ecosystems even if their presumed size (Luef et al., 2015) is below the 0.2  $\mu\text{m}$  pore-size of the membranes that we used to collect cells. Therefore, their relative proportion might have been underestimated in our study, since many cells unattached to their hosts might have escaped detection. At any rate, the relatively high abundance of CPR bacteria in these small water bodies opens intriguing questions about their ecological role. Despite their likely dependence on other organisms, CPR bacteria exhibit a wide range of metabolic activities. For example, among the CPR lineages detected in this study, Microgenomates have the potential to degrade a large range of carbon substrates (Danczak et al., 2017), Parcubacteria can be involved in the nitrogen cycle through denitrification and Saccharibacteria (TM7) and Dojkabacteria (WS6) have been reported as parasites (Castelle and Banfield, 2018). The same situation might apply to the potentially parasitic Woesearchaeota (Liu et al., 2018), which reach relative high abundance in some of our shallow freshwater ecosystems. The importance of parasitism in aquatic plankton communities is well known (Hudson et al., 2006; Jephcott et al., 2016; Lafferty et al., 2006). In particular, since parasites have complex life cycles and are known to infect organisms at various trophic levels, they increase the complexity of planktonic trophic networks by extending the food chains (Grami et al., 2011) and enhancing the level of biodiversity (Hudson et al., 2006). This, in turn, may result in the increase of both the connectivity and efficiency of carbon fluxes, although

these potential consequences are still subject to debate (Grami et al., 2011).

In summary, our study revealed a high diversity of prokaryotes in addition to that previously observed for microbial eukaryotes in small freshwater ecosystems. Despite a microbial community structure characterized by a predominance of rare taxa and a marked OTU turnover rate, we observed marked seasonal patterns and community drift over the two-year survey for both prokaryotes and microbial eukaryotes of five different microbial communities. These observations suggest that diverse prokaryote and eukaryote communities react similarly to environmental changes over time (seasonal to multiple-year scale). These changes were only partially driven by measured environmental conditions. By contrast, biotic interactions, such as symbiosis, are suspected to play more significant roles. The abundance of CPR bacteria in some of these ecosystems suggests that they may be important ecological players also in oxic environments.

# Supplementary Material

## 4.A Supplementary Tables

**Supplementary Table 4.A.1** Names, numbers of reads, OTUs and diversity estimates obtained for the different samples

Ecosystem	Sample name	Sampling date	Number of removed singletons	Number of reads	Observed number of OTUs	Richness (Standard error)	Evenness
Etang des Vallées	EV11	April 2011	1089	19801	296	251.77 (5.37)	0.49
Etang des Vallées	EV13	May 2011	765	9156	160	157.85 (1.43)	0.48
Etang des Vallées	EV15	June 2011	765	9414	193	189.28 (1.87)	0.49
Etang des Vallées	EV17	July 2011	1320	11258	325	309.01 (3.74)	0.49
Etang des Vallées	EV19	September 2011	1178	10701	280	270.90 (2.85)	0.50
Etang des Vallées	EV21	October 2011	659	11774	182	173.42 (2.72)	0.45
Etang des Vallées	EV23	November 2011	1078	11326	248	234.77 (3.39)	0.47
Etang des Vallées	EV25	December 2011	1550	14347	347	305.75 (5.63)	0.45
Etang des Vallées	EV27	January 2012	1231	11606	234	220.74 (3.37)	0.48
Etang des Vallées	EV29	February 2012	788	8043	234	232.58 (1.18)	0.48
Etang des Vallées	EV31	March 2012	2286	25281	650	440.67 (10.59)	0.42
Etang des Vallées	EV33	April 2012	1420	22989	260	201.78 (5.85)	0.43
Etang des Vallées	EV35	May 2012	1043	21075	195	157.27 (4.88)	0.41
Etang des Vallées	EV37	June 2012	686	13426	186	171.72 (3.38)	0.48
Etang des Vallées	EV39	July 2012	780	13351	244	225.67 (3.86)	0.47
Etang des Vallées	EV41	August 2012	1181	22879	280	217.23 (6.09)	0.44
Etang des Vallées	EV43	September 2012	1122	12838	270	254.32 (3.62)	0.48
Etang des Vallées	EV45	October 2012	476	7106	137	137.00 (0.00)	0.50
Etang des Vallées	EV47	November 2012	626	10432	160	156.23 (1.85)	0.49
Etang des Vallées	EV49	December 2012	1069	11220	259	245.62 (3.42)	0.47
Etang des Vallées	EV51	January 2013	1444	16834	311	258.81 (6.04)	0.40
Etang des Vallées	EV53	February 2013	1435	8385	326	321.55 (2.08)	0.47
Etang des Vallées	EV55	March 2013	879	7882	166	165.41 (0.76)	0.47
Etang des Vallées	EV57	April 2013	735	8256	157	155.97 (1.01)	0.46
La Claye	LC3	April 2011	679	9537	220	160.84 (5.97)	0.44
La Claye	LC5	May 2011	1101	10347	256	177.39 (6.65)	0.42
La Claye	LC7	June 2011	729	8756	141	110.21 (4.48)	0.43
La Claye	LC19	January 2012	1721	5891	490	421.89 (7.23)	0.46
La Claye	LC21	February 2012	897	3492	248	246.49 (1.56)	0.48
La Claye	LC23	March 2012	1988	11862	516	326.85 (9.72)	0.41
La Claye	LC25	April 2012	1352	13723	272	163.45 (7.06)	0.41
La Claye	LC27	May 2012	1085	13897	203	121.26 (6.16)	0.39
La Claye	LC29	June 2012	660	8701	162	127.74 (4.78)	0.41
La Claye	LC31	July 2012	791	11284	154	106.85 (5.11)	0.42
La Claye	LC33	August 2012	764	11253	133	90.54 (4.80)	0.42
La Claye	LC37	October 2012	1633	2832	555	556.00 (0.00)	0.56
La Claye	LC39	November 2012	1530	12505	435	259.39 (9.35)	0.36

*Continued on next page*

Supplementary Table 4.A.1 – *Continued from previous page*

Ecosystem	Sample name	Sampling date	Number of removed singletons	Number of reads	Observed number of OTUs	Richness (Standard error)	Evenness Evenness
La Claye	LC41	December 2012	1884	15124	537	295.71 (10.17)	0.39
La Claye	LC43	January 2013	1545	9379	366	262.08 (7.84)	0.44
La Claye	LC45	February 2013	975	7887	200	154.73 (5.51)	0.45
La Claye	LC47	March 2013	740	9056	178	137.30 (5.06)	0.45
La Claye	LC49	April 2013	963	19006	164	92.92 (5.19)	0.43
Mare Gabard	MG3	April 2011	3475	4190	63	49.03 (2.92)	0.42
Mare Gabard	MG5	May 2011	3496	1204	90	91.00 (0.00)	0.56
Mare Gabard	MG7	June 2011	774	4952	148	104.21 (4.76)	0.49
Mare Gabard	MG9	July 2011	956	7409	151	91.26 (4.83)	0.47
Mare Gabard	MG11	September 2011	943	4692	236	165.47 (6.13)	0.49
Mare Gabard	MG13	October 2011	1060	10752	198	102.56 (5.47)	0.45
Mare Gabard	MG15	November 2011	453	10451	77	33.30 (3.62)	0.36
Mare Gabard	MG17	December 2011	1298	9055	350	160.49 (7.95)	0.42
Mare Gabard	MG19	January 2012	1640	11962	313	109.41 (7.29)	0.38
Mare Gabard	MG21	February 2012	1163	8045	426	194.05 (8.95)	0.40
Mare Gabard	MG23	March 2012	823	12218	127	50.15 (4.45)	0.38
Mare Gabard	MG25	April 2012	1446	18702	469	154.83 (8.18)	0.40
Mare Gabard	MG27	May 2012	1637	17200	279	90.37 (6.30)	0.37
Mare Gabard	MG29	June 2012	1012	11175	157	65.47 (5.07)	0.34
Mare Gabard	MG31	July 2012	624	5201	94	61.96 (4.02)	0.46
Mare Gabard	MG33	August 2012	943	9711	195	92.13 (5.93)	0.35
Mare Gabard	MG35	September 2012	1114	9944	198	87.69 (5.81)	0.39
Mare Gabard	MG37	October 2012	668	7931	172	91.33 (5.47)	0.41
Mare Gabard	MG41	December 2012	414	5493	90	64.63 (3.59)	0.50
Saint Robert	SR3	April 2011	3849	7439	256	166.83 (6.67)	0.45
Saint Robert	SR5	May 2011	1308	15499	248	115.62 (6.40)	0.42
Saint Robert	SR7	June 2011	1361	13413	274	131.13 (6.90)	0.42
Saint Robert	SR9	July 2011	1298	10148	235	133.98 (6.51)	0.42
Saint Robert	SR11	September 2011	1400	11442	232	137.55 (6.01)	0.47
Saint Robert	SR13	October 2011	2033	1882	91	92.00 (0.00)	0.52
Saint Robert	SR15	November 2011	706	6649	147	106.98 (4.78)	0.36
Saint Robert	SR17	December 2011	1133	4246	256	209.09 (5.78)	0.48
Saint Robert	SR19	January 2012	756	6801	140	98.38 (4.80)	0.42
Saint Robert	SR21	February 2012	953	11088	295	163.03 (7.34)	0.41
Saint Robert	SR23	March 2012	823	12081	241	132.97 (6.36)	0.43
Saint Robert	SR25	April 2012	1500	11675	316	165.85 (7.44)	0.45
Saint Robert	SR27	May 2012	1312	15174	316	156.22 (7.29)	0.40
Saint Robert	SR29	June 2012	1030	17177	242	111.95 (6.25)	0.40
Saint Robert	SR31	July 2012	580	6355	127	90.55 (4.58)	0.45

*Continued on next page*

Supplementary Table 4.A.1 – *Continued from previous page*

Ecosystem	Sample name	Sampling date	Number of removed singletons	Number of reads	Observed number of OTUs	Richness (Standard error)	Evenness Evenness
Saint Robert	SR33	August 2012	1385	11398	339	196.84 (7.45)	0.47
Saint Robert	SR35	September 2012	1127	10454	281	165.08 (6.94)	0.42
Saint Robert	SR37	October 2012	845	9394	199	125.80 (5.76)	0.45
Saint Robert	SR41	December 2012	620	4439	129	109.75 (3.79)	0.49
Saint Robert	SR43	January 2013	890	10916	129	83.06 (4.41)	0.43
Saint Robert	SR45	February 2013	827	6663	123	87.02 (4.51)	0.45
Saint Robert	SR47	March 2013	330	5054	69	59.76 (2.60)	0.48
Saint Robert	SR49	April 2013	576	6264	92	68.55 (3.71)	0.47
Ru Sainte Anne	RSA3	April 2011	3779	5033	938	885.99 (6.75)	0.54
Ru Sainte Anne	RSA5	May 2011	4011	6273	1058	938.34 (9.61)	0.54
Ru Sainte Anne	RSA7	June 2011	3703	10396	1328	992.68 (13.99)	0.53
Ru Sainte Anne	RSA9	July 2011	2595	6723	721	637.14 (8.08)	0.53
Ru Sainte Anne	RSA11	September 2011	3214	6606	898	789.84 (9.12)	0.53
Ru Sainte Anne	RSA13	October 2011	2524	11242	663	485.17 (10.24)	0.46
Ru Sainte Anne	RSA15	November 2011	1841	3662	627	627.01 (0.99)	0.51
Ru Sainte Anne	RSA17	December 2011	3726	5612	975	898.04 (7.99)	0.54
Ru Sainte Anne	RSA19	January 2012	3925	8822	1182	920.39 (12.89)	0.52
Ru Sainte Anne	RSA21	February 2012	2917	7219	1105	940.35 (10.89)	0.53
Ru Sainte Anne	RSA23	March 2012	3187	12007	1412	981.92 (15.11)	0.51
Ru Sainte Anne	RSA25	April 2012	3796	8953	1169	908.65 (12.84)	0.51
Ru Sainte Anne	RSA27	May 2012	4630	14919	1597	998.99 (16.59)	0.51
Ru Sainte Anne	RSA29	June 2012	4314	12101	1779	1 195.60 (17.18)	0.53
Ru Sainte Anne	RSA31	July 2012	1864	3464	585	586.00 (0.00)	0.54
Ru Sainte Anne	RSA37	October 2012	2870	7429	1065	905.79 (10.65)	0.55
Ru Sainte Anne	RSA39	November 2012	3286	5746	979	894.28 (8.32)	0.54
Ru Sainte Anne	RSA41	December 2012	5472	14362	2187	1 352.19 (19.29)	0.53
Ru Sainte Anne	RSA43	January 2013	4908	11201	1742	1 205.56 (16.83)	0.53
Ru Sainte Anne	RSA45	February 2013	5247	11468	1694	1 159.55 (16.69)	0.52
Ru Sainte Anne	RSA47	March 2013	3360	6195	1018	903.55 (9.44)	0.53
Ru Sainte Anne	RSA49	April 2013	4987	14204	1906	1 188.11 (18.14)	0.52

**Supplementary Table 4.A.2** Number of abundant and all OTUs affiliated to prokaryotes and microbial eukaryotes in the five ecosystems and in the five ecosystems considered together. Abundant OTUs were detected in each ecosystem in at least 25% of samples, and with <0.1% mean relative abundance per sample.

	Prokaryotes		Microbial eukaryotes	
	Abundant	All	Abundant	All
Etang des Vallées	103	2463	108	886
La Claye	57	2938	65	1063
Mare Gabard	51	2074	36	530
Saint Robert	75	2326	110	1946
Ru Sainte Anne	133	11347	52	648
All ecosystems	271	16573	300	3583



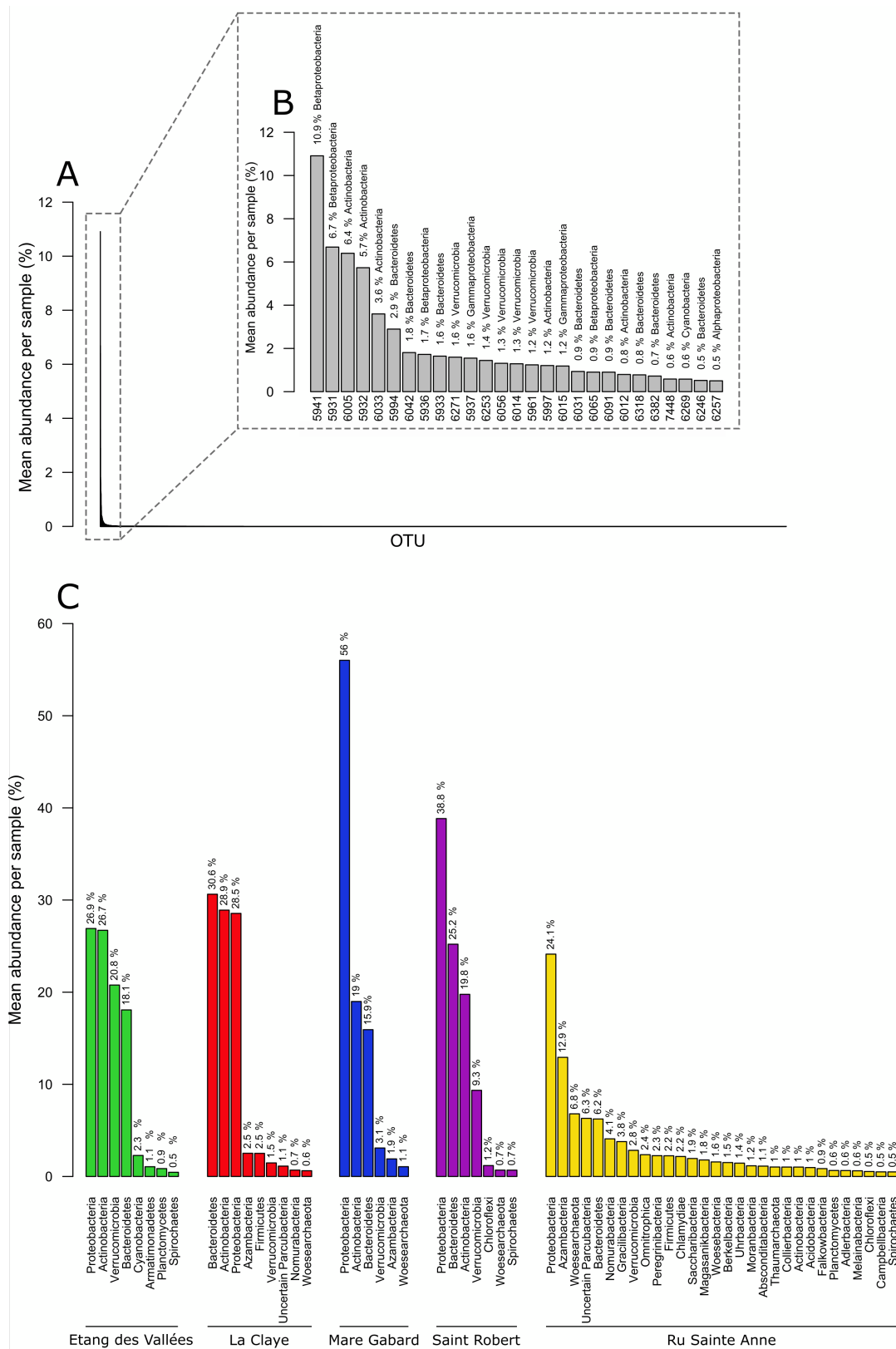
**Supplementary Table 4.A.3** Linear and polynomial models were computed based on time-lag analyses (TLA) between sampling dates over increasing time-lags over the 2-years survey (from one-month to 23-months) for the five ecosystems. Wisconsin standardized Bray-Curtis dissimilarities were used for all OTUs (separated into prokaryotes, microbial eukaryotes and whole microbial communities) while Euclidean distances were applied with environmental parameters. Non-significant P-values are highlighted in red.

			Prokaryotes	Microbial eukaryotes	Whole microbial communities	Physico-chemical parameters
Etang des Vallées	Linear model	Slope	0.10	0.22	0.16	0.02
		R <sup>2</sup>	0.02	0.05	0.05	0.00
		Pvalue	2.6 x 10 <sup>-4</sup>	2.2 x 10 <sup>-16</sup>	1.1 x 10 <sup>-7</sup>	0.13
	Polynomial model	R <sup>2</sup>	0.23	0.41	0.41	0.14
		Pvalue	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>
La Claye	Linear model	Slope	0.09	0.18	0.15	-0.01
		R <sup>2</sup>	0.02	0.05	0.05	0.00
		Pvalue	4.6 x 10 <sup>-4</sup>	2.2 x 10 <sup>-5</sup>	2.7 x 10 <sup>-5</sup>	0.45
	Polynomial model	R <sup>2</sup>	0.32	0.25	0.41	0.09
		Pvalue	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>	1.7 x 10 <sup>-6</sup>
Mare Gabard	Linear model	Slope	0.19	0.21	0.18	0.03
		R <sup>2</sup>	0.07	0.07	0.08	0.01
		Pvalue	2.2 x 10 <sup>-7</sup>	1.6 x 10 <sup>-7</sup>	3.8 x 10 <sup>-7</sup>	0.01
	Polynomial model	R <sup>2</sup>	0.23	0.34	0.38	0.05
		Pvalue	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>	7.8 x 10 <sup>-6</sup>
Saint Robert	Linear model	Slope	0.15	0.23	0.19	0
		R <sup>2</sup>	0.09	0.13	0.15	0
		Pvalue	2.2 x 10 <sup>-12</sup>	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>	0.77
	Polynomial model	R <sup>2</sup>	0.23	0.33	0.44	0.04
		Pvalue	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>	1.0 x 10 <sup>-4</sup>
Ru Sainte Anne	Linear model	Slope	0.17	0.19	0.18	0.04
		R <sup>2</sup>	0.07	0.06	0.09	0.03
		Pvalue	1.8 x 10 <sup>-9</sup>	7.2 x 10 <sup>-8</sup>	4.7 x 10 <sup>-10</sup>	1.5 x 10 <sup>-4</sup>
	Polynomial model	R <sup>2</sup>	0.17	0.34	0.32	0.14
		Pvalue	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>	2.2 x 10 <sup>-16</sup>	9.9 x 10 <sup>-16</sup>

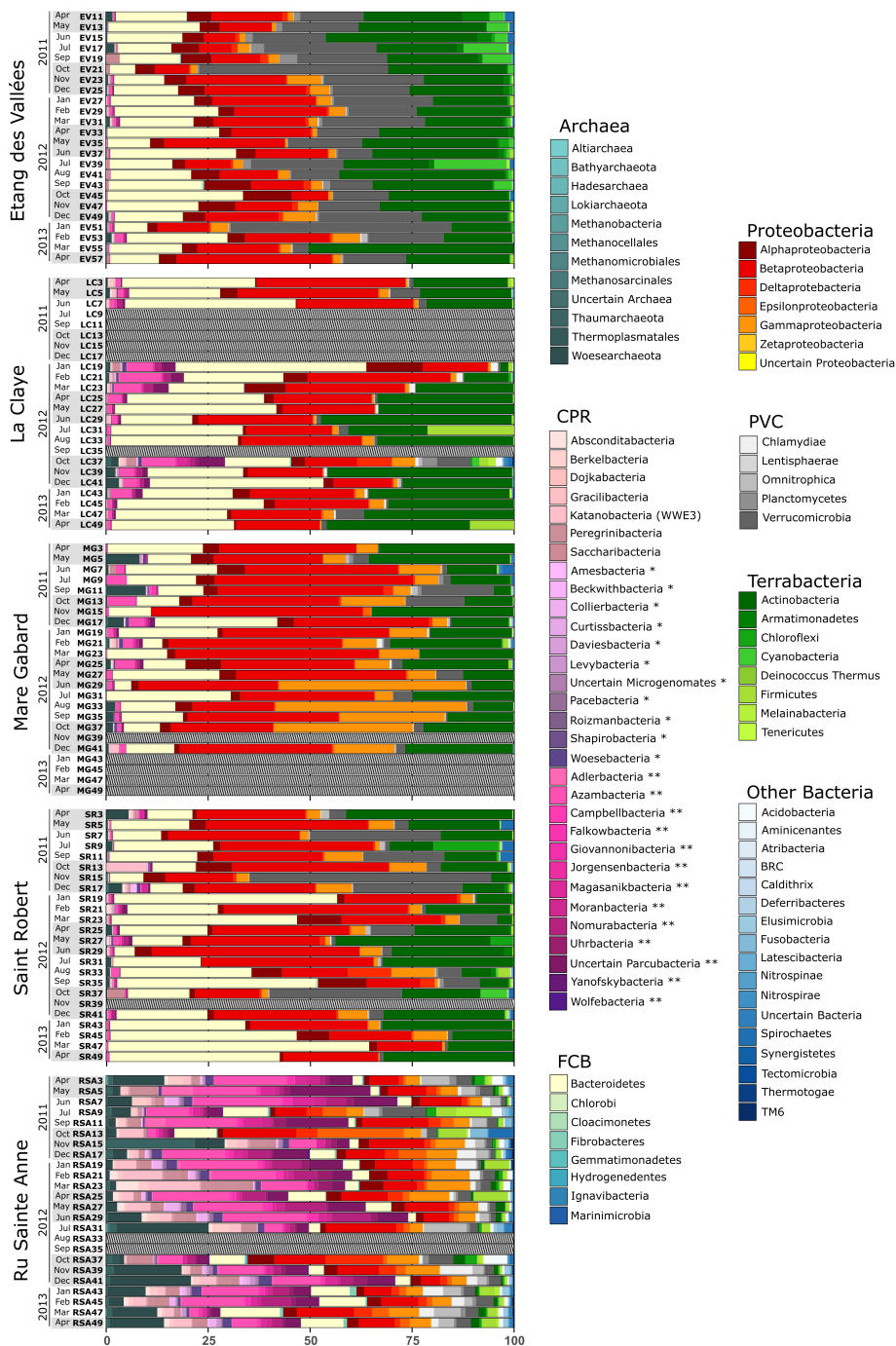
**Supplementary Table 4.A.4** RV index were calculated based on co-inertia analyses to assess co-dynamic between prokaryotes and microbial eukaryotes in the five ecosystems and in all ecosystems considered together. Values in red indicate P-values higher than 0.05, corresponding to non-significant RV index.

	RV index	Pvalue
Etang des Vallées	0.896	0.015
La Claye	0.726	0.487
Mare Gabard	0.730	0.194
Saint Robert	0.880	0.004
Ru Sainte Anne	0.855	0.240
All ecosystems	0.745	0.001

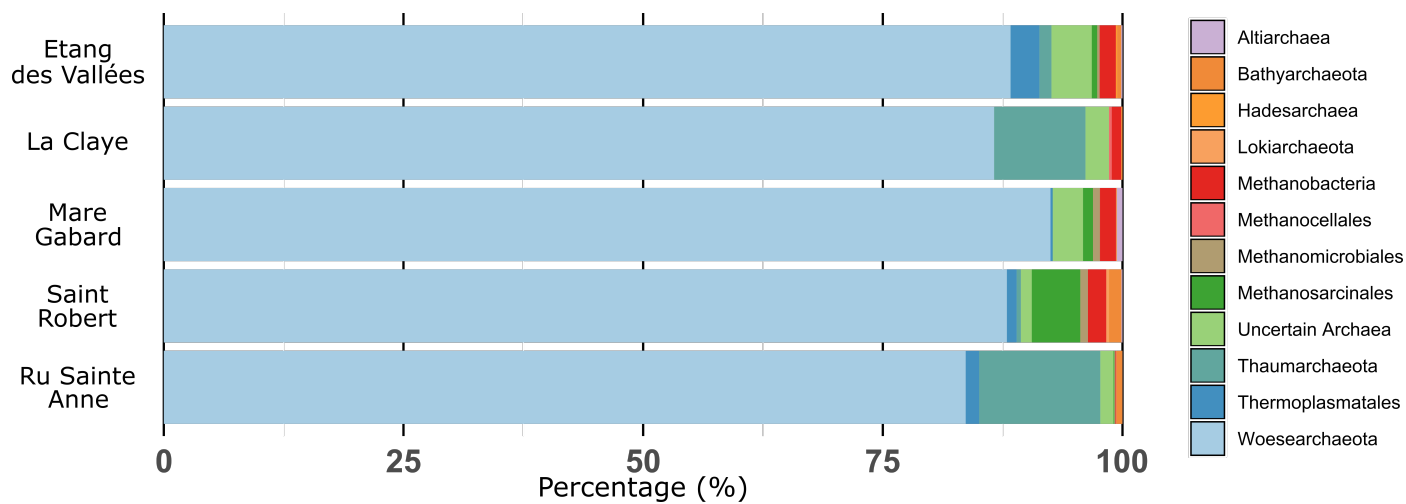
## 4.B Supplementary Figures



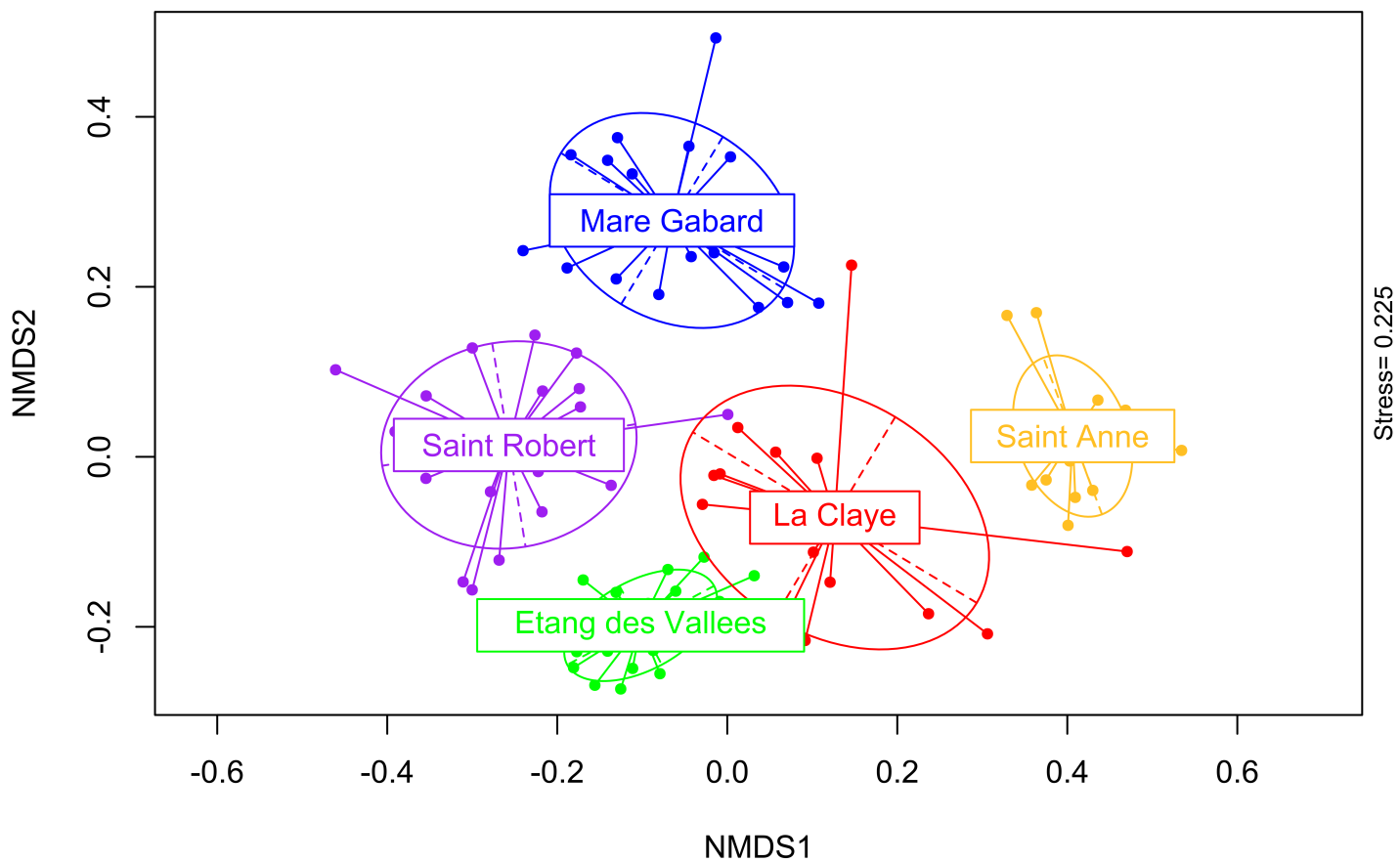
**Supplementary Figure 4.B.1** Rank-abundance histograms of all OTUs detected in the five ecosystems (A), OTUs with a mean relative abundance of at least 0.5% (B) and rank-abundance histograms of taxonomic groups in each ecosystem (Etang des Vallées (EV), La Claye (LC), Mare Gabard (MG), Saint Robert (SR) and Ru Sainte Anne (RSA)) (C). Taxonomic groups and OTUs are sorted based on their mean percentage of reads that are indicated on the top of the bars. In B, the identification number of the OTUs and their taxonomic affiliation are shown below and on the top of the bar, respectively.



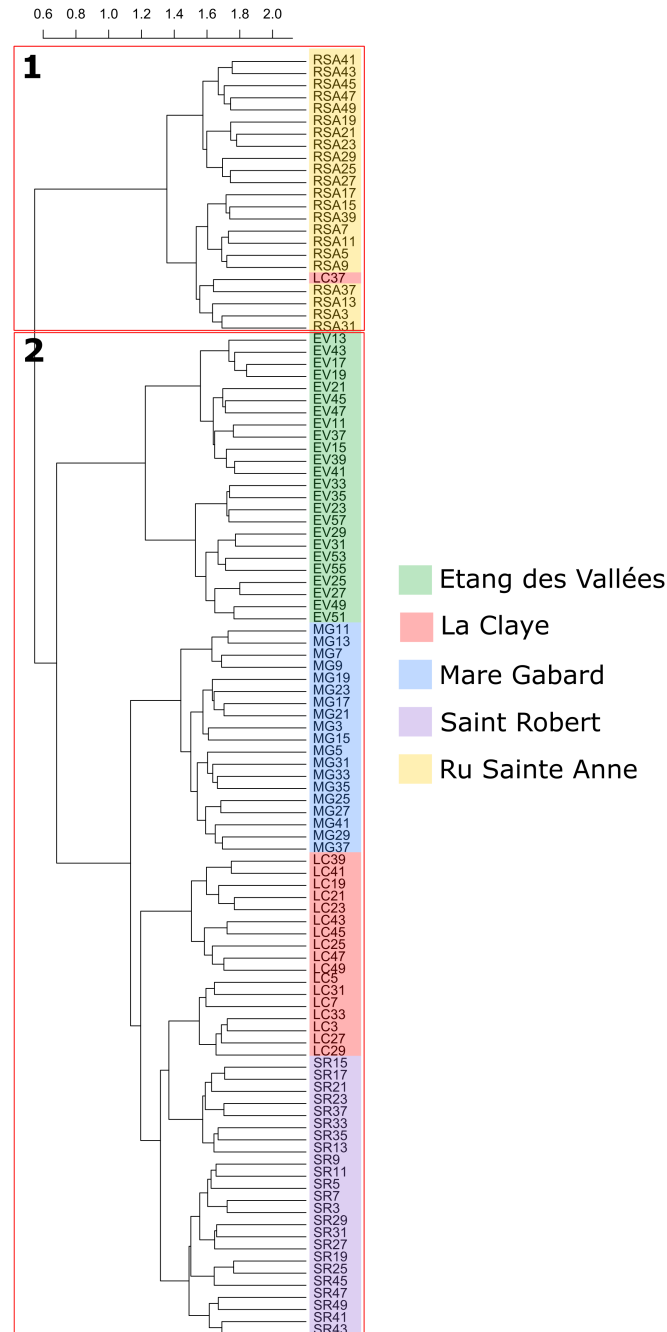
**Supplementary Figure 4.B.2** Histograms of monthly relative proportions of 16S rDNA amplicon reads assigned to high-rank taxa (CPR: Candidate Phyla Radiation; FCB: Fibrobacteres-Chlorobi-Bacteroidetes; PVC: Planctomycetes-Verrucomicrobia-Chlamydiae), from the five ecosystems over the 2-years survey. Hatched bars correspond to missing data because of the drought periods, putative cross-contamination and missing samples: Ru Sainte Anne was dried in August and September 2012 and La Claye from the end of July to December 2011 and in September 2012 while we suspected contaminations in samples in Mare Gabard from January to April 2013 and samples collected in November 2012 in Mare Gabard and in November 2012 in Saint Robert were missing. One star (\*) indicates Microgenomates and two stars (\*\*) shows Parcubacteria.



**Supplementary Figure 4.B.3** Histograms of average relative proportions of 16S rDNA amplicon reads assigned to archaeal high-rank taxa, from the five ecosystems over the 2-years survey.

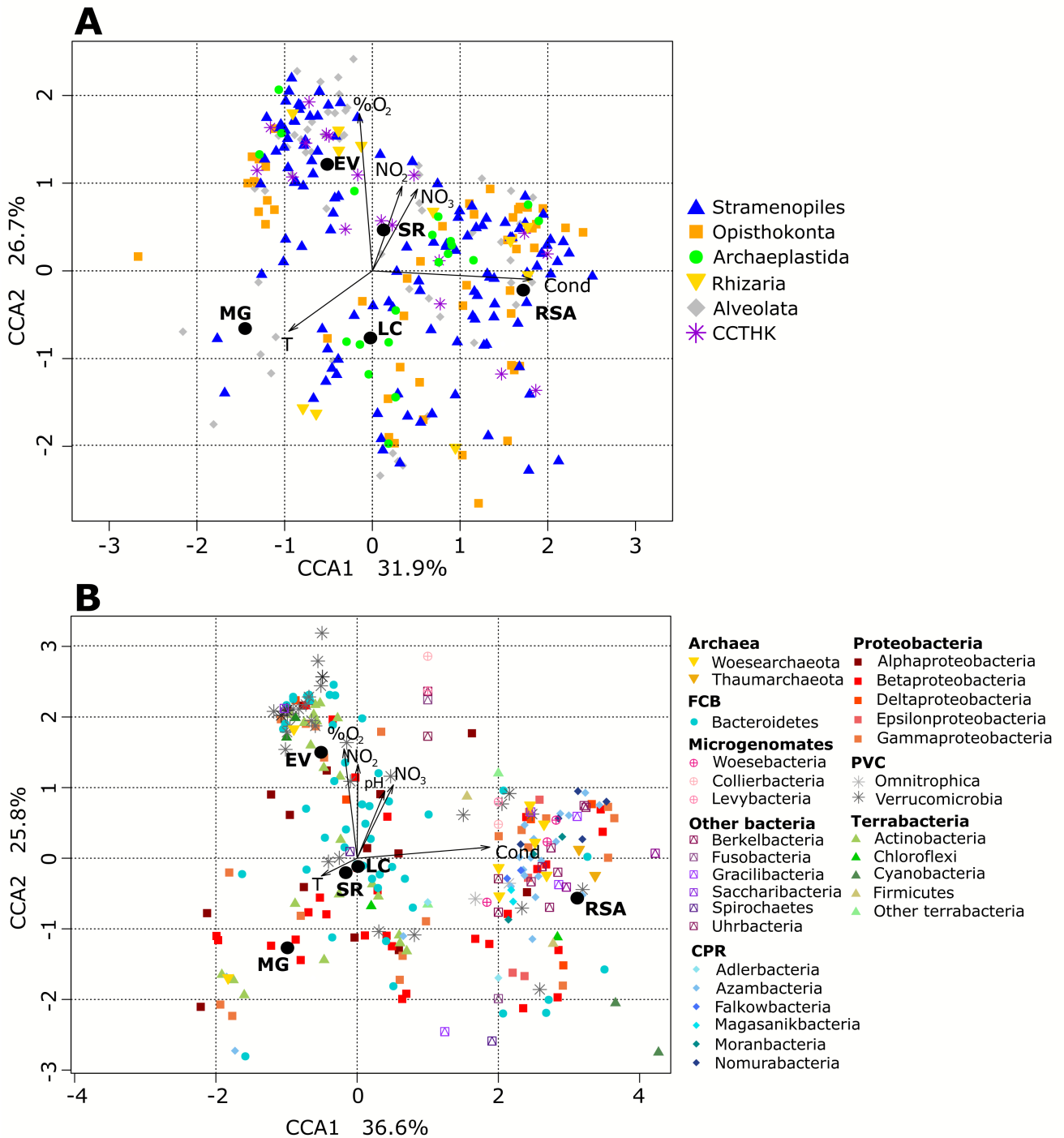


**Supplementary Figure 4.B.4** Non-metric multidimensional scaling (NMDS) plot, built on Wisconsin standardized Bray–Curtis dissimilarities between combined microbial eukaryote and prokaryote OTUs of all samples collected in the five ecosystems over the 2-years survey. Stress value is displayed on the right border of the plot.

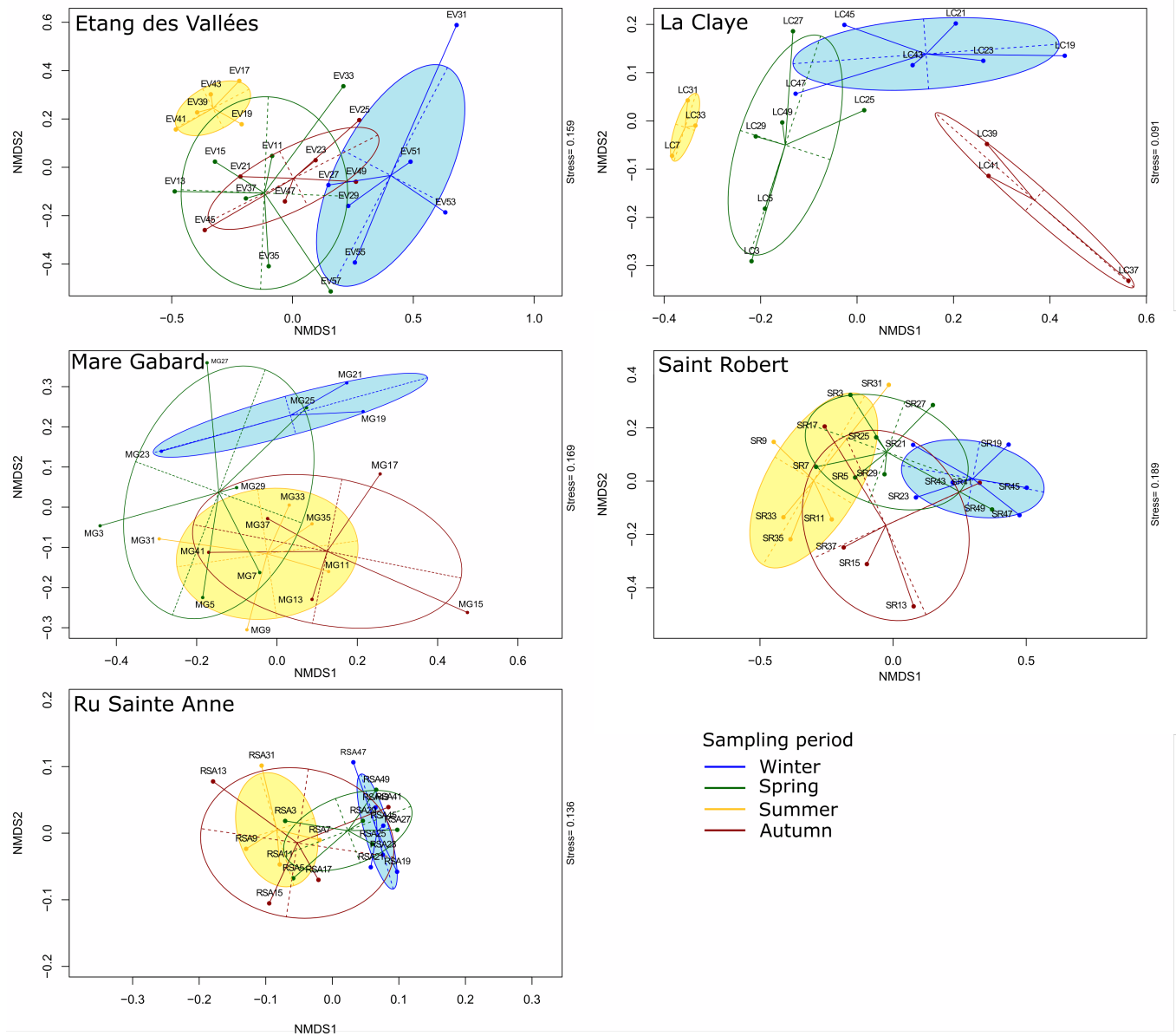


**Supplementary Figure 4.B.5** Hierarchical cluster analysis build using Wisconsin standardized Bray-Curtis dissimilarities between every samples collected over the 2-years survey for the five ecosystems (Etang des Vallées (EV), La Claye (LC), Mare Gabard (MG), Saint Robert (SR) and Ru Sainte Anne (RSA)). The two main clusters are differentiated by red lines and numbered 1 (Ru Sainte Anne alone) and 2 (comprising the four other systems).

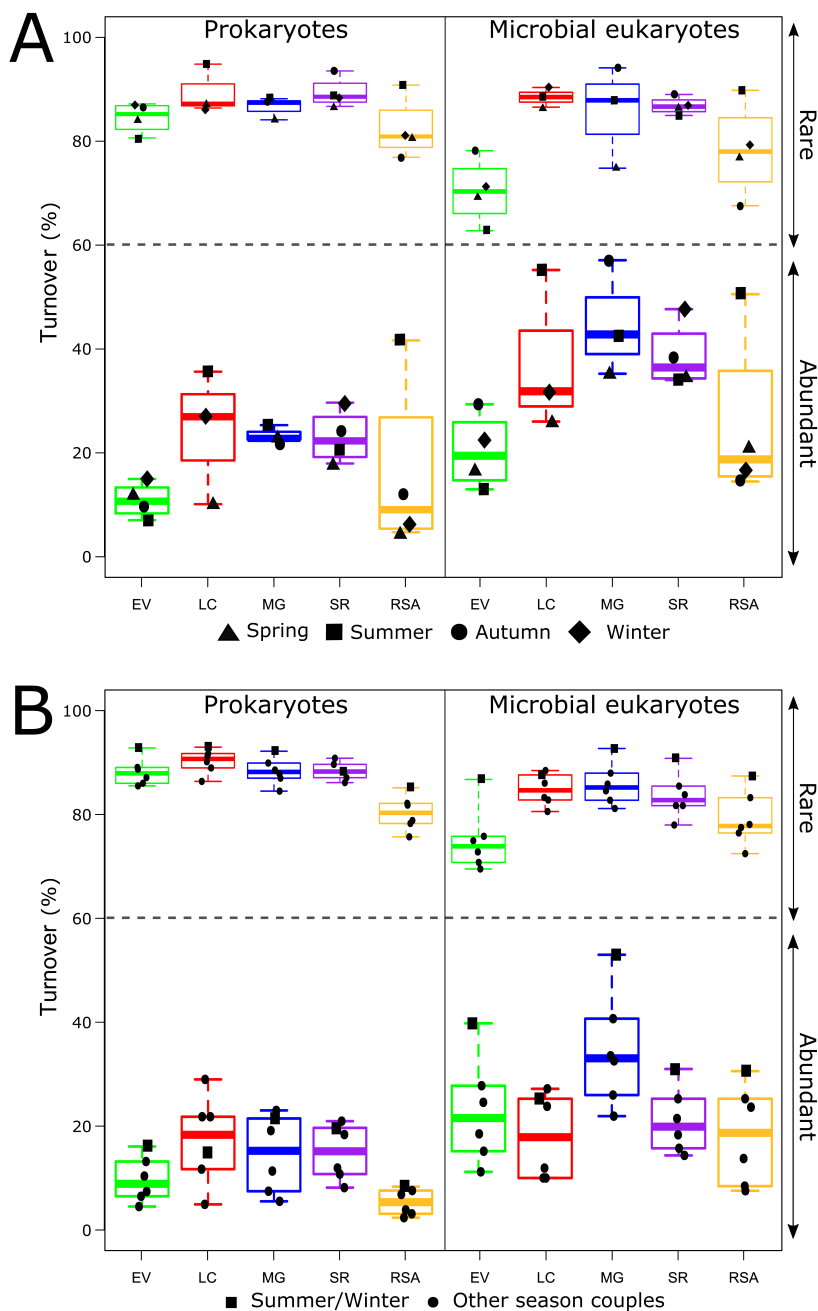




**Supplementary Figure 4.B.6** Canonical correspondence analysis (CCA) conducted on abundant microbial eukaryote (A) and prokaryote (B) Wisconsin-standardized OTUs retrieved in the five systems over the 2-years survey (CCTHK: Cryptophyta, Centrohelioczoa, Telonema, Haptophyta, Katablepharida; CPR: Candidate Phyla Radiation; FCB: Fibrobacteres-Chlorobi-Bacteroidetes; PVC: Planctomycetes-Verrucomicrobia-Chlamydiae). Abundant OTUs were detected in each ecosystem in at least 25% of samples, and with <0.1% mean relative abundance per sample. Forward selection was used to sort environmental parameters explaining significantly community composition variation for CCA model. Dots represent OTUs, with color and form indicating their taxonomic affiliations. Cond, conductivity; T, temperature; pH, pH; %O<sub>2</sub>, concentration of dissolved oxygen in%; NO<sub>2</sub>, nitrite concentration; NO<sub>3</sub>, nitrate concentration.



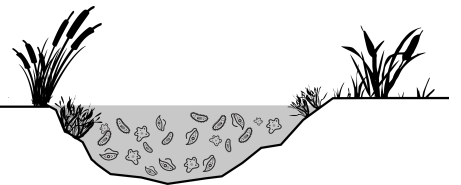
**Supplementary Figure 4.B.7** Non-metric MultiDimensional Scaling (NMDS) plots, based on Bray-Curtis distances computed after Wisconsin standardization of all OTU frequencies in samples from the five ecosystems. Samples are indicated in grey, and grouped by season. Stress value of each NMDS is displayed on the right border of the corresponding plot.



**Supplementary Figure 4.B.8** Box plots of turnover rates computed for rare and abundant OTUs affiliated to microbial eukaryotes and prokaryotes in the five ecosystems (Etang des Vallées (EV), La Claye (LC), Mare Gabard (MG), Saint Robert (SR) and Ru Sainte Anne (RSA)) over the 2-years survey. Abundant OTUs were detected in each ecosystem in at least 25% of samples, and with  $>0.1\%$  mean relative abundance per sample while rare OTUs were characterised by the opposite patterns. The thickest line inside each box represents the median on the distribution, bottom and top borders of boxes correspond to the first and third quartiles and whiskers extend to minimal and maximal distances. Turnover were calculated between the samples collected at the same season one year apart (A) or between seasons (B). Dots represent turnover values with forms indicating seasons.

## Chapter 5

Dissimilar protist and prokaryote communities display similar intra- and inter-annual temporal patterns in small freshwater ecosystems





## 5.1 Context and objective

Following the monthly survey conducted during two years, a seasonal survey was conducted in the same five small freshwater ecosystems. Long term surveys are essential to describe community composition and temporal dynamics. Indeed, microbial communities and ecosystems are changing in time, and are subject to perturbations. Therefore, long surveys allow to catch more variability than studies lacking of temporal dimension. As the detection of a consistent seasonal pattern requires more than two years of observations, it was necessary to continue the sampling (Fuhrman et al., 2015). It is even more important in our small freshwater ecosystems, because they respond rapidly to environmental perturbation. Indeed, their small volume of water reduces their buffering capacity. Therefore, a long survey is needed to cover a wide range of potential environmental conditions. We performed the seasonal sampling for eight years, which represents about 30 samples for each ecosystem. The data were obtained using Illumina MiSeq sequencing of 16S and 18S rDNA amplicons. In this new approach, I targeted a larger fraction of the microbial communities (*i.e.* 0.22 to 30  $\mu\text{m}$  instead of 0.22 to 5  $\mu\text{m}$ ), to consider pico- to micro-planktonic microorganisms. In this study, my aims were to:

1. Extend our description of the microbial composition of the microorganisms of the three domains of life, as we did not reach a saturation plateau of diversity in the previous studies.
2. Confirm or infirm the temporal patterns observed at the two-year scale with the data of the eight-year survey.
3. Investigate specific temporal dynamics to detect OTUs with clear recurrent patterns.
4. Infer a global interaction network to study the community structure shared between the ecosystems.

## 5.2 Manuscript

### 5.2.1 Abstract

Small and shallow freshwater ecosystems are widely distributed and known to be responsible for significant greenhouse gases emission and adsorption ( $\text{CO}_2$  and  $\text{CH}_4$ ). They are sensitive to any environmental change because of their small volume that induces a low buffering capacity. However, the knowledge of the diversity, ecology, and roles of microbial communities inhabiting these systems are still scarce. Hence, the functioning of small freshwater ecosystems is poorly understood. In this study, we used Illumina MiSeq sequencing data of 16S and 18S rDNA amplicons from five small freshwater ecosystems located in northwestern France sampled during eight years (2011-2019) at a seasonal frequency. As observed in a previous study, small freshwater ecosystems harbored very diverse microbial communities (respectively 22 764 and 25 536 operational taxonomic units -OTUs- at 95% identity). Moreover, they were characterised by specific microbial communities despite their geographic proximity. Microbial eukaryotes and prokaryotes communities exhibited seasonal patterns featuring a marked seasonality (intra-annual) and an increase of dissimilarities between communities over the years (inter-annual). A seasonal recurrence was detected at the OTU level for less than 2% of the microbial community. In each ecosystem, the measured physico-chemical parameters had only a low influence on microbial community variance. This suggests that biotic factors, such as predation, mutualisms, or parasitism play a major role in community composition. The global structure over the five ecosystems indicated that the most abundant OTUs were grouped in the network following their taxonomic affiliation, suggesting that they share similar ecological niches.

### 5.2.2 Introduction

Aquatic ecosystems have a major importance in climate regulation and in carbon cycle, participating actively to its sequestration (Williamson et al., 2008; Riebesell et al., 2008; Cavicchioli et al., 2019). Microorganisms (microbial eukaryotes, bacteria and archaea) are known to play essential roles in all aquatic ecosystems, as key players in the microbial loop (Pomeroy et al., 2007). Microbial eukaryotes and prokaryotes act at different levels: phytoplankton performs photosynthesis, bacteria and archaea are highly involved in the degradation of dissolved and particulate organic matter, heterotrophic and mixotrophic protists are active predators of prokaryotes (Sherr and Sherr, 2002; Kritzberg et al., 2006; Zubkov and Tarran, 2008; Falkowski

et al., 2008). Prokaryotes are also highly involved in all biogeochemical cycles (Falkowski et al., 2008; Jetten, 2008; Worden et al., 2015). Prokaryotes and microbial eukaryotes feature diverse biological traits (Massana and Logares, 2013). For example, microbial eukaryotes are able to phagocytose other cells, an almost exclusive characteristic to this domain of life (Shiratori et al., 2019). On the other side, prokaryotes present higher growth rates, phylogenetic and metabolic diversities.

Microbes are at the base of trophic webs and are thus major actors in ecosystem functioning, in both marine and freshwater systems. However, the vast majority of microbial ecology studies have mainly been conducted in large ecosystems, while small ecosystems were barely studied because of their supposed little importance at the global scale (Mullins and Doyle, 2019). Freshwater ponds (surface area  $<0.001\text{km}^2$ ), are shallow, widespread on continents, and be numerically the most abundant lentic freshwater bodies (Downing et al., 2006; Downing, 2010; Bartout et al., 2015; Biggs et al., 2017). These ecosystems are very sensitive and are often facing drought periods (Simon et al., 2015; Biggs et al., 2017). Because of their small volumes, they present low buffering capacities towards disturbances. They are also actively involved in biogeochemical cycles, such as the carbon cycle through  $\text{CO}_2$  fixation and  $\text{CH}_4$  emissions (with contributions up to 15% and 40% respectively) (Céréghino et al., 2014; Holgerson and Raymond, 2016; Yvon-Durocher et al., 2017). However, despite their ecological importance, there are still major gaps in our knowledge on the functioning of these ecosystems.

Understanding the functioning of an aquatic ecosystem of any kind, marine or freshwater, can be addressed through spatial or temporal surveys (Gonzalez et al., 2012; Widder et al., 2016). Several spatial surveys have been conducted in marine systems and lakes (Kirkham et al., 2011; Lepère et al., 2013; Sunagawa et al., 2015), but long temporal surveys are scarce and often restricted to coastal marine systems. The majority of the temporal surveys are less than one year long (Yannarell and Triplett, 2005; Gilbert et al., 2009) but few studies are going from several years to decade-long surveys (Fuhrman et al., 2006; Chow et al., 2013; Linz et al., 2017). Because of the highly variable environmental conditions among small freshwater ecosystems and the little characterisation of these systems there is a need to conduct spatiotemporal surveys to better appreciate their functioning (Pätzig et al., 2020). Long term temporal studies focusing on both microbial eukaryotes and prokaryotes are rarely conducted in marine systems (Chénard et al., 2019; Lambert et al., 2019) and freshwaters (Su et al., 2017; Bock et al., 2018; Obertegger et al., 2019). Temporal studies are crucial to decipher structuring patterns of microbial communities. Indeed, surveys conducted at diverse time scales revealed that microbial communities are changing over short and long periods, going from hours to years (Mangot et al., 2013; Fuhrman et al., 2015). Communities are sometime presenting seasonal temporal patterns (Gilbert et al., 2012; Linz et al., 2017; Giner et al., 2018) and when studied



together, microbial eukaryotes and prokaryotes can be synchronous (Bock et al., 2018; Obertegger et al., 2019).

The main objective of this study was to decipher the composition of microbial communities (microbial eukaryotes, bacteria and archaea) of a large size-range (0.22–30  $\mu\text{m}$ ) over multiple years, in the overlooked small freshwater ecosystems. Therefore, we conducted an eight-years survey at a seasonal frequency in five small freshwater ecosystems (four ponds and one brook) located in North-Western France, using Illumina MiSeq sequencing of 18S/16S rDNA amplicons. In parallel, the measurements of environmental conditions was used to estimate the impact of abiotic parameters on the community structuration. Previous works realised in the same ecosystems for two-year revealed a high diversity of small microbial eukaryotes and prokaryotes ( $<5 \mu\text{m}$ ) (Simon et al., 2015; David et al. 2020, submitted). Despite their marked differences in composition in the five ecosystems, microbial eukaryotes and prokaryotes had similar temporal patterns. Here, we investigated the composition and its variations over time in the same ecosystems microbial but at a longer time scale, using the latest sequencing technology, and considered a larger size fraction of microorganisms. We aimed at confirming the temporal patterns observed previously at and at deciphering the temporal pattern of recurrences in specific operational taxonomic units (OTUs). Even if the global community presented a recurrent seasonal pattern, we found that only a small proportion of the detected OTUs displayed a marked recurrence over the eight-year survey.

### 5.2.3 Material and Methods

#### 5.2.3.1 Sampling and measurement of physico-chemical parameters

Five small and shallow freshwater ecosystems (ponds Etang des Vallées (EV), La Claye (LC), Mare Gabard (MG), Saint Robert (SR), and brook Ru Sainte Anne (RSA)) are located in a restricted area (less than 9 km apart from each other) in the Natural Regional Park of the Chevreuse Valley (Ile-de-France, France). A total of 144 samples were collected at surface every 3 months between October 2011 and January 2019, using sterile plastic carboys. Data were missing for one sampling date in La Claye, Mare Gabard and Saint Robert due to insufficient amount of DNA and in one sampling date in Saint Robert and two in Ru Sainte Anne due to dry event. Once collected, water samples were serially filtered immediately back to the laboratory through 100, 30, 5 and 0.22  $\mu\text{m}$  pore-size Nuclepore membranes (Whatman). Membranes were immediately stored frozen at  $-80^{\circ}\text{C}$  until DNA extraction.

Water temperature, dissolved oxygen concentration, pH and conductivity were measured *in situ* using a multiparameter probe Multi 350i (WTW, Weilheim, Germany). Concentrations of dissolved nutrients, i.e. total phosphorus ( $P_{tot}$ ), orthophosphate ( $PO_4^{3-}$ ), total nitrogen ( $N_{tot}$ ), nitrite ( $NO_2^-$ ), nitrate ( $NO_3^-$ ) and ammonia ( $NH_3$ ) were measured on the sampling day using colorimetric tests following manufacturer instructions (Hach-Lange, Düsseldorf, Germany). Organic carbon concentrations (DOC) were measured using a TOC VCSH analyzer (Shimadzu, Tokyo, Japan) according to Rochelle-Newall et al. (2014) protocol. Chlorophyll a was measured from samples collected on GF/F filters (Whatman) by spectrometry, following the protocol described in Simon et al. (2015).

### 5.2.3.2 DNA extraction, amplification and sequencing of 16S rDNA fragments

DNA was extracted using the PowerSoil DNA extraction kit (MoBio, Carlsbad, CA, USA) from cells collected in two size-fractions 0.22–5  $\mu m$  and 5–30  $\mu m$ . DNA from both size fraction were mixed proportionally to the volume of water filtered. Eukaryotic 18S rRNA gene fragments including the V4 hypervariable region were amplified using primers EK-565F (5'-35 GCAGTTAAAAAGCTCGTAGT) and EK-1134-R-UNonMet (5'-TTTAAGTTTCAGCCTTGCG), biased against metazoa (Bower et al., 2004). To target both bacterial and archaeal, 16S rRNA gene fragments encompassing the V4 hypervariable region were PCR-amplified using U515F (5'-GTGCCAGCMGCCGCGGTAA) and U806R (5'-GGACTACVSGGGTATCTAAT) primers. These primers were tagged with 10 bp Molecular IDentifiers (MIDs) to allow differentiation of PCR products from 89 distinct samples pooled together for multiplex sequencing. PCR were conducted in a total volume of 25  $\mu l$  using 1.5 mM  $MgCl_2$ , 0.2 mM dNTPs, 0.3  $\mu M$  primers, 0.3–3  $\mu l$  of DNA sample and 0.5 U HotStart Taq polymerase (Taq Platinum, Invitrogen, Carlsbad, CA, USA). The amplifications were performed in a thermocycler. After a 2 min denaturation step at 94°C, 35 cycles of amplification were performed (94°C for 15s, 55°C or 58°C for 30s (for eukaryotes and prokaryotes respectively), 72°C for 90s). The process ended with a final extension set at 72°C for 10 min. For each sample, five PCR products were pooled together. Pools were then purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany). Equivalent amounts of purified amplicons of 89 different samples were multiplexed and sequenced using paired-end (2x300 bp) MiSeq Illumina technology (Eurofins Genomics, Ebersberg, Germany).

### 5.2.3.3 MiSeq Illumina analysis

A total of 7 007 571 raw merged reads was obtained for the eukaryotes, and 5 737 732 for the prokaryotes. Reads were filtered through a modified local pipeline to remove spurious sequences and to regroup reads in operational taxonomic units (OTUs). Briefly, reads are merged using FLASH (Magoč and Salzberg, 2011), primers and MIDs are trimmed with ‘cutadapt’ (Martin, 2011), and dereplication of the clean merged reads was realised using ‘vsearch’ (Rognes et al., 2016) with a ‘uchime\_denovo’ option to eliminate potential chimeras.

Cleaned reads were gathered at 95% identity cut-off for both microbial eukaryotes and prokaryotes using CD-HIT-TEST (Fu et al., 2012). Singletons, which are OTUs composed of only one single sequence, were excluded. OTUs reference sequences were assigned to taxonomic groups based on similarity with sequences of a local database including sequences from cultured organisms and environmental datasets retrieved from SILVAv128 and PR2v4. To avoid wrong affiliations, OTU representative sequences with a similarity best-it lower than 80% were verified by phylogenetic placement in reference phylogenetic trees. Wrong affiliations were modified manually. Reads affiliated to mitochondria and chloroplast were excluded from the dataset. Finally, filtered sequences from all OTUs were attributed to samples according to their MIDs. At the end, 5 630 664 correct reads were retained for microbial eukaryotes, and 4 576 603 for prokaryotes.

### 5.2.3.4 Statistical analyses

All statistical analyses were conducted with the R-software (<http://cran.r-project.org>) (R Core Team, 2019). All analyses with no further indications were performed with the ‘vegan’ R-software package (Oksanen et al., 2007). Bray-Curtis dissimilarities were conducted to determine distances between and within microbial communities. Wisconsin standardization of the reads was performed prior to dissimilarities calculations to limit putative biases due to the influence of very abundant OTUs. The abiotic parameters were scaled and centred in all multivariate analysis. To see if the ecosystems were different in terms of abiotic parameters, we performed a principal correspondence analysis (PCA) on abiotic parameters. Rarefaction, i.e. a randomly selected subset of sequences in each sample based on the lowest number of sequence, was applied to the global sequencing dataset to provide a more accurate comparison of the five ecosystems. Microbial eukaryotic and prokaryotic communities were rarefied to their smallest level of reads (respectively 6 355 and 5 574 reads per sample). The potential influence of rarefaction depths (i.e. from 5000 to 20000 reads) was tested on the statistical analyses

performed afterwards. All analyses were conducted using a rarefied table based on the smallest level of reads, referred as the “global OTU dataset”. The different statistical analyses based on microbial eukaryotes and prokaryotes were conducted either using the global OTU dataset, or using the two datasets representing the abundant and rare OTUs. Abundant OTUs had a mean relative abundance superior to 0.1% and detected in at least 20% of the samples within one ecosystem while rare OTUs were represented by the remaining OTUs. Richness was determined as the number of OTUs detected per sample, and evenness was calculated following the formula described in Pielou (1966) with the ‘vegan’ R package. The number of OTUs shared by the five ecosystem was determined using the ‘gplots’ of the R-software package (Warnes et al., 2016) and represented in Venn diagrams.

The three following analyses were based on Bray-Curtis dissimilarities. Non-metric multi-dimensional scaling (NMDS) were performed on the global OTU dataset or on each ecosystem to visualise the differences in community composition between sites or seasons. An analysis of similarities (ANOSIM) was performed between pairs of groups, with 9 999 permutations, to statistically test and quantify the differences between the NMDS groups (Clarke and Green, 1988). We performed permutational analyses of variance (PERMANOVA), to assess whether the habitat (pond or stream), the ecosystem (localization), the season or the year explained significantly community variance.

Turnovers between two seasons over the 8-year survey (e.g. summer versus winter) were calculated considering one season as the sum of all samples collected in this particular season. The turnover between samples collected at the same season but during different years was also estimated. Turnover was determined through the comparison of the number of OTUs gained and lost between two samples, using the ‘codyn’ R-software package (Hallett et al., 2016). Time-lag analysis (TLA) was performed to identify patterns in temporal dynamics of microbial communities and environmental variables (Collins et al., 2000). TLA were performed on (i) Wisconsin standardized Bray-Curtis dissimilarities for microbial communities and on (ii) Euclidean distances for environmental data. Linear regression was fitted on the dissimilarity distributions using a linear model using the ‘stats’ package of R-software. Permutations were performed using the ‘lmPerm’ package to test for the robustness of regressions (Wheeler et al., 2016).

The Recurrence Index (RI) developed by (Giner et al., 2018) was applied for the detection of OTUs or group of OTUs showing a seasonal pattern. This index is based on auto-correlation functions that are calculated over 1 000 randomized occurrences. The recurrence index was applied on all OTUs with a mean frequency  $> 0.01$ . All the OTUs occurring less than 3 times were considered as “Uncertain”. An OTU or a group of OTUs were considered seasonal if RI

was significant. To estimate the contribution of each environmental parameter in community variance, we performed canonical correspondence analyses (CCA). Putative multicollinearity between environmental parameters was checked by computing the variance inflation factors (VIF). A VIF value above 10 indicated a strong multicollinearity and resulted in the exclusion of the environmental factors. Then, forward selection was used to identify the most important environmental parameters in the CCA model (Monte Carlo permutation tests, 999 permutations).

The global community network was built based on co-occurrence analysis. The co-occurrence relationships between OTUs were inferred using a multivariate Poisson lognormal model with the ‘PLNmodels’ package of R-software (Chiquet et al., 2019). This model infers direct statistical covariations by focusing on the inverse of the covariance matrix between OTU abundances, also called precision matrix. The precision matrix allows to avoid false connections between nodes, e.g. a covariation between two nodes induced by their direct association to a third node (Kurtz et al., 2015). Moreover, the model allows to take into account the differences in sampling depth between samples, as well as various covariables such as environmental parameters. We conducted the analysis on the OTUs present in 35% of the samples, and with a mean abundance higher than 0.01%. The sequencing depth was added as an offset of the model. The sampling site and the temperature were considered as covariables, to avoid spurious associations. The selection of the optimal model was conducted on a 50-size grid of penalties, using the BIC criteria. The visualization of the network was conducted using the ‘ggnet’ R-software package (Tyner et al., 2017). We obtained an undirected network, with the OTUs represented by the nodes, and the relationship between them by the edges. Network topology was investigated with the calculation of classical network properties as well as a block model (BM) analysis. The network properties were calculated using the ‘igraph’ R-software package (Csardi et al., 2006). We estimated the number of positive and negative edges, the total number of nodes and the number of connected nodes for the network. The mean degree of the network was also calculated, corresponding to the average number of edges established by each node. The average path length, representing the mean number of edges that are needed to link two random nodes was determined. The complexity of the network was estimated with two indicators. The connectance, which quantifies the proportion of edges established in the network compared to all the potential edges, was calculated as  $c = \frac{2E}{N \times (N-1)}$ , where E is the number of edges and N the number of nodes (Barrat et al., 2008). The clustering coefficient was calculated to estimate the probability that two nodes sharing the same neighbour are connected with another (Delmas et al., 2019). The block model analysis was conducted using a stochastic block model approach, developed in the ‘blockmodel’ R-software package, and based on the binary network reporting the presence-absence of significant direct associations (Leger, 2016). With this method, we were able to gather nodes according to similar modes of interactions. The detection of modules

in the networks, i.e. groups of OTUs that constitute a sub-network, was conducted using the ‘igraph’ R-software package.

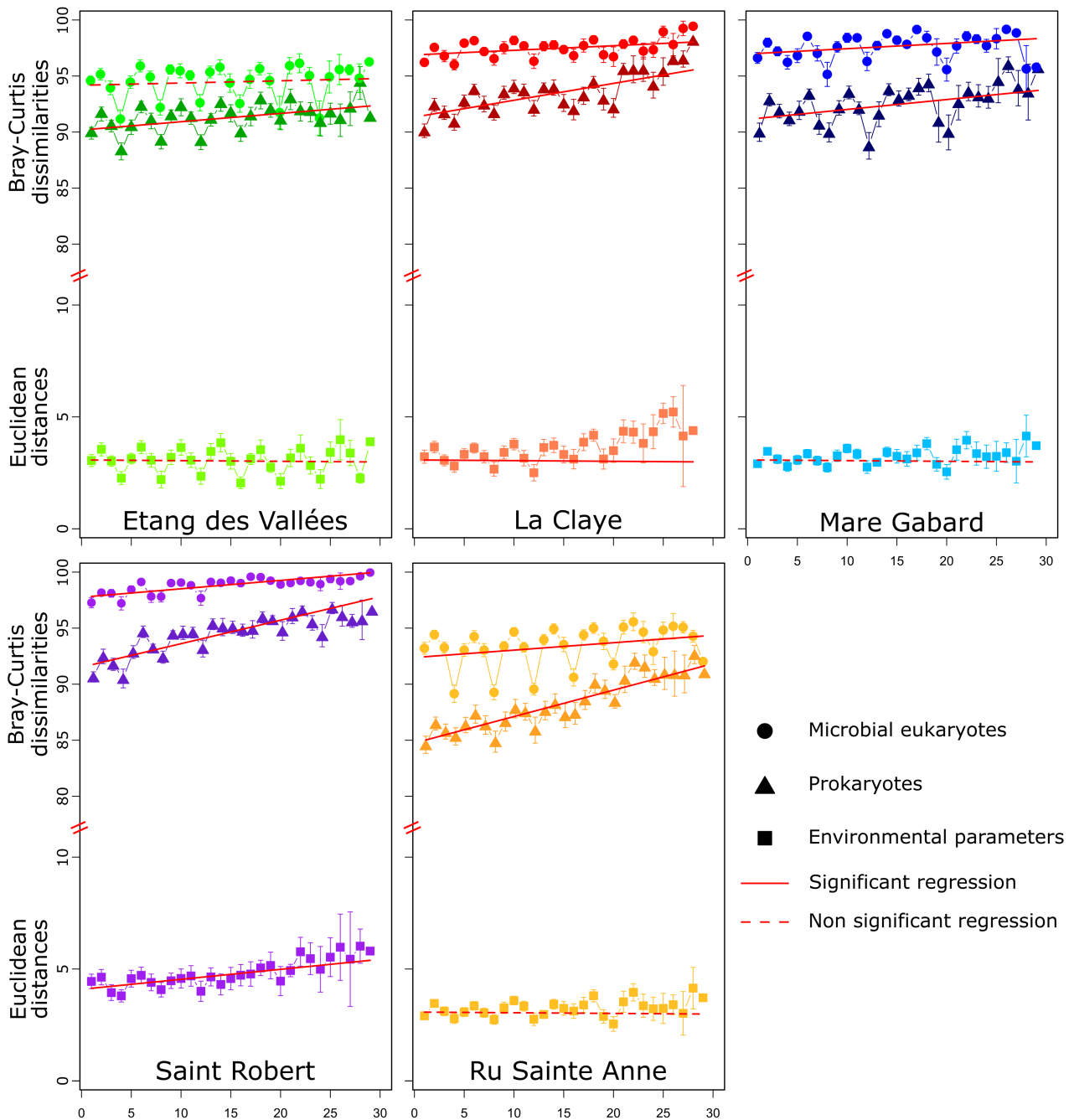
## 5.2.4 Results

### 5.2.4.1 Environmental ecosystem characterisation

Five ecosystems were sampled seasonally (autumn in October, winter in January, Spring in April and Summer in July) from October 2011 to January 2019. They were selected because they were representative of small freshwater ecosystems in the “Parc Naturel Régional de la Haute Vallée de Chevreuse”. Etang des Vallées (EV) is a large pond, Ru Sainte Anne (RSA) is a stream, Saint Robert (SR) is a village pond, La Claye (LC) is a peat bog, Mare Gabard (MG) is a forest pond (see Simon et al., 2015 for more details). Over the 8-year survey, they experienced marked variations in their abiotic environment (Supplementary Table 5.A.1 and Figure 5.B.1).

The PCA (Supplementary Figure 5.B.1) revealed that RSA was characterised by higher conductivity, EV by higher DO concentrations and MG by lower pH and conductivity values. Etang des Vallées had the highest mean dissolved oxygen (78.1%; DO), nitrate (3.94 mg/l) and nitrite (0.05 mg/l) concentrations and pH (7.3) but lower ammonia (0.17 mg/l), orthophosphate (0.05 mg/l) and dissolved organic carbon concentration (3410  $\mu\text{M C}$ ; DOC). Mare Gabard was characterised by the highest average DOC concentration (6239  $\mu\text{M C}$ ) and the lowest pH (6.55), conductivity (67.4  $\mu\text{S/cm}$ ), DO (37.2%) and nitrite concentrations (0.01 mg/l). Saint Robert featured the highest mean chlorophyll *a* (72.9 mg/l), orthophosphate (0.23 mg/l) and ammonia (0.62 mg/l) concentrations and the lowest nitrate (1.51 mg/l) concentration. Ru Sainte Anne had the highest average conductivity (654.6  $\mu\text{S/cm}$ ) and lowest chlorophyll *a* concentration (1.34 mg/l) (Supplementary Table 5.A.1).

Time Lag Analyses (TLA) revealed that environmental conditions were globally stable at the 8-year scale in Etang des Vallées, Mare Gabard and Ru Sainte Anne but changed significantly over time in La Claye and Saint Robert (Figure 1). However, there was significant changes of particular parameters, such as conductivity that decreased in Etang des Vallées, La Claye and Ru Sainte Anne when it increased in Mare Gabard over the 8-year survey (Table S1). Also, ammonia concentration reached higher values in Etang des Vallées and Mare Gabard over time. Chlorophyll *a* concentrations decreased in Mare Gabard and Saint Robert during the study. Nitrate concentrations were lower in La Claye over the survey. DOC concentration increased



**Figure 5.2.1** Time-lag analyses (TLA) were computed for each ecosystem between sampling dates over increasing time lags over the 8-year survey (from one-month until 29-month). Wisconsin standardized Bray-Curtis dissimilarities were calculated on the rarefied microbial eukaryotic and prokaryotic datasets. Euclidean distances were calculated on measured physico-chemical parameters. Each point represents the mean value of pairwise Bray-Curtis dissimilarities or Euclidean distances for a given time-lag and is plotted with associated standard errors. Triangles indicate the prokaryotic data, circles indicate the eukaryotic data and square indicate the physico-chemical data. Linear regression was calculated on the raw dissimilarities and distances, previously to the mean transformation. Significant regressions ( $p < 0.05$ ) are indicated with a star.

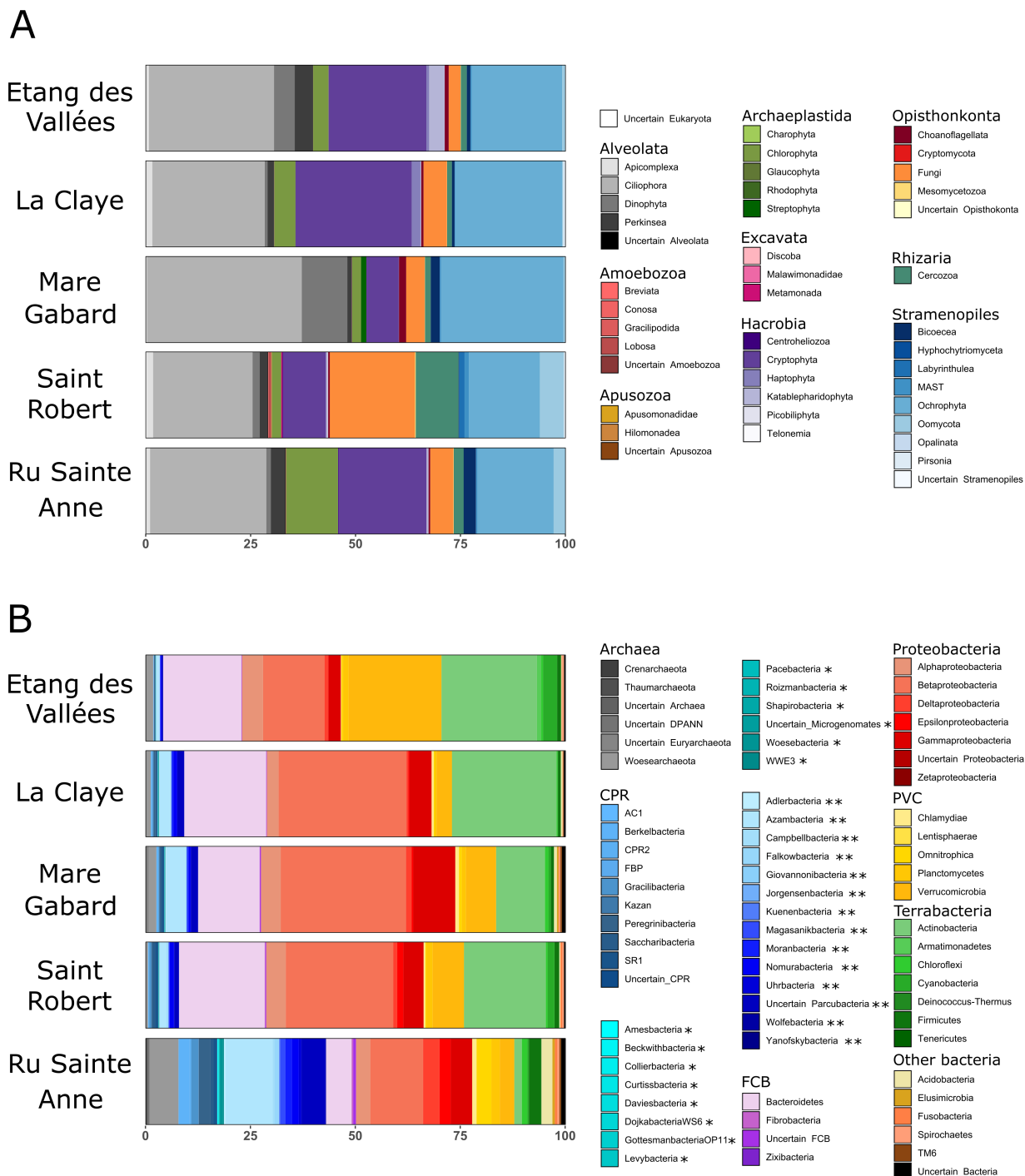
markedly in Etang des Vallées over time.

#### 5.2.4.2 Overall structure of microbial communities in and within the five ecosystems

**Diversity of microbial eukaryotes** The 5 630 664 clean merged reads obtained for the eukaryotes were rarefied at a minimum of 6 355 reads per sample. They were clustered into 22 764 OTUs (Figure 5.2.2 ). The mean richness per sample was 469.6, and the most diverse community was harboured by Ru Sainte Anne with a mean richness of 941 (Supplementary figure 5.B.3, Supplementary table 5.A.2 and 5.A.3). Evenness varied strongly in the ecosystems over the survey, with a maximal difference of 0.71 in Mare Gabard (Supplementary figure 5.B.3). The highest mean value of evenness, 0.73, was founded in Ru Sainte Anne. The detected OTUs belonged to 37 different phyla. The most abundant phyla were Ciliophora, Ochrophyta and Cryptophyta with a total abundance of 29, 22.3 and 18%, respectively. The more diversified phyla were Ciliophora, Cryptophyta, Fungi and Ochrophyta, each composed of 144 different OTUs in the dataset. A low proportion of OTUs (less than 2.0%) were detected as abundant in each ecosystem (i.e. mean abundance > 0.1% and present in 20% of the ecosystem samples), but they represented between 51.9% (Ru Sainte Anne) and 75.4% (Etang des Vallées and Mare Gabard) of the total number of reads. Only 13 OTUs were present at every time point in one of the ecosystems.

**Diversity of prokaryotes** For the prokaryotes, 4 576 603 clean merged reads were rarefied to 5 574 reads per sample. Reads were clustered into 25 536 OTUs (Figure 5.2.2). The mean richness per sample was 471.8. As for eukaryotic communities, the mean richness in Ru Sainte Anne was higher compared to other ecosystems, with a mean number of 2 184 OTUs per sample (Supplementary figure 5.B.3, Supplementary table 5.A.2 and 5.A.3). Evenness indices were higher for prokaryotes than for microbial eukaryotes, but Ru Sainte Anne still presented the highest mean value over the five ecosystems (0.88) (Supplementary figure 5.B.3). OTUs were affiliated to 38 phyla, belonging to all currently recognized prokaryotic super-groups. Considering all ecosystems, the diversity was dominated by Proteobacteria, Actinobacteria, and Bacteroidetes (respective abundance of 34.8, 16.2 and 15.9%). These three phyla dominated in all ecosystem except Ru Sainte Anne, where Proteobacteria, Parcubacteria and Woesearchaota dominated the communities over the eight-year survey (Supplementary figure 5.B.4). Archaea represented a small proportion of the reads (13 801 OTUs representing 0.49% of the reads), including DPANN members. A large diversity of bacteria of the Candidate Phyla Radiation (32 phyla, 48 076 OTUs) was also detected. The most abundant OTUs (i.e. mean abundance





**Figure 5.2.2** Histograms showing the relative abundance of 16S and 18S rDNA amplicon reads assigned to high-rank taxa average over the 8-year survey in the five ecosystems.

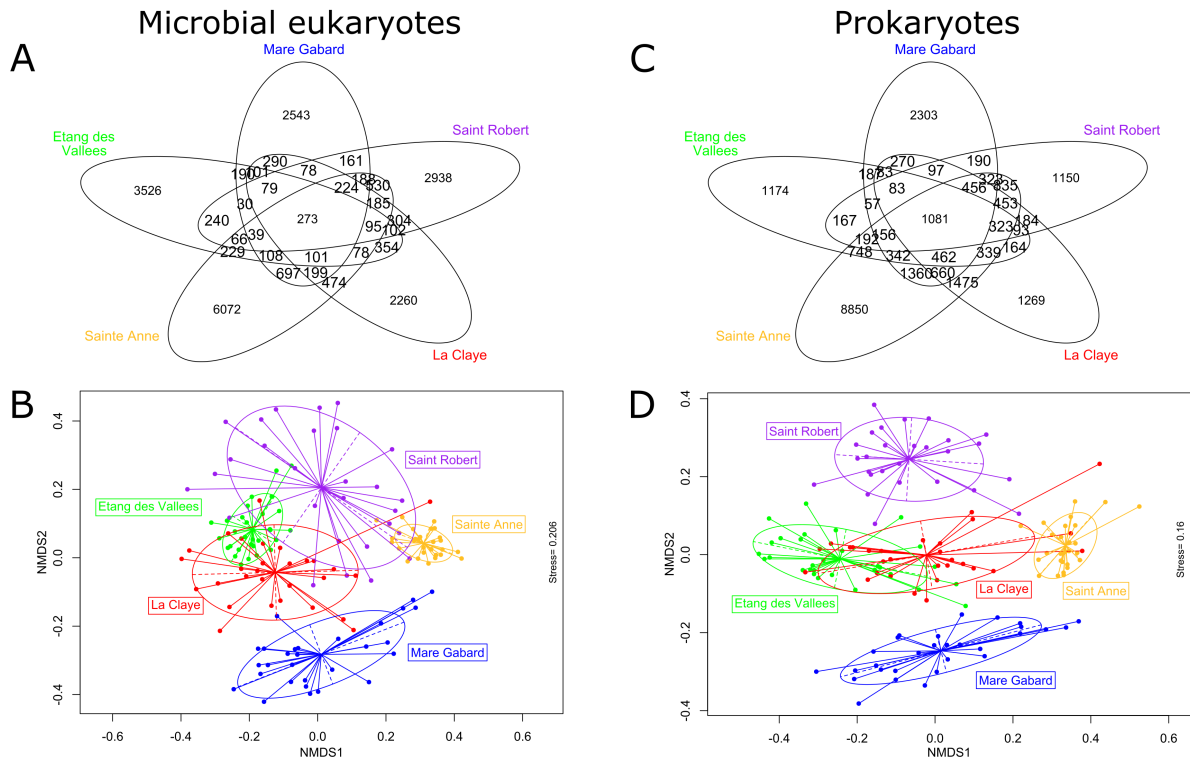
>0.1% and present in 20% of the ecosystem samples) were composed of less than 2.0% of the total number of OTUs, but they represented between 31.5% (Ru Sainte Anne) and 84.7% (Etang des Vallées) of the reads. Only 71 OTUs were detected in all samples in one ecosystem.

**Comparison of microbial communities between ecosystems** The differences between communities were analysed based on the presence and absence of OTUs (Figure 5.2.3). A low percentage of OTUs were shared between the five ecosystems (1.2% for eukaryotes, 4.2% for prokaryotes). However, they represented up to 50.3% of the reads for the eukaryotes, and 69.6% for the prokaryotes. The differences between samples and ecosystems were investigated based on community dissimilarities. NMDS revealed that microbial communities were rather distinct from each other. According to the overlapping of the ellipse grouping the samples of each ecosystem, prokaryotic communities seemed more dissimilar than their microbial eukaryote counterparts (Figure 5.2.3). These observations were confirmed by significant and high pairwise ANOSIM analyses between each ecosystem (Supplementary table 5.A.4). For both communities, the highest values of dissimilarities were found between Ru Sainte Anne and the other ecosystems, and also between Etang des Vallées and Mare Gabard.

Using PERMANOVA analysis, we tested factors to explain Bray–Curtis variances among samples (Supplementary Table 5.A.5). Only a small proportion of the variance of microbial eukaryote communities was explained, mainly by differences between ecosystems (2.6% of the variance explained by the type of habitat (e.g. lentic or lotic) and 4.4% explained by the ecosystem,  $p$ -value = 0.0002). If the season played also a role, the year was not a significant parameter (Supplementary Table 5.A.5). For the prokaryote communities, all the variable played a significant role. In particular, the ecosystem and the year explained the most prokaryotic community variance (respectively 6.0 and 5.9%).

### 5.2.4.3 Global description of the temporal dynamics

Along the time survey, community structure can change drastically between two season; e.g. for the eukaryotes in Mare Gabard, community diversity changed from a community dominated by Ciliophora (Abundance of 93.0% in MG77, Autumn 2017) to a community dominated by Ochrophyta (Abundance of 83.9% in MG79, Winter 2017). Prokaryotes were less represented by phyla dominating the community in one ecosystem, at the exception of Proteobacteria which reached 79.1% in Mare Gabard (MG59, Summer 2014). This observation is coherent with the evenness values (0.52 for eukaryotic and 0.69 for prokaryotic communities).



**Figure 5.2.3** Venn diagram showing the number of rarefied OTUs over the 8-year survey shared by the different ecosystems or specific to one ecosystem for microbial eukaryotes and prokaryotes (EV Etang des Vallées, LC La Claye, MG Mare Gabard, SR Saint Robert, RSA Ru Sainte Anne). (B) Non-metric multidimensional scaling plot constructed with Wisconsin-standardized Bray-Curtis dissimilarities applied on rarefied data between all samples, for eukaryotic and prokaryotic communities. Stress value is displayed on the right border of the plot.

Turnover analysis on abundant and rare OTUs between seasons, independently from the year, revealed lower turnover rates within prokaryotes than within microbial eukaryotes. Turnovers of abundant prokaryotes were globally lower (mean value of 4.6%) than for microbial eukaryotes (15.0%). In most of the ecosystems, we found the highest turnover values when comparing microbial communities in summer and winter (Supplementary figure 5.B.5). The same observation was realized for the turnover analysis on abundant and rare OTUs collected at the same season between following years. For both communities, SR presented the highest mean turnover value (43% for abundant, 89.6% for rare) in all ecosystems and RSA the lowest (14.1% for abundant, 83.6% for rare) (Supplementary figure 5.B.6). This can be linked to the number of OTUs present only between 1 and 10 time in the ecosystem (minimum of 95.7% for eukaryotes and 93.2% for prokaryotes, both in Ru Sainte Anne and maximum of 99.6% for eukaryotes in Saint Robert and 97.1% for prokaryotes in La Claye, 96.8% in Saint Robert). Prokaryote communities in winter and summer were characterized by lower turnovers, according to the expansion of the boxplot.

Temporal evolution of communities at the OTU level was examined using Bray-Curtis dissimilarities. Microbial communities were characterised by seasonal patterns, as showed by lowest dissimilarities when comparing the same season of different years and highest when comparing “opposite” seasons, e.g. summer versus winter (Figure 5.2.1). This pattern was strongly marked for the microbial eukaryotes of Etang des Vallées and Ru Sainte Anne, and less in Saint Robert. Dissimilarities over time were however strong (minimal dissimilarity obtained for the prokaryotes in Ru Sainte Anne, with a value of 79.3). As for the turnover, we observed that microbial communities in Ru Sainte Anne were less dissimilar than in the other ecosystems. In addition to the intra-annual seasonal pattern, the microbial community composition also featured a global inter-annual pattern. Dissimilarities raised over increasing time lags in the time lag analyses in all ecosystems and for prokaryotes and microbial eukaryotes. This points to continuous changes of the microbial communities during the 8-year survey (Figure 5.2.1). Eukaryotes in the Etang des Vallées were the only exception to this general pattern (non-significant linear regression). Eukaryotes experienced lower changing rates than prokaryotes (average slope of 5.4% and 14.4%, respectively) (Supplementary table 5.A.6). The same analysis was performed taking into account separately abundant and rare OTUs. If the same seasonal pattern was observed, there were differences in the inter-annual pattern (Supplementary table 5.A.6). The eukaryotic abundant OTUs of La Claye, as well as the eukaryotic rare OTUs and prokaryotic abundant OTUs of Mare Gabard did not present a significant variation of their dissimilarities over time.

#### 5.2.4.4 Factors explaining community composition changes

To link microbial communities and measured environmental parameters, the global dataset was constrained with abiotic parameters. The measured abiotic parameters explained a small proportion of the community variance (3.8% for the eukaryotes and 6.8% for the prokaryotes) (Supplementary figure 5.B.7). Abiotic parameters that explained this variance were the same for eukaryotes and prokaryotes (i.e. conductivity, dissolved oxygen and temperature), at the exception of pH which played a significant role only for prokaryotes.

The same analysis was conducted considering the ecosystems separately (Supplementary table 5.A.7). For the eukaryotes, temperature explained about 6.0% of the variance in Etang des Vallées, La Claye and Mare Gabard. The explained variance was much higher in Ru Sainte Anne (15.6%) and the community composition was driven by dissolved oxygen and nitrate concentrations as well as conductivity. In Saint Robert, variance of the community composition was not explained by the abiotic parameters. For the prokaryotes, the explained variance was higher than for eukaryotes, with a mean value for the five ecosystems of 14.2%. Dissolved

oxygen concentration was a significant parameter in all ecosystems. We did not find pH as a parameter explaining significantly community variance in any ecosystem.

#### 5.2.4.5 Seasonal changes at a finer scale

**Dissimilarities between communities sampled at different seasons** A NMDS analysis was conducted on each ecosystem and each community to confirm the seasonal patterns observed with the time-lag analysis. The samples were then grouped according to their season (Supplementary figure 5.B.8). Ellipses defining seasons only overlapped slightly for the prokaryotes of Saint Robert. An ANOSIM analysis determined the significance and the strength of the dissimilarities between pairs of seasons. For each ecosystem except Saint Robert and Ru Sainte Anne for the eukaryotes and Mare Gabard for the prokaryotes, the highest dissimilarity was observed for the comparison of winter and summer.

**Identification of recurrent OTUs over time** The Recurrence Index (Giner et al., 2018) determines the proportion of recurrent OTUs in the different ecosystems. We applied this analysis for each ecosystem separately. OTUs present less than 3 times in the dataset were excluded because they could not be considered as recurrent. The results of Saint Robert are not presented since the OTUs detected as recurrent in this ecosystem did not present clear seasonal dynamics.

Overall, recurrent OTUs only represented a small proportion of the communities. They corresponded to a mean value of 2.23% of the OTUs affiliated to microbial eukaryotes and 1.11% for the prokaryotes (Table 5.2.1). They represented 13.2% and 10.3% of the total number of reads (respectively for microbial eukaryotes and prokaryotes, mean value for the four ecosystems). Not all these OTUs featured a clear seasonal pattern, that is an occurrence each year at one specific season, and an absence the rest of the year. They represented respectively a mean value of 1.15 and 0.48% of the total number of OTUs for microbial eukaryotes and prokaryotes, and a mean value of 11.3 and 5.0% of the total number of reads, respectively. OTUs with a clear recurrent pattern were mostly detected in summer and in winter .

**Table 5.2.1** Repartition of all the rarefied OTUs in 4 categories, with the percentage of OTUs and their total relative abundance within the categories for microbial eukaryotes and prokaryotes. First, OTUs with a mean abundance lower than 0.01%. Then, within the remaining OTUs, the OTUs present less than 3 times in the ecosystem. Finally, according to the recurrence index, the recurrent or non-recurrent OTUs.

	Eukaryotes											
	EV			LC			MG			RSA		
	Nb OTU	%OTU	Nb Reads	%Read	Nb OTU	%OTU	Nb Reads	%Read	Nb OTU	%OTU	Nb Reads	%Read
All	5611		190650	10.7	5197	184295	5301	184295	9558	177940		
Uncertain (n = 3)	4756	84.8	20473	10.7	4806	33634	18.3	4760	89.8	24581	13.3	27101
Recurrent OTUs	204	3.6	23478	12.3	45	17467	9.5	42	0.8	27861	15.1	28294
Seasonal	98	1.7	19865	10.4	24	11892	6.5	26	0.5	27519	14.9	23837
Non recurrent OTUs	651	11.6	146699	76.9	346	133194	72.3	499	9.4	131853	71.5	122545
	Prokaryotes											
	EV			LC			MG			RSA		
	Nb OTU	%OTU	Nb Reads	%Read	Nb OTU	%OTU	Nb Reads	%Read	Nb OTU	%OTU	Nb Reads	%Read
All	5651		167220	5.1	7492	161646	8110	161646	18055	156072		
Uncertain (n = 3)	4700	83.2	8541	5.1	6478	15023	9.3	6364	78.5	15489	9.6	24967
Recurrent OTUs	64	1.1	21296	12.7	41	3151	1.9	115	1.4	32556	20.1	11064
Seasonal	34	0.6	16788	10.0	16	1771	1.1	50	0.6	8237	5.1	6578
Non recurrent OTUs	887	15.7	137383	82.2	973	143472	88.8	1631	20.1	113601	70.3	120041

For microbial eukaryotes, Ru Sainte Anne presented the highest proportion of seasonal recurrent OTUs, with 1.9% of the OTUs corresponding to 13.4% of the reads. The proportion of OTUs had similar rates in Etang des Vallées (1.7%). In Mare Gabard, recurrent OTUs were not abundant (0.5%) but their proportion in terms of reads was the highest (14.9%). The lowest proportion was detected in La Claye, with only 0.5% of the OTUs that were characterised by a recurrent pattern, corresponding to 6.5% of the reads. For the prokaryotes, the highest proportion of seasonal recurrent OTUs was detected in Etang des Vallées (0.6% of the OTUs, 10.0% of the reads). Similarly to the microbial eukaryotes, La Claye harbored the lowest amount of recurrent OTUs (0.2% of the OTUs, 1.1% of the reads). Ochrophytes represented a large proportion of the recurrent OTUs in all the ecosystems, in particular in Mare Gabard (13.2% of the total number of reads, corresponding to 13 OTUs) (Supplementary figure 5.B.9). One OTU was responsible for this domination of the community, OTU X788345. This OTU was affiliated to *Gonyostomum semen* (Raphidophyceae) and represented 12.8% of the reads in Mare Gabard. In Etang des Vallées and Ru Sainte Anne, Ciliophora were among the most abundant recurrent OTUs (respectively 5.90% and 5.89% of the OTUs reads). In La Claye, recurrent OTUs were dominated by the Ochrophytes (5.86% of the recurrent OTUs reads). The highest proportion of seasonal recurrent prokaryote OTUs was detected for Proteobacteria in La Claye, Mare Gabard and Ru Sainte Anne (1.44, 16.8, 3.94% of the total number of reads respectively). Members of the Candidate Phyla Radiation were recurrent in all ecosystems, but at higher proportion for the super-group Parcubacteria (respectively 0.13, 0.21, 0.97 and 1.29% of the reads in Etang des Vallées, La Claye, Mare Gabard and Ru Sainte Anne. (Supplementary figure 5.B.9).

#### 5.2.4.6 Global structure of the microbial communities

The global network presented only 0.4% of positive links (3 over 673 edges) (Figure 5.2.4). Over the 874 selected OTUs, only 40% were connected in the network. The connectance of the network was 0.003, with a clustering coefficient of 0.20. The global mean degree of the network was 1.54. Over the 351 nodes, 50.1% were affiliated to bacteria, 46.7% to eukaryota and 3.13% to archaea. There was a majority of interactions between the same domains of life (40.9% of interactions within bacteria, 47.8% with eukaryote). Microbial eukaryotes had a higher mean degree than the two other domains (mean degree of 4.28 for the eukaryotes). The majority of interactions was conducted within heterotrophs (44.9% of the interactions). There were also many interactions happening between autotrophs and heterotrophs (18.6% of the interactions). Putative parasites were hardly connected to autotrophs and heterotrophs (2.1% of the interactions between putative parasites and autotrophs, and 3.7% between putative parasites and

heterotrophs). They were more connected between each other (13.7% of interactions between parasites). Globally, autotrophs and heterotrophs had a similar degree (mean degree of 3.98 for the autotrophs and 4.15 for the heterotrophs). Putative parasites were less connected, with a mean degree of only 2.90. The 22 nodes with a degree higher than 10 were mostly heterotrophs (63.6%) and were not dominated by a specific phylum.

The block model analysis differentiated eight groups in the networks (Supplementary Figure 5.B.10). Several block model groups (BMG) were mainly composed of microbial eukaryote nodes (BMG1, 4, 5 and 8). The block model groups were often dominated by one specific taxonomic group (e.g. BMG3 and BMG7 by Parcubacteria, or the BMG4 by Ochrophytes). BMG4 was dominated by autotrophic microbial eukaryotes. Overall, the different BMG seemed to gather nodes with the same taxonomy or the same trophic mode.

Seven modules were detected in the network (Supplementary Figure 5.B.11). As for the block model analysis, nodes were gathered in the module according to the taxonomic affiliation. For example, module 3 was dominated by Ochrophyta and module 4 by Fungi. Modules 6 and 7 were clearly isolated from the rest of the network. They were mainly composed of OTUs affiliated to parasites. A majority of Woesearchaeota and Parcubacteria composed module 6 while module 7 was dominated by Parcubacteria. OTUs (31) that did not belong to a specific module were separated from the global network and interacted by pair or triplets.

## 5.2.5 Discussion

### 5.2.5.1 Overview of microbial communities in the five ecosystems

The analyses of the 16S and 18S rDNA amplicon sequences revealed a high diversity of microbial eukaryotes and prokaryotes in the five small freshwater ecosystems. Microbial communities were dominated by a low number of very abundant taxa. They belonged to typical freshwater phyla, i.e. Proteobacteria, Actinobacteria, Bacteroidetes, Verrucomicrobia for prokaryotes, and Ciliophora, Ochrophyta and Cryptophyta for microbial eukaryotes (Zwart et al., 2002; Pernthaler et al., 2013; Salcher, 2014; Debroas et al., 2017). Compared to our previous monthly survey, we detected new phyla of microbial eukaryotes and prokaryotes, e.g. Hilomonadea, Picobiliphyta for microbial eukaryotes, Kuenenbacteria, Kazanbacteria for prokaryotes (David et al., submitted, see Chapter 4). The microbial diversity in the present study differs from the one described in David et al. 2020 because of the longer survey duration (8 vs 2 years), large size fraction targeted (0.22–30  $\mu\text{m}$  vs 0.22–5  $\mu\text{m}$ ) and the sequencing technology (Illumina



MiSeq vs 454-pyrosequencing). For example, a higher number of OTUs affiliated to ciliates was detected in this study (mean relative abundance of 29.0% in this study in contrast with 18.2% in David et al., submitted, see Chapter 4) though this could also be due to the higher number of rDNA genes copy within alveolates (Prescott, 1994). Similar results were obtained for the global structure of microbial communities. The same phyla dominated the communities, namely Cryptophyta, Ochrophyta and Ciliophora for the microbial eukaryotes, and Proteobacteria, Actinobacteria and Verrucomicrobia for the prokaryotes (Supplementary figure 5.B.4). The wide majority of the microbial diversity was composed of low-abundant OTUs, i.e. rare OTUs, as in David et al., submitted (see Chapter 4).

A wide diversity of members of the Candidate Phyla Radiation was observed (mean abundance between 2.4% in Etang des Vallées and 35.2% in Ru Sainte Anne). Members of the DPANN (Diapherotrites-Parvarchaeota-Aenigmarchaeota-Nanoarchaeota-Nanohaloarchaeota) group were also detected (mean abundance between 0.5% in Saint Robert and 7.0% in Ru Sainte Anne). These bacteria and archaea are suspected to be potential parasites or symbionts because of their restricted genomes (Castelle and Banfield, 2018). They could also be important players in small freshwater ecosystem functioning because of their potential role in CO<sub>2</sub> fixation (Lannes et al., 2019). Cyanobacteria were not abundant (mean abundance between 0.14% in La Claye and 3.4% in Etang des Vallées) and not recurrent in summer, which was surprising knowing the eutrophic status of these small freshwater ecosystems.

Microbial communities were specific to each ecosystem despite their close geographical localisation, suggesting a strong filtering effect of the environment. The global community composition differed significantly between ecosystems (Figure 5.2.3B and 5.2.3D, Supplementary Table 5.B.5) and few OTUs were common to the five ecosystems (respectively 1.2 and 4.2% of the microbial eukaryotes and prokaryotes communities, Figure 5.2.3A and 5.2.3C). Time series analysis revealed that a large proportion of the diversity was composed of OTUs occurring only one time in the ecosystems (54.8% of the microbial eukaryotic OTUs appearing a single time during the eight-year, and 39.3% for the prokaryotes). They could be always present in background at too low abundances to be detected or they could be result from immigration processes (e.g. shore, banks, animals, rain, wind). However, dispersing taxa may not have profound impacts on composition of communities likely because they do not succeed to settle in the ecosystem because of mass effect for example (Jones and McMahon, 2009; Comte et al., 2017).

### 5.2.5.2 Inter-annual patterns of community composition

Despite specific communities in each ecosystem, a similar inter-annual temporal pattern of microbial community composition was observed in the five ecosystems, for both prokaryotes and microbial eukaryotes, at the exception of microbial eukaryotes in Etang des Vallées. This temporal dynamic, which has been previously observed in the two-year survey, is thus validated on a longer time scale. However, we observed different rates of change in microbial eukaryotes and prokaryotes. Linear regressions applied to the microbial eukaryotes showed lower slopes than those obtained for prokaryotes. The global evolution of the prokaryotes is therefore stronger, and resulted in more changes in the community composition over time. This evolution pattern was inversed in the two-year survey. Logares et al. (2018) discussed that both prokaryotes and microbial eukaryotes had different structuring processes, with bacteria mainly structured by selection and microbial eukaryotes by ecological drift. However, longer time surveys, even if they were only focusing on prokaryotic communities, indicated that prokaryotes were characterized by strong inter-annual changes of community composition (Shade et al., 2007; Tammert et al., 2015; Linz et al., 2017).

In Etang des Vallées, no inter-annual evolution of microbial eukaryotes was observed, and prokaryotes harbored the lowest rate of change. Therefore, microbial communities appeared to be more stable in this ecosystem compared to the other four. The Etang des Vallées is the largest ecosystem of this study, in terms of surface, depth and water volume, resulting in a higher buffering capacity toward environmental disturbances. One hypothesis is that in larger ecosystems, the abundant microbial communities in place are well established, with a high number of organisms. In this situation, it could be more difficult for immigration process to happen, because of a stronger competition (Comte et al., 2017).

### 5.2.5.3 Intra-annual patterns of community composition

Strong intra-annual patterns were detected in the five ecosystems. The seasonal survey revealed strong similarities between microbial communities of the same season over the eight-year, whereas communities at opposite seasons (e.g. summer versus winter) were the most dissimilar (Figure 5.2.1). Intra-annual temporal patterns (i.e. seasonality) in aquatic microbial communities have been observed multiple times at different time scale at surface (Kim et al., 2014; Yan et al., 2017). Seasonal patterns were detected when taking into account the whole microbial communities (microbial eukaryotes or prokaryotes) but also at the scale of abundant and rare taxa. However, the temporal patterns were less marked for rare OTUs. The rhythmic-

ity of rare microbial taxa has been already evidenced in marine bacterioplankton (Alonso-Sáez et al., 2015). Not surprisingly, environmental conditions were also experiencing marked seasonal patterns, which lead to variations in microbial habitat over time that can, in return, impact microbial communities. When conditions are not optimal, some microorganisms are able to enter in dormant stages (e.g. cysts, spores, reduced metabolism) (Finlay, 2002; Lennon and Jones, 2011; Locey et al., 2020). Habitat variations may also favor constant immigration.

Even if environmental conditions harbored similar intra-annual patterns than community composition, a low proportion of the community variance was explained by environmental parameters in the five ecosystems, as previously observed in marine or freshwater systems (Kara et al., 2013; Ward et al., 2017) (Supplementary figure 5.B.8). Microbial communities are often more related together than to the environmental factors (Steele et al., 2011; Lima-Mendez et al., 2015). The variance was slightly higher when considering ecosystems separately (Supplementary table 5.A.7). However, variance explained by abiotic parameters has already reached up to 30% for microbial eukaryotic communities in marine systems (Genitsaris et al., 2016; Giner et al., 2018). Interestingly, the significant abiotic parameters explaining microbial community variance in our five ecosystems were those related to seasonal variations (i.e. conductivity, temperature and pH). Dissolved oxygen and pH were positively correlated which could underline the role played by photosynthetic activity in the strong modification of its environment in these ecosystems.

Within the five ecosystems, Etang des Vallées and Ru Sainte Anne presented more clear seasonal patterns, with regular variations of the dissimilarities over the increasing time-lags and high differences between the mean minimum and maximum dissimilarities (Figure 5.2.1). Microbial community were stable at the intra- and inter-annual scales in Etang des Vallées, probably because of its higher buffering capacity. Ru Sainte Anne was the only lotic system of this study. The first hypothesis was that it would be the most variable ecosystem: it was characterised by the highest number of OTUs and its diversity was constantly rising over time (Supplementary Figure 5.B.2). Moreover, running water can induce a mixing of the banks and the streambed communities (Crump et al., 2012). Ru Sainte Anne was also more subject to perturbations of the banks by wild animals (as frequently observed on site), and the ecosystem experienced several drought periods. However, we observed marked seasonality patterns and low turnover rates in this ecosystem. Also, the lowest proportion of OTUs present in less than 10 samples during the eight-year survey was recorded in this stream (minimum 95.7% of the OTUs for RSA and maximum of 99.6% for SR for microbial eukaryotes, and 93.2% of the OTUs for RSA and maximum of 97.1% for SR for the prokaryotes).

#### 5.2.5.4 Patterns of recurrence in OTUs

The identification of recurrent OTUs revealed that even if the global communities featured an intra-annual pattern, only a low number of OTUs were characterized by clear seasonal recurrence patterns. Recurrence mostly happened in summer and winter. This could be explained by the environmental conditions of these seasons, when the minimum and maximal values of temperature were recorded for example. They were detected in all ecosystems at the exception of Saint Robert. La Claye was the ecosystem with the lowest percentage of recurrent OTUs. Over the five ecosystems, only Saint Robert and La Claye experienced an inter-annual variation of the abiotic parameters. The abundance of recurrent OTUs could therefore be linked to environmental parameter stability. The percentage of recurrent OTUs detected in these freshwater ecosystems was lower compared to a previous study on planktonic micro-algae in marine environments (Giner et al., 2018). Only a few phyla harboured OTUs with recurrent patterns (Supplementary figure 5.B.9). Even if the ecosystems harboured all intra-annual patterns, they did not present the same recurrent phyla. In Ru Sainte Anne, we found mostly OTUs characterized by heterotrophic trophic modes, whereas in Mare Gabard, the recurrent OTUs were dominated by autotrophic trophic modes (Supplementary figure 5.B.9).

As in marine systems (Giner et al., 2018), we found recurrent OTUs in non-recurrent phylogenetic groups. The measure of the recurrent index on phylogenetic classes revealed that only a low proportion were recurrent (data not showed). For example, Berkelbacteria were recurrent in Etang des Vallées, but no single OTU belonging to this class was recurrent. However, in all ecosystems, phyla belonging to the super-phyla Parcubacteria were recurrent, as well as single OTUs. Therefore, several members of the CPR presented a seasonality, with peaks of abundances during winter.

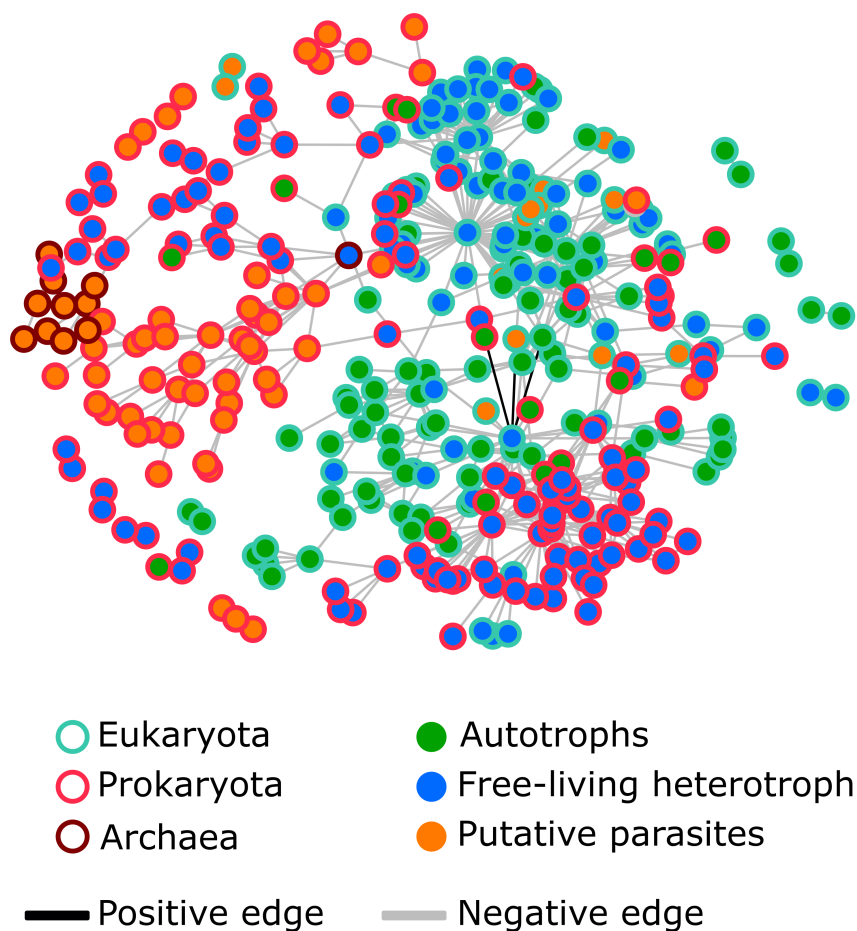
Finally, recurrence analysis revealed OTUs of potential interest for further investigations the functioning in the five freshwater ecosystems. In Etang des Vallées and La Claye, one of the phyla harboring the highest abundance of recurrent OTUs was Cryptophyta. This phylum is known to be characterized by bacterivorous members, implying that it can have major implication in the control of microbial abundance and potentially composition (Grujicic et al., 2018). Moreover, recurrence analysis confirmed the observation made during the 2-year survey that a toxic algae was occurring almost every summer in Mare Gabard. This OTU is a flagellated algae, affiliated to *Gonyostomum semen*. *G. semen* is known to be invasive and to release toxic trichocytes in its environment (Hagman et al., 2015). It was only found in Mare Gabard, which can be explained by the lower pH of its water (Cronberg et al., 1988). The OTU X789795 and X827462 were also recurrent, but with lower abundances (respectively 1.6 and 0.05%). They have been detected in summer 2013, 2014, 2016 and 2017. Members of the same

class had been detected over the two-year survey, but not as toxic recurrent organisms. These dinophytes have been affiliated to *Azadinium*, a group releasing toxic compounds in the environment (Tillmann et al., 2009). These toxic algae may control not only other microorganism but also macroorganisms occurrences.

#### 5.2.5.5 Shared co-occurrence interactions between small freshwater ecosystems

The study of the modules and the block model analysis revealed that OTUs of the same phyla or the same trophic groups were clustered together (highly connected to each other and less with the rest of the network), or presented the same type of connection (i.e. in the same block model groups) (Supplementary Figures 5.B.10 and 5.B.11). The more frequent co-occurrence of OTUs belonging to the same taxonomic groups had already been observed in other networks (Hou et al., 2020). These clusters could be linked to turnovers of OTUs with the same trophic mode, but adapted to different conditions over time.

In summary, our study in five small freshwater ecosystems during an 8-year survey confirmed that microbial communities living in geographic close environments were specific to their ecosystem. Microbial eukaryotes and prokaryotes followed the same global community structure over time, with a majority of the diversity composed of rare OTUs. Microbial communities presented a strong intra-annual pattern (seasonal variations) but were also characterized by an ecological drift at the inter-annual scale. Environmental parameters explained only a restricted proportion of community variance over time. At the OTUs scale, strong turnovers were observed for abundant and rare communities, and a low percentage of OTUs harboured recurrent seasonal dynamics. Moreover, the abundance, the taxonomic affiliation and the trophic mode of recurrent microorganisms varied depending on the ecosystem, suggesting that the ecosystems have functioning particularities.



**Figure 5.2.4** Co-occurrence network inferred by the PLN approach and representing the global structure of microorganisms over the five ecosystems. The network is inferred based on the OTUs present in 35% of the samples, and with a mean abundance higher than 0.01%. The sequencing depth was added as an offset of the model, and the sampling site and the temperature were considered as covariables.

# Supplementary Material

## 5.A Supplementary Tables

**Supplementary Table 5.A.1** Summary of the physico-chemical composition of the five ecosystems. Mean, minimal and maximal value of the different measured abiotic parameters over the eight-year survey are presented. In complement, the results of linear regressions performed for each parameters over the eight years are indicated.

	Etang des Vallées				La Claye				Mare Gabard						
	Linear regression		Linear regression		Linear regression		Linear regression		Linear regression		Linear regression				
	Mean	Min	Max	Slope	pvalue	Mean	Min	Max	Slope	pvalue	Mean	Min	Max	Slope	pvalue
Water temperature (°C)	13.2	3.30	24.0	-0.01	0.39	11.1	1.7	22.0	-0.01	0.67	11.3	1.10	20.2	-0.02	0.42
pH	7.30	6.70	8.30	0.01	0.17	7.25	6.65	7.7	-0.01	0.61	6.55	5.40	8.36	0.05	0.17
Dissolved oxygen (%)	78.1	31.0	150	0.00	0.91	38.9	4.20	91.2	-0.01	0.59	37.2	6.10	78.9	-0.01	0.60
Conductivity (µS/cm)	286	234	346	0.00	0.90	297	131	520	-0.02	0.01	67.4	43.5	209	0.01	0.06
Nitrate (mg/l)	3.94	1.00	8.11	0.00	0.97	2.73	0.49	9.49	-0.04	0.03	1.69	1.00	3.75	0.00	0.80
Nitrite (mg/l)	0.05	0.00	0.21	-0.01	0.67	0.01	0.00	0.11	-0.03	0.28	0.01	0.00	0.02	0.00	0.28
Ammonia (mg/l)	0.17	0.00	0.80	0.01	0.04	0.33	0.01	3.40	0.04	0.12	0.20	0.00	1.68	0.03	0.02
Orthophosphate (mg/l)	0.05	0.00	0.20	0.00	0.80	0.12	0.02	0.49	-0.01	0.77	0.11	0.02	0.38	0.00	0.70
Dissolved Organic Carbon (µM/C)	3410	499	12136	0.04	0.01	5597	1018	19246	0.04	0.11	6239	823.2	17543	0.04	0.11
Chlorophyll a (mg/l)	16.0	7.70	28.9	0.00	0.94	23.7	1.87	239	0.04	0.088	41.6	0.22	149.2	-0.05	0.01
	Saint Robert				Ru Sainte Anne										
	Linear regression		Linear regression		Linear regression		Linear regression		Linear regression		Linear regression		Linear regression		
	Mean	Min	Max	Slope	pvalue	Mean	Min	Max	Slope	pvalue	Mean	Min	Max	Slope	pvalue
Water temperature (°C)	12.2	3.20	21.5	-0.02	0.39	9.52	1.50	16	-0.02	0.17					
pH	7.25	6.50	8.15	-0.02	0.17	7.26	6.55	8.03	0.00	0.91					
Dissolved oxygen (%)	46.3	0.60	130	0.04	0.08	47.5	0.00	85.6	0.01	0.64					
Conductivity (µS/cm)	445	278	615	0.00	0.66	655	396	1247	-0.04	0.03					
Nitrate (mg/l)	1.51	0.02	6.54	0.02	0.08	2.89	0.51	10.7	-0.04	0.16					
Nitrite (mg/l)	0.03	0.00	0.16	-0.02	0.45	0.03	0.00	0.08	-0.02	0.26					
Ammonia (mg/l)	0.62	0.00	4.25	-0.05	0.16	0.23	0.03	1.36	0.01	0.39					
Orthophosphate (mg/l)	0.23	0.02	1.26	0.07	0.94	0.11	0.01	0.32	0.00	0.07					
Dissolved Organic Carbon (µM C)	4516	982.8	19537	0.03	0.21	3851	957.3	10185	0.01	0.33					
Chlorophyll a (mg/l)	72.9	1.80	418	-0.07	0.07	1.34	0.29	2.60	0.00	0.44					



**Supplementary Table 5.A.2** Description of the samples used in the analyses. Names, number of rarefied microbial eukaryotic and prokaryotic OTUs, evenness for microbial eukaryotes and prokaryote and number of reads before rarefaction for microbial eukaryotes and prokaryotes. The final number of reads was the same in all sample because of the rarefaction of the data (6355 for the microbial eukaryotes, 5574 for the prokaryotes).

Ecosystem	Sample name	Sampling date	Observed number of rarefied OTUs		Evenness		Number of reads before rarefaction	
			Microbial eukaryotes	Prokaryotes	Microbial eukaryotes	Prokaryotes	Microbial eukaryotes	Prokaryotes
Etang des Vallées	EV21	Autumn 2011	314	387	0.54	0.69	18145	30281
Etang des Vallées	EV27	Winter 2012	504	455	0.56	0.66	59548	29181
Etang des Vallées	EV33	Spring 2012	447	256	0.66	0.59	17245	17838
Etang des Vallées	EV39	Summer 2012	616	422	0.61	0.68	111831	40565
Etang des Vallées	EV45	Autumn2012	431	549	0.6	0.74	26567	24414
Etang des Vallées	EV51	Winter 2013	551	860	0.62	0.68	37214	21019
Etang des Vallées	EV57	Spring 2013	394	398	0.57	0.58	29608	33870
Etang des Vallées	EV59	Summer 2013	683	248	0.66	0.63	19744	109886
Etang des Vallées	EV61	Autumn2013	425	379	0.77	0.58	36234	102453
Etang des Vallées	EV63	Winter 2014	548	627	0.63	0.67	27144	76322
Etang des Vallées	EV65	Spring 2014	426	391	0.57	0.65	14846	66358
Etang des Vallées	EV67	Summer 2014	541	444	0.69	0.75	23324	88235
Etang des Vallées	EV69	Autumn2014	637	444	0.62	0.65	26184	32412
Etang des Vallées	EV71	Winter 2015	353	1063	0.52	0.75	32571	72473
Etang des Vallées	EV73	Spring 2015	289	336	0.51	0.71	27944	110056
Etang des Vallées	EV75	Summer 2015	271	279	0.47	0.7	23712	112734
Etang des Vallées	EV77	Autumn2015	414	560	0.66	0.73	23494	75869
Etang des Vallées	EV79	Winter 2016	369	596	0.58	0.7	30739	116208
Etang des Vallées	EV81	Spring 2016	404	798	0.67	0.74	31314	84705
Etang des Vallées	EV83	Summer 2016	397	491	0.57	0.7	26524	69645
Etang des Vallées	EV85	Autumn2016	216	537	0.33	0.67	39584	75731
Etang des Vallées	EV87	Winter 2017	559	852	0.67	0.73	22961	63152
Etang des Vallées	EV89	Spring 2017	543	696	0.67	0.71	24952	36879
Etang des Vallées	EV91	Summer 2017	416	415	0.63	0.69	29082	85530
Etang des Vallées	EV93	Autumn2017	581	487	0.65	0.67	17942	49119
Etang des Vallées	EV95	Winter 2018	545	1019	0.67	0.73	34825	79286
Etang des Vallées	EV97	Spring 2018	539	661	0.66	0.72	15457	50173
Etang des Vallées	EV99	Summer 2018	641	331	0.71	0.63	57537	23059
Etang des Vallées	EV101	Autumn2018	406	284	0.57	0.64	103792	38041
Etang des Vallées	EV103	Winter 2019	390	345	0.56	0.69	66247	24524
La Claye	LC19	Winter 2012	231	1389	0.4	0.72	28769	14309
La Claye	LC25	Spring 2012	190	677	0.33	0.66	29423	11325

*Continued on next page*

Supplementary Table 5.A.2 – *Continued from previous page*

Ecosystem	Sample name	Sampling date	Microbial eukaryotes	Prokaryotes	Microbial eukaryotes	Prokaryotes	Microbial eukaryotes	Prokaryotes
La Claye	LC31	Summer 2012	505	515	0.62	0.59	32211	15866
La Claye	LC37	Autumn2012	796	1327	0.67	0.79	46746	37733
La Claye	LC43	Winter 2013	143	1066	0.32	0.67	36155	37400
La Claye	LC49	Spring 2013	366	605	0.42	0.56	38697	40415
La Claye	LC51	Summer 2013	319	266	0.34	0.58	36646	14093
La Claye	LC53	Autumn2013	246	554	0.32	0.61	39862	13236
La Claye	LC55	Winter 2014	227	401	0.35	0.65	25860	7339
La Claye	LC57	Spring 2014	436	379	0.41	0.65	19111	10725
La Claye	LC59	Summer 2014	301	363	0.31	0.65	23906	7701
La Claye	LC61	Autumn2014	138	461	0.18	0.6	34773	10435
La Claye	LC63	Winter 2015	215	458	0.38	0.62	51898	8690
La Claye	LC65	Spring 2015	212	239	0.27	0.66	60283	17091
La Claye	LC67	Summer 2015	583	439	0.61	0.66	55519	19194
La Claye	LC69	Autumn2015	169	763	0.31	0.66	65171	8646
La Claye	LC71	Winter 2016	203	544	0.37	0.61	45526	10264
La Claye	LC73	Spring 2016	512	772	0.47	0.7	40017	12946
La Claye	LC75	Summer 2016	516	733	0.42	0.69	35912	15614
La Claye	LC77	Autumn2016	132	658	0.25	0.61	36085	5827
La Claye	LC79	Winter 2017	133	643	0.24	0.59	42357	11439
La Claye	LC81	Spring 2017	329	547	0.4	0.6	18993	8362
La Claye	LC83	Summer 2017	383	602	0.43	0.72	43407	16490
La Claye	LC85	Autumn2017	342	430	0.5	0.68	29459	17426
La Claye	LC87	Winter 2018	304	482	0.56	0.65	41283	14587
La Claye	LC89	Spring 2018	252	387	0.43	0.64	29156	13325
La Claye	LC91	Summer 2018	462	495	0.57	0.68	95850	36894
La Claye	LC93	Autumn2018	105	979	0.24	0.61	78077	25943
La Claye	LC95	Winter 2019	93	180	0.36	0.65	78743	25804
Mare Gabard	MG13	Autumn 2011	273	300	0.56	0.64	29570	33251
Mare Gabard	MG19	Winter 2012	353	801	0.46	0.65	30513	27841
Mare Gabard	MG31	Summer 2012	89	306	0.12	0.6	35999	13458
Mare Gabard	MG37	Autumn2012	249	376	0.44	0.61	13083	14304
Mare Gabard	MG43	Winter 2013	615	1331	0.74	0.77	28658	28110
Mare Gabard	MG49	Spring 2013	803	1707	0.83	0.89	35772	29403
Mare Gabard	MG51	Summer 2013	157	642	0.32	0.63	33005	11059
Mare Gabard	MG53	Autumn2013	194	451	0.4	0.64	42706	14541
Mare Gabard	MG55	Winter 2014	126	593	0.43	0.59	40263	15257
Mare Gabard	MG57	Spring 2014	326	405	0.61	0.51	31948	11837
Mare Gabard	MG59	Summer 2014	242	287	0.41	0.42	25177	11850
Mare Gabard	MG61	Autumn2014	168	528	0.41	0.68	32786	10611

*Continued on next page*

Supplementary Table 5.A.2 – *Continued from previous page*

Ecosystem	Sample name	Sampling date	Microbial eukaryotes	Prokaryotes	Microbial eukaryotes	Prokaryotes	Microbial eukaryotes	Prokaryotes
Mare Gabard	MG63	Winter 2015	698	1686	0.62	0.83	31646	31487
Mare Gabard	MG65	Spring 2015	178	706	0.35	0.58	33847	9694
Mare Gabard	MG67	Summer 2015	115	433	0.34	0.62	19380	5574
Mare Gabard	MG69	Autumn2015	396	889	0.53	0.73	13969	16743
Mare Gabard	MG71	Winter 2016	252	578	0.47	0.62	22180	38507
Mare Gabard	MG73	Spring 2016	298	1571	0.39	0.84	41631	39392
Mare Gabard	MG75	Summer 2016	299	567	0.51	0.59	34167	15802
Mare Gabard	MG77	Autumn2016	234	768	0.36	0.66	29030	26367
Mare Gabard	MG79	Winter 2017	110	549	0.21	0.65	39891	8107
Mare Gabard	MG81	Spring 2017	209	362	0.44	0.58	22560	24603
Mare Gabard	MG83	Summer 2017	313	463	0.48	0.61	24271	25315
Mare Gabard	MG85	Autumn2017	505	546	0.62	0.66	22219	43143
Mare Gabard	MG87	Winter 2018	1175	1789	0.81	0.9	22371	33327
Mare Gabard	MG89	Spring 2018	542	1317	0.62	0.81	29749	7966
Mare Gabard	MG91	Summer 2018	369	714	0.55	0.69	85575	27294
Mare Gabard	MG93	Autumn2018	454	1628	0.51	0.87	36701	42646
Mare Gabard	MG95	Winter 2019	266	901	0.57	0.6	135979	25057
Saint Robert	SR13	Autumn 2011	285	445	0.53	0.63	25111	16425
Saint Robert	SR19	Winter 2012	85	355	0.31	0.59	38135	20340
Saint Robert	SR31	Summer 2012	185	507	0.17	0.58	30169	21676
Saint Robert	SR37	Autumn2012	553	445	0.62	0.59	42586	19642
Saint Robert	SR43	Winter 2013	58	709	0.26	0.58	32971	17410
Saint Robert	SR49	Spring 2013	309	360	0.43	0.52	23796	7167
Saint Robert	SR51	Summer 2013	100	555	0.26	0.54	24579	16250
Saint Robert	SR53	Autumn2013	297	494	0.33	0.65	28059	94042
Saint Robert	SR55	Winter 2014	184	425	0.33	0.61	31205	83285
Saint Robert	SR57	Spring 2014	380	506	0.57	0.72	34121	108027
Saint Robert	SR59	Summer 2014	234	392	0.3	0.7	43521	165410
Saint Robert	SR61	Autumn2014	180	576	0.29	0.68	10244	9514
Saint Robert	SR63	Winter 2015	115	496	0.41	0.59	37928	15635
Saint Robert	SR65	Spring 2015	297	501	0.61	0.61	41989	17776
Saint Robert	SR67	Summer 2015	237	488	0.43	0.64	17798	13340
Saint Robert	SR69	Autumn2015	634	553	0.69	0.71	25476	22615
Saint Robert	SR71	Winter 2016	559	508	0.58	0.49	14963	28902
Saint Robert	SR73	Spring 2016	688	402	0.75	0.57	13086	28785
Saint Robert	SR75	Summer 2016	278	251	0.37	0.45	13515	12356
Saint Robert	SR79	Winter 2017	161	542	0.18	0.62	31125	21703
Saint Robert	SR81	Spring 2017	440	519	0.42	0.62	35902	30775
Saint Robert	SR83	Summer 2017	63	285	0.19	0.43	28026	19945

*Continued on next page*

Supplementary Table 5.A.2 – *Continued from previous page*

Ecosystem	Sample name	Sampling date	Microbial eukaryotes	Prokaryotes	Microbial eukaryotes	Prokaryotes	Microbial eukaryotes	Prokaryotes
Saint Robert	SR85	Autumn2017	427	477	0.48	0.55	42133	27381
Saint Robert	SR87	Winter 2018	707	925	0.62	0.66	34798	23653
Saint Robert	SR89	Spring 2018	191	481	0.47	0.65	29093	21339
Saint Robert	SR91	Summer 2018	314	807	0.37	0.64	135749	44918
Saint Robert	SR93	Autumn2018	257	502	0.47	0.63	94514	27426
Saint Robert	SR95	Winter 2019	357	777	0.45	0.66	35114	74629
Ru Sainte Anne	RSA13	Autumn 2011	844	1051	0.68	0.7	40896	16332
Ru Sainte Anne	RSA19	Winter 2012	772	1549	0.74	0.89	30543	33105
Ru Sainte Anne	RSA25	Spring 2012	685	2047	0.64	0.9	97799	32438
Ru Sainte Anne	RSA31	Summer 2012	989	1794	0.74	0.85	71297	27848
Ru Sainte Anne	RSA37	Autumn2012	989	1743	0.74	0.87	106188	30901
Ru Sainte Anne	RSA43	Winter 2013	861	1895	0.79	0.93	31039	22101
Ru Sainte Anne	RSA49	Spring 2013	1183	1963	0.8	0.93	37867	11333
Ru Sainte Anne	RSA51	Summer 2013	914	1970	0.67	0.84	30327	16914
Ru Sainte Anne	RSA53	Autumn2013	838	2196	0.67	0.89	38809	8232
Ru Sainte Anne	RSA55	Winter 2014	893	2392	0.74	0.91	35840	9974
Ru Sainte Anne	RSA57	Spring 2014	1037	2583	0.89	0.92	36751	22664
Ru Sainte Anne	RSA59	Summer 2014	890	2607	0.62	0.92	28706	10301
Ru Sainte Anne	RSA61	Autumn2014	840	2173	0.69	0.86	45434	9673
Ru Sainte Anne	RSA63	Winter 2015	1097	2566	0.88	0.92	52190	21515
Ru Sainte Anne	RSA65	Spring 2015	1076	2408	0.8	0.89	79086	27093
Ru Sainte Anne	RSA69	Autumn2015	948	2062	0.66	0.87	22604	24762
Ru Sainte Anne	RSA71	Winter 2016	1014	2592	0.81	0.93	23844	30946
Ru Sainte Anne	RSA73	Spring 2016	1013	2437	0.84	0.91	27393	20606
Ru Sainte Anne	RSA75	Summer 2016	1007	2662	0.78	0.92	34334	11365
Ru Sainte Anne	RSA77	Autumn2016	805	2478	0.67	0.9	60866	12470
Ru Sainte Anne	RSA79	Winter 2017	1039	2372	0.74	0.9	35978	23461
Ru Sainte Anne	RSA81	Spring 2017	720	2209	0.62	0.87	6355	16243
Ru Sainte Anne	RSA85	Autumn2017	1031	2488	0.7	0.91	40405	16153
Ru Sainte Anne	RSA87	Winter 2018	1282	2717	0.86	0.93	32834	23249
Ru Sainte Anne	RSA89	Spring 2018	1284	2245	0.81	0.86	14129	8952
Ru Sainte Anne	RSA91	Summer 2018	535	1987	0.59	0.83	114424	25774
Ru Sainte Anne	RSA93	Autumn2018	710	2133	0.73	0.89	69341	41063
Ru Sainte Anne	RSA95	Winter 2019	1049	1842	0.79	0.88	69053	39177

**Supplementary Table 5.A.3** Total number of OTUs, mean number of OTUs and number of abundant OTUs per ecosystem for microbial eukaryotes and prokaryotes.

	Total number of OTUs		Mean number of OTUs		Number of abundant OTUs	
	Microbial eukaryotes	Prokaryotes	Microbial eukaryotes	Prokaryotes	Microbial eukaryotes	Prokaryotes
EV	5611	5651	462	520	101	115
LC	5197	7492	305	598	40	74
MG	5301	8110	345	800	65	95
SR	5532	5840	306	510	44	103
RSA	9558	18055	941	2184	129	101

**Supplementary Table 5.A.4** ANOSIM analysis between each pair of ecosystem based on Bray-Curtis dissimilarities computed after Wisconsin standardization of rarefied OTUs (9999 permutations). \*\*\* pvalue < 0.001

	Microbial eukaryotes		Prokaryotes	
	R	pvalue	R	pvalue
EV-LC	0.65	***	0.61	***
EV-MG	0.76	***	0.83	***
EV-SR	0.62	***	0.69	***
EV-RSA	0.99	***	0.95	***
LC-MG	0.45	***	0.59	***
LC-SR	0.33	***	0.44	***
LC-RSA	0.75	***	0.74	***
MG-SR	0.49	***	0.73	***
MG-RSA	0.71	***	0.82	***
SR-RSA	0.56	***	0.83	***

**Supplementary Table 5.A.5** Permutational multivariate analyses of variance (PERMANOVA) calculated using Wisconsin standardization on rarefied OTUs for microbial eukaryotes and prokaryotes. The tested factors were Habitat (lentic or lotic), Ecosystem (location), Season and Year.

	Microbial eukaryotes		Prokaryotes	
	r <sup>2</sup> (%)	pvalue	r <sup>2</sup> (%)	pvalue
Habitat	2.64	0.0002	3.90	0.0005
Ecosystem   Habitat	4.41	0.0002	6.00	0.0005
Season	2.78	0.0002	2.74	0.0005
Year	5.45	0.0602	5.95	0.0005

**Supplementary Table 5.A.6** Linear models were computed based on the time-lag (TLA) calculated between sampling dates over increasing time-lags over the 8-year survey (from one-month to 30-month) for the five ecosystems. TLA calculation was based for the microbial communities on Wisconsin Standardized Bray-Curtis dissimilarities of rarefied OTUs, and based on Euclidean distances for environmental parameters. Linear models were applied for physico-chemical parameters, prokaryotes (all community, abundant OTUs only, rare OTUs only), and microbial eukaryotes (all community, abundant OTUs only, rare OTUs only). Abundant OTUs were detected in each ecosystem in at least 20% of samples and with >0.1% of mean relative abundance per sample. Non-significant P values are highlighted in red.

	Abiotic parameters	Microbial eukaryotes	Prokaryotes	Abundant microbial eukaryotes	Rare microbial eukaryotes	Abundant prokaryotes	Rare prokaryotes
Etang des Vallées	lag	-0.003	0.022	0.055	0.087	0.014	0.122
	r <sup>2</sup>	-0.002	0.000	0.013	0.006	0.000	0.01
	pvalue	0.760	0.299	0.010	0.065	0.355	0.021
La Claye	lag	0.04	0.046	0.151	0.128	0.045	0.615
	r <sup>2</sup>	0.034	0.023	0.107	0.005	0.042	0.144
	pvalue	2.2e-16	0.001	2.2e-16	0.075	2.2e-16	2.2e-16
Mare Gabard	lag	0.013	0.04	0.084	0.113	0.025	0.044
	r <sup>2</sup>	0.005	0.009	0.023	0.01	0.006	-0.001
	pvalue	0.075	0.034	0.001	0.022	0.065	0.427
Saint Robert	lag	0.045	0.067	0.193	0.092	0.045	0.213
	r <sup>2</sup>	0.030	0.079	0.218	0.013	0.067	0.034
	pvalue	2.2e-16	2.2e-16	2.2e-16	0.016	2.2e-16	2.2e-16
Ru Sainte Anne	lag	0.016	0.064	0.235	0.211	0.045	0.687
	r <sup>2</sup>	0.006	0.022	0.199	0.039	0.022	0.283
	pvalue	0.065	0.002	2.2e-16	2.2e-16	0.002	2.2e-16

**Supplementary Table 5.A.7** Canonical correspondence analyses (CCA) conducted on all rarefied OTUs affiliated either to microbial eukaryotes or prokaryotes and on physico-chemical parameters in the samples collected during the 8-year survey in each ecosystem. A forward selection process was applied to only select the parameters explaining significantly community variance.

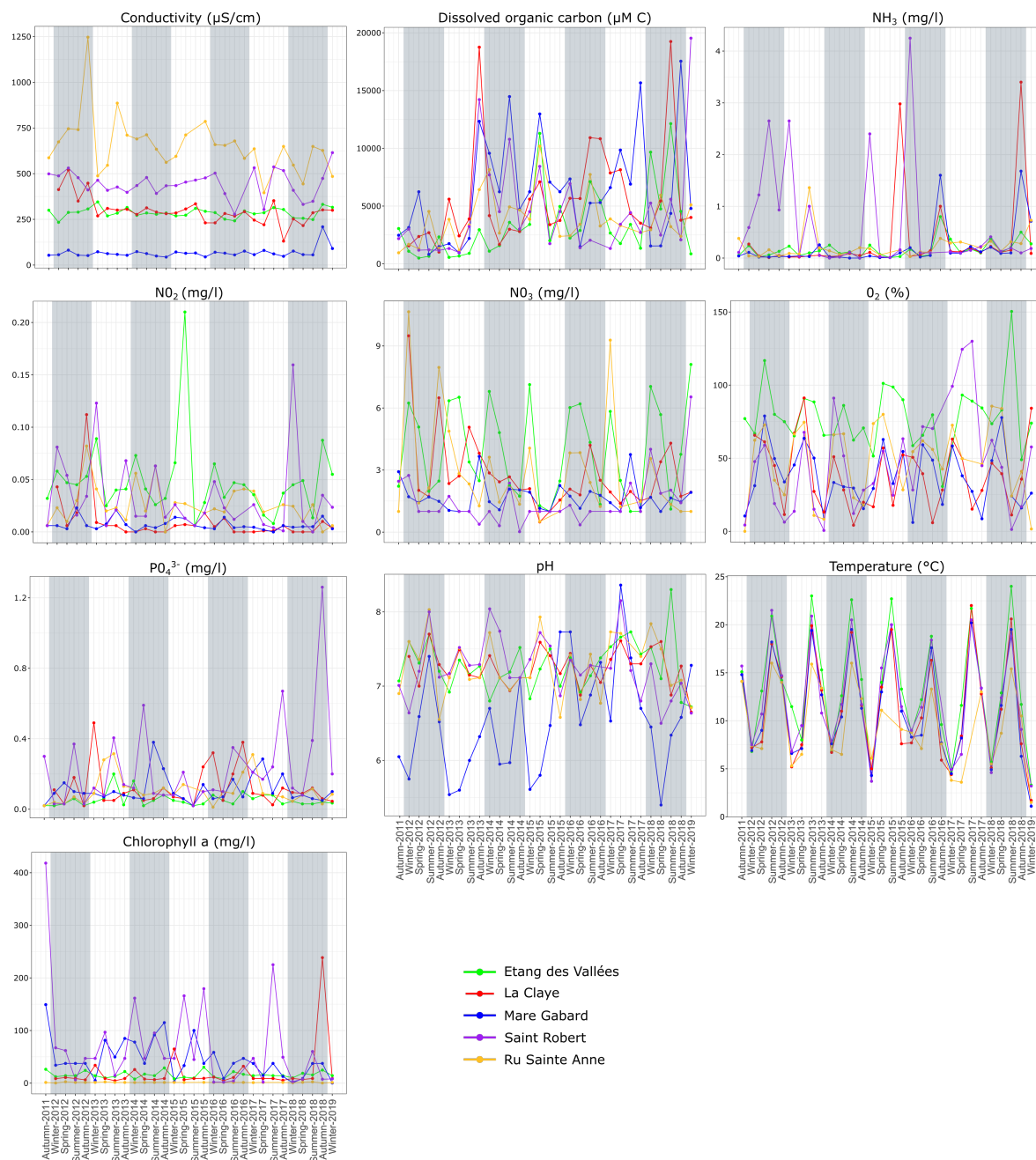
	Microbial eukaryotes		Prokaryotes	
	Explained variance	Parameters	Explained variance	Parameters
EV	6.8%	Temperature	12.9%	Temperature, O <sub>2</sub>
LC	5.1%	Temperature	20.0%	NO <sub>2</sub> , O <sub>2</sub>
MG	6.0%	Temperature	16.5%	Conductivity, O <sub>2</sub> , Temperature
SR	none	none	5.6%	O <sub>2</sub>
RSA	15.6%	O <sub>2</sub> , NO <sub>3</sub> , Conductivity	15.9%	Conductivity, O <sub>2</sub> , Temperature



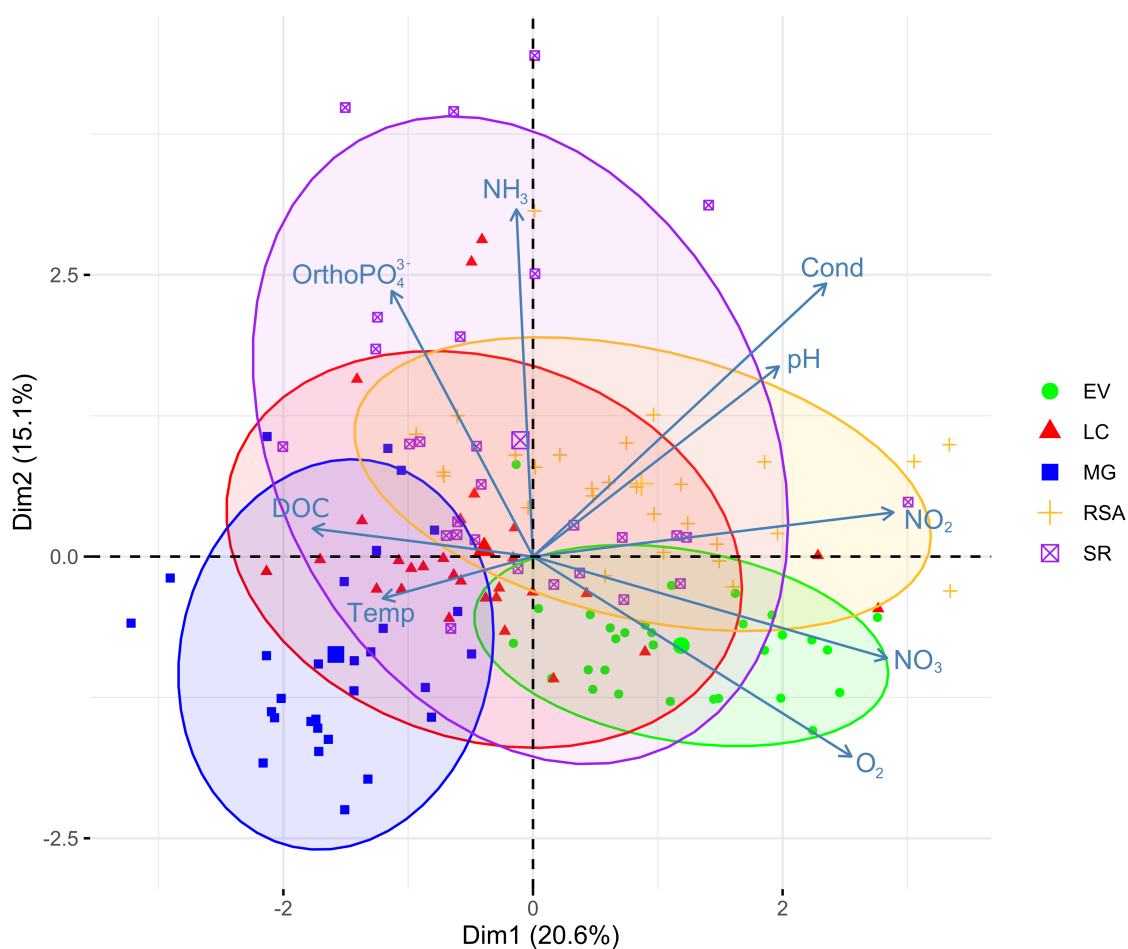
**Supplementary Table 5.A.8** ANOSIM analysis conducted for each ecosystems between each pair of season, based on Bray-Curtis dissimilarities computed after Wisconsin standardization of rarefied OTUs (9999 permutations). \*\*\* p-value < 0.001, \*\* p-value < 0.01, \* p-value < 0.05

	EV		LC		MG		SR		RSA	
	R	pvalue	R	pvalue	R	pvalue	R	pvalue	R	pvalue
Microbial eukaryotes										
Winter-Spring	0.58	***	0.17		0.12		-0.11		0.55	***
Winter-Summer	0.88	***	0.44	***	0.62	***	0.39	**	0.76	***
Winter-Autumn	0.64	***	0.15		0.39	**	0.23	*	0.53	***
Spring-Summer	0.68	**	0.40	***	0.46	**	0.41	***	0.85	***
Spring-Autumn	0.40	***	0.23	**	0.40	**	0.24	*	0.71	***
Summer-Autumn	0.33	***	0.17	*	0.17		0.30	*	0.21	*
Prokaryotes	R	pvalue	R	pvalue	R	pvalue	R	pvalue	R	pvalue
Winter-Spring	0.27	*	0.12		0.13		0.07		-0.04	
Winter-Summer	0.83	***	0.37	**	0.4	**	0.44	***	0.3	*
Winter-Autumn	0.64	***	0.32	**	0.3	**	0.32	*	0.11	
Spring-Summer	0.48	**	0.27	*	0.34	*	0.1		0.24	*
Spring-Autumn	0.32	***	0.3	**	0.58	***	0.28	*	0.13	
Summer-Autumn	0.15		0.07		0.05		0.14		-0.1	

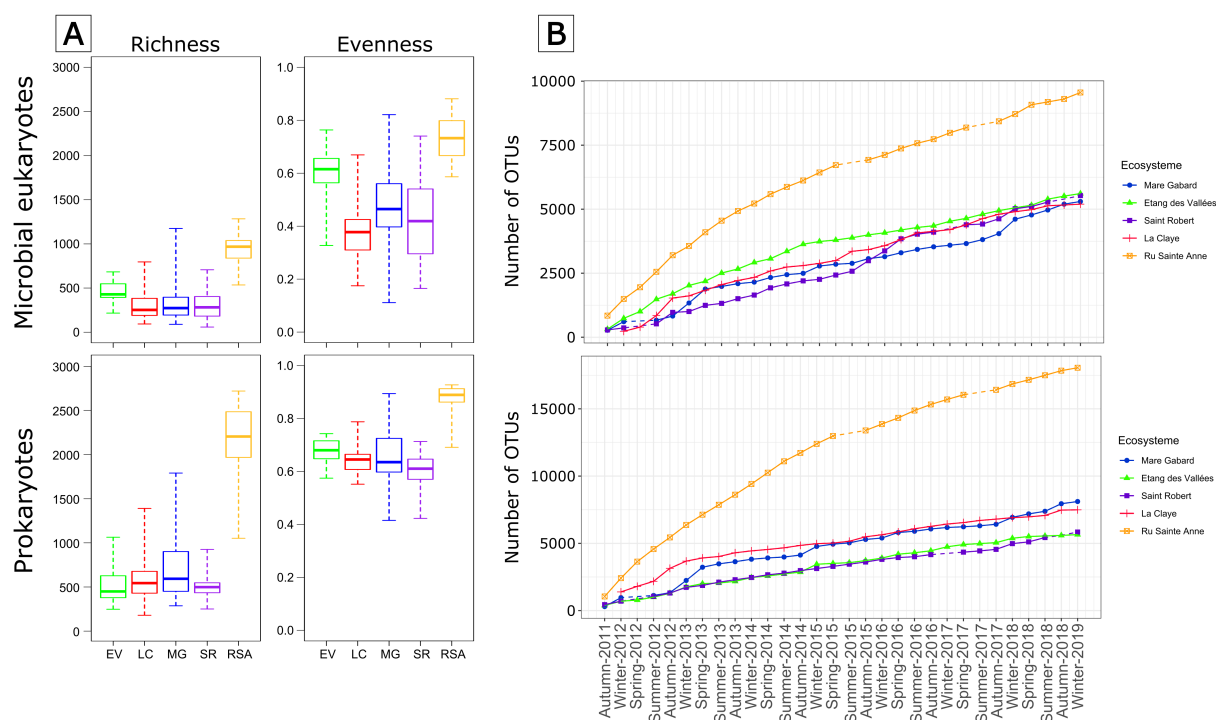
## 5.B Supplementary Figures



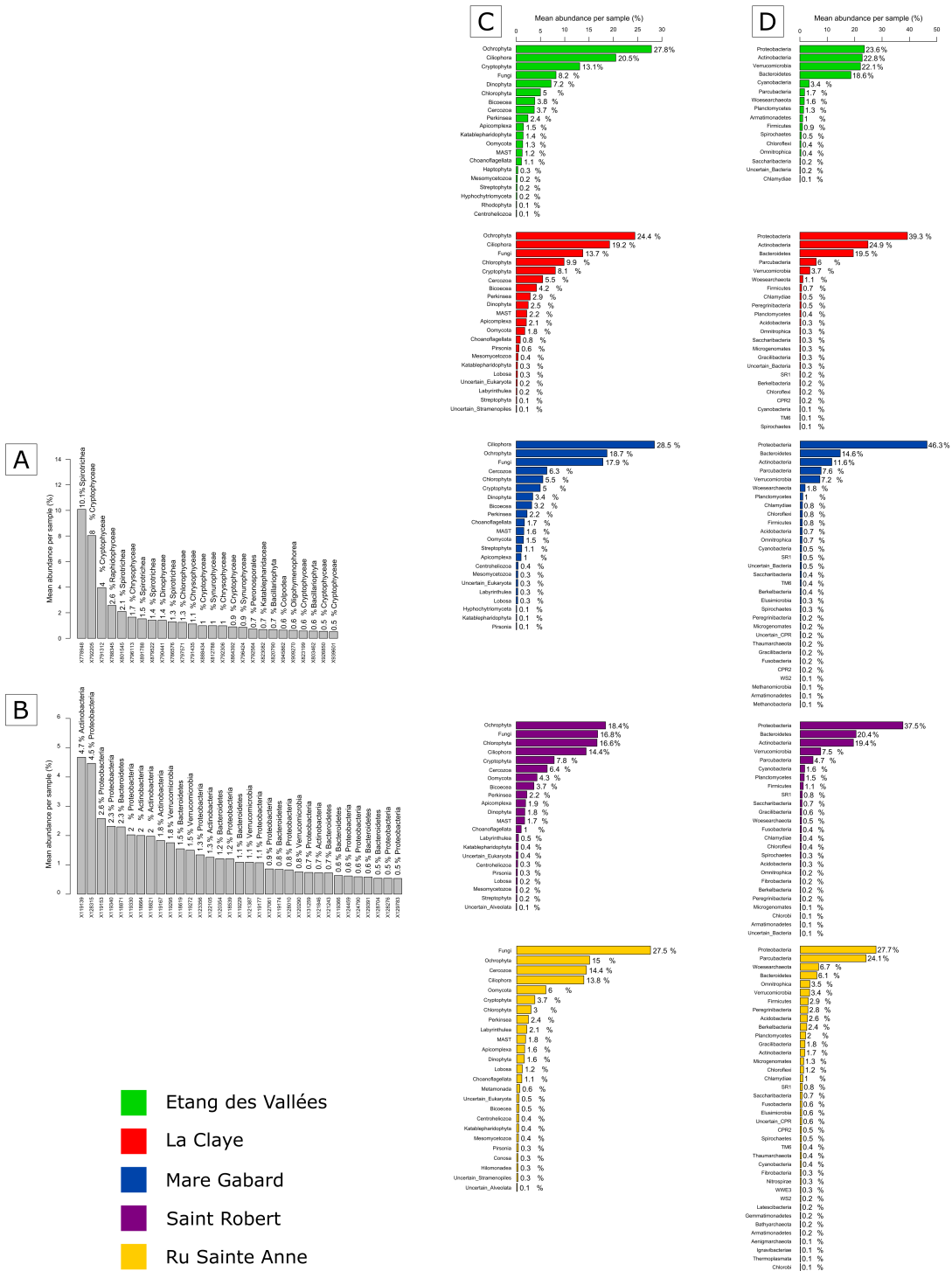
**Supplementary Figure 5.B.1** Temporal dynamics of the physico-chemical parameters and the chlorophyll a concentration measured over the 8-year survey. Temperature, conductivity, pH and  $\text{O}_2$  concentration were measured in situ. Nutrient concentrations ( $\text{NO}_3^-$ ,  $\text{NO}_2^-$ ,  $\text{NH}_3$ ,  $\text{PO}_4^{3-}$ ) were estimated on the sampling day with water pre-filtered through  $0.22 \mu\text{m}$  pore-sized filters. Dissolved organic carbon was estimated using water prefiltered through  $0.22 \mu\text{m}$  pore-sized filters. Chlorophyll a concentration was measured by spectroscopy on pigments extracted from cells harvested on filters. Dashed lines correspond to missing samples. Grey and white zones indicate the different sampling years.



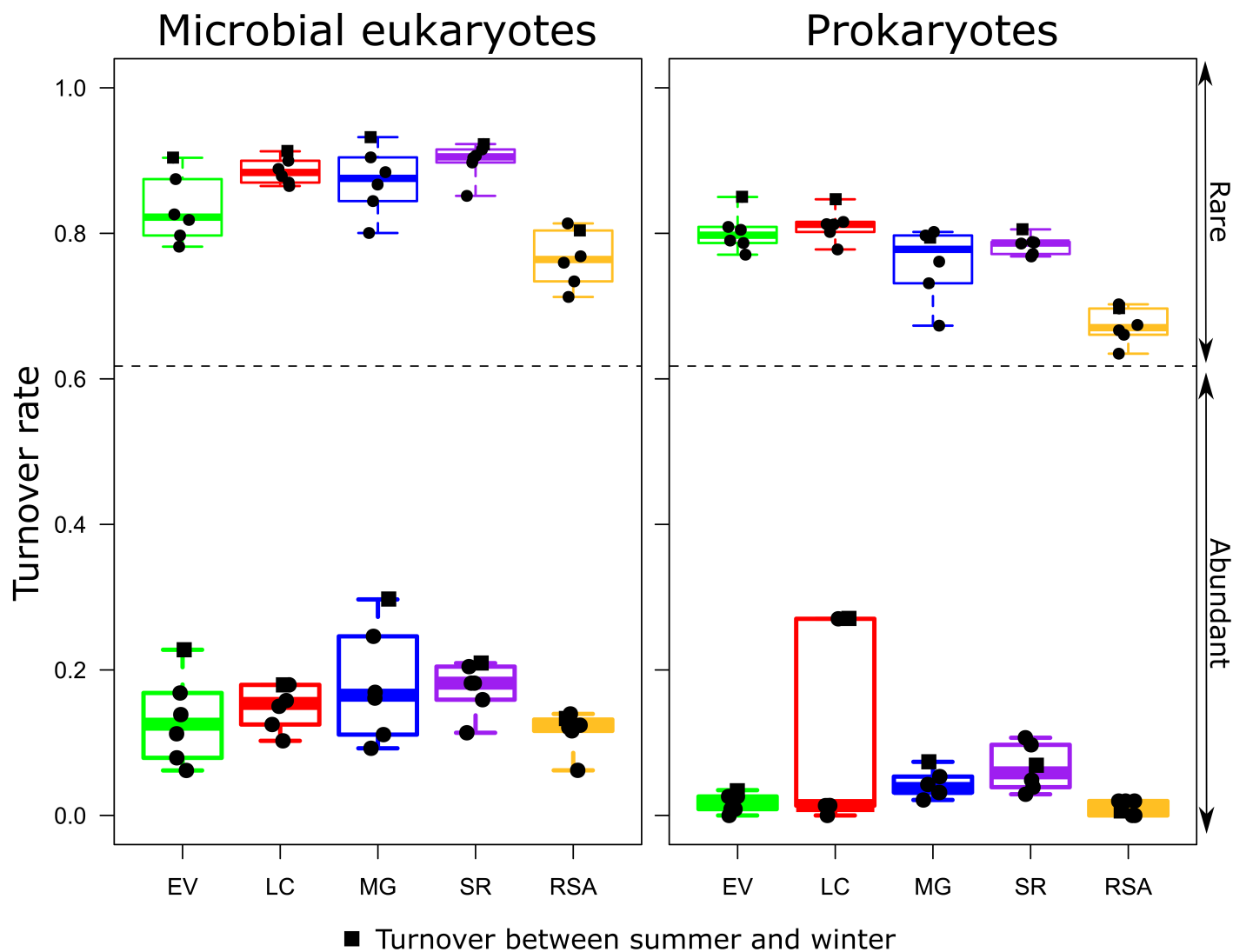
**Supplementary Figure 5.B.2** Principal Component Analysis (PCA) on all physico-chemical parameters measured in the five ecosystems during the 8-year survey. The shape and the color of the points indicate their sampling origin. The biggest points represent the barycentre of each ecosystem. Ellipses were drawn to show the clustering of samples according to their sampling ecosystem. DOC: Dissolved Organic Carbon; Temp: Temperature; Cond: Conductivity.



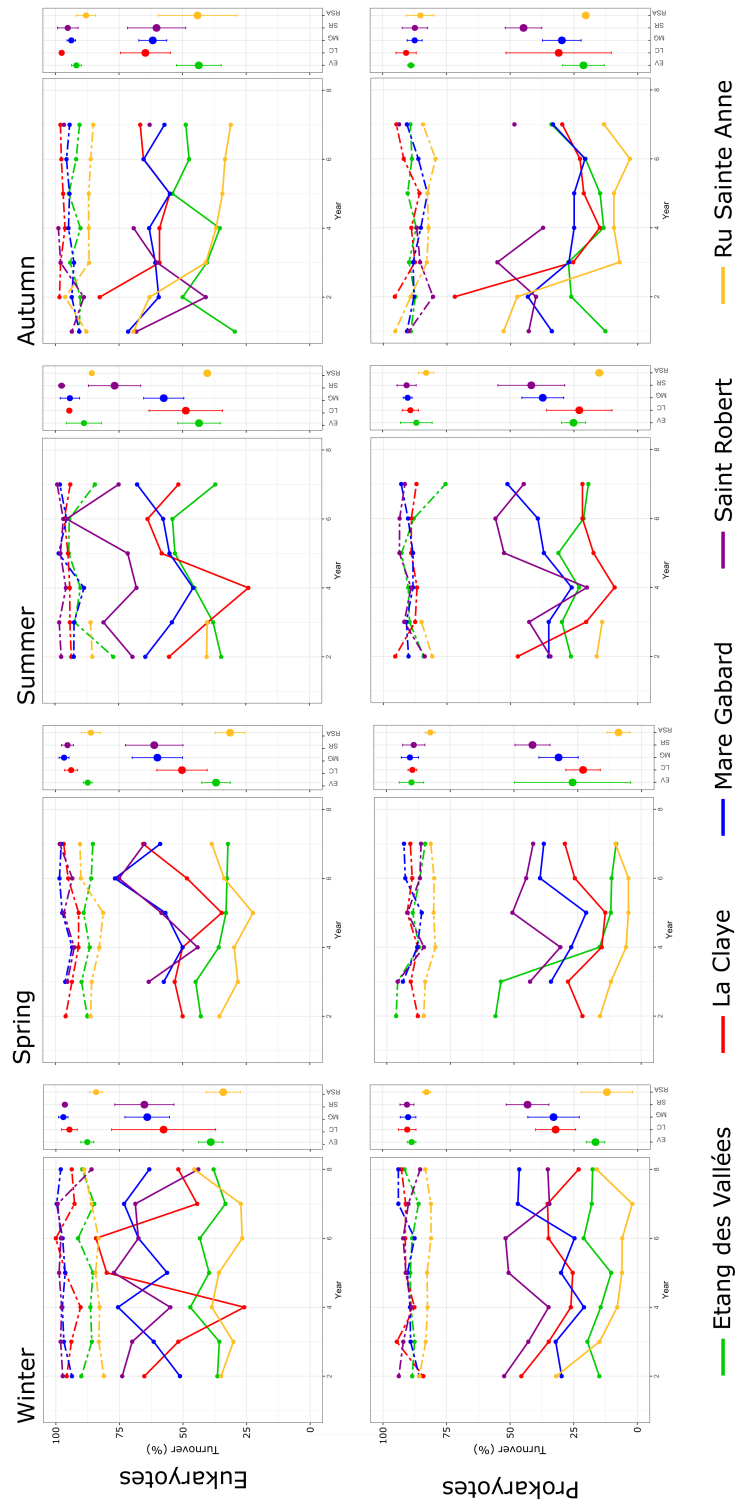
**Supplementary Figure 5.B.3** (A) Richness and evenness were computed on the rarefied microbial eukaryotic and prokaryotic communities in the samples collected during the 8-year survey in the five ecosystems (EV Etang des Vallées, LC La Claye, MG Mare Gabard, SR Saint Robert, RSA Ru Sainte Anne). The thickest line inside each box represents the median on the distribution, bottom and top borders of boxes correspond to the first and the third quartiles and whiskers extend to minimal and maximal distances. (B) Cumulative curves showing the total number of eukaryotic and prokaryotic rarefied new OTUs detected in each ecosystems from October 2011 (Autumn 2011) until January 2019 (Winter 2019).



**Supplementary Figure 5.B.4** Rank-abundance histograms of all rarefied OTUs in the entire dataset for microbial eukaryotes and prokaryotes, (B) with a zoom on the rarefied OTUs with a mean relative abundance of at least 0.5%. Their affiliation and their mean abundances are indicated at the top of the bar. (C) Rank-abundance histograms of taxonomic groups in each ecosystem presenting a mean relative abundance higher than 0.1% (indicated on the top of the bar) for eukaryotic and (D) prokaryotic communities of each ecosystems over the 8-year survey.

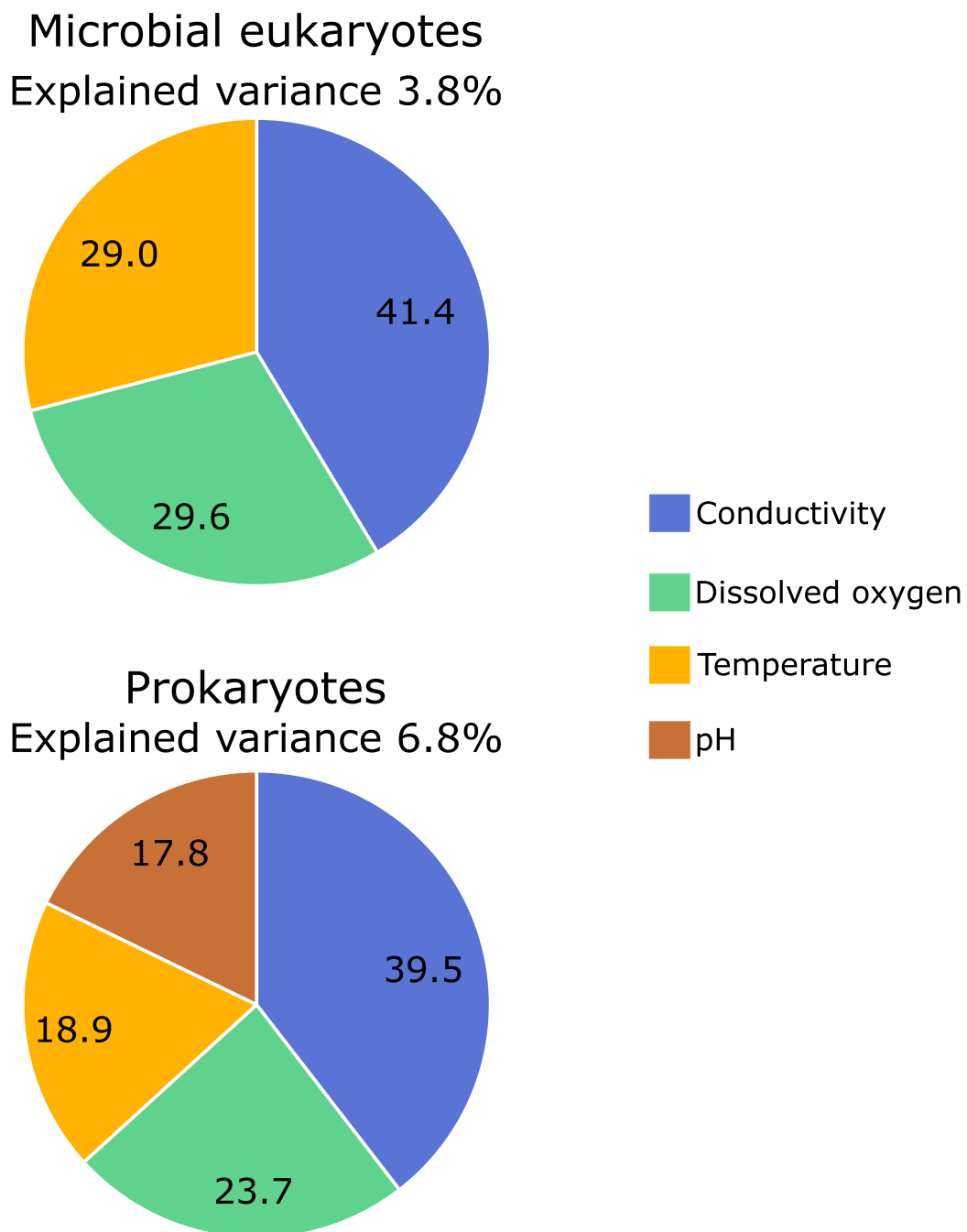


**Supplementary Figure 5.B.5** Box plots of turnover rates computed for rare and abundant rarefied OTUs, affiliated to microbial eukaryotes and prokaryotes in the five ecosystems between samples regrouped by season. Abundant OTUs were detected in each ecosystem in at least 20% of samples and with  $>0.1\%$  of mean relative abundance per sample. Rare OTUs were the rest of the community. The thickest line inside each box represents the median on the distribution, bottom and top borders of boxes correspond to the first and third quartiles and whiskers extend to minimal and maximal distances.



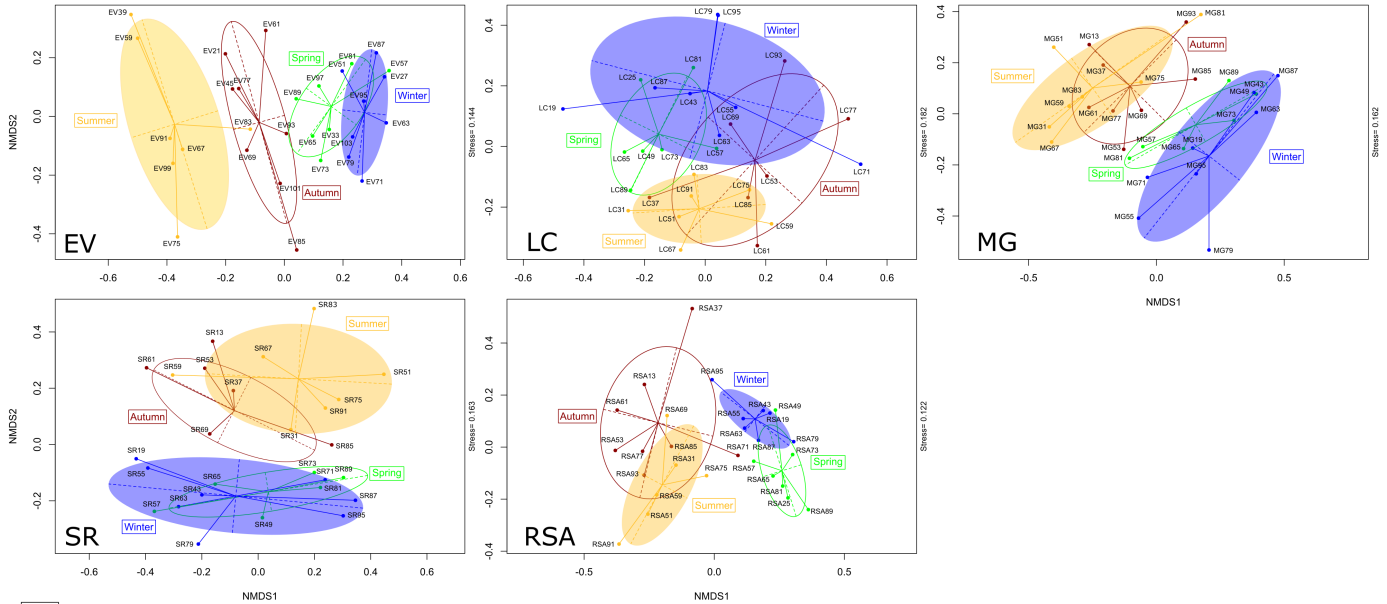
**Supplementary Figure 5.B.6** (A) Successive turnover rates computed for rare and abundant rarefied OTUs, affiliated to microbial eukaryotes and prokaryotes in the five ecosystems between samples regrouped by sampling years. Full lines correspond to abundant OTUs, and dashed lines to rare OTUs. (B) Mean value of the turnover rate for each ecosystem. The biggest points correspond to abundant OTUs. Abundant OTUs were detected in each ecosystem in at least 20% of samples and with >0.1% of mean relative abundance per sample. Rare OTUs were the rest of the community. Green elements correspond to samples collected in Etang des Vallées, red in La Claye, blue in Mare Gabard, purple in Saint Robert and yellow in Ru Sainte Anne.



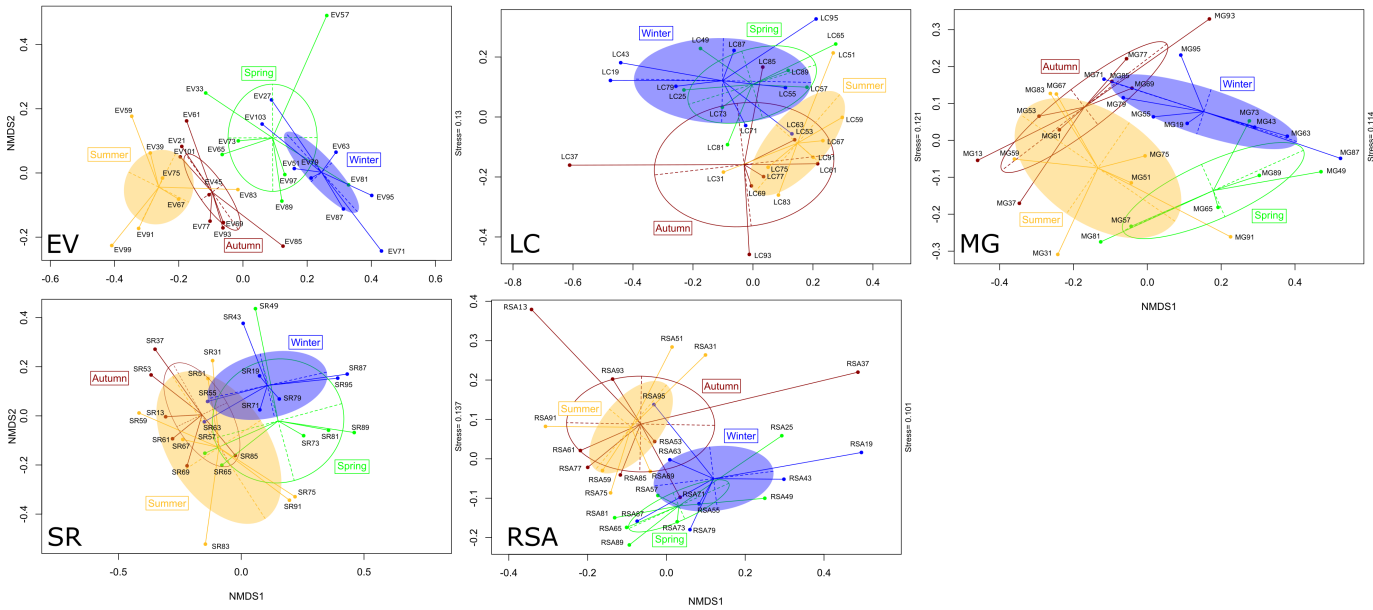


**Supplementary Figure 5.B.7** Pie-chart summarizing the results of Canonical correspondence analyses (CCA) conducted on all rarefied OTUs affiliated either to microbial eukaryotes or prokaryotes in all samples collected during the 8-year survey. A forward selection of parameters revealed that among measured environmental parameters, only Conductivity, dissolved oxygen, temperature were explained significantly both community variance. pH was only significant for the prokaryotic community. Values within each section of the pie charts indicate the proportion of the total explained variance represented by each parameter.

**A Microbial eukaryotes**



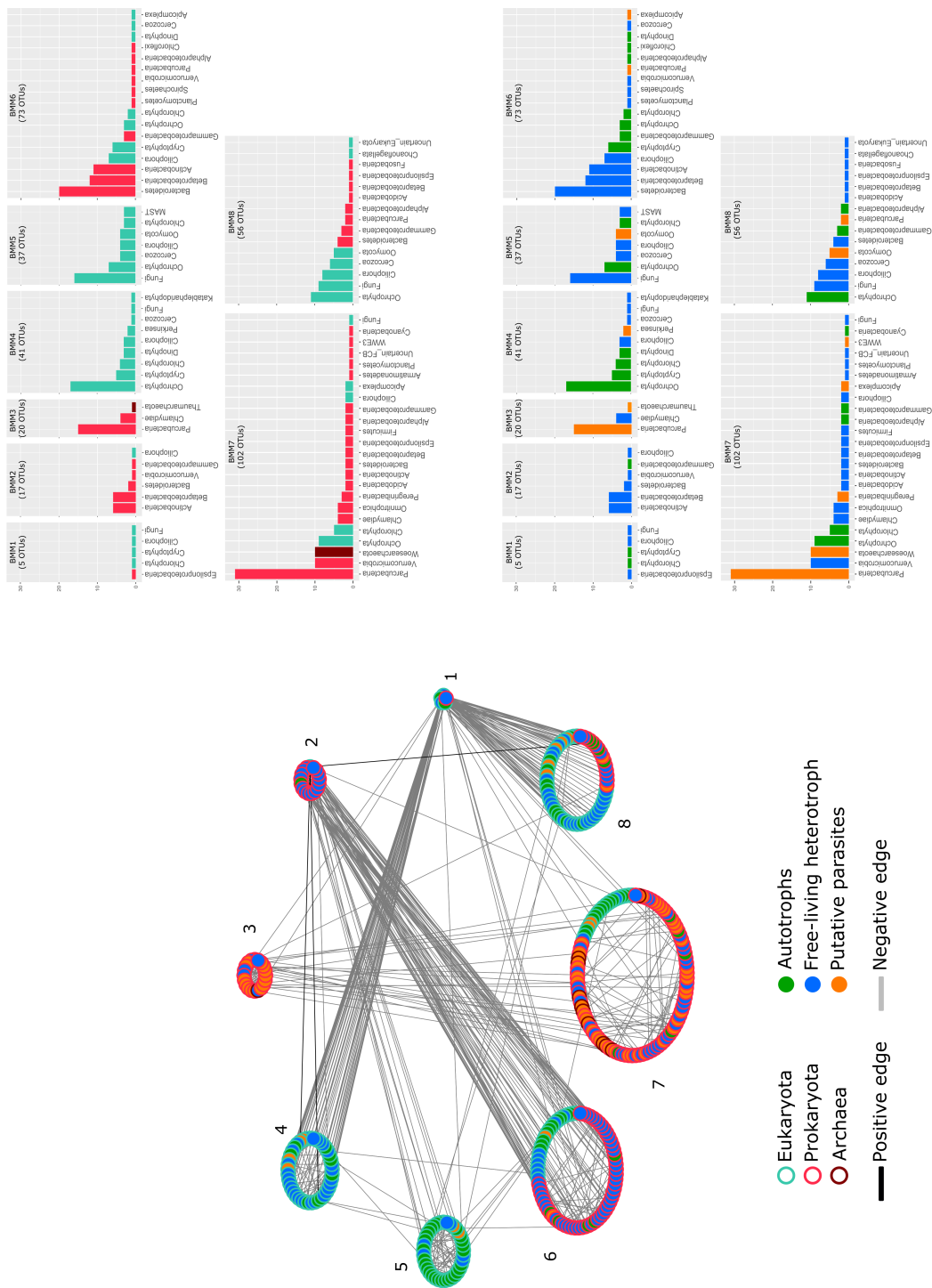
**B Prokaryotes**



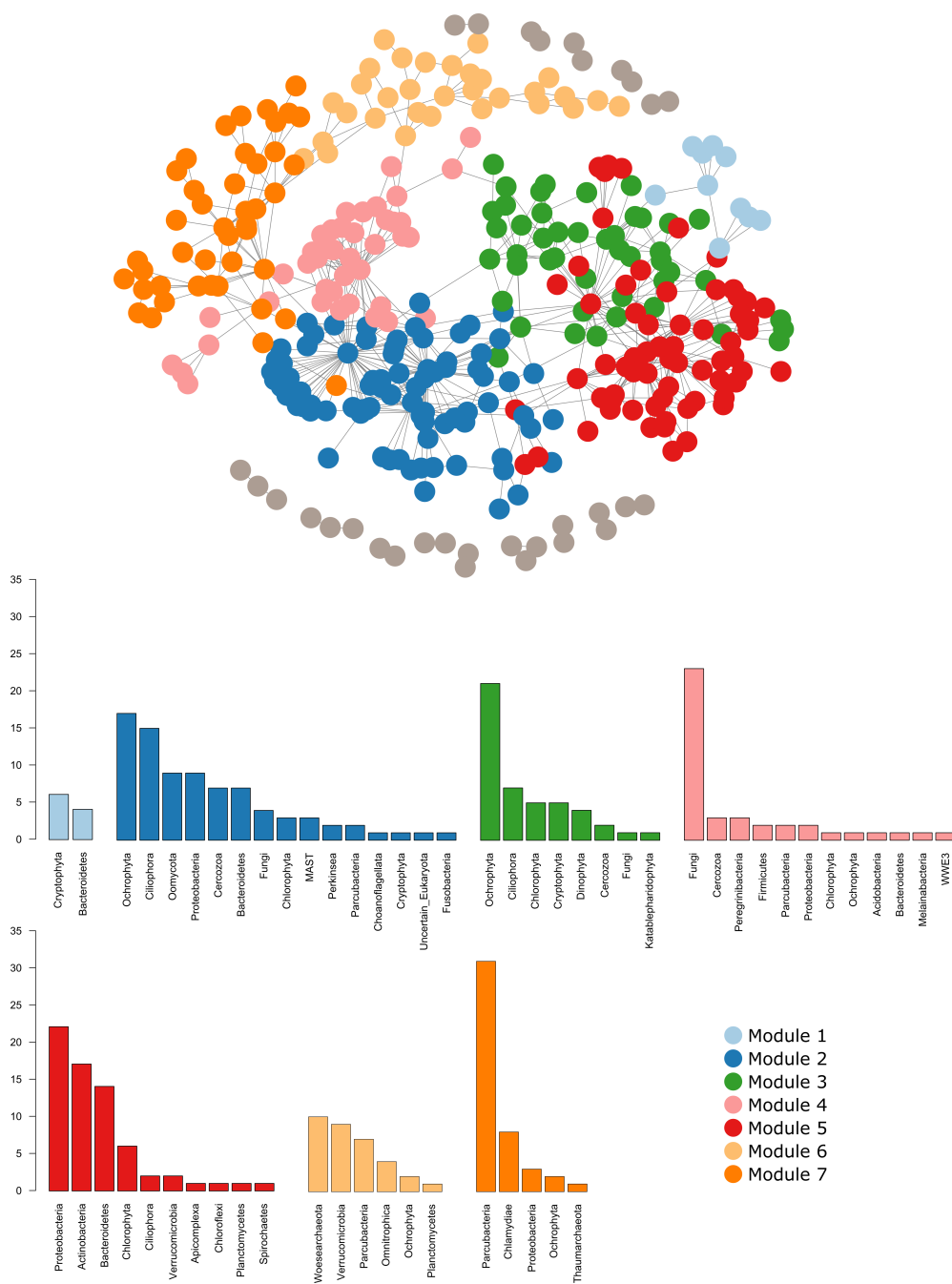
**Supplementary Figure 5.B.8** Non-metric Multi-Dimensional Scaling (NMDS) plots, based on pair-wise Bray-Curtis dissimilarities computed after Wisconsin standardization of rarefied OTU in samples for each of the five ecosystems for the microbial eukaryotes (A) and the prokaryotes (B). Samples are group by sampling season. Stress value of each NMDS is displayed on the right border of the corresponding plot. The table indicated the pairwise ANOSIM results with 9999 permutations between sampling season, i.e. the strength of the differences between seasons, and their significances.



**Supplementary Figure 5.B.9** Histograms of the recurrent OTUs per ecosystem in terms of diversity and abundance. Recurrence index was conducted on all rarefied OTUs, at the exception of OTUs with a mean relative abundance  $<0.01\%$  or present less than 3 times in the ecosystem survey. The recurrent OTUs were then regroup according to their taxonomic affiliation. The bars indicate the relative abundance in the entire dataset of the recurrent OTUs within a taxonomic group. The points represent the number of recurrent OTUs represented within a taxonomic group. (A) Microbial eukaryotic recurrent OTUs. The color of the bar represents the affiliation of the taxonomic group to potential parasites, heterotrophs and primary producers. (B) Prokaryotic recurrent OTUs. The color of the bar indicates the affiliation of the taxonomic group to a higher super-group. CPR: Candidate Phyla Radiation; FCB: Fibrobacteres-Chlorobi-Bacteroidetes; PVC: Planctomycetes-Verrucomicrobia-Chlamydiae.



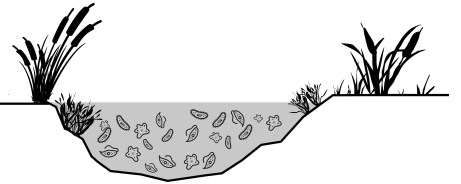
**Supplementary Figure 5.B.10** Block-model representation of the co-occurrence network inferred by the PLN approach and representing the global structure of microorganisms over the five ecosystems. The network is build based on the OTUs present in 35% of the samples, and with a mean abundance higher than 0.01%. The sequencing depth was added as an offset of the model, and the sampling site and the temperature were considered as covariables. The color of the nodes indicates the trophic mode of the OTU, and the outline indicates the affiliation of the OTU. Barplots present the OTU composition of each block model group (BMG), with colors indicating either trophic mode or taxonomic affiliation.



**Supplementary Figure 5.B.11** Module representation of the co-occurrence network inferred by the PLN approach and representing the global structure of microorganisms over the five ecosystems. Barplots are presenting the composition of the modules (number of OTUs for each taxonomic group). Grey nodes are not included in a particular module.

## Chapter 6

Topology of seasonal co-occurrence networks reveals similar community structuration in different small freshwater ecosystems





## 6.1 Context and objective

The study of microbial communities in five small freshwater ecosystems with different temporal frequency and duration allowed me to discover a large diversity of microorganisms. These inhabitants of these ecosystems presented recurrent seasonal patterns as well as constant directional changes of community composition. To explain these dynamics, I previously focused on the abiotic parameters that I measured in these ecosystems. However, these parameters were not enough to explain most of the variation of microbial community composition. Therefore, I decided to focus on biotic parameters, that is the interactions between microorganisms. The idea was to investigate the structure of communities at each season, in each ecosystem. To do so, I inferred co-occurrence networks based on the most abundant microorganisms, using the PLNmodels approach (Chiquet et al., 2019). The objectives of these exploratory studies were to:

1. Determine the structure of co-occurrence networks in each ecosystem during the four seasons.
2. Establish if the seasonal structuration of communities is similar between seasons despite dissimilar communities.
3. Detect the topological properties that are specific to each seasonal network.



## 6.2 Manuscript

Gwendoline M. David<sup>1</sup>, Purificación López García<sup>1</sup>, Ludwig Jardillier<sup>1</sup>, David Moreira<sup>1</sup>, and Elisa Thébaud<sup>2</sup>

<sup>1</sup>Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, 91405, Orsay, France

<sup>2</sup>Sorbonne Université, UPEC, CNRS, IRD, INRAE, Institute d'Ecologie de des Sciences de l'Environnement de Paris, iEES-Paris, F-75005 Paris, France

Article in preparation.

### 6.2.1 Abstract

Microbial communities in small freshwater ecosystems present strong seasonal variations in their composition. At the same time, microorganisms are known to be in constant interactions with their environment and other organisms, through beneficial or antagonistic relationships. In the present study, we used an eight-years survey to investigate the seasonal structure of microbial communities in five small freshwater ecosystems. We previously observed that abiotic factors did not fully explain variations in community composition. Therefore, we decided to focus on the interactions among microorganisms based on metabarcoding of 16S and 18S rRNA genes. We inferred interaction networks for each season in each ecosystem, using the most abundant microbial operational taxonomic units (OTUs, occurring at an abundance  $> 0.05\%$  in more than 70% of the samples). We also compared the topological properties of each network to find structuring patterns of microbial communities. Summer networks were more complex, whereas their winter counterparts suggested a preponderant role of parasitic microorganisms. This study provides a first insight into the seasonal structuring of microbial communities and underlines the potential importance of particular groups in small freshwater ecosystems.

### 6.2.2 Introduction

Microorganisms are highly diverse, widely distributed and play key roles in all biogeochemical cycles (Salcher, 2014; Worden et al., 2015). Similar to macroorganisms, the composition of microbial communities is controlled by bottom-up and top-down factors as well as environmental

parameters (Pernthaler, 2005; Worden et al., 2015). Microbial communities succeed to each other over seasons leading to permanent reassembling (Gilbert et al., 2012) even at time scales as short as 2-3 days (Mangot et al., 2013), with communities becoming dominant thanks to optimal conditions of growth (Treusch et al., 2009). Because of their important functional roles, any change in microbial community composition can have in return profound impacts on the functioning of their ecosystem. Identifying structuring patterns at different time and spatial scales is thus of primary importance. It is now well acknowledged that resources, mortality factors (predators and parasites, including viruses) and environmental parameters control significantly both the biomass and the composition of microbial communities (Jones and McMahon, 2009). However, the variation in microbial composition that remains unexplained is surprisingly large even when taking into account several of these factors (Muylaert et al., 2002; Steele et al., 2011; Lima-Mendez et al., 2015). This may result from the difficulties to identify the multiple and complex interactions between microorganisms that may drive microbial taxa temporal dynamics (Steele et al., 2011). It is therefore crucial to investigate further biotic interactions between microorganisms to better decipher the community structuring over space and time. These interactions are diverse and include for example competition for resources, predation, parasitism and cooperation. For instance, algae are grazed by zooplankton and release dissolved organic carbon used by prokaryotes (Azam and Malfatti, 2007), heterotrophic protists are known to be major predators of prokaryotes (Sherr and Sherr, 1994) and bacteria of the Candidate Phyla Radiation (CPR) are suspected to have a symbiosis lifestyle because of their reduced genome (Castelle and Banfield, 2018). These specific ecological relationships between microbial taxa remain mostly unknown because it is much more difficult to identify them *in vivo* for microorganisms than macroorganisms. Natural microbial communities are very complex, composed of cells small in size that make their identification difficult and belong to a tremendous number of taxa that can interact between each other in multiple ways. A culture based approach could allow identifying interactions between microorganisms and revealing the precise mechanisms involved in these interactions. Such an approach is highly limited by the low diversity of cultivated microorganisms (<1%) and results would not necessarily reflect what happens in the natural environment. Recently developed single cell genomic based approaches, based on the sequencing of genes or transcripts, considerably expanded our understanding of the processes taking place in the interactions between microorganisms. However, the advances brought by these two approaches are restricted to decipher mechanisms of a limited number of microbial interactions. A first step screening at the community scale for particular interactions of potential evolutionary or ecological interest was missing. Co-occurrence networks, originally developed for macroorganisms to identify interactions based on direct observations, gained in interest in microbial ecology over the last years thanks to the rapid development of massive sequencing technologies. Applied to metabarcoding dataset, it is now possible to infer all type of

microbial interactions at multiple scales (e.g. community versus populations, intra versus inter-annual). In combination with the analysis of the topology of co-occurrence networks, it is now possible to explore structuring of complex microbial communities over time and space. Over the last decade, diverse methods based on co-occurrence analyses allowed better apprehending community structuring in various types of environments, from soil to marine systems (Barberán et al., 2012; Genitsaris et al., 2016; Röttjers and Faust, 2018). In lakes for instance, the impact of phytoplankton blooms on community composition is now better understood (Li et al., 2015; Escalas et al., 2019). Usually, particular groups of microbes attracted higher interest, such as Actinobacteria (e.g. Eiler et al., 2012), to the detriment of a global understanding of microbial community structuring (e.g. Bock et al., 2020). Also, the study of the structure of microbial communities is essentially conducted in large ecosystems, such as Oceans and lakes. Ecosystems smaller in size are neglected although they may contribute significantly to the cycles of CO<sub>2</sub> and CH<sub>4</sub> for example (Holgerson and Raymond, 2016; Yvon-Durocher et al., 2017). These ecosystems have the particularity to be more sensitive to environmental changes due to their small volumes. They can also experience algae blooms, which can compromise their water quality and human safety at the same time pressure of human activities on surface freshwaters is increasing. There is therefore an urgent need to unravel microbial community structuring in these ecosystems to, in return, improve our understanding of their functioning. The microbial community composition in five small freshwater ecosystems located in the “Parc Naturel Régional de la Haute Vallée de Chevreuse”, near Paris (France), was described thanks to pluri-annual surveys (David et al. in prep., see Chapter 5). Microbial eukaryote and prokaryote communities were highly diverse, and specific to each ecosystem despite their close localization. These microbial communities experienced similar temporal patterns, such as marked seasonal dynamics and a global evolution of community composition over years. The variations in community composition over time and among ecosystems were hardly explained by a set of environmental parameters. Biotic interactions were thus suspected to play important roles in the structuring of these microbial communities (Steele et al., 2011; Karimi et al., 2017). In this study, we constructed co-occurrence networks for each season combining the data of a 8-years survey. We aimed at unravelling particular temporal structuring due to microbial interactions. To do so, we investigated seasonal patterns in the topology of microbial networks. We hypothesized that the seasonal structuring observed in the community composition was the result of intense biotic interactions. For instance, we expected that communities of the warmer seasons, i.e. spring and summer, featured more active communities compared to winter (Wilhelm et al., 2014). This would result in return to more complex networks in summer compared to winter (Zhu et al., 2019). To test this hypothesis, we analysed the topology of seasonal co-occurrence networks in four small freshwater ecosystems, constructed using the Poisson Log-Normal model. With this study we aimed at identifying (*i*) structural patterns

within the seasonal co-occurrence networks, (*ii*) topological characteristics that distinguished the different seasons, and (*iii*) key interactions among microbial groups.

### 6.2.3 Material and methods

#### 6.2.3.1 Study sites and sampling design

In this study, five small freshwater ecosystems located in the “Parc Naturel Regional de la Haute Vallée de Chevreuse” (PNR-HVC), and located less than 10 km apart from each other, were sampled at surface. They were composed of four ponds (Etang des Vallées (EV), La Claye (LC), Mare Gabard (MG) and Saint Robert (SR)) and one brook (Ru Sainte Anne (RSA)). Sampling took place every 3 months between October 2011 and January 2019. Water was collected using sterile plastic carboys. Once back to the laboratory, the water was immediately serially filtered using 100, 30, 5 and 0.22  $\mu\text{m}$  pore-size Nucleopore membranes (Whatman) and then stored frozen at  $-80^{\circ}\text{C}$  until DNA extraction. Four physico-chemical parameters (water temperature, dissolved oxygen concentration, pH and conductivity) were directly measured in situ using a multiparameter probe Multi 350i (WTW, Weilheim, Germany).

#### 6.2.3.2 DNA extraction and sequencing

DNA extraction was performed using the PowerSoil DNA extraction kit (MoBio, Carlsbad, CA, USA) on cells of 0.22–30  $\mu\text{m}$  in size (see Simon et al., 2015 for more details). To analyze the composition of eukaryote and prokaryote communities, we amplified the V4 hyper-variable region of the 18S and the 16S rRNA gene, respectively. For eukaryotes, we used the primer sets EK-565F (5'-GCAGTTAAAAAGCTCGTAGT) and EK-1134-R-UNonMet (5'-TTTAAGTTTCAGCCTTGCG), the latter being biased against metazoa (Bower et al., 2004). For the prokaryotes, we used the primers U515F (5'-GTGCCAGCMGCCGCGGTAA) and U806R (5'-GGACTACVSGGGTATCTAAT). In order to affiliate the reads to their sample in the sequencing runs, primers were tagged using 10 bp Molecular Identifiers (MIDs). PCR reactions were performed in a mix containing 1.5 mM  $\text{MgCl}_2$ , 0.2 mM dNTPs, 0.3  $\mu\text{M}$  primers, 0.3–3  $\mu\text{l}$  of DNA sample and 0.5 U HotStart Taq polymerase (Taq Platinum, Invitrogen, Carlsbad, CA, USA), for a total volume of 25  $\mu\text{l}$ . In the thermocycler, samples were exposed to a 2 min denaturation step at  $94^{\circ}\text{C}$ , then to 35 cycles of amplification (at  $94^{\circ}\text{C}$  for 15s for both eukaryotes and prokaryotes, then at  $55^{\circ}\text{C}$  and  $58^{\circ}\text{C}$  for 30s for eukaryotes and prokaryotes, respectively, and at  $72^{\circ}\text{C}$  for 90s for both eukaryotes and prokaryotes), and finally to a 10 min

extension step at 72°C. Five PCR products were pooled for each sample, and then purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany). Purified pools were then multiplexed, by adding equivalent amounts of DNA. Amplicons were sequenced by paired-end MiSeq Illumina technology (2x300 bp, Eurofins Genomics, Ebersberg, Germany).

### 6.2.3.3 Network construction

All statistical analyses were performed under the R-software (R Core Team, 2019). Co-occurrence analyses were conducted using the package ‘PLNmodels’ in R-software (Chiquet et al., 2019). This method uses a multivariate Poisson LogNormal model to infer associations between OTUs while taking into account effects of environmental covariables that might explain the covariations in OTUs counts between samples. As SPiEC-Easi, PLNmodels infer associations as direct statistical covariations by focusing on the precision matrix, which is the inverse of the covariance matrix between OTU abundances (Kurtz et al., 2015). It avoids the detection of associations resulting from indirect covariations (e.g. covariations between two OTUs due to their direct association with a third OTU). This method uses raw data reads and accounts for differences in sampling depth between samples through an offset in the model. It thus avoids building co-occurrence networks from compositional data, which may lead to spurious associations (Faust and Raes, 2012; Gloor et al., 2017). Wrong associations can also result from OTUs covarying concomitantly with the environmental parameters (Deutschmann et al., 2020). Therefore, three abiotic variables were included in the model. More precisely, we added the effect of conductivity, dissolved oxygen, and pH because of their influence on community structure in the five ecosystems (David et al, in prep, see Chapter 5). Model selection was performed using a 50-size grid of penalties, following the same methodology as Chiquet et al. (2019).

To investigate seasonal community structuring in each ecosystem, samples collected at the same season were gathered. For each ecosystem, the dataset was split into four groups (one per season). Each group was composed of 8 time points in autumn and winter and 7 time points in spring and summer (sampling period from October 2011 to January 2019), with three exceptions (7 time points in autumn in La Claye, 6 times points in spring in Mare Gabard and 5 times points in summer in Ru Sainte; due to droughts or missing data). Because of the sparsity of the dataset (large proportion of zeros), for each season, only OTUs present in at least 6 times over the 8 years of survey and having a mean abundance higher than 0.05% were taken into account.

#### 6.2.3.4 Network visualization

Networks were plotted using the ‘ggnet’ R-software package (Tyner et al., 2017). Nodes in the networks represent OTUs, while edges represent relationships between OTUs. The networks were undirected. Each OTU was categorized by its taxonomic affiliation (Eukaryota, Bacteria, Archaea) and by its lifestyle, i.e. its functional group (Autotroph, Heterotroph, Putative parasite).

#### 6.2.3.5 Network topology analysis

For each network, several properties were calculated using the ‘igraph’ R-software package (Csardi et al., 2006). Networks were characterized by their numbers of positive and negative edges as well as nodes and connected nodes. The network mean degree corresponds to the average number of edges established by each node. The average path length is an indicator of the mean number of edges that are necessary to link a given node to another one. Network complexity was estimated using two indicators: connectance and clustering coefficient. The connectance was calculated as  $c = \frac{2E}{N \times (N-1)}$ , where E is the number of edges and N the number of nodes (Barrat et al., 2008). The connectance quantifies the proportion of links established in the network and compared to the number of possible potential links. The connectance is equal to one when all possible links are realised. The clustering coefficient is the probability that two nodes having similar neighbor are connected to each other (Delmas et al., 2019). The clustering coefficient varies between 0 and 1, low values indicating globally poorly connected neighborhood.

A block model analysis was performed to analyze further the topology of the networks. To do so, a stochastic block model approach was used, implemented in the ‘blockmodel’ R-software package (Leger, 2016), on the binary network reporting the presence-absence of significant direct associations. This approach allows to synthetize the overall network structure by gathering the nodes into groups with similar modes of interactions (e.g. group of strongly connected OTUs, group of OTUs associated with the same set of OTUs, groups of OTUs with few significant associations).

Principal Component Analysis were conducted on centered and scaled network properties using the ‘FactomineR’ R-software package (Husson et al., 2016) to identify any clustering within the sampling seasons or ecosystems.

Linear regressions were performed to detect relationships between network properties and

the OTU diversity, without taking into account ecosystems or seasons. The diversity was estimated using either the richness, the evenness or the Simpson index. These three indicators were computed with the ‘vegan’ R-software package on the basis of the dataset selected for the network construction (Oksanen et al., 2007).

Two generalized linear models were used to assess the effects of several variables on network properties. The first model, called the “degree” model, quantifies the effect of both the node’s lifestyle (i.e. the trophic mode of the OTU) and the season on the normalized degree of the nodes in the networks. The effect of the ecosystem and of its statistical interaction with the season were also considered in the model. The second model, called the “parasite” model, was designed to quantify the effect of the season on the percentage of edges established with putative parasitic OTUs in networks. It also takes into account the initial proportion of putative parasitic OTUs in the dataset selected for the network construction. Both models were ran using the ‘glm’ function of the R-software. The variables were fitted with quasi-binomial distributions to account for overdispersion. An ANOVA analysis was used to estimate the significance of the different variables in the models. The estimated marginal means of each qualitative level was estimated using the ‘emmeans’ R-software package to conduct multiple pairwise comparisons between seasons and between lifestyles (Lenth et al., 2018).

## 6.2.4 Results

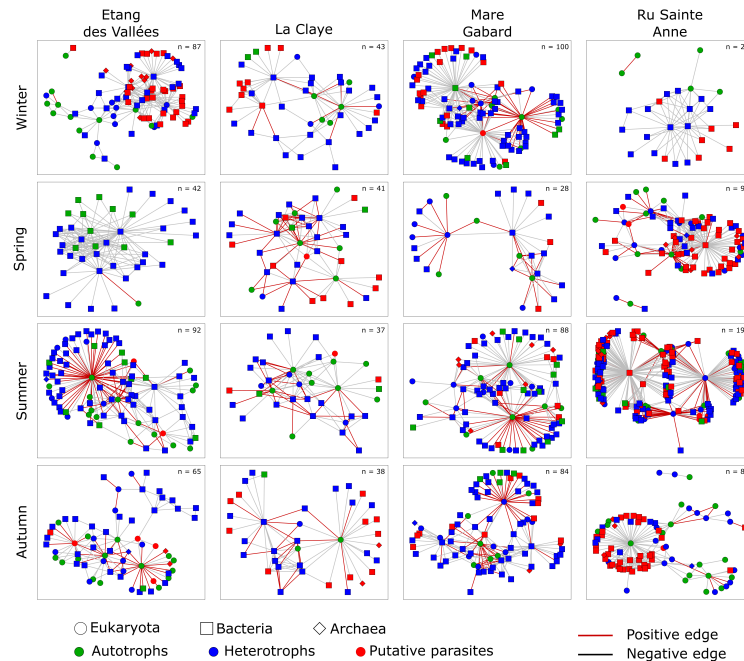
### 6.2.4.1 General network properties

We inferred networks and calculated properties for each ecosystem at each season (Figure 6.2.1, Table 6.2.1). For Saint Robert, the PLN model did not converge in summer and autumn. This ecosystem was thus excluded from this study.

In Etang des Vallées, the number of selected OTUs for network construction was the highest, varying between 378 and 1118. The final percentage of connected nodes was low, ranging between 7.8 and 24.3%. The percentage of positive links varied strongly between seasons, from less than 1% to 39.1%. Connectance revealed that only about 5% of the possible connections between nodes were inferred, at the exception of spring (17.4%). The mean degree varied between 3.8 and 7.1.

La Claye was characterized by the lowest number of OTUs (46 to 97) implemented in the networks. OTUs were highly interacting in the networks of this ecosystem, as revealed by the marked proportions of connected nodes (44.3-82.6%) and the highest connectances (11.7% of

the possible connections were realized on average). In total, 48.9% of the connections were positive. Overall, the mean degree in La Claye varied between 3.6 and 5.6.



**Figure 6.2.1** Seasonal co-occurrence networks over the eight-year survey in the four studied ecosystems. The networks were built on OTUs present in at least 6 seasons over the 8 years and presenting a relative abundance higher than 0.05%. For each network, the number of connected nodes  $n$  is indicated. The shape and the color of the nodes indicate respectively the kingdom and the lifestyle of the OTU. Red and grey edges indicate respectively positive and negative relationships.

Mare Gabard networks were constructed using between 118 and 534 OTUs. About 21% of these OTUs were connected, except at summer when 74.6% of the OTUs were connected. The proportion of potential links established in the network was low, of about 4.6% except in spring (12.4%). Globally, a large proportion (45.7%) of the links were positive. The mean degree varied between 3.4 and 4.2.

In Ru Sainte Anne, networks were constructed using between 282 and 636 OTUs. The percentage of connected nodes was low (mean value of 15.6% in winter, spring and autumn), except in summer where it reached 68.8%. The percentage of positive links was low, of 23.6% in total over the four seasons. The connectance was low (about 4.0%), except in winter where it reached 17.1%. The mean degree of the nodes varied between 3.6 and 4.9.

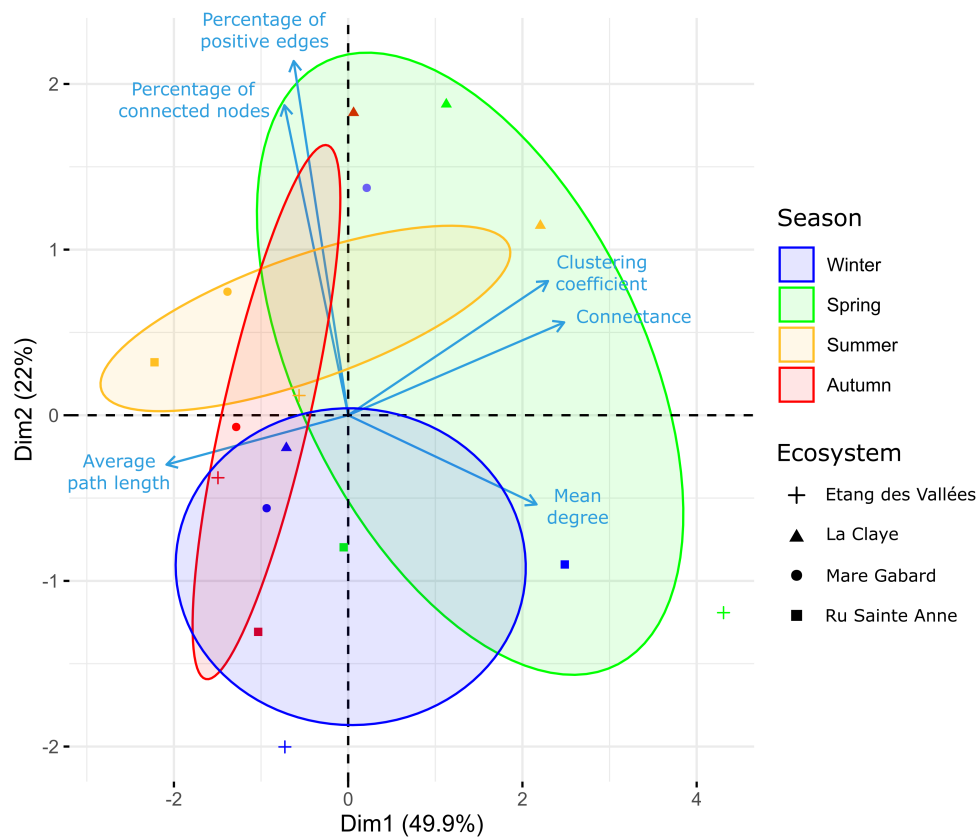
Overall, seasonal specificities were observed in each ecosystem. In Etang des Vallées and Ru Sainte Anne, the highest percentage of positive links and connected nodes were found in summer. The lowest proportion of positive nodes was detected in winter in all ecosystems. In



**Table 6.2.1** Summary of the seasonal co-occurrence networks properties in the four ecosystems. The networks were built on the OTUs present in at least 6 seasons over the 8 years and presenting a relative abundance higher than 0.05%.

	EV				IC			
	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn
Number of positive edges	1	1	81	41	10	50	35	22
Number of negative edges	194	149	126	82	67	51	68	53
Total number of edges	195	150	207	123	77	101	103	75
Percentage positive edges	0.5	0.67	39.13	33.3	13.0	49.5	33.98	29.3
Percentage negative edges	99.5	99.33	60.87	66.7	87.0	50.5	66.02	70.7
Total number of nodes	1118	478	378	571	97	85	82	46
Total number of nodes - eukaryotes	278	151	112	154	12	19	18	3
Total number of nodes - prokaryotes	840	327	266	417	85	66	64	43
Number of connected nodes	87	42	92	65	43	41	37	38
Percentage of connected nodes	7.8	8.79	24.34	11.4	44.3	48.2	45.1	82.6
Mean degree	4.5	7.10	4.5	3.8	3.6	4.9	5.6	3.9
Connectance	0.05	0.17	0.05	0.06	0.12	0.12	0.15	0.11
Clustering coefficient	0.21	0.45	0.17	0.22	0.42	0.42	0.39	0.35
Average path length	2.9	1.90	2.36	3.1	2.4	2.4	2.2	2.5
	MG				RSA			
	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn
Number of positive edges	58	24	51	58	1	51	83	10
Number of negative edges	149	23	134	112	59	168	266	157
Total number of edges	207	47	185	170	60	219	349	167
Percentage positive edges	28.0	51.06	27.6	34.1	1.7	23.3	23.8	6.0
Percentage negative edges	72.0	48.94	72.4	65.9	98.3	76.7	76.2	94.0
Total number of nodes	534	139	118	346	317	636	282	346
Total number of nodes - eukaryotes	66	16	16	41	103	126	84	119
Total number of nodes - prokaryotes	468	123	102	305	214	510	198	227
Number of connected nodes	100	28	88	84	27	90	194	84
Percentage of connected nodes	18.7	20.14	74.6	24.3	8.5	14.2	68.8	24.3
Mean degree	4.1	3.40	4.2	4	4.4	4.9	3.6	4
Connectance	0.04	0.12	0.05	0.05	0.17	0.05	0.02	0.05
Clustering coefficient	0.12	0.34	0.11	0.2	0.33	0.2	0.06	0.17
Average path length	2.36	2.42	2.49	2.8	1.88	2.42	2.54	2.7

La Claye and Mare Gabard, the lowest percentage of connected nodes also occurred in winter. The clustering coefficient was higher in spring for all ecosystems, except for Ru Sainte Anne where it was higher in winter. It reached up to 0.45 in Etang des Vallées. Along with the high connectance, it suggests a higher node interconnection in spring. These seasonal features were underlined in the PCA analysis based on the network properties (Figure 6.2.2). The two first axes of the PCA explained up to 71.9% of the variations of network properties. According to the point distribution, summer and winter appeared to be different, with summer networks characterized by a higher percentage of positive links and connected OTUs. Spring showed the largest variabilities in terms of network properties.



**Figure 6.2.2** Principal Component Analysis (PCA) of the topological properties of the seasonal network. The shape of the point indicates the ecosystem and the color the season of sampling. Ellipses highlight the seasonal clusters.

#### 6.2.4.2 Complexity level in the networks

The nodes with similar patterns of connections in the networks (i.e. highly connected to the rest of the network) were gathered into Block Model Groups (BMG). Overall, this analysis

revealed that the networks were always structured by a restricted number (2–5) of BMG. The more BMG are present in networks, the more complex the network is. They often comprised one or two small groups of OTUs that were characterized by very high numbers of connections with other OTUs in the network and a few larger groups of weakly connected OTUs.

In Etang des Vallées, all seasons were characterized by 4 BMG, except in summer (5 BMG) (Supplementary figure 6.B.1 and 6.B.2). In winter, 10 OTUs were highly connected together and to the other OTUs in the network. They were dominated by Ciliophora (50% of the nodes in the BMG). The majority (61.5%) of the OTUs affiliated to autotrophic taxa were linked to each other, in a specific BMG. In summer, a BMG was composed of 6 OTUs that were highly connected to each other but had low connections with the other OTUs of the network. One BMG gathered 8 OTUs that were highly and positively connected to the other OTUs of the network. They comprised five autotrophs which belonged to Chlorophyta, Cryptophyta and Ochrophyta. The three others were affiliated to Betaproteobacteria, Verrucomicrobia and Fungi. In autumn, one BMG was represented by only 2 OTUs, one autotroph (Ochrophyta) and one putative parasite (Apicomplexa), which were highly connected to other BMGs of the model.

In La Claye, all networks featured 3 BMG (Supplementary figure 6.B.1 and 6.B.3). In winter, the network was dominated by one BMG constituted of 3 OTUs, including two autotrophic Cryptophyta and one Gammaproteobacteria. They were highly connected to the two other BMGs. In summer, the BMG containing the most connected nodes was composed of 10 OTUs, 6 of them being affiliated to autotrophic taxa. The four others being heterotrophic taxa. As for the network obtained in winter, the autumn network was composed of one BMG composed of only 2 OTUs including one autotroph (Cryptophyta) and one heterotroph (Bacteroidetes). They were highly connected to the two other BMGs.

In Mare Gabard, the season that showed the most complex structure was summer, with 5 BMG (Supplementary figure 6.B.1 and 6.B.4). In winter, one of the BMGs was composed of only 3 OTUs, belonging to the three lifestyle classes (autotroph, heterotroph, putative parasite). It was connected to the other BMGs, mostly through negative edges. However, the winter network was also characterized by a BMG composed of 8 OTUs, including five autotrophs (Ochrophyta) and three heterotrophs (two Actinobacteria and one Ciliophora), most of them being only linked with positive edges to the other BMGs. The spring network was only composed of two BMGs, including one constituted of two heterotrophic bacteria (Verrucomicrobia and Epsilonproteobacteria). In summer, one BMG was composed of two highly connected autotrophs, one dinophyta and one ochrophyta (*Gonyostomum*), already described in the ecosystem as potential toxic algae (Simon et al., 2015; David et al, in prep see Chapter

5). This latter OTU was characterized by the highest degree of positive edges.

In Ru Sainte Anne, spring and autumn networks were the most complex, featuring 5 BMG (Supplementary figure 6.B.1 and 6.B.5). In winter, only 3 BMG were found, including one with only two OTUs. In spring, one of the BMGs was only represented by OTUs affiliated to Parcubacteria, presenting 65 edges, mostly negative. Two BMGs were composed of a majority of nodes affiliated to bacteria of the CPR, negatively and highly interconnected. In summer, the network was also constituted of one BMG comprising only 2 OTUs, affiliated to Ciliophora and Parcubacteria. These OTUs were highly interconnected to the rest of the BMG. A large number of OTUs (157) were poorly connected (with a degree of 1 or 2). Similar to spring, the network in autumn was characterised by the presence of a BMG that was constituted of only one node. In this case, it was an Ochrophyta (Chrysophyceae). Also, in autumn, BMG5 was poorly connected to the rest of the network, and was composed of half autotrophic nodes and half heterotrophic nodes.

#### **6.2.4.3 Relationships between ecosystem diversity and network topology**

Among diverse network properties measured, only the percentage of connected nodes, the percentage of positive edges and the percentage of edges established with putative parasites were significantly linked to one or several diversity indicators (Supplementary figure 6.B.6, Supplementary table 6.A.1). The percentage of positive edges was negatively linked to the richness calculated for each ecosystem. The percentage of connected nodes significantly decreased with diversity as found for the richness, the evenness and the Simpson index. The percentage of edges including putative parasites was positively linked to evenness.

#### **6.2.4.4 Global taxonomic composition of the nodes in the networks**

The initial number of OTU used to infer the co-occurrence networks varied among ecosystems (Table 6.2.1, Supplementary figure 6.B.7 and 6.B.8). Indeed, OTUs were arbitrarily selected based on marked occurrence and abundance to construct the networks (detected with a mean abundance  $>0.05\%$  in 70% of the samples per ecosystem and season). For both eukaryotes and prokaryotes, the lowest numbers of OTUs were detected in La Claye in autumn (3 and 43 OTUs, respectively) and the highest is in Etang des Vallées in winter (278 and 840 OTUs, respectively) (Supplementary table 6.A.2).

The composition of the connected nodes varied among the ecosystems, but also within

seasons (Supplementary table 7 and 8). Bacteria dominated among the connected nodes in all networks. They represented a minimum of 49.2% in autumn and a maximum of 97.6% in spring in Etang des Vallées (Supplementary table 6.A.2).

In Etang des Vallées, microbial eukaryotes were highly interacting in autumn. They represented 46.2% of the total number of connected nodes. Globally, the number of nodes affiliated to putative parasites in the networks was low, except in winter when they represented up to 43.8% of the connected nodes. Nodes affiliated to autotrophic taxa occurred in high proportions, with a mean value of 19.9% of the connected nodes over the four seasons. In La Claye, archaea were never found in the networks, except in autumn when two DPANN nodes were retrieved. Nodes affiliated to autotrophic taxa were mostly present in summer and represented globally 29.7% of the connected nodes. Prokaryotic OTUs affiliated to parasitic taxa (only prokaryotes) reached their highest proportions in autumn (34.2% of the connected nodes). Autumn was also characterized by a low proportion of nodes affiliated to microbial eukaryotes (7.9% of the connected nodes). In Mare Gabard, microbial eukaryote nodes represented a mean value of 16.7% of the connected nodes, except in spring when they reached 35.7%. Nodes affiliated to autotrophic taxa were mostly abundant in summer (10.8% of the connected nodes). The proportions of nodes affiliated to putative parasites were relatively stable over the seasons (average percentage of 15.3% of the connected nodes). In Ru Sainte Anne, nodes of putative parasites dominated the networks, ranging from 35.7% in autumn to 55.6% in winter. Contrary to the other ecosystems, nodes affiliated to autotrophic taxa reached their highest abundances in autumn (14.3% of the connected nodes).

Globally, nodes affiliated to putative parasites were more abundant in the networks in winter in all ecosystems, except in La Claye. In addition, La Claye and Mare Gabard featured the highest proportions of nodes affiliated to autotrophic taxa.

Among the different ecosystems and seasons, the composition of unconnected nodes varied greatly (Supplementary figure 6.B.7 and 6.B.8), although they were often affiliated to Proteobacteria and CPR members. Proteobacteria constituted 16.8% to 36.7% of the unconnected nodes, respectively in winter in Etang des Vallées and in summer in Mare Gabard., while CPR represented 5.2 to 39.8% of these nodes, respectively in summer in Etang des Vallées and spring in Ru Sainte Anne.

#### 6.2.4.5 Focus on the central nodes in the networks

Among the five most connected nodes inferred at each season, those affiliated to putative parasites were only observed in winter, at the exception of Ru Sainte Anne where parasitic OTUs were observed in each season (Supplementary table 6.A.3). The mean degree of these five most connected nodes was higher in summer (30.2 in Etang des Vallées, 16 in La Claye, and 59.6 in Ru Sainte Anne), except in Mare Gabard where the winter nodes were the most connected (mean degree of 39.6 in winter and 29 in summer). As observed in the BM analyses, the most connected nodes in summer were dominated by OTUs affiliated to autotrophs. In particular, this was the case of an OTU affiliated to *Gonyostomum semen* in Mare Gabard, that was central in the network. This OTU was characterised by the highest degree (52 interactions, Supplementary table 6.A.3). Among these highly connected nodes, 28 were positive and connected *G. semen* to 12 different phyla, while 24 were negative and connecting 8 different phyla. Parcubacteria were well represented in the most connected nodes despite global low degree, especially in Ru Sainte Anne (Supplementary table 6.A.3).

#### 6.2.4.6 Variations of the node degree among seasons and lifestyles

All the parameters (i.e. lifestyles, seasons, ecosystems and statistical interactions between seasons and ecosystems) had a significant effect on the node degree (Supplementary Table 6.A.4). There was no significant effect of statistical interactions between lifestyles and ecosystems or functional groups and season, suggesting that the differences in the proportions of lifestyles do not depend on the ecosystem or the season.. The degree of the nodes affiliated to autotrophic taxa was, overall, significantly higher than that of nodes of heterotrophs and parasites (Supplementary Table 6.A.4, Tukey test, estimate = 0.491, z.ratio = 3.418, p = 0.0018 and estimate = 0.428, z.ratio = 2.512, p = 0.0321 respectively). Moreover, nodes in summer and spring were found to be globally significantly more connected than nodes in winter (Supplementary Table 6.A.4, Tukey test, estimate = 0.742, z.ratio = 4.130, p = 0.0002 and estimate = 0.564, z.ratio = 2.708, p = 0.0342 respectively). However, the statistical interaction between ecosystems and seasons had a significant effect. Taken individually, a clear distinction between summer/spring and winter was only observed in Etang des Vallées and Mare Gabard (Supplementary Table 6.A.5), while node degrees did not significantly depend on the season in La Claye and Ru Sainte Anne.

#### 6.2.4.7 Inferred biotic interactions depend on the functional group of the node

In order to better describe the network topology, we summarized the nature of the edges between the three lifestyles (autotrophs, heterotrophs or putative parasites) (Supplementary Table 6.A.6). We observed that the proportion of edges established with nodes affiliated to putative parasites depended on seasons (“parasite” model, glm,  $F = 4.07$ ,  $p = 0.036$ ). It was significantly higher in winter than in summer or spring for all ecosystems (Tukey test, estimate = 1.556, z.ratio = 2.848,  $p = 0.023$  and estimate = 1.518, z.ratio = 2.865 and  $p = 0.022$  respectively), whatever was the proportion of parasites initially present in the network (Supplementary table 6.A.7). The number of OTUs affiliated to putative parasites in the network significantly increased the proportion of edges established with parasites (“parasite” model, glm,  $F = 35.3$ ,  $p < 0.001$ ). Autumn and winter, as well as spring and summer, did not differ significantly in their proportions of edges established with nodes affiliated to parasites (Supplementary Table 6.A.7). A majority of positive edges happened between nodes affiliated to autotrophic and heterotrophic taxa (Supplementary table 6.A.6). The only exceptions were found in autumn in Etang des Vallées, La Claye and Mare Gabard, as well as in Ru Sainte Anne in winter and summer.

### 6.2.5 Discussion

Important efforts have been invested in the understanding of microbial community structuring over time and space since the beginning of microbial ecology. Because of their small size and supposed low importance at the global scale, small freshwater ecosystems have been overlooked for the profit of their large counterparts. Their importance in greenhouse gases fluxes such as  $\text{CO}_2$  and  $\text{CH}_4$  has been revealed over the last years, increasing the interest in their functioning (Holgerson and Raymond, 2016; Yvon-Durocher et al., 2017). Previous investigations showed microbial communities differing among the five small freshwater ecosystems investigated that all were characterised by seasonal patterns (Simon et al., 2015; David et al., submitted, David et al., in prep., see Chapters 4 and 5). The differences in communities between ecosystems and their temporal structuring were little explained by environmental parameters, suggesting that biotic interactions could strongly shape microbial communities over space and time. This spatio-temporal structuration could result in contrasting co-occurrence network topologies, which was the case between seasons and ecosystems (Figure 6.2.1). However, we also detected features common to the different seasons in these ecosystems, as already observed in lakes (Kara et al., 2013; Zhao et al., 2016). This is in line with the hypothesis that ecosystems have the same global functioning processes, but conducted by different organisms (Louca et al., 2018).

### 6.2.5.1 A few common seasonal network properties across ecosystems

The global structure of the networks varied between and within ecosystem, without obvious clear seasonal pattern. Networks topology revealed common features between seasons, with a distinction between warm (summer, spring) and cold (winter) seasons.

Summer networks were characterized by a higher percentage of positive edges and of connected nodes (Figure 6.2.2). Globally, the degree of nodes was significantly higher in spring and summer compared to autumn and winter (Supplementary Table 6.A.4). Therefore, warmer season networks tended to be larger in size, and to present more positive interactions. More complex networks in summer were also observed in bacterial networks in eutrophic urban lakes (Zhao et al., 2016). Similarly, in Lake Mendota, a large freshwater ecosystem, spring networks presented more co-occurrences than at other seasons (Kara et al., 2013). Moreover, we obtained higher connectance values in spring and summer in all ecosystems at the exception of Ru Sainte Anne where it was observed in winter. Connectance is often linked to network stability, but it has contrasted effects on stability depending on the types of interactions involved (e.g. mutualistic or trophic) or on the measures of stability considered (Thébault and Fontaine, 2010; Landi et al., 2018). While the relationship between the connectance of co-occurrence networks and community stability remains to be solved, our results suggest that network stability might differ between seasons. In other aquatic ecosystems studies, the connectance is often low (around 0.02 for lakes and marine systems, Peura et al., 2015; Hernández-Ruiz et al., 2018). These connectance values were lower than those obtained in the seasonal network of our small ecosystems (Table 6.2.1). It can indicate either different stability of the small ecosystems at the seasonal level, or more likely a strong variation of connectance when comparing a specific season or the entire time-serie. Indeed, the connectance of a global network gathering all ecosystems and sampling point was lower than those obtained by ecosystems and seasons (David et al., in prep, see Chapter 5). However, these differences could also be the result of the network construction method used. In our case, we based our networks on direct associations through the use of the inverse covariance matrix instead of correlations (Kurtz et al., 2015). We also added the effect of environmental parameters, which remove spurious correlations, likely decreasing the connectance value.

On the opposite, winter periods exhibited a significantly higher percentage of edges involving putative parasitic nodes (Supplementary Table 6.A.7). In colder season, the parasite hosts can be under stress, either due to the lower nutrient concentration or the lower light level. Therefore, parasites can be more efficient in infecting their host, which could result in a higher proportion of interactions with parasites (Cleary and Durbin, 2016).



The pattern of higher node degree in warm compared to cold seasons was observed only for the Etang des Vallées and Mare Gabard (Supplementary Table 6.A.5). This is coherent with the complexity of the BMG, which were more numerous in Etang des Vallées and Mare Gabard in Summer (Supplementary Figure 6.B.2). Therefore, the pool of OTUs selected in La Claye and Ru Sainte Anne did not reflect clear seasonal patterns in terms of mean degree and network complexity. For instance, the level of complexity revealed by the number of BMG in La Claye was stable over time (Supplementary Figure 6.B.2). Even if we observed previously common patterns between ecosystems in the global community composition, this is not fully reflected by the network structures (David et al., in prep, see Chapter 5).

#### 6.2.5.2 Autotrophic OTUs are central in the co-occurrence networks

Autotrophic primary producers (i.e. phytoplankton) are at the basis of the trophic webs, producing organic carbon through photosynthesis that is intensively used by heterotrophic prokaryotes and being grazed by microbial or metazoan predators. The released organic matter generates micro-areas with higher concentration of organic matter compared to the rest of the ecosystem (Worden et al., 2015). Therefore, this region surrounding phytoplankton, called the phycosphere, attracts heterotrophic organisms such as free-living prokaryotes (Bell and Mitchell, 1972; Seymour et al., 2017). In our networks, we observed that primary producers were more connected than heterotrophs or putative parasites, independently from the ecosystem or the season (Supplementary Table 6.A.3 and 6.A.4). Moreover, we detected that over all the ecosystem and season, the higher amount of positive interactions happened between autotrophic and heterotrophic nodes (Supplementary Table 6.A.6). The higher number of positive edges underlines the importance of the interactions between phytoplankton and various heterotrophic microorganisms (Paver et al., 2013; Linz et al., 2020). The diversity of the heterotrophs attracted by the phycosphere is high, probably because of the variety of organic compounds released by the phytoplankton.

Harmful algae were also detected in the ecosystems (Simon et al., 2015; David et al. in prep, see Chapter 5). In Mare Gabard, the Raphidophyceae *Gonyostomum semen* (Stramenopiles) was detected at high abundance in summer. Temperature is known to be one of the drivers of the development of *G. semen*, which can explain their presence in summer (Rengefors et al., 2012). No phyla interacted specifically with *G. semen*, but both types of interactions (positive and negative) were mainly conducted with heterotrophic bacteria, underlying the potential importance of the phycosphere around the algae.

### 6.2.5.3 Importance of predator-prey interactions

Ciliates are major grazers in aquatic ecosystems, feeding on phytoplankton and bacteria (Seong et al., 2006; Pernthaler, 2005; Simek et al., 2019). In our networks, members of the ciliophora phylum were not present in high abundances in terms of nodes in the networks (Supplementary Figure 6.B.7). They represented at maximum 12.3% of the nodes (autumn in Etang des Vallées). However, when they were present, they had a high degree (mean degree of ciliates over all ecosystems and seasons of 7.7). They were also well represented in the most connected nodes, and found in the BMG presenting the highest connectivity with the rest of the network (Supplementary Table 6.A.3, Supplementary Figures 6.B.2 to 6.B.5). Despite a weak representativeness, this reveals the importance of the ciliates in the construction of the networks. Predation by ciliates impacts bacterial community composition, either by directly reducing specific taxa, or indirectly through the excretion of nutrients useful for other microbes (Sherr and Sherr, 2002; Jardillier et al., 2005).

### 6.2.5.4 Parcubacteria as the dominant putative parasites

The majority of the parasitic nodes involved in the networks were affiliated to bacteria of the Candidate Phyla Radiation, and more precisely to the Parcubacteria. Most of the Parcubacteria were connected to a low number of phyla, as underlined by their presence in the poorly connected BMG. However, among the most connected Parcubacteria, the majority of the links were established with those poorly connected Parcubacteria. Intra-connections are often observed in co-occurrence networks, and can indicate a competition for the same resources (Zelezniak et al., 2015; Bock et al., 2020). For the case of Parcubacteria, it could be a competition for the same hosts (Nelson and Stegen, 2015).

## 6.2.6 Conclusion

In this study, we used four small freshwater ecosystems to decipher the seasonal structure of microbial communities. Despite the apparent dissimilarities between the seasonal networks between the ecosystems, several network properties showed common seasonal patterns. Summer and spring networks were often characterized by larger networks, with more nodes, higher degree and a higher proportion of positive interactions. On the opposite, winter networks had the particularity to favor interactions with parasitic nodes. Finally, some expected freshwater interactions were observed in the networks. The position of primary producers and ciliates

in the networks validates the ecological coherence of our method. Further highlights on key interactions in the seasons will help to improve our understanding of the functioning of small freshwater ecosystems over time.

# Supplementary Material

## 6.A Supplementary Tables

**Supplementary Table 6.A.1** Linear regressions conducted between network properties and diversity indexes (richness, evenness and Simpson index).

		Richness	Evenness	Simpson
Percentage of positive edges	lag	-704.0	-0.2692	-0.2243
	r <sup>2</sup>	0.167	0.1034	0.0522
	pvalue	0.06511	0.1207	0.198
Number of connected nodes	lag	1.781	0.0004154	0.0005878
	r <sup>2</sup>	0.02907	-0.04401	-0.01551
	pvalue	0.2486	0.5539	0.3948
Percentage of connected nodes	lag	-6.569	-0.0029616	-0.0036425
	r <sup>2</sup>	0.4084	0.418	0.6825
	pvalue	0.004582	0.004052	4.884e-05
Clust coef	lag	-530.8	-0.008382	-0.10935
	r <sup>2</sup>	-0.001035	-0.07134	-0.05616
	pvalue	0.3379	0.9734	0.6597
Average path length	lag	233.2	-0.08394	0.00474
	r <sup>2</sup>	0.02734	-0.007231	-0.07122
	pvalue	0.253	0.3609	0.9591
Percentage of links involving a parasitic node	lag	3.213	0.002171	0.001491
	r <sup>2</sup>	0.0571	0.2233	0.0701
	pvalue	0.1886	0.037	0.1664
Connectance	lag	-157.6	0.78081	0.47885
	r <sup>2</sup>	-0.07037	0.05897	-0.02148
	pvalue	0.908	0.1854	0.4219
Mean degree	lag	45.18	0.02697	0.02523
	r <sup>2</sup>	-0.03991	-0.01505	-0.02119
	pvalue	0.5253	0.3928	0.4205

**Supplementary Table 6.A.2** Summary of the seasonal co-occurrence networks properties in the four ecosystems. The networks were built on the OTUs present in at least 6 seasons over the 8 years and presenting a relative abundance higher than 0.05%.

	EV						LC										
	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn					
	%	%	%	%	%	%	%	%	%	%	%	%					
Eukaryotes	22	25.3	1	2.38	23	25	30	46.2	8	18.6	9	22.0	14	37.8	3	7.89	
Autotrophs	13	14.9	1	2.4	15	16.3	15	23.1	5	11.6	7	17.1	9	24.3	2	5.3	
Heterotrophs	9	10.3	0	0.0	6	6.5	11	16.9	3	7.0	1	2.4	3	8.1	1	2.6	
Parasites	0	0	0	0	2	2.2	4	6.2	0	0	1	2.4	1	2.7	0	0	
Uncertain	0	0	0	0	0	0	0	0	0	0	0	0	1	2.7	0	0	
Prokaryotes	65	74.7	41	97.6	69	75	35	53.8	35	81.4	32	78.0	23	62.2	35	92.1	
Autotrophs	0	0.0	11	26.2	9	9.8	1	1.5	5	11.6	2	4.9	2	5.4	2	5.3	
Heterotrophs	27	31.0	30	71.4	59	64.1	30	46.2	21	48.8	20	48.8	19	51.4	20	52.6	
Parasites	0	0	0	0	1	1.1	1	1.5	0	0	0	0	0	0.0	0	0.0	
	27	31.0	0	0	0	0	1	1.5	9	20.9	10	24.4	2	5.41	11	28.9	
	11	12.6	0	0	0	0	2	3.1	0	0	0	0	0	0	2	5.3	
MG																	
	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	
	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	
Eukaryotes	16	16	10	35.7	16	18.2	16	19.0	5	18.5	25	27.8	52	26.8	36	42.8	
Autotrophs	8	8.0	5	17.9	8	9.1	7	8.3	5	18.5	11	12.2	18	9.3	17	20.2	
Heterotrophs	7	7.0	5	17.9	8	9.1	9	10.7	0	0.0	12	13.3	28	14.4	18	21.4	
Parasites	1	1	0	0	0	0	0	0	0	0	2	2.2	6	3.09	1	1.19	
Uncertain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Prokaryotes	84	84	18	64.3	72	81.8	67	79.7	22	81.5	65	72.2	142	73.2	48	57.1	
Autotrophs	7	7.0	0	0.0	11	12.5	4	4.8	1	3.7	0	0.0	14	7.2	7	8.3	
Heterotrophs	59	59.0	14	50.0	47	53.4	52	61.9	6	22.2	22	24.4	36	18.6	12	14.3	
Parasites	0	0	1	3.57	0	0	0	0	0	0	1	1.1	0	0	0	0	
	18	18	3	10.7	7	7.95	11	13.1	14	51.9	36	40	92	47.4	29	34.5	
	0	0	0	0	7	8.0	2	2.4	1	3.70	6	6.67	0	0	0	0	



**Supplementary Table 6.A.4** Output of the “degree” model, quantify the effect of both the node’s lifestyle and the season. The results of the model section indicates the pairwise comparison of the different variable implemented in the model (Tukey test). The results of the anova section indicates the global effect of the different variable on the model.

<b>Results of the model</b>				
<b>Variable</b>	<b>Pairwise comparison</b>	<b>estimate</b>	<b>z.ratio</b>	<b>pvalue</b>
Lifestyle	Autotroph-Heterotroph	0.491	3.418	0.0018
	Autotroph-Parasite	0.428	2.512	0.0321
	Heterotroph-Parasite	-0.063	-0.391	0.9194
Season	Summer-Spring	0.179	0.940	0.7835
	Summer-Autumn	0.349	1.978	0.1964
	Summer-Winter	0.742	4.130	0.0002
	Spring-Autumn	0.170	0.830	0.8402
	Spring-Winter	0.564	2.708	0.0342
	Autumn-Winter	0.393	2.019	0.1810
<b>Results anova</b>				
<b>Variable</b>	<b>Deviance</b>	<b>F</b>		<b>pvalue</b>
NULL	8600.2			
Lifestyle	245.69	8.366		2.5x10-4
Season	460.54	10.45		9.0x10-7
Ecosystem	1438.09	32.65		2.2x10-16
Season:Ecosystem	404.2	3.058		1.3x10-3



**Supplementary Table 6.A.5** Details of the pairwise comparison of the “degree” model between the season within the different ecosystems (Tukey test).

<b>Etang des Vallées</b>	Estimate	z.ratio	p.value
Summer-Spring	-0.190	-0.643	0.9180
Summer-Autum	0.646	1.936	0.2129
Summer-Winter	1.176	3.985	0.0004
Spring-Autumn	0.836	2.376	0.0818
Spring-Winter	1.366	4.322	0.0001
Autumn-Winter	0.531	1.521	0.4248

<b>Mare Gabard</b>	Estimate	z.ratio	p.value
Summer-Spring	0.397	0.785	0.8614
Summer-Autum	1.218	3.776	0.0009
Summer-Winter	1.526	5.201	0.0001
Spring-Autumn	0.821	1.59	0.3842
Spring-Winter	1.129	2.263	0.1067
Autumn-Winter	0.308	0.987	0.7567

<b>La Claye</b>	Estimate	z.ratio	p.value
Summer-Spring	0.058	0.141	0.9990
Summer-Autum	-0.433	-0.969	0.7673
Summer-Winter	0.483	1.104	0.6870
Spring-Autumn	-0.490	-1.129	0.6713
Spring-Winter	0.426	0.998	0.7504
Autumn-Winter	0.916	1.983	0.1945

<b>Ru Sainte Anne</b>	Estimate	z.ratio	p.value
Summer-Spring	0.451	1.699	0.342
Summer-Autum	-0.034	-0.119	0.999
Summer-Winter	-0.216	-0.552	0.9462
Spring-Autumn	-0.485	-1.57	0.3959
Spring-Winter	-0.667	-1.626	0.3638
Autumn-Winter	-0.182	-0.428	0.9737

**Supplementary Table 6.A.6** Summary of percentage of the type of interactions developed in the seasonal co-occurrence networks for each ecosystem. All the edges in the network were categorized following the lifestyle of the nodes interacting, and the type of the interaction (*i.e.* positive or negative).

<b>Etang des Vallées</b>	Pos. edges	Neg. edges	Pos. edges	Neg. edges	Pos. edges	Neg. edges
Winter	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	0.6	5.5	0	20	0	1.82
Heterotrophs			0	34.5	0	23
Put. Parasites					0	14.5
Spring	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	0	7.3	0.67	35.3	0	0
Heterotrophs			0	56.7	0	0
Put. Parasites					0	0
Summer	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	5.31	6.28	29	21.3	0	1.45
Heterotrophs			2.9	30	1.93	1.93
Put. Parasites					0	0
Autumn	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	4.88	7.31	17.1	16.3	4.88	4.88
Heterotrophs			4.88	26	1.63	10.6
Put. Parasites					1.63	1.62
<b>La Claye</b>	Pos. edges	Neg. edges	Pos. edges	Neg. edges	Pos. edges	Neg. edges
Winter	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	0	10.4	9.1	28.6	0	2.6
Heterotrophs			1.3	20.8	2.6	13
Put. Parasites					0	11.7
Spring	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	3.96	2.97	18.8	14.9	2.97	8.9
Heterotrophs			14.8	16.8	8.91	6.93
Put. Parasites					0	0
Summer	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	1.94	16.5	15.5	29.1	0.97	3.88
Heterotrophs			15.5	16.5	0	0
Put. Parasites					0	0

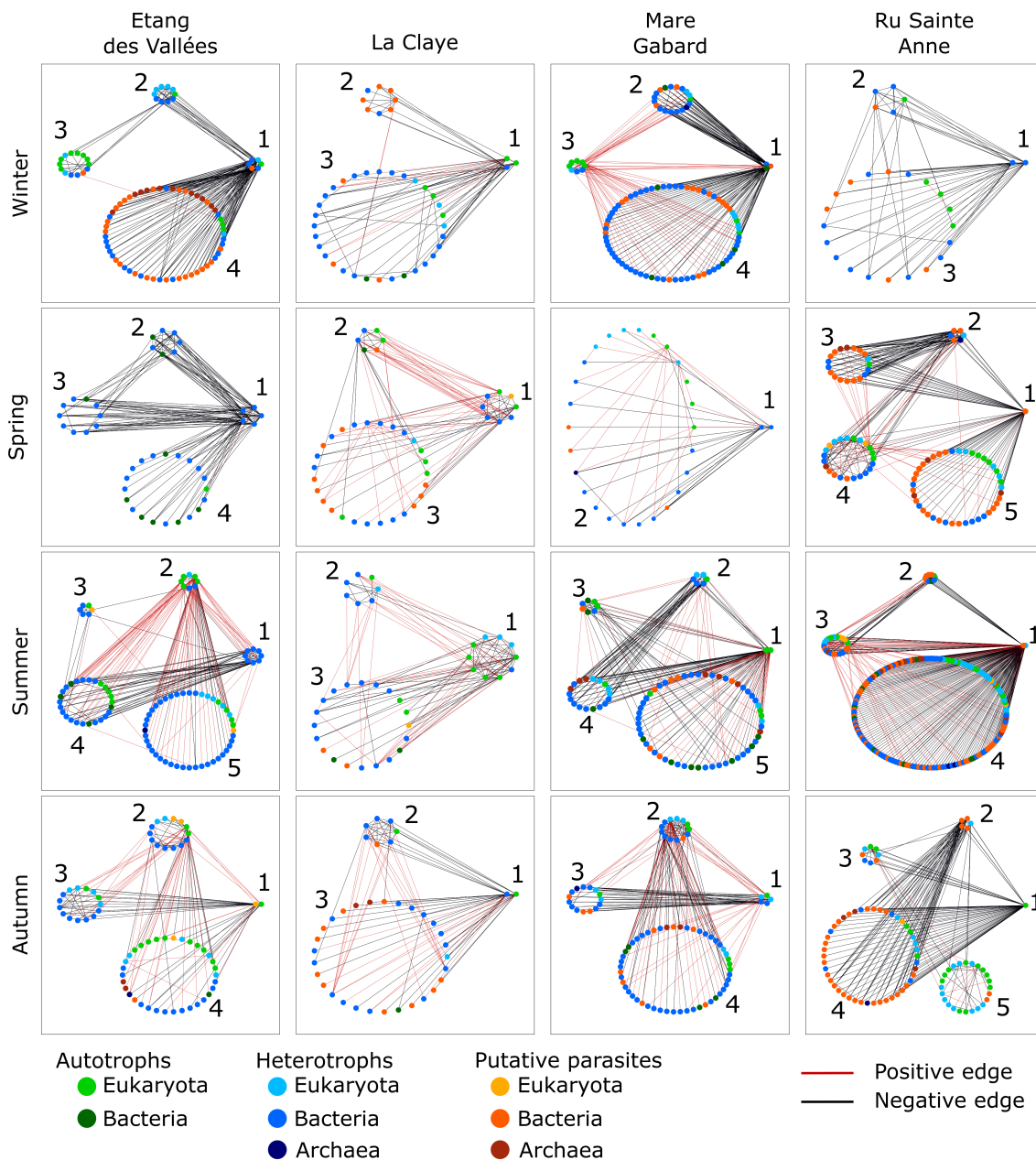
Autumn	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	0	1.33	6.67	21.3	0	12
Heterotrophs			21.3	24	1.33	12
Put. Parasites					0	0
<b>Mare Gabard</b>	Pos. edges	Neg. edges	Pos. edges	Neg. edges	Pos. edges	Neg. edges
Winter	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	2.42	4.37	17	20.9	1.94	12.1
Heterotrophs			6.31	10.7	0	22.3
Put. Parasites					0	1.94
Spring	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	0	2.13	36.2	0	2.13	0
Heterotrophs			12.8	38.3	0	8.51
Put. Parasites					0	0
Summer	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	4.3	6.45	15.1	39.2	3.22	9.68
Heterotrophs			4.3	14.5	0.54	2.69
Put. Parasites					0	0
Autumn	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	1.76	1.18	11.8	13.5	0.59	1.18
Heterotrophs			15.9	41.8	4.12	8.23
Put. Parasites					0	0
<b>Ru Sainte Anne</b>	Pos. edges	Neg. edges	Pos. edges	Neg. edges	Pos. edges	Neg. edges
Winter	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	1.7	3.33	0	13.3	0	1.67
Heterotrophs			0	50	0	30
Put. Parasites					0	0
Spring	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	0.46	0.91	7.76	2.28	8.22	3.2
Heterotrophs			3.65	11.9	3.2	28.31
Put. Parasites					0	30.1
Summer	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	0.29	1.15	3.15	7.16	6.88	9.17
Heterotrophs			10	10.3	3.44	30.4

Put. Parasites					0	18.1
Autumn	Autotrophs		Heterotrophs		Put. Parasites	
Autotrophs	0	9	4.8	15.6	0	25.7
Heterotrophs			0.6	4.2	0.6	8.4
Put. Parasites					0	31.1

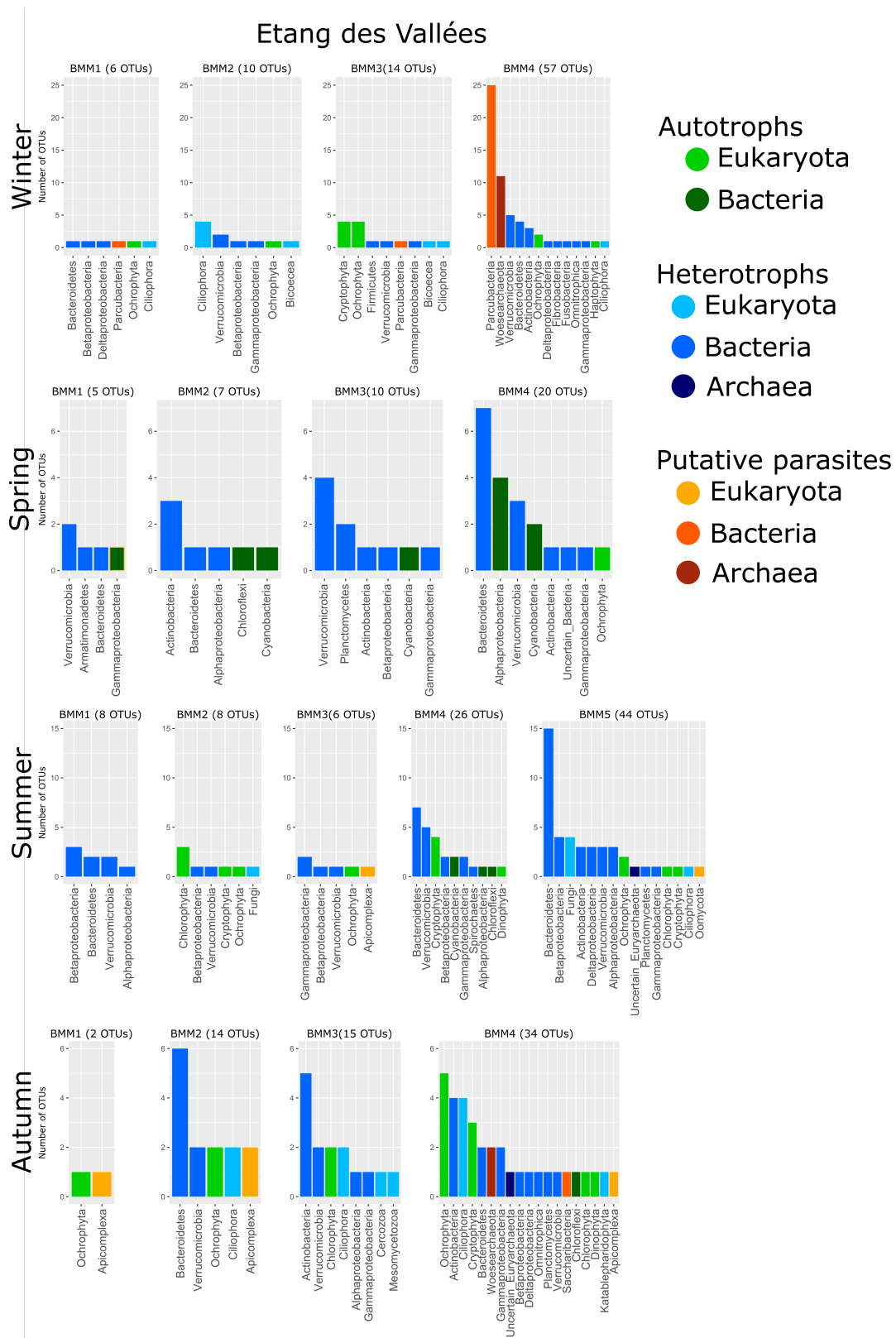
**Supplementary Table 6.A.7** Output of the “parasite” model, quantifying the effect of the season on the percentage of edges established with putative parasitic OTUs in networks. The pairwise comparisons of the different variable implemented in the model are indicated (Tukey test). The effect of the covariable “initial proportion of parasitic nodes” is indicated.

<b>Variable</b>	<b>Pairwise comparison</b>	<b>estimate</b>	<b>z.ratio</b>	<b>pvalue</b>
<b>Season</b>	Winter-Spring	1.556	2.848	0.023
	Winter-Summer	1.518	2.865	0.022
	Winter-Autumn	0.277	0.607	0.930
	Spring-Summer	-0.038	-0.065	0.999
	Spring-Autumn	1.278	-2.343	0.089
	Summer-Autumn	-1.240	-2.357	0.086
<b>Initial proportion of parasitic nodes</b>				0.00023

## 6.B Supplementary Figures

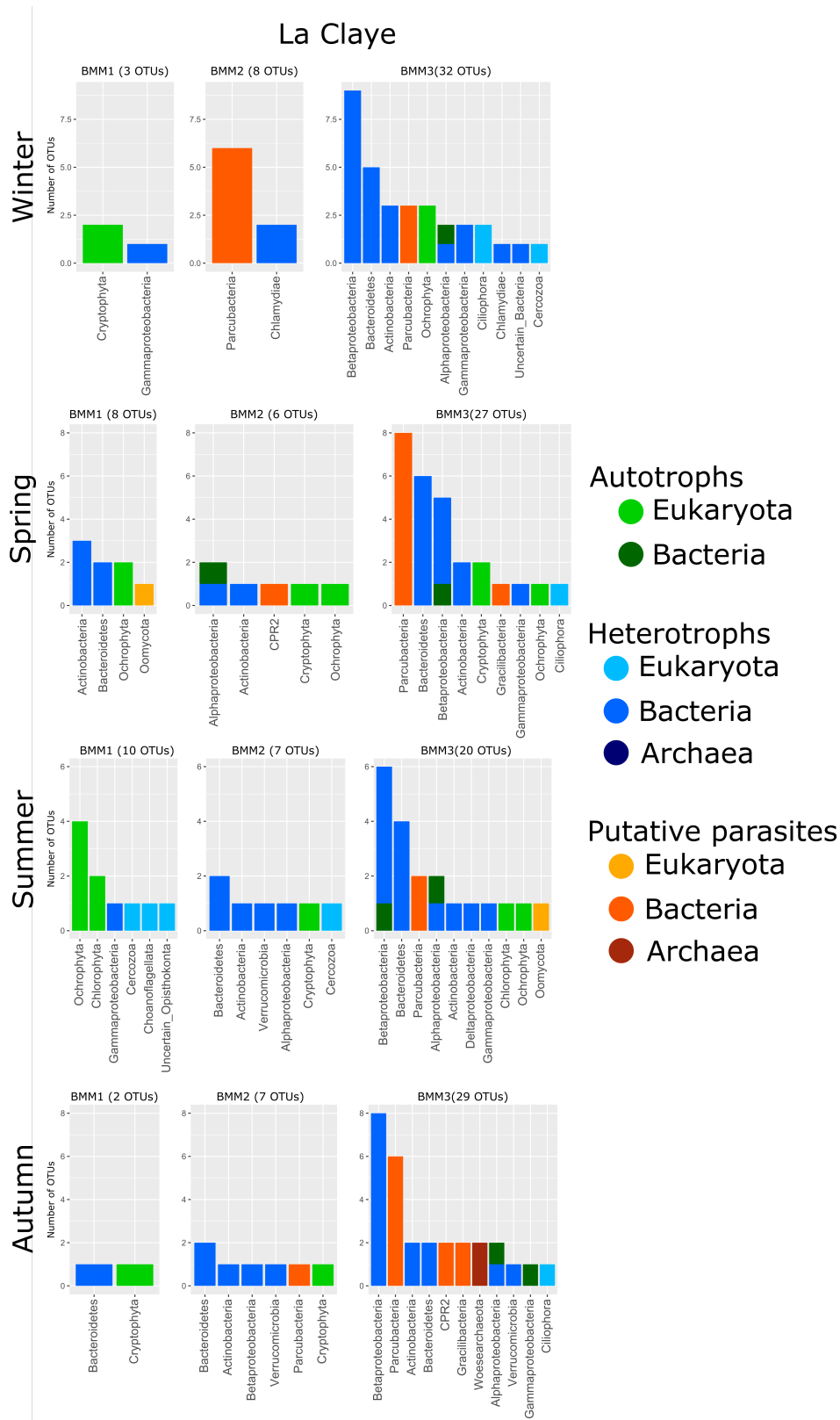


**Supplementary Figure 6.B.1** Block-model representation of the seasonal networks over the eight-year survey in the four ecosystems. The networks were built on OTUs present in at least 6 seasons over the 8 years and presenting a relative abundance higher than 0.05%. The color of the nodes indicates the kingdom and the lifestyle of the OTUs. Green shades correspond to OTUs affiliated to autotrophs, blue shades correspond to OTUs affiliated to heterotrophs and orange shades correspond to OTUs affiliated to putative parasites. Positive co-occurrences are indicated in red whereas negative interaction are represented in black.

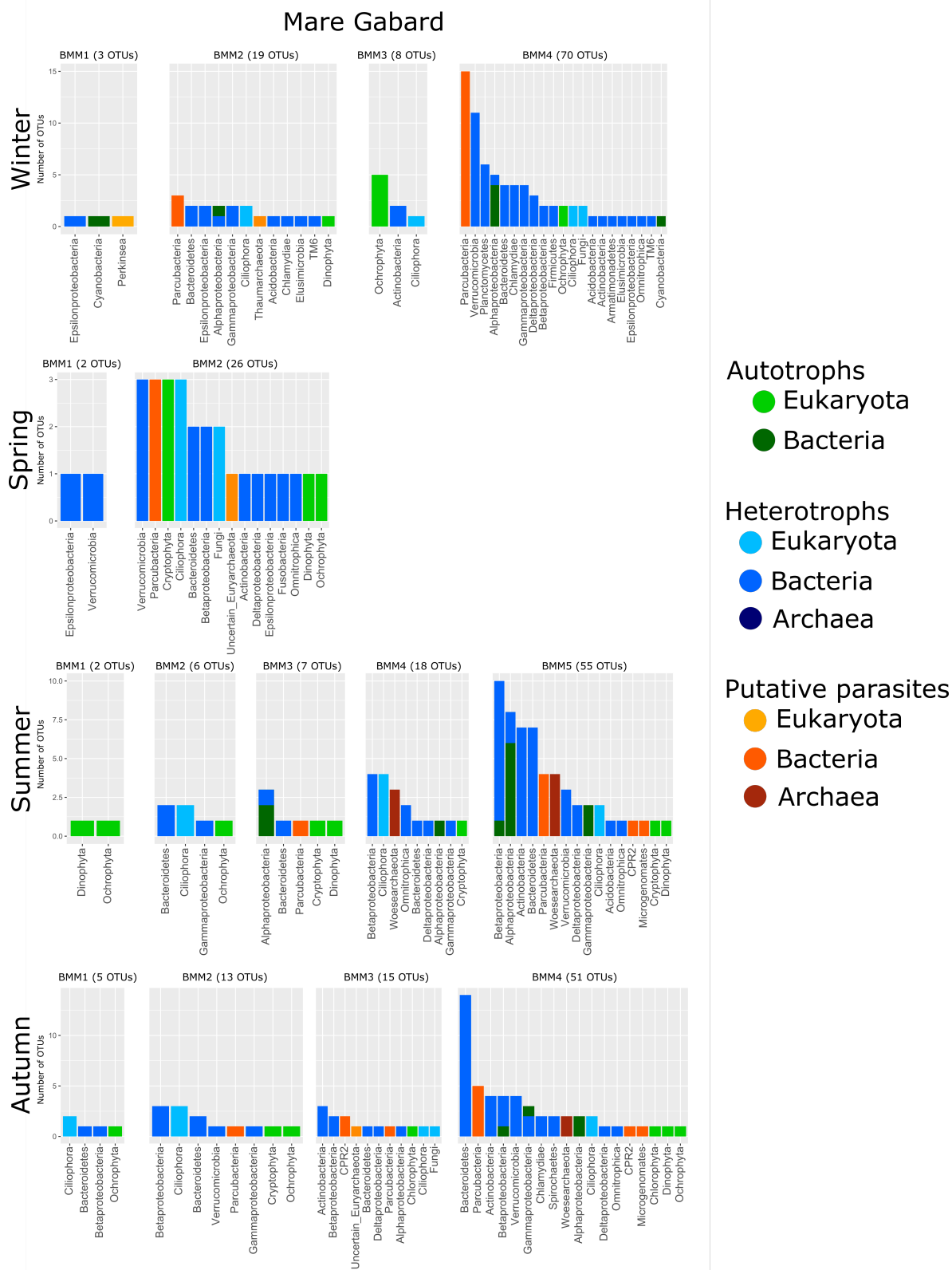


**Supplementary Figure 6.B.2** Barplots presenting the OTU composition of the block model groups (BMG) for the season networks of the Etang des Vallées. Near to the BMG number is indicated the total number of OTU involved in the BMG. Taxonomic affiliation is indicated for each bar. The color of the bar indicates the kingdom and the lifestyle of the taxa. Shade of green correspond to autotrophic OTUs, shade of blue correspond to heterotrophic OTUs and shade of orange correspond to putative parasite OTUs.

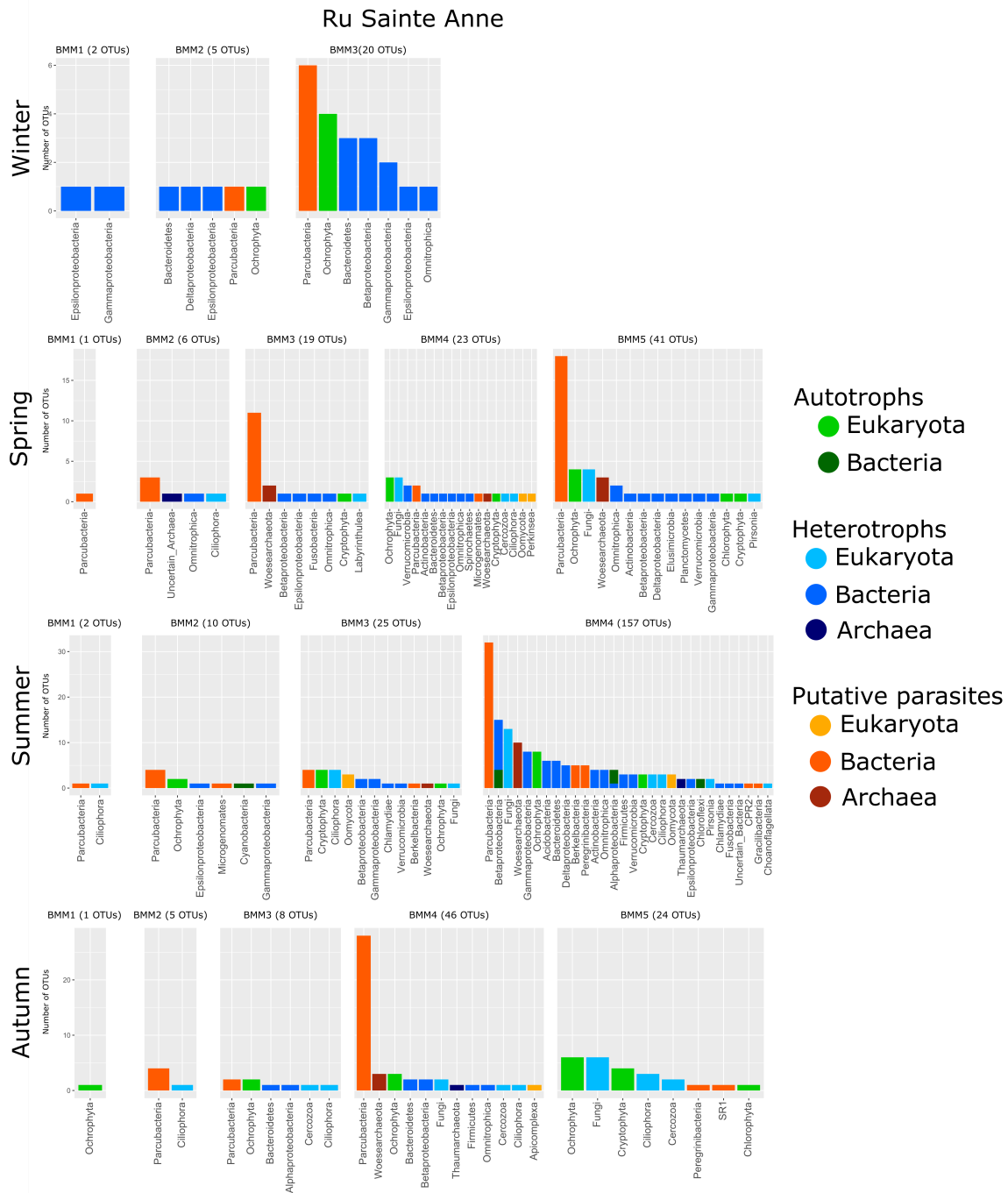




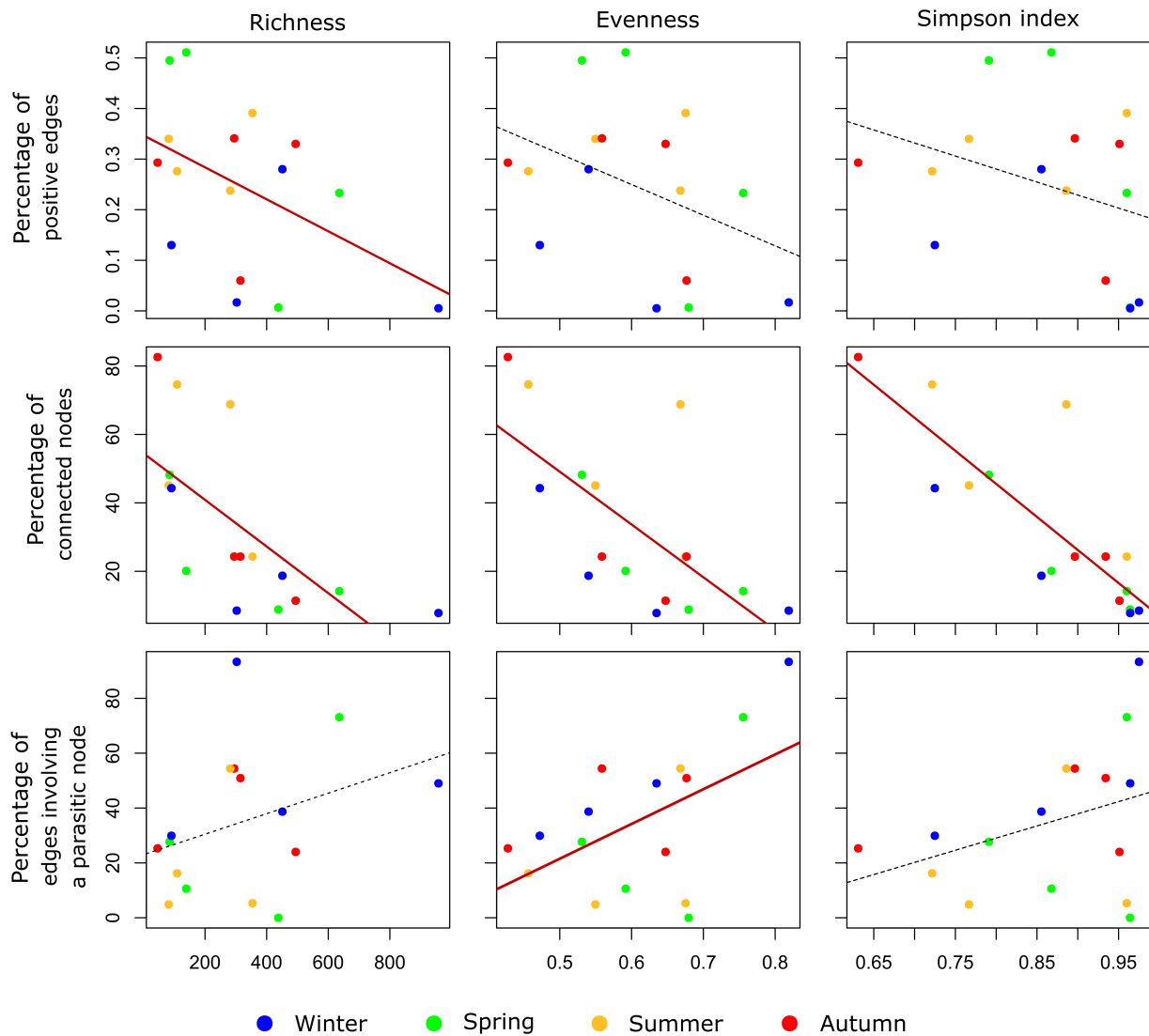
**Supplementary Figure 6.B.3** Barplots presenting the OTU composition of the block model groups (BMG) for the season networks of La Claye. Near to the BMG number is indicated the total number of OTU involved in the BMG. Taxonomic affiliation is indicated for each bar. The color of the bar indicates the kingdom and the lifestyle of the taxa. Shade of green correspond to autotrophic OTUs, shade of blue correspond to heterotrophic OTUs and shade of orange correspond to putative parasite OTUs.



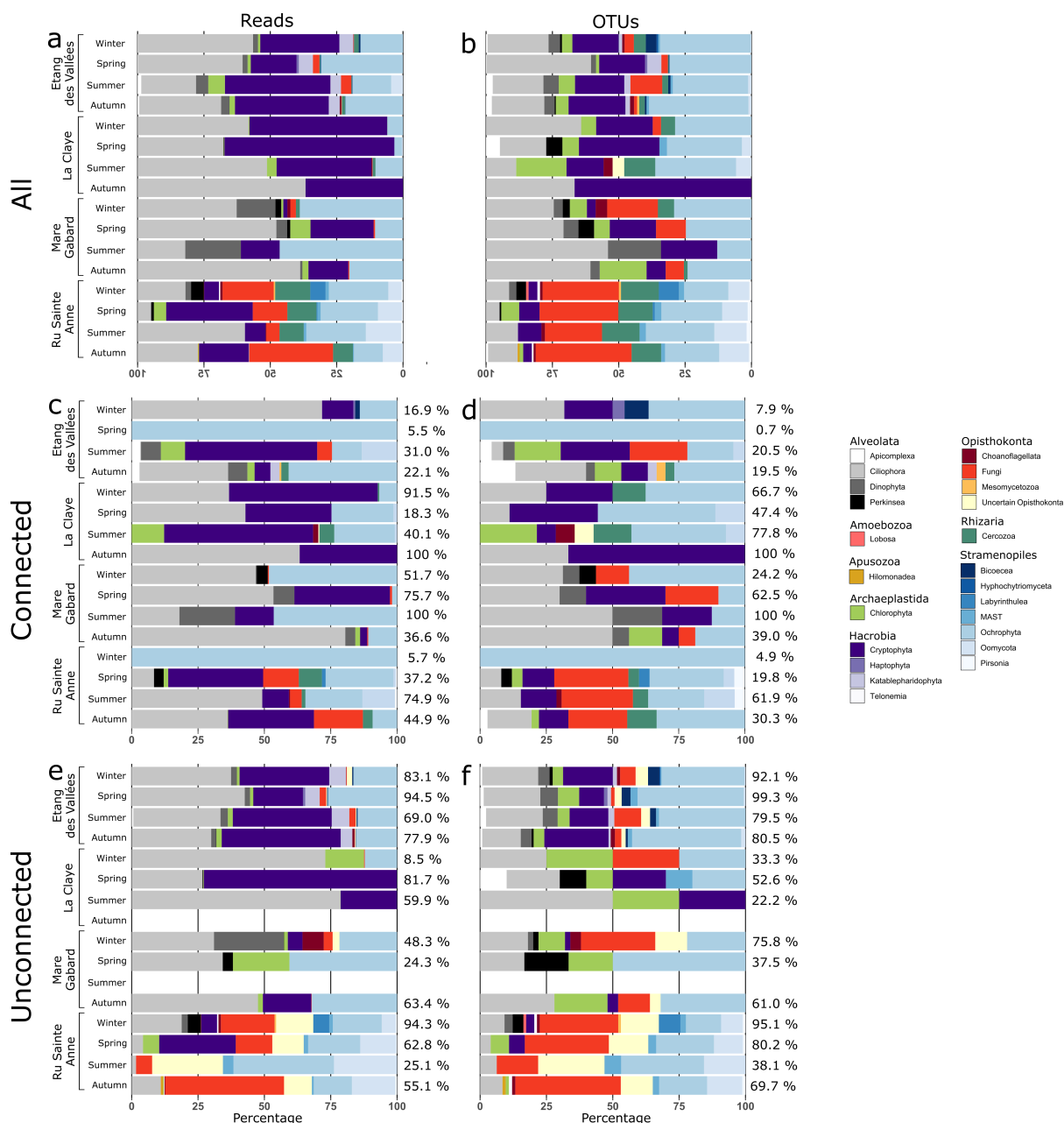
**Supplementary Figure 6.B.4** Barplots presenting the OTU composition of the block model groups (BMG) for the season networks of Mare Gabard. Near to the BMG number is indicated the total number of OTU involved in the BMG. Taxonomic affiliation is indicated for each bar. The color of the bar indicates the kingdom and the lifestyle of the taxa. Shade of green correspond to autotrophic OTUs, shade of blue correspond to heterotrophic OTUs and shade of orange correspond to putative parasite OTUs.



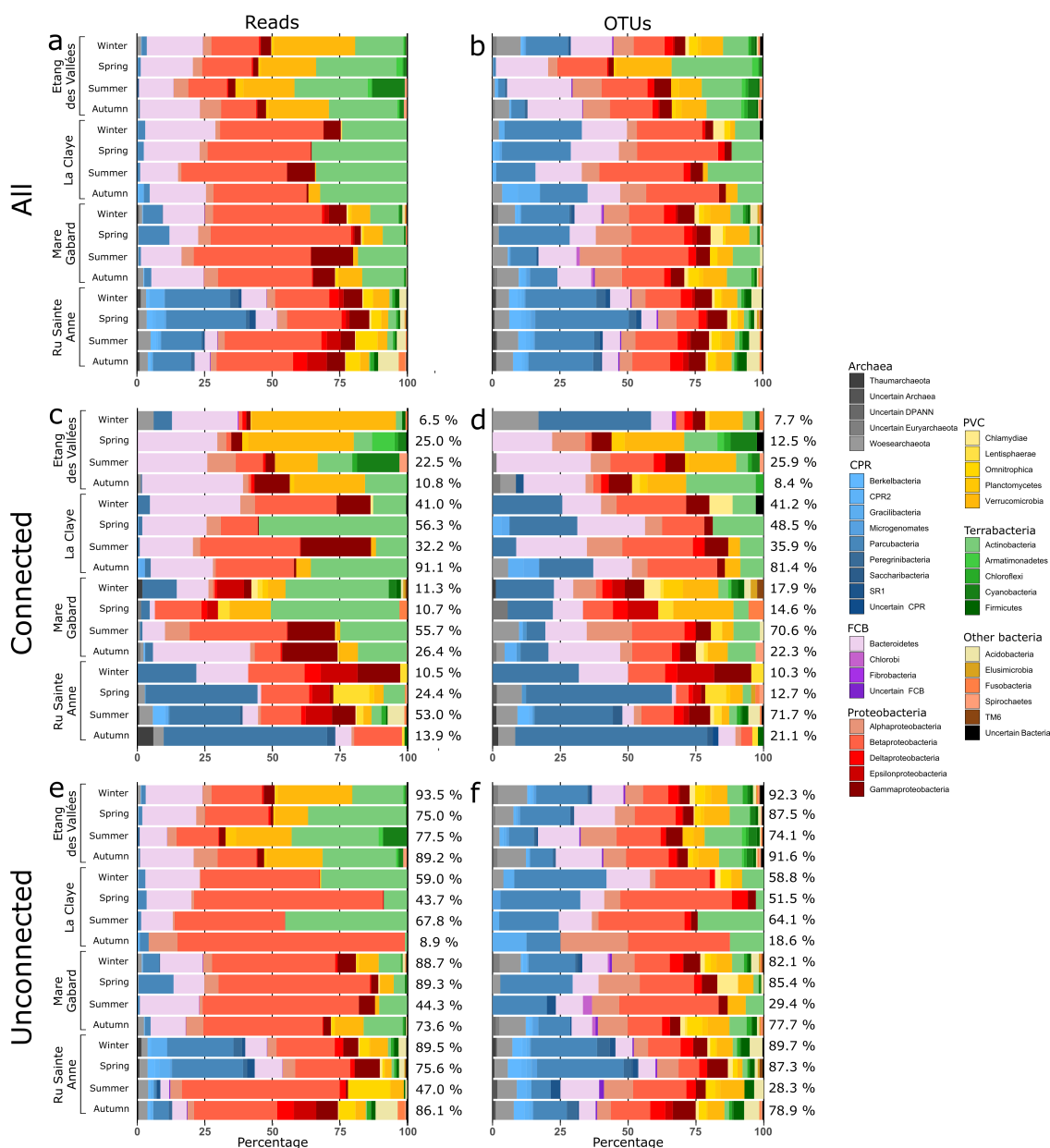
**Supplementary Figure 6.B.5** Barplots presenting the OTU composition of the block model groups (BMG) for the season networks of Ru Sainte Anne. Near to the BMG number is indicated the total number of OTU involved in the BMG. Taxonomic affiliation is indicated for each bar. The color of the bar indicates the kingdom and the lifestyle of the taxa. Green shades correspond to OTUs affiliated to autotrophs, blue shades correspond to OTUs affiliated to heterotrophs and orange shades correspond to OTUs affiliated to putative parasites.



**Supplementary Figure 6.B.6** Representation of the links between network topology and diversity indexes. The significant regressions are indicated in thick and red lines, whereas the non-significant are indicated with dashed lines. The color of the points corresponds to the season of the corresponding network.



**Supplementary Figure 6.B.7** Histogram showing the composition of the eukaryotic nodes in the networks. The networks were built on the OTUs present in at least 6 seasons over the 8 years and presenting a relative abundance higher than 0.05%. They are represented here in terms of (a) percentage of reads and (b) number of OTUs for each season in each ecosystem. The composition of the connected nodes in the networks is represented in terms of (c) percentage of reads and (d) number of OTUs for each season in each ecosystem. The composition of the unconnected nodes in the networks is represented in terms of (e) percentage of reads and (f) number of OTUs for each season in each ecosystem. The percentage of reads or number OTUs represented by the connected or the unconnected nodes on the total number of nodes is indicated near each histogram.

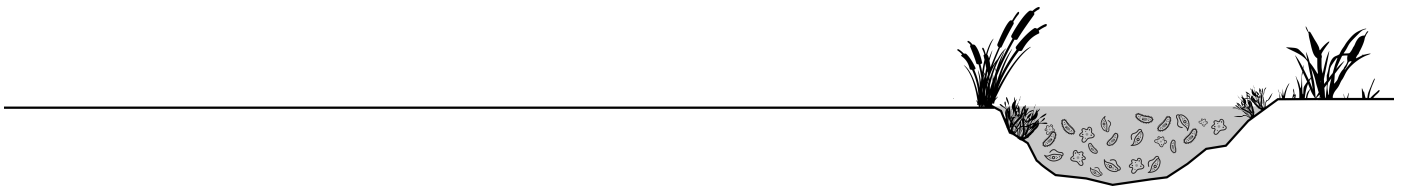


**Supplementary Figure 6.B.8** Histogram showing the composition of the prokaryotic nodes in the networks. The networks were built using the OTUs present in at least 6 seasons over the 8 years and presenting a relative abundance higher than 0.05%. They are represented here in terms of (a) percentage of reads and (b) number of OTUs for each season in each ecosystem. The composition of the connected nodes in the networks is represented in terms of (c) percentage of reads and (d) number of OTUs for each season in each ecosystem. The composition of the unconnected nodes in the networks is represented in terms of (e) percentage of reads and (f) number of OTUs for each season in each ecosystem. The percentage of reads or number OTUs represented by the connected or the unconnected nodes on the total number of nodes is indicated near each histogram.



# Chapter 7

## General Discussion







The main aim of this thesis was to understand microbial community structure in the more poorly studied freshwater ecosystems. Because of their important roles in biogeochemical cycles and trophic webs, any change in microbial communities may affect the functioning of freshwater ecosystems. To reach this objective, I considered *(i)* two types of freshwater ecosystems (the largest lake in the world and five representatives of small ecosystems), *(ii)* two spatial surveys (horizontal and vertical), and *(iii)* two temporal surveys (at various frequencies and duration). Moreover, I focused on different techniques, with *(iv)* two size-fractionation (0.22 to 5 or 0.22 to 30  $\mu\text{m}$ ), *(v)* two massive sequencing technologies (Pyrosequencing 454 and Illumina MiSeq), and *(vi)* two types of analyses (multivariate analyses and network inferences). The first study focused on a spatial survey of the community of microbial eukaryotes in Lake Baikal, along both latitudinal and depth gradients (Chapter 3). This exploration allowed me to investigate the spatial variability of microbial community composition punctually in a large aquatic environment. Then, I explored another component of spatial variability by studying five contrasted small ecosystems. First, I explored the small microbial communities (with sizes in the range 0.2 - 5  $\mu\text{m}$ ) using a 2-years monthly survey, with a particular focus on the small prokaryotic communities (Chapter 4). Then I explored the temporal variability of microbial communities composition at both fine and large scales with respect to different organisms size (Chapter 5). Finally to better understand the functioning of these communities I characterized the interactions between microorganisms in Lake Baikal and in the small freshwater ecosystems by inferring co-occurrence networks (Chapters 3 and 6).

## 7.1 Microbial community composition in freshwater ecosystems

The description of the organisms inhabiting an ecosystem is one of the prerequisite steps to take before addressing its functioning. Microbial communities in Lake Baikal are sporadically depicted in the literature, where only particular areas of the lake were investigated and rarely along the entire water column. For instance, Kurilkina et al. (2016) explored the prokaryotic communities in water columns of the central basin whereas Annenkova et al. (2020a) studied the microbial eukaryotic communities in the entire latitudinal gradient of the lake, but the sampling was not deeper than 50 m. Therefore, integrative studies of the microbial communities in both vertical and horizontal dimensions are missing. In small freshwater ecosystems, despite their clear ecological interest and practical advantages, microbial communities remain poorly described today (see Chapter 1). Therefore, for both our ecosystem models, I first started by describing the composition of their microbial communities.

### 7.1.1 Global composition of microbial communities

The eukaryotic microbial communities of the five small freshwater ecosystems and of Lake Baikal present similarities in terms of phyla composition. The most important phyla are Ciliophora, Ochrophyta, Fungi and Cryptophyta for both types of ecosystems, at the exception of Dinophyta which were found to be abundant in Lake Baikal only. Such a community structure is consistent with what was found in other freshwater ecosystems, such as lakes (Debroas et al., 2017). Likewise, Humbert et al. (2009) compared the bacterial composition of temperate and tropical lakes, and found that they presented similar communities in terms of relative abundance of bacterial phyla. Therefore, it appears that freshwater aquatic ecosystems feature a similar global microbial community composition. However, there are large differences in the community composition at the OTU scale. Comparative analysis of the small freshwater ecosystems revealed that a large proportion of the OTUs were not shared between ecosystems, despite their geographical proximity (Simon et al., 2015, Chapter 4 and 5). This observation is consistent with other studies on freshwater ecosystems, such as the comparison of three Austrian lakes or of eight bog lakes in Wisconsin, USA (Linz et al., 2017; Bock et al., 2018). Moreover, compared to Lake Baikal and to other studies on microbial communities using Illumina MiSeq sequencing, we detected a higher richness in the small freshwater ecosystems (Bock et al., 2018; Obertegger et al., 2019). Therefore, small freshwater ecosystems seem to harbour more diverse communities than larger ones. However, this observation can be the result of different biases induced during library construction (see the paragraph below). Moreover, we conducted the survey over eight-years, which allow us to capture the largest part of the rare biosphere (Pedrós-Alió, 2006). Another hypothesis is that the pelagic area of small ecosystems is more subject to disturbances and to variations of environmental conditions, which generates higher turnovers of communities at the intra- and inter-annual scales.

In both Lake Baikal and small ecosystems surveys, the representation of the communities that we obtained has to be handled carefully. Indeed, the metabarcoding approach based on 16S and 18S rDNA can generate a number of biases. Among them, we can underline the fact that the data are compositional (Gloor et al., 2017). This means that we do not have a precise quantitative image of the communities. Some microorganisms such as ciliates present a higher abundance of gene copies for example (Prescott, 1994; Ward et al., 1997). This can result in an amplification of their detection rate and can make them appear more important in the community than they really are. Also, as we are targeting rDNA, we do not differentiate the metabolic states of the microorganisms. For instance, we can easily hypothesise that the photosynthetic organisms detected at the bottom of Lake Baikal, *i.e.* in the euphotic zone, are inactive cells. The proportion of inactive cells can vary between ecosystems (Lennon and Jones,

2011). Therefore, it is important to take into account the presence of both active and inactive cells in consideration in the interpretations. Finally, the comparison of communities described in previous studies has to be handled carefully. Indeed, various methodological aspects can generate differences between the metabarcoding libraries. Among others, the primer sets used can generate biases in the amplification, which therefore affects the richness estimation (Debroas et al., 2017).

#### 7.1.1.1 Frontier between freshwater and marine systems

Transitions of microorganisms between marine and freshwater systems have long been considered to be rare (Logares et al., 2009). The main difference between the two types of ecosystems is their salinity, which constitutes a strong frontier for microorganisms. Indeed, freshwater microorganisms are adapted to a low salinity level: they have to maintain a constant osmotic pressure, the salinity inside the cell being higher than the one in its environment (Logares et al., 2009). Salinity being one of the major drivers of microbial communities, freshwater and marine systems harbour different inhabitants (Lozupone and Knight, 2007). For instance, haptophytes are very abundant in marine systems, but are found in low proportion in freshwaters. Moreover, some marine groups (*e.g.* diplomonads, radiolarians) have never (or rarely) been observed in freshwater systems. Nowadays, some studies are reevaluating the frequency of these marine/freshwater transitions (Simon et al., 2014; Paver et al., 2018). In the case of our small freshwater ecosystems and Lake Baikal, several taxa known as typically marine lineages have been detected. Indeed, several lineages of Marine Stramenopiles have been identified in our freshwater ponds, whereas some of them had only been detected in marine systems previously (Simon et al., 2014). It may be less surprising for Lake Baikal, which has the properties of a freshwater sea: it is the deepest lake and the largest volume of freshwater in the world, and it is highly oligotrophic. However, it presents a very low salinity level (0.0‰ in Lake Baikal versus an average value of 35‰ in marine systems; Madsen, 2015). Several marine lineages have been detected in Lake Baikal, in the water column and in sediment samples (Annenkova et al., 2020a; Yi et al., 2017; Reboul et al., submitted). Furthermore, members of the microbial eukaryotic communities of Lake Baikal are phylogenetically close to Arctic lineages, indicating that they might share similar ancestors (Annenkova et al., 2020a,b).

## 7.2 Spatial changes in microbial community composition

### 7.2.1 Effect of depth on microbial communities

The strongest environmental effect driving microbial eukaryotic communities in Lake Baikal appears to be the sampling depth, that can be considered as a proxy for light variation, nutrient concentration or pressure for instance. However, we did not detect clear changes in the stratification with depth of the measured environmental parameters, which remain rather stable whatever the depth is. As we did not measure a large panel of abiotic parameters, we did not find which ones affected the most the communities. Microbial communities in the different depth classes (*i.e.* epi-, meso- and bathypelagic) present dissimilarities in terms of community composition and structure. This depth-related fractionation has been also observed in Lake Baikal for prokaryotic communities (*e.g.* Mikhailov et al., 2019b, and is common in freshwater ecosystems (Kurt, 2019). Lake stratification and light variation lead to different environmental conditions along the water column, than can explain variations in community composition. The community composition of bathypelagic sample may be also influenced by the proximity with the sediments at the bottom of the lake, even if the sampling has been conducted carefully in order to avoid any possible disturbance. Sediment communities in freshwater ecosystems are different from water column communities (Jiang et al., 2006). The mixing of the sediment with the planktonic communities can explain part of the variation of the microbial communities along the water column. In fact, the detection of photosynthetic organisms below the epipelagic area, likely inactive, could be an indicator of the exchanges at the water-sediment interface (Zakharova et al., 2013). In small freshwater ecosystems, the water column is too shallow to conduct correctly a sampling at different depth. However it is likely that there is also a variation of community composition with depth, for instance between the communities close to the sediment and the surface communities. However, the frequent mixing of the water column may attenuate the differences, with sediment being able to reach surface more often (Blottière et al., 2017). This implies a strong temporal variability in the ecosystems.

### 7.2.2 Variation of microbial communities depending of the sampling station

The horizontal variations of community composition can be estimated with two complementary approaches. We can focus either on different ecosystems and calculate the dissimilarities between their respective communities, or on one specific ecosystem, often large, and estimate

the differences of community composition at different sampling points within it.

The comparison of several ecosystems often reveals strong differences in community composition. The segregation of samples according to the ecosystem of origin is observed in several studies, for both eukaryotes and prokaryotes (*e.g.* Chase (2007); Lear et al. (2014); Simon et al. (2015); Linz et al. (2017); Bock et al. (2018)). It underlines the fact that communities are influenced by a large number of drivers. On the contrary, we found that the choice of the sampling basin has a small effect on the microbial communities inventoried in Lake Baikal. At a given depth from North to South, microbial eukaryotic communities remained similar (Chapter 3). The same result was obtained for prokaryotic communities by Mikhailov et al. (2019b) and for eukaryotic communities in the southern basin by Yi et al. (2017). These differences can be explained by possible changes in environmental conditions. Indeed, the southern basin is more impacted by human activities than the northern basin which is a preserved area. However, it has been shown that the dissimilarities in the composition of microbial communities are in general lower within ecosystems than between them (Yannarell and Triplett, 2004). In small freshwater ecosystems, for practical reasons and due to their limited sizes, we chose to focus on only one sampling point, at the middle of the ecosystems in the pelagic area. A preliminary check for the homogeneity in community composition within ecosystems was conducted previously by Simon et al. (pers. comm.). The comparison of the microbial communities sampled at three different locations in the Etang des Vallées revealed a high similarity within the ecosystem. However, it was a punctual comparison, and surface communities might not be homogeneous. For instance, Lear et al. (2014) established in alpine ponds that there were significant dissimilarities of bacterial communities within aquatic ecosystems, especially for samples separated by more than 20 meters. Moreover, even if we did not detect a coastal effect in Lake Baikal, we can hypothesize that in small freshwater ecosystems the littoral zone offers more different environmental conditions than the pelagic area does. In particular, the littoral zone hosts denser populations of macroorganisms, is adjacent to the potential terrestrial perturbation, and receives all the ecosystem inflow (Lewis, 2009; Obertegger et al., 2018; Søndergaard et al., 2005).

### 7.3 Temporal evolution of microbial communities

Our temporal surveys of the small freshwater ecosystems have two major advantages: (*i*) we focused on the three domains of life and (*ii*) during a long term temporal survey. Indeed, the temporal surveys often treat only eukaryotic, bacterial, or prokaryotic communities, and the average number of samples collected is only 15 time points per time series (Buckley et al., 2019; Lambert et al., 2019).

Using two temporal surveys differing in frequency and duration, I detected two types of temporal patterns in the five small freshwater ecosystems, for both eukaryotic and prokaryotic communities. The first pattern is a seasonal pattern. Intra-annual seasonal patterns have been widely observed at the surface of marine and freshwater aquatic ecosystems (*e.g.* Hullar et al. (2006); Lara et al. (2011); Gilbert et al. (2012); Yan et al. (2017); Bunse et al. (2019)). In most studies, communities presented clear differences between warm (summer) and cold (winter) periods. Autumn and spring often act as transitional states. However, seasonal patterns are not observed in all ecosystems. For instance, Linz et al. (2017) did not observe repeatable annual patterns of bacterioplankton communities in freshwater bog lakes. Neither did Lepere et al. (2006) in their study of small microbial eukaryotes in Lake Pavin. Given the multiple environmental perturbations that face small freshwater ecosystems during the year and their small buffering effect, the seasonal patterns of microbial communities I found in small freshwater ecosystems were rather unexpected. Moreover, even if I obtained clear seasonal patterns, dissimilarities between samples remained high. This underlines that, globally, there is a significant change in the whole microbial community composition at the monthly and the seasonal scale, for both eukaryotes and prokaryotes. The second pattern is an inter-annual rise of the dissimilarities between microbial community composition in the five small freshwater ecosystems. Dissimilarities between communities are rising with time lag between samples. Such a temporal evolution was depicted in other ecosystems, such as subtropical reservoirs or marine systems (Chow et al., 2013; Cram et al., 2015; Liu et al., 2015; Bock et al., 2018). Both patterns were observed for eukaryotic and prokaryotic communities, underlining concurrent temporal variations between prokaryote and microbial eukaryote communities (Liu et al., 2015; Bock et al., 2018; Obertegger et al., 2019). This temporal synchrony between prokaryotes and eukaryotes indicates that they have similar responses over time despite their ecological differences (Masana and Logares, 2013). It can be driven indirectly by the influence of abiotic factors, or directly because of biotic interactions such as predation or parasitism (Kent et al., 2007; Bock et al., 2018). Moreover, these patterns have also been detected for both abundant and rare microbial communities. For the abundant community, it was surprising to observe a directional change. Indeed, abundant taxa can be considered as a core community, well adapted to the ecosystems (Pedrós-Alió, 2006). As the measured physico-chemical parameters remained stable in time, the abundant community could have followed the same pattern. It was also surprising to observe a seasonality of the rare community, because it is composed of low abundance and low frequency taxa and it presented high turnover rate at the intra- and inter-annual scale. However, seasonality in the occurrence of rare communities has already been described in marine bacterioplankton (Alonso-Sáez et al., 2015). This underlines the fact that rare communities are not only dormant cells, but play an active role in microbial communities (Logares et al., 2015; Debross et al., 2015).

One hypothesis to explain these gradual changes over time in community composition could be that physico-chemical conditions in ecosystems changed over time. Therefore, I investigated how abiotic parameters could explain the temporal dynamics of microbial communities. However, we did not detect strong directional changes over the eight-year survey in the measured environmental parameters. On the other hand, the seasonal variation of the parameters is clearly marked in all ecosystems, similarly to what we observed for microbial communities. Over the entire survey, the measured abiotic parameters explain less than 10% of the community composition variation. Even if several abiotic parameters such as temperature or pH are known to significantly influence microbial community composition (see Chapter 1, section 1.3.1), the total proportion of the community variance explained by measured abiotic parameters remains low (Kent et al., 2007; Lima-Mendez et al., 2015; Linz et al., 2017; Berdjeb et al., 2018). According to the models predator-prey based on phytoplanktonic and zooplanktonic communities of Dakos et al. (2009), inter-annual changes in communities can arise without an inter-annual variability of the abiotic parameters. Dakos et al. emit the hypothesis that seasonal dynamics generates a form of chaos, that favors an inter-annual variability.

At the taxa scale, several patterns could be distinguished. OTUs had either hectic, recurrent, or sporadic dynamics. These taxa dynamics are globally poorly known, because they are overlooked (Lambert et al., 2019). Indeed, microbial communities are complex and diverse, and it is difficult to investigate the dynamic of each taxon. Despite the clear seasonal patterns we observed at the whole community scale, only a small proportion of OTUs presented a clear recurrence at a specific season. Several OTUs have been frequently detected at various abundances along the seasonal surveys times, encompassing several seasons, without distinct pattern. Recurrent OTUs were composed of abundant and rare taxa. Their proportions (in terms of both number of reads and OTUs) were similar to what has been observed in marine communities (Giner et al., 2018; Lambert et al., 2019). Recurrent OTUs occurred either at a seasonally or without any identified occurrence periods. The seasonality of a taxa was characterised by a peak of abundance during one specific season, and an absence or a very low abundance the rest of the year. Non-seasonal recurrence corresponded to OTUs that peaked for instance during two consecutive seasons several times. Finally, in our study, a large number of OTUs have been detected only once over the whole survey (representing respectively 54.7% and 39.3% of the eukaryotic and prokaryotic OTUs). It is consistent with other surveys, with for instance the study of Shi et al. (2020) where 50% of the richness was represented by OTUs detected in less than three samples over 12. Similarly, Giner et al. (2018) observed that almost 90% of the OTUs were detected less than 10 times over 120 sampling points. These infrequent OTUs are probably detected during an immigration process, but they do not succeed in colonizing the ecosystem and are not detected later during the survey Jia et al. (2018).



## 7.4 Exploration of the interactions in the communities

### 7.4.1 Spatio-temporal structure of co-occurrence networks

In aquatic ecosystems, organisms are in constant interaction, which results in either beneficial or negative impacts on the protagonists. Studying these interactions is a major computational and experimental challenge. In this thesis, I used co-occurrence networks to have a first insight into the organisation of the microbial communities inhabiting different ecosystems.

The study of the co-occurrence networks in the three depth classes (*i.e.* epi-, meso- and bathy-pelagic) of Lake Baikal revealed different microbial community interaction patterns. Planktonic communities at surface appeared to be more complex, showing highly interconnected OTUs. As I did not find large differences in microbial communities between the three basins (they only differ with depth) with multivariate analyses, I analysed samples from the same depth zone simultaneously. It allowed me to have available a network based on a larger number of samples. In order to consider the slight effect of the sampling basin, I added it as a covariable in the networks. However, in 2019, Mikhailov et al. showed a variation of network topology as a function of geographical location in Lake Baikal during spring blooms in the pelagic zone. Contrarily to our study, they focused on both microbial eukaryote and prokaryote communities, which provided them a punctual view of planktonic communities at surface of Lake Baikal. In terms of depth variation of community structure, the differences in network complexity (*e.g.* more connected nodes, edges or higher connectance) between surface and bottom samples have already been observed in other ecosystems (Jiao et al., 2020). Communities at surface may be more complex compared to deepest communities. Indeed, the constant effect of external disturbances such as rain, wind, and the mixing of the water column induces more perturbations, as well as more dispersal events (Zeng et al., 2019; Jiao et al., 2020).

To study the interactions between microbial communities over time, the datasets are often summarized in one network taking into account all the samples (*e.g.* Steele et al., 2011; Chow et al., 2013; Parada and Fuhrman, 2017). The comparison of the structure of the community over several years with one network per season remains scarce (*e.g.* Jones et al., 2018; Zhu et al., 2019; Jiao et al., 2020). Moreover, co-occurrence patterns are rarely compared between different ecosystems (Williams et al., 2014). Here, we described microbial communities inhabiting ecosystems differing in their abiotic conditions and sizes, but also compared them one to each other by building one network per season per ecosystem. Because of the seasonality of the sampling design, I obtained a snapshot of the community structure at the seasonal scale but I was not able to infer specific interactions between two microorganisms into details. I

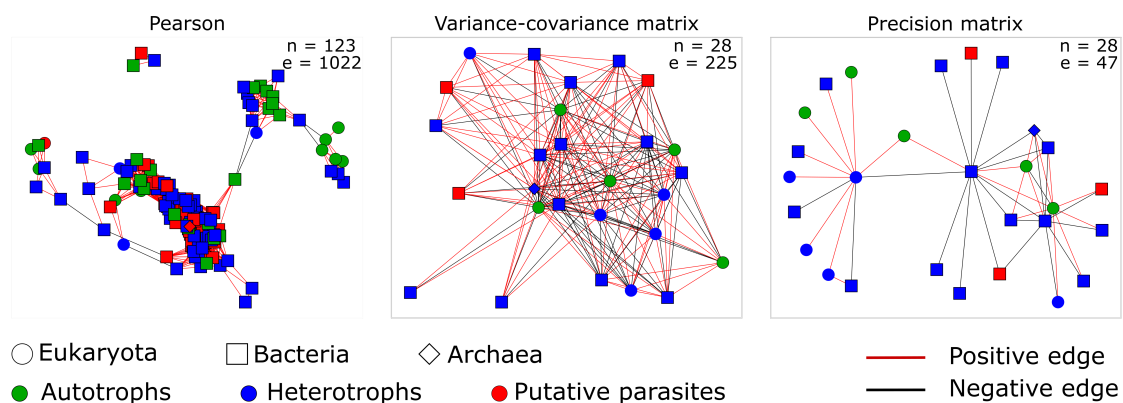
highlighted different network structures depending on the ecosystem considered. Differences in community structure between ecosystems could be driven by their nutrient status. Indeed, a study on lake communities has highlighted that surface networks are more complex in oligotrophic systems than in eutrophic ones (Cao et al., 2018; Zeng et al., 2019). The five small freshwater ecosystems that I considered were all eutrophic, but I did not rank them following an eutrophication level in this work. However, even if the topology of the networks were globally different, I detected some common seasonal features between ecosystems. For instance, we found that networks in spring and summer (warm period) networks were globally more complex in term of number of connections than those in autumn and winter (cold period). The positive link between warmer conditions and the increase of interaction had already been observed in bacterioplankton communities in lakes (Kara et al., 2013; Zhao et al., 2016). Warm seasons are characterized by a higher primary productivity, which could explain that there are more interactions within the ecosystems, especially direct or indirect interactions with the autotrophic microorganisms (Cotner and Biddanda, 2002).

Finally, the interactions that I inferred in the networks were biologically meaningful. More specifically, typical features of microbial community structuring, such as the tight link between autotrophs and heterotrophs, are expected to be retrieved in co-occurrence networks. Autotrophic organisms appear to be central in the networks, often corresponding to the most connected nodes. These highly connected microorganisms can be considered as keystone species (Berry and Widder, 2014). Further investigations on the interactions in the five small freshwater ecosystems could be based on these highly connected taxa. Indeed, they have a key role in the functioning of the ecosystems and can help to predict changes in community composition (Herren and McMahon, 2018).

### **Large variety of methods to infer co-occurrence networks**

The construction of co-occurrence networks faces different methodological issues when using metabarcoding data. It is important to be aware of pros and cons of the many construction methods that are available (Röttjers and Faust, 2018). The most standard way to investigate interactions between organisms is done using pairwise correlations, such as the Pearson's correlation coefficient (Woodhouse et al., 2016). When using this method, correlations are calculated between each pair of organisms and only the most significant and important interactions are kept for the network constructions. However, this method can lead to spurious correlations resulting from compositional effects for instance, and should not be used alone (Weiss et al., 2016). Alternatively, several approaches such as SparCC are taking into account compositional effects to produce more robust variance-covariance matrix of the taxa (Friedman

and Alm, 2012). However, they often do not allow to include environmental covariates and they do not take into account the existence of indirect interactions. In this thesis, I decided to build co-occurrence networks based on a method recently developed by Chiquet et al. (2019). For now, only few studies are based on this method, but it presents several advantages (*e.g.* Sassenhagen et al., 2020). The PLNmodel method is based on the precision matrix, that is the inverted variance-covariance matrix. This transformation allows to detect and delete associations resulting from indirect covariations, such as covariations between two OTUs due to their direct association to a third OTU (Kurtz et al., 2015). A comparison of three constructions methods revealed that networks based on simple Pearson's correlation calculations and variance-covariance matrices were more complex than those based on the precision matrix produced by PLNmodels (see Figure 7.4.1). The PLNmodel approach is more stringent and avoids spurious correlations. This method has many assets, that were important for me. Indeed, it is robust regarding compositionality, it uses raw count data, it allows to take into account both eukaryotic and prokaryotic dataset in the same network, and to take into account both quantitative and qualitative covariables (*e.g.* physico-chemical parameters or sampling season). Finally, once the method to analyse the data is chosen, the interpretation must be conducted carefully. Indeed, the association networks do not represent true associations (Derocles et al., 2018). Positive and negative edges can be interpreted by various types of interactions, and sometimes the same in the case of a parasitism for instance, because of the co-presence of both host and parasite.



**Figure 7.4.1** Co-occurrence network of the microbial communities in Mare Gabard in spring using three different methods. *Left panel:* network based on Pearson's correlation, with a selection of the correlation  $\geq 0.7$  and with a p-value  $\leq 0.01$ . *Middle panel:* network based on PLNmodel but without an inversion of the variance-covariance matrix. *Right panel:* network based on PLNmodel. For each network, the number of nodes (n) and edges (e) is indicated.

### 7.4.2 Distribution of the functional groups in the ecosystems

After the description of the composition of microbial communities over space and time, of their link to environmental parameters, and the investigation of the structure of the communities, I explored the temporal dynamics and interactions of the main microbial functional groups to address the functioning of an ecosystem.

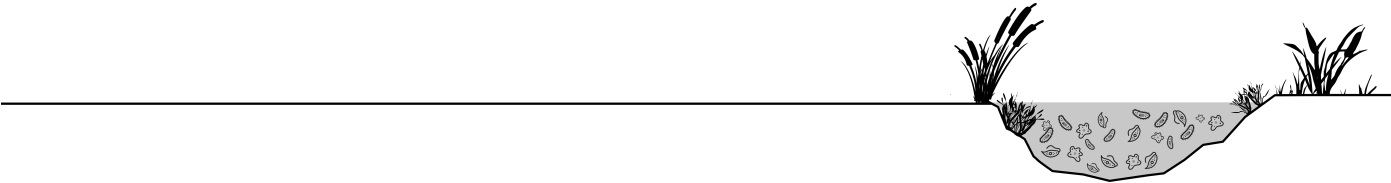
A classical way to address this question using metabarcoding data is to attribute to each phyla or genus a trophic mode or lifestyle (*i.e.* autotroph, heterotroph, putative parasite) based on the literature. These attributions are not highly precise because the known microorganisms groups can harbour different types of trophic strategies (Genitsaris et al., 2016). For instance, gammaproteobacteria encompass bacteria that are heterotrophic and others that are autotrophic. Microorganisms can also harbour mixotrophic lifestyle and are therefore capable of both autotrophy or heterotrophy. I established a corresponding list between taxonomic groups and lifestyle. For the taxonomic groups, I used different levels (*e.g.* phyla, class) in order to be as accurate as possible.

In both small freshwater ecosystems and Lake Baikal, I detected a majority of heterotrophic organisms (respectively about 53.9% and 60.3% of the total number of reads of the microbial community in Baikal and in the freshwater ponds). With the seasonal networks analysis in the small freshwater ecosystems, I observed that the majority of the positive links were established between autotrophs and heterotrophs. These relationships can be representative of predator-prey interactions, such as for instance the grazing of phytoplankton by heterotrophic predators (Pernthaler, 2005). The putative parasitic organisms were more present in the networks in winter in small freshwater ecosystems, whereas in Lake Baikal, the global putative parasitic load tended to be lower in deep samples. In winter, putative parasitic organisms could take advantage of a certain stress of other organisms due to lower nutrient and light level (Cleary and Durbin, 2016). Deeper, the low abundance of putative parasites could be the effect of deep microbial communities constituted mostly of inactive cells that sink to the bottom of the ecosystem (Jones and Lennon, 2010).



# Chapter 8

## Perspectives and Conclusions





## 8.1 Perspectives

### 8.1.1 Explore the missing pieces in ecosystem functioning

This thesis gave me a comprehensive view of the microbial ecology in diverse freshwater ecosystems, at different scales: spatio-temporal dynamics, three domains of life, and two types of ecosystems (*i.e.* the largest on Earth, and common small freshwater ecosystems). However, precisions on physico-chemical conditions, or on the actors in the trophic networks are still needed to go further in the understanding the functioning of aquatic ecosystems.

#### 8.1.1.1 Ameliorate the survey of physico-chemical variations

In this thesis, I took into account several physico-chemical parameters to explore the drivers of microbial community composition. However, there is a very large variety of physico-chemical parameters that can be measured in aquatic ecosystems, and I only considered few of them. As for Lake Baikal, even if it is a highly oligotrophic lake, it could have been interesting to have more data on nutrient concentrations, for instance. Indeed, the chemical composition of the lake is complex, and varies as a function of horizontal and vertical stratifications (Khodzher et al., 2017). Therefore, adding data on phosphorus or nitrogen loads in the lake could give better insights into microbial community spatial structure. Moreover, the impact of nutrient concentrations on bacterial communities is already documented in Lake Baikal (Kurilkina et al., 2016). In the case of small freshwater ecosystems, it would be interesting to have a better view on the meteorological context (*e.g.* rain events, wind speed), and on the short temporal variations of abiotic parameters (*e.g.* daily scale). For instance, a meteorological station placed in the middle of an ecosystem and recording daily the temperature, the precipitations, the light intensity, and the wind speed could help to have a better view of the meteorological events that the ecosystems are facing and therefore help results interpretation. Indeed, a strong rain event the day before the sampling may affect temporary the communities through the inflow of water, with for instance changes a dilution effect in the relative proportion of the taxa and of the abiotic parameters (Jones et al., 2013; Adams et al., 2014; Staley et al., 2015). Similarly, the wind speed can have an impact on the microbial communities, by inducing the mixing of the water, as it has been showed on phytoplankton biomass (Blottière et al., 2017; Rusak et al., 2018). Moreover, wind can generate waves in the ecosystems, and therefore passively transport microorganisms (Garneau et al., 2013). Combined to the survey of wind speed to identified wind-induced mixing, we could monitor the water temperature over a vertical profile in the ecosystems in order to define precisely the stratification in the water column over time



(Martinsen et al., 2019).

Finally, a regular survey of the water level would allow to have a precise idea of the drought events that ecosystems are facing, and to investigate their resilience over time (Simon et al., 2016). Moreover, low water levels can lead to more interactions with the sediments and also more turbidity in the ecosystems, resulting in lower light availability for photosynthesis. Sediment resuspension has a strong impact on microbial communities through the seed bank and their role as vehicle for microbial dispersion in the ecosystems (*e.g.* bacteria associated to sediment particles) (see section 8.1.1.3 and Amalfitano et al., 2017).

### 8.1.1.2 Incorporating other actors of the trophic web

The integration of the three domains of life in our study and the analyses of a large size fraction allowed me to construct networks including many actors of the trophic web, from primary producers to predators. However, considering the web structure (see Figure 1.5), my dataset missed other major actors, such as zooplankton and viruses which are known to be mortality factors of microbial communities. These missing protagonists could explain a part of the currently unexplained variance. On the one hand, viruses are known to have a direct impact on community composition, for instance by regulating prokaryotic communities (Jardillier et al., 2005; Weinbauer et al., 2003). They can control the most abundant communities, a theory known as the “kill the winner rule” (Thingstad, 2000). On the other hand, zooplankton plays a major predating role. Moreover, they release dissolved organic matter in the ecosystem, and therefore provide food for heterotrophic organisms (De Corte et al., 2018). They also move microbial communities through the water column, in particular in the vertical dimension (Grossart et al., 2010). Therefore, to better tackle the interactions which take place within ecosystems, I could determine the composition and abundance of the zooplanktonic communities, and incorporate them in the co-occurrence networks. For both viruses and zooplankton, samples are available and should be considered for future analyses.

There can also be an effect of indirect players on microbial communities. Fishes, amphibians such as tadpoles, and macroinvertebrates generate bioturbation by moving the sediments (*e.g.* while they are searching for food, Meysman et al., 2006). We never investigated the fish communities, but they are at least present in the Etang des Vallées, where fishing activity is supervised by the “Office National des Forêts”. The bioturbation of the sediments can play a role in the resuspension of currently inactive cells that can therefore be remobilized, and it can also affect nutrient concentration (Rugenski et al., 2012; Adámek and Maršálek, 2013). Moreover, studying fishes and amphibians may also be the opportunity to investigate their microbiome.

Indeed, these microorganisms have not been investigated during our study, but they host part of the microbial communities of the ecosystems. Fishes and other macroorganisms can be a habitat for microorganisms, especially for rare taxa, and they maintain the microbial diversity of ecosystems by “sustaining the rare” (Troussellier et al., 2017). For instance, amphibians are well represented in our ecosystems, and they can also be considered as a putative immigration source. A comparative study between the microbial communities of our small freshwater ecosystems and of the microbiome of different amphibians revealed that they share about 40% of OTUs (Dupuis and Pollet, unpublished).

Finally, we are missing a large part of the microbial communities in the interaction networks. For methodological reasons, we had to build networks only based on the highly abundant members of the communities (Röttjers and Faust, 2018). However, rare OTUs are known to have a major importance in the functioning of ecosystems (see Chapter 1, Jousset et al., 2017). Incorporating infrequent taxa would allow us to investigate their important potential role in the community structure. This modelling issue is a whole current field of research. For instance, Cougoul et al. (2019) offer a model integrating the associations between abundant and rare organisms, but not yet between pairs of rare ones.

### 8.1.1.3 Identifying and quantifying external imports of microorganisms

Freshwater ecosystems are not closed systems and are thus in constant interaction with the surrounding environment. In particular, there is a wide range of dispersal pathways for microorganisms, from the surroundings. To better understand community structuring over time, a sampling of the communities inhabiting the potential immigration sources (*e.g.* rain, atmospheric deposition, and ecosystem banks) would allow determining their contribution in community changes (Jones and McMahon, 2009; Comte et al., 2017). For example, a survey could be conducted in Mare Gabard because it is easily accessible and there is only one water inlet that could be sampled. However, even if they have already been identified as potential vector of immigration, animal vectors (*e.g.* birds, wild boars) would be more difficult to investigate (Incagnone et al., 2015). At the same time, it would be of great interest to sample water in the littoral areas of the freshwater ecosystems. Even if there were no differences observed in Lake Baikal between littoral and pelagic areas, it is known that littoral areas are hydrologically unstable (Obertegger et al., 2018). Indeed, they receive the water inflow from the precipitations, and they are an area of interactions with external perturbations, such as animals coming to the shore and mixing the sediments (Hairston Jr and Kearns, 2002). This area is therefore more subject to allochthonous inputs, and may harbor more diverse communities.

More generally, it would be interesting to explore sediment communities in the ecosystems. The benthic communities can contribute significantly to pelagic community composition, through the regular mixing events and the bioturbation by macroorganisms. Moreover, sediments are characterised by a high microbial diversity (Lozupone and Knight, 2007). Even if sediment taxa are mostly anaerobic and not able to survive at the surface, a part of the sediment community is composed of taxa representing a seed bank of microorganisms. Indeed, dormant and inactive cells sink to the bottom of the ecosystems, and can develop in the case of changes in the environmental conditions (Jones and Lennon, 2010; Lennon and Jones, 2011). They constitute a “reservoir of genetic and phenotypic diversity” that can contribute to ecosystem functioning (Shoemaker and Lennon, 2018). Therefore, we could explore sediment communities using sediment traps to have a better view of the communities that can be remobilized in the water column. In Lake Baikal, the effect of mixing on water column communities is not as important as the one in the small freshwater systems. Previous analyses in Lake Baikal have revealed that the lake harboured different microbial eukaryotic communities between the sediments and the water column in the south basin in July (Yi et al., 2017). In our case, sediments were collected and analysed at several stations along the lake (Reboul et al., submitted). Therefore, we could conduct a comparative study of sediment and the pelagic samples to check if these differences are observed in the entire lake. This will also allow to identify taxa that are shared between the sediment and the water-column sample, underlying potential interactions.

### 8.1.2 Determine the proper sampling frequency

One of the main issues that we are facing when handling a long time survey sampled with a large frequency is: do the observed variations correspond to biological variations? I worked at two frequencies, namely monthly and seasonally. Monthly surveys are the most often used, as seen for instance with the long-term microbial observatories in marine systems (reviewed in Buttigieg et al., 2018). Indeed, it is considered as a suitable frequency for the study of microbial ecology (Schauer et al., 2003). Seasonal survey was a good compromise to obtain reliable data on the ecosystem over a long time period without being too much time consuming and costly. It allowed me to explore in depth the dynamics of the whole microbial community (<30 $\mu$ m in size) and specific OTUs at a medium time scale of eight-years. Also, it confirmed the temporal trends observed at the monthly scale. However, the seasonal survey missed some periods, such as drought events. For instance, according to the monthly survey, Ru Sainte Anne had no water in August and September 2012, but it is not revealed by the seasonal survey. Similarly, La Claye was dry during 5 months in 2011 and it only impacted one time point at the beginning of the seasonal survey. Therefore, important events (*e.g.* drought) are not taken into account

in the interpretation of microbial community composition.

Large frequency surveys allow to investigate the global patterns of microbial community composition, but they are not enough to understand the fine-scale of variability of microorganisms (Bunse et al., 2019). To answer this latter question, we could conduct surveys at higher frequencies which is something that has never been done in small freshwater ecosystems. Microbial communities are known to change over very short periods of time (Mangot et al., 2013; Fuhrman et al., 2015). With a more frequent sampling (*e.g.* weekly), I could for instance catch more precisely the variations of abundance of the conditionally rare taxa (i.e. rare taxa that are become occasionally very abundant) and understand the mechanisms leading to these dynamics (Lindh et al., 2015). A survey of day and night variations could also be conducted. Indeed, small freshwater ecosystems present daytime stratification, which could induce changes in composition along the water column within the day (Martinsen et al., 2019). This time lag is principally used to detect day/night cycles and day-to-day changes in metabolic processes and gene expression of microbes, using metatranscriptomic (Needham et al., 2018; Linz et al., 2020; Trench-Fiol and Fink, 2020). Moreover, when the objective of a study is to infer precise interactions between microorganisms using co-occurrence networks, low frequency surveys are not efficient because they do not reflect the lifespan of individuals. Therefore, we would need a biologically adapted time-lag to resolve specific microbial biotic interactions.

In conclusion, it is important to study ecosystems at various sampling frequencies and related durations that are complementary because they address biological questions occurring at diverse time scales (Ladau and Eloë-Fadrosch, 2019). The combination of the different results will provide an integrative view of the structure of microbial communities in time, and of the processes that take place in the ecosystems.

### 8.1.3 Focus on Candidate Phyla Radiation

During this thesis, I conducted a global analysis of the ecosystems. I described the communities and investigated the spatio-temporal modifications of their structure. This approach allowed me to identify some organisms playing potential important roles in ecosystem functioning, such as bacteria of the Candidate Phyla Radiation (CPR). These microorganisms are well represented in the co-occurrence networks, often being at the central position. As presented in the introduction (Chapter 1), CPR members have no cultured representatives, they present a cryptic life cycle, and their functions in the ecosystems remain mostly unknown (Solden et al., 2016). They are suspected to be in interactions with other microorganisms, through parasitism for instance. However, they are also suspected to be producers of molecules (*e.g.* volatile fatty

acids, hydrogen) for methanogens and methanotrophs (Vigneron et al., 2020). CPR members were detected in all our ecosystems sometimes in high abundances, in particular in Ru Sainte Anne. Therefore, it could be interesting to have a specific focus on their dynamics and their role as potential symbionts, in order to identify their role in this ecosystems.

## 8.2 Conclusions

My thesis project was integrated in the global ecological question of the functioning of freshwater ecosystems. A first step to answer this question is to better understand the composition and the structure of all the microbial communities (bacteria, archaea and microbial eukaryotes) inhabiting freshwater ecosystems. To do so, I used spatio-temporal surveys and I focused on two types of freshwater ecosystems: large lakes with the study of Lake Baikal, and small freshwater ecosystems, with four ponds and one brook. The dataset for Lake Baikal consisted in a spatial survey of the microbial communities over horizontal and vertical gradients. The dataset of the small freshwater systems consisted in a temporal survey of the three domains of life, with two durations and frequencies. I can draw the following conclusions:

- **Freshwater ecosystems present a wide diversity of microorganisms.** In both Lake Baikal and small freshwater ecosystems, we detected all the eukaryotic supergroups. In addition, we detected all the prokaryotic supergroups in small freshwater ecosystems. The most abundant phyla in both type of ecosystems were emblematic of freshwater microbial communities. Communities in the ecosystems were divided into abundant taxa and diverse rare taxa.
- **Salinity barrier seems to be less selective than what was previously thought.** Several typical marine lineages were discovered over the entire Lake Baikal, concordantly with other studies. Marine affiliated groups were also observed, although to a lesser extent, in small freshwater systems. The salinity barrier crossing may not be a rare event.
- **The composition of microbial communities is highly influenced by depth.** Microbial composition is stratified over the water column, with surface communities being different from their benthic counterparts. Depth can be considered as a proxy for various environmental parameters (*e.g.* light), and is a strong factor governing community composition.
- **Microbial communities are dissimilar among different ecosystems, and similar within a given one.** The microbial communities seem rather homogeneous inside a single ecosystem, even a large one. However, microbial communities are specific to one ecosystem, and present strong differences from one ecosystem to another, even if they are close to each other.
- **Seasonality is a marked pattern in microbial communities observed in small freshwater ecosystems.** In terms of community composition and structuration, each season presents specificities. The strongest differences are observed between winter and summer.

- **Microbial community composition is in constant evolution over time.** Along the eight-year survey, I observed a global divergence of the community composition over time. Therefore, microbial communities are not stable over time in the five small freshwater ecosystems studied in this thesis.
- **Eukaryotic and prokaryotic communities share similar features.** Despite dissimilarities in cell organisation, metabolisms and ecological behaviour, microbial eukaryotes and prokaryotes share common temporal dynamics.
- **Abiotic parameters have a low explanation power.** The physico-chemical parameters that we measured in Lake Baikal and small freshwater ecosystems did not explain most of the variations observed in microbial community composition. Therefore, they are not the only factors driving community composition.
- **Co-occurrence networks in our studies reflect spatio-temporal variations.** The inference of interactions at the different depth classes in Lake Baikal revealed different types of structuration with, in particular, a more important connection of communities living in surface water. The seasonal networks built in the small freshwater ecosystems revealed common features between the same season despite different global structuration.

The analysis conducted in Lake Baikal provided first insights into the diversity and composition of microbial eukaryotic communities over the entire lake, which had never been investigated before. However, it provided an overview of the microorganisms inhabiting the lake during the summer period only. Comparative analysis over time (e.g. at different season over several years) in the same stations should be conducted later, especially in the global context of climate change and its implication for the communities. The analyses in small freshwater ecosystems confirmed the ecological importance of these overlooked systems. They are good models to explore the temporal dynamics of microbial communities. Among others, they are easy to access and to sample, and they are largely distributed which generates a wide range of environmental characteristics. Microorganism are key players in aquatic ecosystems and this work allowed to better understand their composition and structuration over time. Therefore, it can help to improve the management of these common ecosystems in order to preserve their biodiversity.

To conclude, I would say that at first sight, studying in the same project the biggest lake in the world and small freshwater systems could have appeared contradictory. However, even with ecosystems of completely different scales and characteristics, I highlighted similarities in the composition and structuration of pelagic communities. At a broader scale, similarities can also be observed between marine and freshwater systems (Stockner, 1988). For instance, the microbial loop is a model initially developed in marine systems, but which also fits in

---

freshwater ones (Fenchel, 2008). At the opposite, the Plankton Ecology Group model was established in freshwater systems, and it also find applications in marine ones (Sommer et al., 2012). Therefore, aquatic ecosystems seem to share a similar functioning in terms of microbial community composition and dynamics. Despite strong variations in the spatiotemporal scales and the speed of biological process inherent to the ecosystem size, the use of small freshwater ecosystems as model of aquatic ecosystems to understand microbial communities functioning can help future researches in aquatic ecology.





# Bibliography

- Aarnio S, Teittinen A, Soininen J (2019) High diatom species turnover in a baltic sea rock pool metacommunity. *Marine Biodiversity* 49(6):2887–2899
- Abnizova A, Siemens J, Langer M, Boike J (2012) Small ponds with major impact: The relevance of ponds and lakes in permafrost landscapes to carbon dioxide emissions. *Global Biogeochemical Cycles* 26(2)
- Adámek Z, Maršálek B (2013) Bioturbation of sediments by benthic macroinvertebrates and fish and its implication for pond ecosystems: a review. *Aquaculture International* 21(1):1–17
- Adams HE, Crump BC, Kling GW (2014) Metacommunity dynamics of bacteria in an arctic lake: the impact of species sorting and mass effects on bacterial production and biogeography. *Frontiers in Microbiology* 5:82
- Adams JK, Martins CC, Rose NL, Shchetnikov AA, Mackay AW (2018) Lake sediment records of persistent organic pollutants and polycyclic aromatic hydrocarbons in southern siberia mirror the changing fortunes of the russian economy over the past 70 years. *Environmental pollution* 242:528–538
- Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, et al. (2012) The revised classification of eukaryotes. *Journal of eukaryotic microbiology* 59(5):429–514
- Akaike H (1973) Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika* 60(2):255–265
- Alonso-Sáez L, Díaz-Pérez L, Morán XAG (2015) The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Environmental microbiology* 17(10):3766–3780
- Amalfitano S, Corno G, Eckert E, Fazi S, Ninio S, Callieri C, Grossart HP, Eckert W (2017) Tracing particulate matter and associated microorganisms in freshwaters. *Hydrobiologia* 800(1):145–154

- Andersen MR, Kragh T, Sand-Jensen K (2017) Extreme diel dissolved oxygen and carbon cycles in shallow vegetated lakes. *Proceedings of the Royal Society B: Biological Sciences* 284(1862):20171,427
- Anderson MJ (2014) *Permutational multivariate analysis of variance (permanova)*. Wiley statref: statistics reference online pp 1–15
- Angélibert S, Marty P, Céréghino R, Giani N (2004) Seasonal variations in the physical and chemical characteristics of ponds: implications for biodiversity conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems* 14(5):439–456
- Annenkova NV, Lavrov DV, Belikov SI (2011) Dinoflagellates associated with freshwater sponges from the ancient lake baikal. *Protist* 162(2):222–236
- Annenkova NV, Giner CR, Logares R (2020a) Tracing the origin of planktonic protists in an ancient lake. *Microorganisms* 8(4):543
- Annenkova NV, Hansen G, Rengefors K (2020b) Closely related dinoflagellate species in vastly different habitats—an example of a marine–freshwater transition. *European Journal of Phycology* pp 1–12
- Anneville O, Ginot V, Druart JC, Angeli N (2002) Long-term study (1974–1998) of seasonal changes in the phytoplankton in lake geneva: a multi-table approach. *Journal of plankton research* 24(10):993–1008
- Azam F, Malfatti F (2007) Microbial structuring of marine ecosystems. *Nature Reviews Microbiology* 5(10):782–791
- Azam F, Fenchel T, Field JG, Gray J, Meyer-Reil L, Thingstad F (1983) The ecological role of water-column microbes in the sea. *Marine ecology progress series* pp 257–263
- Baas-Becking LGM (1934) *Geobiologie; of inleiding tot de milieukunde*. WP Van Stockum & Zoon NV
- Baker G, Smith JJ, Cowan DA (2003) Review and re-analysis of domain-specific 16s primers. *Journal of microbiological methods* 55(3):541–555
- Bar-On YM, Phillips R, Milo R (2018) The biomass distribution on earth. *Proceedings of the National Academy of Sciences* 115(25):6506–6511
- Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, Stamatakis A (2019) Epa-ng: massively parallel evolutionary placement of genetic sequences. *Systematic biology* 68(2):365–369

- Barberán A, Casamayor EO (2010) Global phylogenetic community structure and  $\beta$ -diversity patterns in surface bacterioplankton metacommunities. *Aquatic Microbial Ecology* 59(1):1–10
- Barberán A, Bates ST, Casamayor EO, Fierer N (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME journal* 6(2):343–351
- Barrat A, Barthelemy M, Vespignani A (2008) *Dynamical processes on complex networks*. Cambridge university press
- Bartout P, Touchart L, Terasmaa J, Choffel Q, Marzecova A, Koff T, Kapanen G, Qsair Z, Maleval V, Millot C, et al. (2015) A new approach to inventorying bodies of water, from local to global scale. *DIE ERDE–Journal of the Geographical Society of Berlin* 146(4):245–258
- Bashenkhaeva MV, Zakharova YR, Petrova DP, Khanaev IV, Galachyants YP, Likhoshway YV (2015) Sub-ice microalgal and bacterial communities in freshwater lake baikal, russia. *Microbial ecology* 70(3):751–765
- Bass D, Czech L, Williams BA, Berney C, Dunthorn M, Mahé F, Torruella G, Stentiford GD, Williams TA (2018) Clarifying the relationships between microsporidia and cryptomycota. *Journal of Eukaryotic Microbiology* 65(6):773–782
- Belikov S, Belkova N, Butina T, Chernogor L, Martynova-Van Kley A, Nalian A, Rorex C, Khanaev I, Maikova O, Feranchuk S (2019) Diversity and shifts of the bacterial community associated with baikal sponge mass mortalities. *PloS one* 14(3):e0213,926
- Belkova N, Parfenova V, Kostopnova T, Denisova L, Zauichikov E (2003) Microbial biodiversity in the lake baikal water. *Mikrobiologiya* 72(2):239
- Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK (2005) The contribution of species richness and composition to bacterial services. *Nature* 436(7054):1157–1160
- Bell W, Mitchell R (1972) Chemotactic and growth responses of marine bacteria to algal extracellular products. *The Biological Bulletin* 143(2):265–277
- Berdjeb L, Parada A, Needham DM, Fuhrman JA (2018) Short-term dynamics and interactions of marine protist communities during the spring–summer transition. *The ISME journal* 12(8):1907–1917
- Bergkemper V, Weisse T (2018) Do current european lake monitoring programmes reliably estimate phytoplankton community changes? *Hydrobiologia* 824(1):143–162

- Berry D, Widder S (2014) Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in microbiology* 5:219
- Biggs J, Von Fumetti S, Kelly-Quinn M (2017) The importance of small waterbodies for biodiversity and ecosystem services: implications for policy makers. *Hydrobiologia* 793(1):3–39
- Biller SJ, Berube PM, Dooley K, Williams M, Satinsky BM, Hackl T, Hogle SL, Coe A, Bergauer K, Bouman HA, et al. (2018) Marine microbial metagenomes sampled across space and time. *Scientific data* 5:180,176
- Bjorbækmo MFM, Evenstad A, Røsæg LL, Krabberød AK, Logares R (2020) The planktonic protist interactome: where do we stand after a century of research? *The ISME journal* 14(2):544–559
- Blottière L, Jaffar-Bandjee M, Jacquet S, Millot A, Hulot FD (2017) Effects of mixing on the pelagic food web in shallow lakes. *Freshwater Biology* 62(1):161–177
- Bock C, Salcher M, Jensen M, Pandey RV, Boenigk J (2018) Synchrony of eukaryotic and prokaryotic planktonic communities in three seasonally sampled austrian lakes. *Frontiers in microbiology* 9:1290
- Bock C, Jensen M, Forster D, Marks S, Nuy J, Psenner R, Beisser D, Boenigk J (2020) Factors shaping community patterns of protists and bacteria on a european scale. *Environmental Microbiology*
- Boenigk J, Arndt H (2002) Bacterivory by heterotrophic flagellates: community structure and feeding strategies. *Antonie Van Leeuwenhoek* 81(1-4):465–480
- Bolle HJ (2013) *The natural environment and the biogeochemical cycles*, vol 1. Springer
- Bower SM, Carnegie RB, Goh B, Jones SR, Lowe GJ, Mak MW (2004) Preferential pcr amplification of parasitic protistan small subunit rDNA from metazoan tissues. *Journal of Eukaryotic Microbiology* 51(3):325–332
- Bråte J, Logares R, Berney C, Ree DK, Klaveness D, Jakobsen KS, Shalchian-Tabrizi K (2010) Freshwater perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *The ISME journal* 4(9):1144–1153
- Bravo I, Figueroa RI (2014) Towards an ecological understanding of dinoflagellate cyst functions. *Microorganisms* 2(1):11–32
- Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern wisconsin. *Ecological monographs* 27(4):326–349

- Briée C, Moreira D, López-García P (2007) Archaeal and bacterial community composition of sediment and plankton from a suboxic freshwater pond. *Research in Microbiology* 158(3):213–227
- Brönmark C, Hansson LA (2017) *The biology of lakes and ponds*. Oxford University Press
- Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH, Banfield JF (2015) Unusual biology across a group comprising more than 15% of domain bacteria. *Nature* 523(7559):208–211
- Buckley HL, Day NJ, Case BS, Lear G, Ellison AM (2019) Multivariate methods for testing hypotheses of temporal community dynamics. *bioRxiv* p 362822
- Bunse C, Pinhassi J (2017) Marine bacterioplankton seasonal succession dynamics. *Trends in microbiology* 25(6):494–505
- Bunse C, Israelsson S, Baltar F, Bertos-Fortis M, Fridolfsson E, Legrand C, Lindehoff E, Lindh MV, Martínez-García S, Pinhassi J (2019) High frequency multi-year variability in baltic sea microbial plankton stocks and activities. *Frontiers in microbiology* 9:3296
- Burki F, Roger AJ, Brown MW, Simpson AG (2020) The new tree of eukaryotes. *Trends in ecology & evolution* 35(1):43–55
- Butina TV, Bukin YS, Krasnopeev AS, Belykh OI, Tupikin AE, Kabilov MR, Sakirko aV, Belikov SI (2019) Estimate of the diversity of viral and bacterial assemblage in the coastal water of lake baikal. *FEMS microbiology letters* 366(9):fnz094
- Buttigieg PL, Fadeev E, Bienhold C, Hehemann L, Offre P, Boetius A (2018) Marine microbes in 4d—using time series observation to assess the dynamics of the ocean microbiome and its links to ocean health. *Current opinion in microbiology* 43:169–185
- Cabello-Yeves PJ, Rodriguez-Valera F (2019) Marine-freshwater prokaryotic transitions require extensive changes in the predicted proteome. *Microbiome* 7(1):117
- Cabello-Yeves PJ, Zemskaya TI, Rosselli R, Coutinho FH, Zakharenko AS, Blinov VV, Rodriguez-Valera F (2018) Genomes of novel microbial lineages assembled from the sub-ice waters of lake baikal. *Applied and environmental microbiology* 84(1)
- Cabello-Yeves PJ, Zemskaya TI, Zakharenko AS, Sakirko MV, Ivanov VG, Ghai R, Rodriguez-Valera F (2020) Microbiome of the deep lake baikal, a unique oxic bathypelagic habitat. *Limnology and Oceanography* 65(7):1471–1488

- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) Dada2: high-resolution sample inference from illumina amplicon data. *Nature methods* 13(7):581–583
- Cantonati M, Füreder L, Gerecke R, Jüttner I, Cox EJ (2012) Crenic habitats, hotspots for freshwater biodiversity conservation: toward an understanding of their ecology. *Freshwater Science* 31(2):463–480
- Cao X, Zhao D, Xu H, Huang R, Zeng J, Yu Z (2018) Heterogeneity of interactions of microbial communities in regions of taihu lake with different nutrient loadings: A network analysis. *Scientific reports* 8(1):1–11
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973
- Caron DA, Worden AZ, Countway PD, Demir E, Heidelberg KB (2009) Protists are microbes too: a perspective. *The ISME journal* 3(1):4–12
- Carpenter SR (1981) Submersed vegetation: an internal factor in lake ecosystem succession. *The American Naturalist* 118(3):372–383
- Castelle CJ, Banfield JF (2018) Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* 172(6):1181–1197
- Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF (2018) Biosynthetic capacity, metabolic variety and unusual biology in the cpr and dpann radiations. *Nature Reviews Microbiology* 16(10):629–645
- Cavaliere M, Feng S, Soyer OS, Jiménez JI (2017) Cooperation in microbial communities and their biotechnological applications. *Environmental microbiology* 19(8):2949–2963
- Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, Behrenfeld MJ, Boetius A, Boyd PW, Classen AT, et al. (2019) Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology* 17(9):569–586
- Céréghino R, Biggs J, Oertli B, Declerck S (2007) The ecology of european ponds: defining the characteristics of a neglected freshwater habitat. In: *Pond conservation in Europe*, Springer, pp 1–6
- Céréghino R, Boix D, Cauchie HM, Martens K, Oertli B (2014) The ecological role of ponds in a changing world. *Hydrobiologia* 723(1):1–6

- Chase JM (2007) Drought mediates the importance of stochastic community assembly. *Proceedings of the National Academy of Sciences* 104(44):17,430–17,434
- Chen J, Chen Z (2008) Extended bayesian information criteria for model selection with large model spaces. *Biometrika* 95(3):759–771
- Chénard C, Wijaya W, Vaultot D, dos Santos AL, Martin P, Kaur A, Lauro FM (2019) Temporal and spatial dynamics of bacteria, archaea and protists in equatorial coastal waters. *Scientific reports* 9(1):1–13
- Cheng FY, Basu NB (2017) Biogeochemical hotspots: Role of small water bodies in landscape nutrient processing. *Water Resources Research* 53(6):5038–5056
- Chiquet J, Robin S, Mariadassou M (2019) Variational inference for sparse network reconstruction from count data. In: *International Conference on Machine Learning*, pp 1162–1171
- Cho BC, Azam F (1988) Major role of bacteria in biogeochemical fluxes in the ocean's interior. *Nature* 332(6163):441–443
- Chow CET, Sachdeva R, Cram JA, Steele JA, Needham DM, Patel A, Parada AE, Fuhrman JA (2013) Temporal variability and coherence of euphotic zone bacterial communities over a decade in the southern california bight. *The ISME journal* 7(12):2259–2273
- Clarke K, Green R (1988) Statistical design and analysis for a 'biological effects' study. *Marine Ecology Progress Series* pp 213–226
- Cleary AC, Durbin EG (2016) Unexpected prevalence of parasite 18s rDNA sequences in winter among antarctic marine protists. *Journal of Plankton Research* 38(3):401–417
- Collins SL, Micheli F, Hartt L (2000) A method to determine rates and patterns of variability in ecological communities. *Oikos* 91(2):285–293
- Comte J, Berga M, Severin I, Logue JB, Lindström ES (2017) Contribution of different bacterial dispersal sources to lakes: population and community effects in different seasons. *Environmental microbiology* 19(6):2391–2404
- Cordero OX, Datta MS (2016) Microbial interactions and community assembly at microscales. *Current opinion in microbiology* 31:227–234
- Corliss JO (1975) Three centuries of protozoology: a brief tribute to its founding father, a. van Leeuwenhoek of delft. *The Journal of Protozoology* 22(1):3–7
- Cotner JB, Biddanda BA (2002) Small players, large role: microbial influence on biogeochemical processes in pelagic aquatic ecosystems. *Ecosystems* 5(2):105–121



- Cotterill FP, Augustin H, Medicus R, Foissner W (2013) Conservation of protists: the krauthügel pond in austria. *Diversity* 5(2):374–392
- Cougoul A, Bailly X, Vourec'h G, Gasqui P (2019) Rarity of microbial species: In search of reliable associations. *PloS one* 14(3):e0200458
- Cram JA, Chow CET, Sachdeva R, Needham DM, Parada AE, Steele JA, Fuhrman JA (2015) Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *The ISME journal* 9(3):563–580
- Cronberg G, Lindmark G, Björk S (1988) Mass development of the flagellate gonyostomum semen (raphidophyta) in swedish forest lakes-an effect of acidification? *Hydrobiologia* 161(1):217–236
- Crump BC, Hobbie JE (2005) Synchrony and seasonality in bacterioplankton communities of two temperate rivers. *Limnology and Oceanography* 50(6):1718–1729
- Crump BC, Adams HE, Hobbie JE, Kling GW (2007) Biogeography of bacterioplankton in lakes and streams of an arctic tundra catchment. *Ecology* 88(6):1365–1378
- Crump BC, Amaral-Zettler LA, Kling GW (2012) Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *The ISME journal* 6(9):1629–1639
- Csardi G, Nepusz T, et al. (2006) The igraph software package for complex network research. *InterJournal, complex systems* 1695(5):1–9
- Dakos V, Benincà E, van Nes EH, Philippart CJ, Scheffer M, Huisman J (2009) Interannual variability in species composition explained as seasonally entrained chaos. *Proceedings of the Royal Society B: Biological Sciences* 276(1669):2871–2880
- Danczak R, Johnston M, Kenah C, Slattery M, Wrighton KC, Wilkins M (2017) Members of the candidate phyla radiation are functionally differentiated by carbon-and nitrogen-cycling capabilities. *Microbiome* 5(1):1–14
- De Batist M, Klerkx J, Van Rensbergen P, Vanneste M, Poort J, Golmshtok AY, Kremlev AA, Khlystov OM, Krinitsky P (2002) Active hydrate destabilization in lake baikal, siberia? *Terra Nova* 14(6):436–442
- De Corte D, Srivastava A, Koski M, Garcia JAL, Takaki Y, Yokokawa T, Nunoura T, Elisabeth NH, Sintès E, Herndl GJ (2018) Metagenomic insights into zooplankton-associated bacterial communities. *Environmental microbiology* 20(2):492–505

- De Meester L, Declerck S, Stoks R, Louette G, Van De Meutter F, De Bie T, Michels E, Brendonck L (2005) Ponds and pools as model systems in conservation biology, ecology and evolutionary biology. *Aquatic conservation: Marine and freshwater ecosystems* 15(6):715–725
- De Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* 348(6237)
- Debroas D, Hugoni M, Domaizon I (2015) Evidence for an active rare biosphere within freshwater protists community. *Molecular Ecology* 24(6):1236–1247
- Debroas D, Domaizon I, Humbert JF, Jardillier L, Lepère C, Oudart A, Taïb N (2017) Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. *FEMS Microbiology Ecology* 93(4):fix023
- Del Giorgio PA, Cole JJ (1998) Bacterial growth efficiency in natural aquatic systems. *Annual Review of Ecology and Systematics* 29(1):503–541
- Delmas E, Besson M, Brice MH, Burkle LA, Dalla Riva GV, Fortin MJ, Gravel D, Guimarães Jr PR, Hembry DH, Newman EA, et al. (2019) Analysing ecological networks of species interactions. *Biological Reviews* 94(1):16–36
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU, Martinez A, Sullivan MB, Edwards R, Brito BR, et al. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311(5760):496–503
- Derocles SA, Bohan DA, Dumbrell AJ, Kitson JJ, Massol F, Pauvert C, Plantegenest M, Vacher C, Evans DM (2018) Biomonitoring for the 21st century: integrating next-generation sequencing into ecological network analysis. In: *Advances in ecological research*, vol 58, Elsevier, pp 1–62
- Deutschmann IM, Lima-Mendez G, Krabberød AK, Raes J, Vallina SM, Faust K, Logares R (2020) Disentangling environmental effects in microbial association networks. *BMC Microbiome*
- Diamond JM (1969) Avifaunal equilibria and species turnover rates on the channel islands of california. *Proceedings of the National Academy of Sciences* 64(1):57–63
- Dolédec S, Chessel D (1994) Co-inertia analysis: an alternative method for studying species–environment relationships. *Freshwater biology* 31(3):277–294
- Downing JA (2010) Emerging global role of small lakes and ponds: little things mean a lot. *Limnetica* 29(1):0009–24

- Downing JA, Prairie Y, Cole J, Duarte C, Tranvik L, Striegl RG, McDowell W, Kortelainen P, Caraco N, Melack J, et al. (2006) The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology and Oceanography* 51(5):2388–2397
- Dray S, Dufour AB, et al. (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software* 22(4):1–20
- Dudgeon D, Arthington AH, Gessner MO, Kawabata ZI, Knowler DJ, Lévêque C, Naiman RJ, Prieur-Richard AH, Soto D, Stiassny ML, et al. (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews* 81(2):163–182
- Eiler A, Bertilsson S (2004) Composition of freshwater bacterial communities associated with cyanobacterial blooms in four swedish lakes. *Environmental Microbiology* 6(12):1228–1243
- Eiler A, Heinrich F, Bertilsson S (2012) Coherent dynamics and association networks among lake bacterioplankton taxa. *The ISME journal* 6(2):330–342
- Escalas A, Catherine A, Maloufi S, Cellamare M, Hamlaoui S, Yéprémian C, Louvard C, Troussellier M, Bernard C (2019) Drivers and ecological consequences of dominance in periurban phytoplankton communities using networks approaches. *Water research* 163:114,893
- Evans S, Martiny JB, Allison SD (2017) Effects of dispersal and selection on stochastic assembly in microbial communities. *The ISME journal* 11(1):176–185
- Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive earth's biogeochemical cycles. *science* 320(5879):1034–1039
- Faust K, Raes J (2012) Microbial interactions: from networks to models. *Nature Reviews Microbiology* 10(8):538–550
- Faust K, Lahti L, Gonze D, De Vos WM, Raes J (2015) Metagenomics meets time series analysis: unraveling microbial community dynamics. *Current opinion in microbiology* 25:56–66
- Fenchel T (2008) The microbial loop—25 years later. *Journal of Experimental Marine Biology and Ecology* 366(1-2):99–103
- Filstrup CT, Hillebrand H, Heathcote AJ, Harpole WS, Downing JA (2014) Cyanobacteria dominance influences resource use efficiency and community turnover in phytoplankton and zooplankton communities. *Ecology Letters* 17(4):464–474
- Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. *Science* 296(5570):1061–1063

- Fisher M, Klug JL, Lauster G, Newton M, Triplett E (2000) Effects of resources and trophic interactions on freshwater bacterioplankton diversity. *Microbial Ecology* 40(2):125–138
- Foley B, Jones ID, Maberly SC, Rippey B (2012) Long-term changes in oxygen depletion in a small temperate lake: effects of climate change and eutrophication. *Freshwater Biology* 57(2):278–289
- Forel FA (1904) *Le leman*. Monographie limnologique
- Fowler D, Coyle M, Skiba U, Sutton MA, Cape JN, Reis S, Sheppard LJ, Jenkins A, Grizzetti B, Galloway JN, et al. (2013) The global nitrogen cycle in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1621):20130,164
- Friedman J, Alm EJ (2012) Inferring correlation networks from genomic survey data. *PLoS Comput Biol* 8(9):e1002,687
- Fu L, Niu B, Zhu Z, Wu S, Li W (2012) Cd-hit: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150–3152
- Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proceedings of the National Academy of Sciences* 103(35):13,104–13,109
- Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL, Brown JH (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proceedings of the National Academy of Sciences* 105(22):7774–7778
- Fuhrman JA, Cram JA, Needham DM (2015) Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology* 13(3):133–146
- Garneau MÈ, Posch T, Hitz G, Pomerleau F, Pradalier C, Siegwart R, Pernthaler J (2013) Short-term displacement of *planktothrix rubescens* (cyanobacteria) in a pre-alpine lake observed using an autonomous sampling platform. *Limnology and Oceanography* 58(5):1892–1906
- Genitsaris S, Monchy S, Breton E, Lecuyer E, Christaki U (2016) Small-scale variability of protistan planktonic communities relative to environmental pressures and biotic interactions at two adjacent coastal stations. *Marine Ecology Progress Series* 548:61–75
- Gilbert JA, Field D, Swift P, Newbold L, Oliver A, Smyth T, Somerfield PJ, Huse S, Joint I (2009) The seasonal structure of microbial communities in the western english channel. *Environmental microbiology* 11(12):3132–3139

- Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Temperton B, Huse S, McHardy AC, Knight R, Joint I, et al. (2012) Defining seasonal marine microbial community dynamics. *The ISME journal* 6(2):298–308
- Giner CR, Balagué V, Krabberød AK, Ferrera I, Reñé A, Garcés E, Gasol JM, Logares R, Ramon M (2018) Quantifying long-term recurrence in planktonic microbial eukaryotes. *Molecular Ecology*
- Glöckner FO, Zaichikov E, Belkova N, Denissova L, Pernthaler J, Pernthaler A, Amann R (2000) Comparative 16s rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria. *Applied and environmental microbiology* 66(11):5053–5065
- Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ (2017) Microbiome datasets are compositional: and this is not optional. *Frontiers in microbiology* 8:2224
- Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, Segrè D, Mehta P, Sanchez A (2018) Emergent simplicity in microbial community assembly. *Science* 361(6401):469–474
- Gomi T, Sidle RC, Richardson JS (2002) Understanding processes and downstream linkages of headwater systems: headwaters differ from downstream reaches by their close coupling to hillslope processes, more temporal and spatial variation, and their need for different means of protection from land use. *BioScience* 52(10):905–916
- Gonzalez A, King A, Robeson II MS, Song S, Shade A, Metcalf JL, Knight R (2012) Characterizing microbial communities through space and time. *Current opinion in biotechnology* 23(3):431–436
- González-Olalla JM, Medina-Sánchez JM, Lozano IL, Villar-Argaiz M, Carrillo P (2018) Climate-driven shifts in algal-bacterial interaction of high-mountain lakes in two years spanning a decade. *Scientific reports* 8(1):1–12
- Grami B, Rasconi S, Niquil N, Jobard M, Saint-Béat B, Sime-Ngando T (2011) Functional effects of parasites on food web properties during the spring diatom bloom in lake pavin: a linear inverse modeling analysis. *PLoS One* 6(8):e23,273
- Graneli E, Weberg M, Salomon PS (2008) Harmful algal blooms of allelopathic microalgal species: the role of eutrophication. *Harmful algae* 8(1):94–102
- Granin N, Aslamov I, Kozlov V, Makarov M, Kirillin G, McGinnis D, Kucher K, Blinov V, Ivanov V, Mizandrontsev I, et al. (2019) Methane hydrate emergence from lake baikal: direct

- observations, modelling, and hydrate footprints in seasonal ice cover. *Scientific reports* 9(1):1–10
- Grossart HP, Dziallas C, Leunert F, Tang KW (2010) Bacteria dispersal by hitchhiking on zooplankton. *Proceedings of the National Academy of Sciences* 107(26):11,959–11,964
- Grossmann L, Jensen M, Heider D, Jost S, Glücksman E, Hartikainen H, Mahamdallie SS, Gardner M, Hoffmann D, Bass D, et al. (2016) Protistan community analysis: key findings of a large-scale molecular sampling. *The ISME journal* 10(9):2269–2279
- Grujic V, Nuy JK, Salcher MM, Shabarova T, Kasalicky V, Boenigk J, Jensen M, Simek K (2018) Cryptophyta as major bacterivores in freshwater summer plankton. *The ISME journal* 12(7):1668–1681
- Guillou L, Viprey M, Chambouvet A, Welsh R, Kirkham A, Massana R, Scanlan DJ, Worden AZ (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to syndiniales (alveolata). *Environmental microbiology* 10(12):3349–3365
- Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, De Vargas C, Decelle J, et al. (2012) The protist ribosomal reference database (pr2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic acids research* 41(D1):D597–D604
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* 52(5):696–704
- Hagman CHC, Ballot A, Hjermmann DØ, Skjelbred B, Brettum P, Ptacnik R (2015) The occurrence and spread of *Gonyostomum semen* (ehr.) diesing (raphidophyceae) in norwegian lakes. *Hydrobiologia* 744(1):1–14
- Hairton Jr NG, Kearns CM (2002) Temporal dispersal: ecological and evolutionary aspects of zooplankton egg banks and the role of sediment mixing. *Integrative and Comparative Biology* 42(3):481–491
- Hallett LM, Jones SK, MacDonald AAM, Jones MB, Flynn DF, Ripplinger J, Slaughter P, Gries C, Collins SL (2016) codyn: An R package of community dynamics metrics. *Methods in Ecology and Evolution* 7(10):1146–1151
- Harding K, Turk-Kubo KA, Sipler RE, Mills MM, Bronk DA, Zehr JP (2018) Symbiotic unicellular cyanobacteria fix nitrogen in the arctic ocean. *Proceedings of the National Academy of Sciences* 115(52):13,371–13,375

- Hernández-Ruiz M, Barber-Lluch E, Prieto A, Álvarez-Salgado XA, Logares R, Teira E (2018) Seasonal succession of small planktonic eukaryotes inhabiting surface waters of a coastal upwelling system. *Environmental microbiology* 20(8):2955–2973
- Herren CM, McMahon KD (2018) Keystone taxa predict compositional change in microbial communities. *Environmental microbiology* 20(6):2207–2217
- Hill MJ, Hassall C, Oertli B, Fahrig L, Robson BJ, Biggs J, Samways MJ, Usio N, Takamura N, Krishnaswamy J, et al. (2018) New policy directions for global pond conservation. *Conservation Letters* 11(5):e12,447
- Hoek TA, Axelrod K, Biancalani T, Yurtsev EA, Liu J, Gore J (2016) Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism. *PLoS biology* 14(8):e1002,540
- Holgerson MA, Raymond PA (2016) Large contribution to inland water CO<sub>2</sub> and CH<sub>4</sub> emissions from very small ponds. *Nature Geoscience* 9(3):222–226
- Hou F, Zhang H, Xie W, Zhou X, Zhu X, Zhang D (2020) Co-occurrence patterns and assembly processes of microeukaryotic communities in an early-spring diatom bloom. *Science of The Total Environment* 711:134,624
- Hudson PJ, Dobson AP, Lafferty KD (2006) Is a healthy ecosystem one that is rich in parasites? *Trends in ecology & evolution* 21(7):381–385
- Hug LA (2018) Sizing up the uncultured microbial majority. *MSystems* 3(5)
- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, HERNSDORF AW, Amano Y, Ise K, et al. (2016) A new view of the tree of life. *Nature microbiology* 1(5):1–6
- Hugoni M, Taib N, Debroas D, Domaizon I, Dufournel IJ, Bronner G, Salter I, Agogué H, Mary I, Galand PE (2013) Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proceedings of the National Academy of Sciences* 110(15):6004–6009
- Hullar MA, Kaplan LA, Stahl DA (2006) Recurring seasonal dynamics of microbial communities in stream habitats. *Applied and Environmental Microbiology* 72(1):713–722
- Humbert JF, Dorigo U, Cecchi P, Le Berre B, Debroas D, Bouvy M (2009) Comparison of the structure and composition of bacterial communities from temperate and tropical freshwater ecosystems. *Environmental microbiology* 11(9):2339–2350

- Hurst CJ (2019) Understanding aquatic microbial communities. In: *The Structure and Function of Aquatic Microbial Communities*, Springer, pp 1–12
- Husson F, Josse J, Le S, Mazet J, Husson MF (2016) Package ‘factominer’. An R package 96:698
- Imhoff JF, Rahn T, Künzel S, Neulinger SC (2018) Photosynthesis is widely distributed among proteobacteria as demonstrated by the phylogeny of pufm reaction center proteins. *Frontiers in microbiology* 8:2679
- Incagnone G, Marrone F, Barone R, Robba L, Naselli-Flores L (2015) How do freshwater organisms cross the “dry ocean”? a review on passive dispersal and colonization processes with a special focus on temporary ponds. *Hydrobiologia* 750(1):103–123
- Indermuehle N, Oertli B, Biggs J, Céréghino R, Grillas P, Hull A, Nicolet P, Scher O (2008) Pond conservation in europe: The european pond conservation network (epcn). *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen* 30(3):446–448
- Jardillier L, Bettarel Y, Richardot M, Bardot C, Amblard C, Sime-Ngando T, Debroas D (2005) Effects of viruses and predators on prokaryotic community composition. *Microbial ecology* 50(4):557–569
- Jardillier L, Zubkov MV, Pearman J, Scanlan DJ (2010) Significant  $CO_2$  fixation by small prymnesiophytes in the subtropical and tropical northeast atlantic ocean. *The ISME journal* 4(9):1180–1192
- Jephcott TG, Alves-De-Souza C, Gleason FH, Van Ogtrop FF, Sime-Ngando T, Karpov SA, Guillou L (2016) Ecological impacts of parasitic chytrids, syndiniales and perkinsids on populations of marine photosynthetic dinoflagellates. *fungus ecology* 19:47–58
- Jetten MS (2008) The microbial nitrogen cycle. *Environmental microbiology* 10(11):2903–2909
- Jia X, Dini-Andreote F, Salles JF (2018) Community assembly processes of the microbial rare biosphere. *Trends in microbiology* 26(9):738–747
- Jiang H, Dong H, Zhang G, Yu B, Chapman LR, Fields MW (2006) Microbial diversity in water and sediment of lake chaka, an athalassohaline lake in northwestern china. *Applied and environmental microbiology* 72(6):3832–3845
- Jiao C, Zhao D, Zeng J, Guo L, Yu Z (2020) Disentangling the seasonal co-occurrence patterns and ecological stochasticity of planktonic and benthic bacterial communities within multiple lakes. *Science of The Total Environment* p 140010



- Jones AC, Liao TV, Najjar FZ, Roe BA, Hambright KD, Caron DA (2013) Seasonality and disturbance: annual pattern and response of the bacterial and microbial eukaryotic assemblages in a freshwater ecosystem. *Environmental Microbiology* 15(9):2557–2572
- Jones AC, Hambright KD, Caron DA (2018) Ecological patterns among bacteria and microbial eukaryotes derived from network analyses in a low-salinity lake. *Microbial ecology* 75(4):917–929
- Jones SE, Lennon JT (2010) Dormancy contributes to the maintenance of microbial diversity. *Proceedings of the National Academy of Sciences* 107(13):5881–5886
- Jones SE, McMahon KD (2009) Species-sorting may explain an apparent minimal effect of immigration on freshwater bacterial community dynamics. *Environmental microbiology* 11(4):905–913
- Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, Küsel K, Rillig MC, Rivett DW, Salles JF, et al. (2017) Where less may be more: how the rare biosphere pulls ecosystems strings. *The ISME journal* 11(4):853–862
- Kara EL, Hanson PC, Hu YH, Winslow L, McMahon KD (2013) A decade of seasonal dynamics and co-occurrences within freshwater bacterioplankton communities from eutrophic lake mendota, wi, usa. *The ISME journal* 7(3):680–684
- Karimi B, Maron PA, Boure NCP, Bernard N, Gilbert D, Ranjard L (2017) Microbial diversity and ecological networks as indicators of environmental quality. *Environmental Chemistry Letters* 15(2):265–281
- Karl DM, Church MJ (2014) Microbial oceanography and the hawaii ocean time-series programme. *Nature Reviews Microbiology* 12(10):699–713
- Karpov S, Mamkaeva MA, Aleoshin V, Nassonova E, Lilje O, Gleason FH (2014) Morphology, phylogeny, and ecology of the aphelids (aphelidea, opisthokonta) and proposal for the new superphylum opisthosporidia. *Frontiers in Microbiology* 5:112
- Kashiwaya K, Ochiai S, Sakai H, Kawai T (2001) Orbit-related long-term climate cycles revealed in a 12-myr continental record from lake baikal. *Nature* 410(6824):71–74
- Katoh K, Standley DM (2013) Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30(4):772–780
- Kelly LW, Nelson CE, Haas AF, Naliboff DS, Calhoun S, Carlson CA, Edwards RA, Fox MD, Hatay M, Johnson MD, et al. (2019) Diel population and functional synchrony of microbial communities on coral reefs. *Nature communications* 10(1):1–9

- Kent AD, Yannarell AC, Rusak JA, Triplett EW, McMahon KD (2007) Synchrony in aquatic microbial community dynamics. *The ISME journal* 1(1):38–47
- Keshri J, Ram AP, Nana P, Sime-Ngando T (2018) Taxonomical resolution and distribution of bacterioplankton along the vertical gradient reveals pronounced spatiotemporal patterns in contrasted temperate freshwater lakes. *Microbial ecology* 76(2):372–386
- Khodzher TV, Domysheva VM, Sorokovikova LM, Sakirko MV, Tomberg IV (2017) Current chemical composition of lake baikal water. *Inland Waters* 7(3):250–258
- Kim DY, Countway PD, Jones AC, Schnetzer A, Yamashita W, Tung C, Caron DA (2014) Monthly to interannual variability of microbial eukaryote assemblages at four depths in the eastern north pacific. *The ISME journal* 8(3):515–530
- Kinnunen M, Dechesne A, Albrechtsen HJ, Smets BF (2018) Stochastic processes govern invasion success in microbial communities when the invader is phylogenetically close to resident bacteria. *The ISME journal* 12(11):2748–2756
- Kirchman DL (2018) *Processes in microbial ecology*. Oxford University Press
- Kirkham AR, Jardillier LE, Tiganeşcu A, Pearman J, Zubkov MV, Scanlan DJ (2011) Basin-scale distribution patterns of photosynthetic picoeukaryotes along an atlantic meridional transect. *Environmental Microbiology* 13(4):975–990
- Kritzberg ES, Cole JJ, Pace MM, Granéli W (2006) Bacterial growth on allochthonous carbon in humic and nutrient-enriched lakes: results from whole-lake 13 c addition experiments. *Ecosystems* 9(3):489–499
- Kuenen JG (2008) Anammox bacteria: from discovery to application. *Nature Reviews Microbiology* 6(4):320–326
- Kurilkina MI, Zakharova YR, Galachyants YP, Petrova DP, Bukin YS, Domysheva VM, Blinov VV, Likhoshway YV (2016) Bacterial community composition in the water column of the deepest freshwater lake baikal as determined by next-generation sequencing. *FEMS microbiology ecology* 92(7)
- Kurt H (2019) Depth distribution of microbial diversity in lakes. In: *Freshwater Microbiology*, Elsevier, pp 225–262
- Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA (2015) Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol* 11(5):e1004226

- Ladau J, Eloë-Fadrosh EA (2019) Spatial, temporal, and phylogenetic scales of microbial ecology. *Trends in microbiology* 27(8):662–669
- Lafferty KD, Dobson AP, Kuris AM (2006) Parasites dominate food web links. *Proceedings of the National Academy of Sciences* 103(30):11,211–11,216
- Lam P, Kuypers MM (2011) Microbial nitrogen cycling processes in oxygen minimum zones. *Annual review of marine science* 3:317–345
- Lambert S, Tragin M, Lozano JC, Ghiglione JF, Vaultot D, Bouget FY, Galand PE (2019) Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations. *The ISME journal* 13(2):388–401
- Landi P, Minoarivelo HO, Brännström Å, Hui C, Dieckmann U (2018) Complexity and stability of ecological networks: a review of the theory. *Population Ecology* 60(4):319–345
- Langenheder S, Lindström ES (2019) Factors influencing aquatic and terrestrial bacterial community assembly. *Environmental microbiology reports* 11(3):306–315
- Lannes R, Olsson-Francis K, Lopez P, Baptiste E (2019) Carbon fixation by marine ultrasmall prokaryotes. *Genome biology and evolution* 11(4):1166–1177
- Lara E, Moreira D, Vereshchaka A, López-García P (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplomonads. *Environmental microbiology* 11(1):47–55
- Lara E, Mitchell EA, Moreira D, García PL (2011) Highly diverse and seasonally dynamic protist community in a pristine peat bog. *Protist* 162(1):14–32
- Larned ST, Datry T, Arscott DB, Tockner K (2010) Emerging concepts in temporary-river ecology. *Freshwater Biology* 55(4):717–738
- Lê S, Josse J, Husson F, et al. (2008) Factominer: an R package for multivariate analysis. *Journal of statistical software* 25(1):1–18
- Lear G, Bellamy J, Case BS, Lee JE, Buckley HL (2014) Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds. *The ISME journal* 8(8):1715–1726
- Lechevalier H (1976) Louis Joblot and his microscopes. *Bacteriological reviews* 40(1):241
- Lee JE, Buckley HL, Etienne RS, Lear G (2013) Both species sorting and neutral processes drive assembly of bacterial communities in aquatic microcosms. *FEMS microbiology ecology* 86(2):288–302

- Lefebvre K (2015) Diagnostic et quantification des flux nappe-rivière: modélisations hydrodynamique et géochimique du bassin versant de l'yvette amont (france). PhD thesis, Université Paris-Saclay (ComUE)
- Leger JB (2016) Blockmodels: A r-package for estimating in latent block model and stochastic block model, with various probability functions, with or without covariates. arXiv preprint arXiv:160207587
- Lennon JT, Jones SE (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature reviews microbiology* 9(2):119–130
- Lenth R, Singmann H, Love J, Buerkner P, Herve M (2018) Emmeans: Estimated marginal means, aka least-squares means. R package version 1(1):3
- Lepere C, Boucher D, Jardillier L, Domaizon I, Debroas D (2006) Succession and regulation factors of small eukaryote community composition in a lacustrine ecosystem (lake pavin). *Applied and environmental microbiology* 72(4):2971–2981
- Lepere C, Masquelier S, Mangot JF, Debroas D, Domaizon I (2010) Vertical structure of small eukaryotes in three lakes that differ by their trophic status: a quantitative approach. *The ISME journal* 4(12):1509–1519
- Lepère C, Domaizon I, Taïb N, Mangot JF, Bronner G, Boucher D, Debroas D (2013) Geographic distance and ecosystem size determine the distribution of smallest protists in lacustrine ecosystems. *FEMS microbiology ecology* 85(1):85–94
- Lewis W (2009) Ecological zonation in lakes. In: Likens GE (ed) *Encyclopedia of Inland Waters*, Academic Press, Oxford, pp 416 – 422, DOI <https://doi.org/10.1016/B978-012370626-3.00209-X>
- Lewis Jr WM (1983) A revised classification of lakes based on mixing. *Canadian Journal of Fisheries and Aquatic Sciences* 40(10):1779–1787
- Li J, Zhang J, Liu L, Fan Y, Li L, Yang Y, Lu Z, Zhang X (2015) Annual periodicity in planktonic bacterial and archaeal community composition of eutrophic lake taihu. *Scientific reports* 5(1):1–14
- Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659
- Lidicker JWZ (1979) A clarification of interactions in ecological systems. *Bioscience* 29(8):475–477

- Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, Chaffron S, Ignacio-Espinosa JC, Roux S, Vincent F, et al. (2015) Determinants of community structure in the global plankton interactome. *Science* 348(6237)
- Lindh MV, Sjöstedt J, Andersson AF, Baltar F, Huggerth LW, Lundin D, Muthusamy S, Legrand C, Pinhassi J (2015) Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. *Environmental microbiology* 17(7):2459–2476
- Lindström ES, Langenheder S (2012) Local and regional factors influencing bacterial community assembly. *Environmental Microbiology Reports* 4(1):1–9
- Lindström ES, Eiler A, Langenheder S, Bertilsson S, Drakare S, Ragnarsson H, Tranvik LJ (2007) Does ecosystem size determine aquatic bacterial richness? comment. *Ecology* 88(1):252–253
- Linz AM, Crary BC, Shade A, Owens S, Gilbert JA, Knight R, McMahon KD (2017) Bacterial community composition and dynamics spanning five years in freshwater bog lakes. *MSphere* 2(3)
- Linz AM, Aylward FO, Bertilsson S, McMahon KD (2020) Time-series metatranscriptomes reveal conserved patterns between phototrophic and heterotrophic microbes in diverse freshwater systems. *Limnology and Oceanography* 65:S101–S112
- Liu L, Yang J, Lv H, Yu X, Wilkinson DM, Yang J (2015) Phytoplankton communities exhibit a stronger response to environmental changes than bacterioplankton in three subtropical reservoirs. *Environmental science & technology* 49(18):10,850–10,858
- Liu X, Li M, Castelle CJ, Probst AJ, Zhou Z, Pan J, Liu Y, Banfield JF, Gu JD (2018) Insights into the ecology, evolution, and metabolism of the widespread woesearchaeotal lineages. *Microbiome* 6(1):102
- Liu Z, Iqbal M, Zeng Z, Lian Y, Zheng A, Zhao M, Li Z, Wang G, Li Z, Xie J (2020) Comparative analysis of microbial community structure in the ponds with different aquaculture model and fish by high-throughput sequencing. *Microbial Pathogenesis* p 104101
- Llirós M, Inceoğlu Ö, García-Armisen T, Anzil A, Leporcq B, Pigneur LM, Viroux L, Darchambeau F, Descy JP, Servais P (2014) Bacterial community composition in three freshwater reservoirs of different alkalinity and trophic status. *PLoS One* 9(12):1–27
- Locey K, Muscarella M, Larsen M, Bray S, Jones S, Lennon JT (2020) Dormancy dampens the microbial distance–decay relationship. *Philosophical Transactions of the Royal Society B* 375(1798):20190,243

- Logares R, Bråte J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, Rengefors K (2009) Infrequent marine–freshwater transitions in the microbial world. *Trends in microbiology* 17(9):414–422
- Logares R, Mangot JF, Massana R (2015) Rarity in aquatic microbes: placing protists on the map. *Research in microbiology* 166(10):831–841
- Logares R, Tesson SV, Canbäck B, Pontarp M, Hedlund K, Rengefors K (2018) Contrasting prevalence of selection and drift in the community structuring of bacteria and microbial eukaryotes. *Environmental microbiology* 20(6):2231–2240
- Logares R, Deutschmann IM, Junger PC, Giner CR, Krabberød AK, Schmidt TS, Rubinat-Ripoll L, Mestre M, Salazar G, Ruiz-González C, et al. (2020) Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* 8:1–17
- Logue JB, Lindström ES (2010) Species sorting affects bacterioplankton community composition as determined by 16s rDNA and 16s rRNA fingerprints. *The ISME journal* 4(6):729–738
- Logue JB, Langenheder S, Andersson AF, Bertilsson S, Drakare S, Lanzén A, Lindström ES (2012) Freshwater bacterioplankton richness in oligotrophic lakes depends on nutrient availability rather than on species–area relationships. *The ISME journal* 6(6):1127–1136
- López-García P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea antarctic plankton. *Nature* 409(6820):603–607
- Louca S, Polz MF, Mazel F, Albright MB, Huber JA, O'Connor MI, Ackermann M, Hahn AS, Srivastava DS, Crowe SA, et al. (2018) Function and functional redundancy in microbial systems. *Nature ecology & evolution* 2(6):936–943
- Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences* 104(27):11,436–11,440
- Luef B, Frischkorn KR, Wrighton KC, Holman HYN, Birarda G, Thomas BC, Singh A, Williams KH, Siegerist CE, Tringe SG, et al. (2015) Diverse uncultivated ultra-small bacterial cells in groundwater. *Nature communications* 6(1):1–8
- MacArthur RH, Wilson EO (1963) An equilibrium theory of insular zoogeography. *Evolution* pp 373–387
- Mackay AW, Seddon AW, Leng MJ, Heumann G, Morley DW, Piotrowska N, Rioual P, Roberts S, Swann GE (2017) Holocene carbon dynamics at the forest–steppe ecotone of southern siberia. *Global change biology* 23(5):1942–1960

- Madigan M, Martinko J, Bender K, Buckley D, Stahl D (2014) Brock Biology of Microorganisms, 14th Edn London. Pearson
- Madsen EL (2015) Environmental microbiology: from genomes to biogeochemistry. John Wiley & Sons
- Magoč T, Salzberg SL (2011) Flash: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21):2957–2963
- Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M (2015) Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* 3:e1420
- Maksimov V, Shchetinina E, Kraikivskaya O, Maksimov V, Maksimova E (2002) The classification and the monitoring of the state of mouth riverine and lacustrine ecosystems in lake baikal based on the composition of local microbiocenoses and their activity. *Microbiology* 71(5):595–600
- Maksimova E, Maksimov V (1972) Vertical distribution of microbial plankton in the southern part of lake baikal in 1969. *Mikrobiologija* 41(5):896
- Mallin MA, Ensign SH, Wheeler TL, Mayes DB (2002) Pollutant removal efficacy of three wet detention ponds. *Journal of Environmental Quality* 31(2):654–660
- Mangot JF, Domaizon I, Taib N, Marouni N, Duffaud E, Bronner G, Debroas D (2013) Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. *Environmental Microbiology* 15(6):1745–1758
- Markovic D, Carrizo SF, Kärcher O, Walz A, David JN (2017) Vulnerability of european freshwater catchments to climate change. *Global Change Biology* 23(9):3567–3580
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17(1):10–12
- Martinsen KT, Andersen MR, Sand-Jensen K (2019) Water temperature dynamics and the prevalence of daytime stratification in small temperate shallow lakes. *Hydrobiologia* 826(1):247–262
- Massana R (2011) Eukaryotic picoplankton in surface oceans. *Annual review of microbiology* 65:91–110
- Massana R, Logares R (2013) Eukaryotic versus prokaryotic marine picoplankton ecology. *Environmental microbiology* 15(5):1254–1261

- Massana R, Castresana J, Balagué V, Guillou L, Romari K, Groisillier A, Valentin K, Pedrós-Alió C (2004) Phylogenetic and ecological analysis of novel marine stramenopiles. *Applied and environmental microbiology* 70(6):3528–3534
- Massana R, Terrado R, Forn I, Lovejoy C, Pedrós-Alió C (2006) Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environmental microbiology* 8(9):1515–1522
- Massana R, Del Campo J, Sieracki ME, Audic S, Logares R (2014) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *The ISME journal* 8(4):854–866
- Mats VD, Perepelova TI (2011) A new perspective on evolution of the baikal rift. *Geoscience Frontiers* 2(3):349–365
- McDonough O, Hosen J, Palmer M (2011) Temporary streams: the hydrology, geography, and ecology of non-perennially flowing waters
- McKee D, Atkinson D, Collings S, Eaton J, Gill A, Harvey I, Hatton K, Heyes T, Wilson D, Moss B (2003) Response of freshwater microcosm communities to nutrients, fish, and elevated temperature during winter and summer. *Limnology and Oceanography* 48(2):707–722
- Meysman FJ, Middelburg JJ, Heip CH (2006) Bioturbation: a fresh look at darwin's last idea. *Trends in Ecology & Evolution* 21(12):688–695
- Mikhailov I, Zakharova YR, Galachyants YP, Usoltseva M, Petrova D, Sakirko M, Likhoshway YV, Grachev M (2015) Similarity of structure of taxonomic bacterial communities in the photic layer of lake baikal's three basins differing in spring phytoplankton composition and abundance. In: *Doklady Biochemistry and Biophysics*, Springer, vol 465, pp 413–419
- Mikhailov IS, Bukin YS, Zakharova YR, Usoltseva MV, Galachyants YP, Sakirko MV, Blinov VV, Likhoshway YV (2019a) Co-occurrence patterns between phytoplankton and bacterioplankton across the pelagic zone of lake baikal during spring. *Journal of Microbiology* 57(4):252–262
- Mikhailov IS, Zakharova YR, Bukin YS, Galachyants YP, Petrova DP, Sakirko MV, Likhoshway YV (2019b) Co-occurrence networks among bacteria and microbial eukaryotes of lake baikal during a spring phytoplankton bloom. *Microbial ecology* 77(1):96–109
- Mitra A, Flynn KJ, Tillmann U, Raven JA, Caron D, Stoecker DK, Not F, Hansen PJ, Hallegraeff G, Sanders R, et al. (2016) Defining planktonic protist functional groups on mechanisms for energy and nutrient acquisition: incorporation of diverse mixotrophic strategies. *Protist* 167(2):106–120



- Monchamp ME, Spaak P, Domaizon I, Dubois N, Bouffard D, Pomati F (2018) Homogenization of lake cyanobacterial communities over a century of climate change and eutrophication. *Nature ecology & evolution* 2(2):317–324
- Moore MV, Hampton SE, Izmet'eva LR, Silow EA, Peshkova EV, Pavlov BK (2009) Climate change and the world's "sacred sea"—lake baikal, siberia. *BioScience* 59(5):405–417
- Moore RD, Richardson JS (2003) Progress towards understanding the structure, function, and ecological significance of small stream channels and their riparian zones. *Canadian Journal of Forest Research* 33(8):1349–1351
- Moreira D, López-García P (2019) Time series are critical to understand microbial plankton diversity and ecology. *Molecular ecology* 28(5):920–922
- Moss B (2010) *Ecology of fresh waters: a view for the twenty-first century*. John Wiley & Sons
- Mukherjee I, Hodoki Y, Okazaki Y, Fujinaga S, Ohbayashi K, Nakano Si (2019) Widespread dominance of kinetoplastids and unexpected presence of diplomonads in deep freshwater lakes. *Frontiers in microbiology* 10:2375
- Müller J, Oberhänsli H, Melles M, Schwab M, Rachold V, Hubberten HW (2001) Late pliocene sedimentation in lake baikal: implications for climatic and tectonic change in se siberia. *Palaeogeography, Palaeoclimatology, Palaeoecology* 174(4):305–326
- Mullins ML, Doyle RD (2019) Big things come in small packages: why limnologists should care about small ponds. *Acta Limnologica Brasiliensia* 31
- Murtagh F, Legendre P (2014) Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *Journal of classification* 31(3):274–295
- Musat N, Musat F, Weber PK, Pett-Ridge J (2016) Tracking microbial interactions with nanosims. *Current opinion in biotechnology* 41:114–121
- Muylaert K, Van der Gucht K, Vloemans N, De Meester L, Gillis M, Vyverman W (2002) Relationship between bacterial community composition and bottom-up versus top-down variables in four eutrophic shallow lakes. *Applied and Environmental Microbiology* 68(10):4740–4750
- Needham DM, Fichot EB, Wang E, Berdjeb L, Cram JA, Fichot CG, Fuhrman JA (2018) Dynamics and interactions of highly resolved marine plankton via automated high-frequency sampling. *The ISME journal* 12(10):2417–2432
- Nelson WC, Stegen JC (2015) The reduced genomes of paracubacteria (od1) contain signatures of a symbiotic lifestyle. *Frontiers in microbiology* 6:713

- Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, Knelman JE, Darcy JL, Lynch RC, Wickey P, et al. (2013) Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews* 77(3):342–356
- Newton RJ, McMahon KD (2011) Seasonal differences in bacterial community composition following nutrient additions in a eutrophic lake. *Environmental Microbiology* 13(4):887–899
- Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011) A guide to the natural history of freshwater lake bacteria. *Microbiology and molecular biology reviews* 75(1):14–49
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ (2015) Iq-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* 32(1):268–274
- Obertegger U, Bertilsson S, Pindo M, Larger S, Flaim G (2018) Temporal variability of bacterioplankton is habitat driven. *Molecular ecology* 27(21):4322–4335
- Obertegger U, Pindo M, Flaim G (2019) Multifaceted aspects of synchrony between freshwater prokaryotes and protists. *Molecular Ecology* 28(19):4500–4512
- Oertli B (2018) Freshwater biodiversity conservation: The role of artificial ponds in the 21st century. *Aquatic conservation: marine and freshwater ecosystems* 28(2):264–269
- Oertli B, Joye DA, Castella E, Juge R, Cambin D, Lachavanne JB (2002) Does size matter? the relationship between pond area and biodiversity. *Biological conservation* 104(1):59–70
- Oertli B, Biggs J, Céréghino R, Grillas P, Joly P, Lachavanne JB (2005) Conservation and monitoring of pond biodiversity: introduction. *Aquatic Conservation: Marine and Freshwater Ecosystems* 15(6):535–540
- Oertli B, Céréghino R, Hull A, Miracle R (2009) Pond conservation: from science to practice. *Hydrobiologia* 634:1–9
- Okamoto N, Chantangsi C, Horák A, Leander BS, Keeling PJ (2009) Molecular phylogeny and description of the novel katablepharid *roombia truncata* gen. et sp. nov., and establishment of the hacrobia taxon nov. *PloS one* 4(9):e7080
- Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M (2007) The vegan package. *Community ecology package* 10(631-637):719
- Paerl HW, Otten TG, Kudela R (2018) Mitigating the expansion of harmful algal blooms across the freshwater-to-marine continuum

- Parada AE, Fuhrman JA (2017) Marine archaeal dynamics and interactions with the microbial community over 5 years from surface to seafloor. *The ISME journal* 11(11):2510–2525
- Pätzig M, Kalettka T, Onandia G, Balla D, Lischeid G (2020) How much information do we gain from multiple-year sampling in natural pond research? *Limnologica* 80:125,728
- Paver SF, Hayek KR, Gano KA, Fagen JR, Brown CT, Davis-Richardson AG, Crabb DB, Rosario-Passapera R, Giongo A, Triplett EW, et al. (2013) Interactions between specific phytoplankton and bacteria affect lake bacterial community succession. *Environmental Microbiology* 15(9):2489–2504
- Paver SF, Muratore D, Newton RJ, Coleman ML (2018) Reevaluating the salty divide: phylogenetic specificity of transitions between marine and freshwater systems. *MSystems* 3(6)
- Pavlichenko VV, Protopopova MV, Timofeyev M, Luckenbach T (2015) Identification of a putatively multixenobiotic resistance related *abcB1* transporter in amphipod species endemic to the highly pristine lake baikal. *Environmental science and pollution research* 22(7):5453–5468
- Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? *Trends in microbiology* 14(6):257–263
- Pernthaler J (2005) Predation on prokaryotes in the water column and its ecological implications. *Nature Reviews Microbiology* 3(7):537–546
- Pernthaler J, Rosenberg E, DeLong E, Lory S (2013) *Freshwater microbial communities*. Springer
- Pester M, Bittner N, Deevong P, Wagner M, Loy A (2010) A ‘rare biosphere’ microorganism contributes to sulfate reduction in a peatland. *The ISME journal* 4(12):1591–1602
- Peura S, Eiler A, Hiltunen M, Nykänen H, Tiirola M, Jones RI (2012) Bacterial and phytoplankton responses to nutrient amendments in a boreal lake differ according to season and to taxonomic resolution. *PLoS One* 7(6):e38,552
- Peura S, Bertilsson S, Jones RI, Eiler A (2015) Resistant microbial cooccurrence patterns inferred by network topology. *Applied and environmental microbiology* 81(6):2090–2097
- Pielou EC (1966) The measurement of diversity in different types of biological collections. *Journal of theoretical biology* 13:131–144
- Pinet PR (2019) *Invitation to oceanography*. Jones & Bartlett Learning

- Pomeroy LR, Williams PJ, Azam F, Hobbie JE (2007) The microbial loop. *Oceanography* 20(2):28–33
- Potapov SA, Tikhonova IV, Krasnopeev AY, Kabilov MR, Tupikin AE, Chebunina NS, Zhuchenko NA, Belykh OI (2019) Metagenomic analysis of viroplankton from the pelagic zone of lake baikal. *Viruses* 11(11):991
- Prescott DM (1994) The dna of ciliated protozoa. *Microbiological reviews* 58(2):233–267
- Prokopenko AA, Karabanov EB, Williams DF (2002) Age of long sediment cores from lake baikal. *Nature* 415(6875):976–976
- Qin Y, Hou J, Deng M, Liu Q, Wu C, Ji Y, He X (2016) Bacterial abundance and diversity in pond water supplied with different feeds. *Scientific reports* 6:35,232
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2012) The silva ribosomal rna gene database project: improved data processing and web-based tools. *Nucleic acids research* 41(D1):D590–D596
- Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC bioinformatics* 12(1):38
- R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>
- Rappé MS, Giovannoni SJ (2003) The uncultured microbial majority. *Annual Reviews in Microbiology* 57(1):369–394
- Reche I, Pulido-Villena E, Morales-Baquero R, Casamayor EO (2005) Does ecosystem size determine aquatic bacterial richness? *Ecology* 86(7):1715–1722
- Rengefors K, Gustafsson S, Ståhl-Delbanco A (2004) Factors regulating the recruitment of cyanobacterial and eukaryotic phytoplankton from littoral and profundal sediments. *Aquatic Microbial Ecology* 36(3):213–226
- Rengefors K, Weyhenmeyer GA, Bloch I (2012) Temperature as a driver for the expansion of the microalga *gonyostomum semen* in swedish lakes. *Harmful algae* 18:65–73
- Richards TA, Bass D (2005) Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Current opinion in microbiology* 8(3):240–252
- Richards TA, Leonard G, Mahé F, del Campo J, Romac S, Jones MD, Maguire F, Dunthorn M, De Vargas C, Massana R, et al. (2015) Molecular diversity and distribution of marine fungi

- across 130 european environmental samples. *Proceedings of the Royal Society B: Biological Sciences* 282(1819):20152,243
- Riebesell U, Bellerby R, Grossart HP, Thingstad F (2008) Mesocosm co 2 perturbation studies: from organism to community level. *Biogeosciences* 5(4):1157–1164
- Riemann L, Leitet C, Pommier T, Simu K, Holmfeldt K, Larsson U, Hagström Å (2008) The native bacterioplankton community in the central baltic sea is influenced by freshwater bacterial species. *Applied and environmental microbiology* 74(2):503–515
- Ritchie RJ (2006) Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynthesis research* 89(1):27–41
- Roberts S, Adams JK, Mackay AW, Swann GE, McGowan S, Rose NL, Panizzo V, Yang H, Vologina E, Sturm M, et al. (2020) Mercury loading within the selenga river basin and lake baikal, siberia. *Environmental Pollution* 259:113,814
- Roberts SL, Swann GE, McGowan S, Panizzo VN, Vologina EG, Sturm M, Mackay AW (2018) Diatom evidence of 20th century ecosystem change in lake baikal, siberia. *PloS one* 13(12):e0208,765
- Rochelle-Newall E, Hulot F, Janeau JL, Merroune A (2014) Cdom fluorescence as a proxy of doc concentration in natural waters: a comparison of four contrasting tropical systems. *Environmental monitoring and assessment* 186(1):589–596
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) Vsearch: a versatile open source tool for metagenomics. *PeerJ* 4:e2584
- Rolls RJ, Heino J, Ryder DS, Chessman BC, Growns IO, Thompson RM, Gido KB (2018) Scaling biodiversity responses to hydrological regimes. *Biological Reviews* 93(2):971–995
- Rösel S, Allgaier M, Grossart HP (2012) Long-term characterization of free-living and particle-associated bacterial communities in lake tiefwaren reveals distinct seasonal patterns. *Microbial ecology* 64(3):571–583
- Röttjers L, Faust K (2018) From hairballs to hypotheses—biological insights from microbial networks. *FEMS microbiology reviews* 42(6):761–780
- Rugenski AT, Murria C, Whiles MR (2012) Tadpoles enhance microbial activity and leaf decomposition in a neotropical headwater stream. *Freshwater Biology* 57(9):1904–1913
- Rusak JA, Tanentzap AJ, Klug JL, Rose KC, Hendricks SP, Jennings E, Laas A, Pierson D, Ryder E, Smyth R, et al. (2018) Wind and trophic status explain within and among-lake variability of algal biomass. *Limnology and oceanography letters* 3(6):409–418

- Salazar G (2018) Ecolutils: Utilities for community ecology analysis. r package version 0.1
- Salcher MM (2014) Same same but different: ecological niche partitioning of planktonic freshwater prokaryotes. *Journal of Limnology* 73(1s):74–87
- Salcher MM, Hofer J, Horňák K, Jezbera J, Sonntag B, Vrba J, Šimek K, Posch T (2007) Modulation of microbial predator-prey dynamics by phosphorus availability: growth patterns and survival strategies of bacterial phylogenetic clades. *FEMS microbiology ecology* 60(1):40–50
- Sarmiento JL, Gruber N (2006) *Ocean biogeochemical dynamics*. Princeton University Press
- Sassenhagen I, Irion S, Jardillier L, Moreira D, Christaki U (2020) Protist interactions and community structure during early autumn in the kerguelen region (southern ocean). *Protist* 171(1):125,709
- Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, Funk K, Ketter A, Kim S, Kimchi A, et al. (2020) Database resources of the national center for biotechnology information. *Nucleic acids research* 48(D1):D9
- Schauer M, Balagué V, Pedrós-Alió C, Massana R (2003) Seasonal changes in the taxonomic composition of bacterioplankton in a coastal oligotrophic system. *Aquatic Microbial Ecology* 31(2):163–174
- Scherer PI, Millard AD, Miller A, Schoen R, Raeder U, Geist J, Zwirgmaier K (2017) Temporal dynamics of the microbial community composition with a focus on toxic cyanobacteria and toxin presence during harmful algal blooms in two south german lakes. *Frontiers in microbiology* 8:2387
- Schindler DW (1988) Effects of acid rain on freshwater ecosystems. *Science* 239(4836):149–157
- Schmid M, Budnev NM, Granin NG, Sturm M, Schurter M, Wüest A (2008) Lake baikal deepwater renewal mystery solved. *Geophysical Research Letters* 35(9)
- Schuur EA, McGuire AD, Schädel C, Grosse G, Harden J, Hayes DJ, Hugelius G, Koven CD, Kuhry P, Lawrence DM, et al. (2015) Climate change and the permafrost carbon feedback. *Nature* 520(7546):171–179
- Semlitsch RD, Peterman WE, Anderson TL, Drake DL, Ousterhout BH (2015) Intermediate pond sizes contain the highest density, richness, and diversity of pond-breeding amphibians. *PLoS one* 10(4):e0123,055

- Seong KA, Jeong HJ, Kim S, Kim GH, Kang JH (2006) Bacterivory by co-occurring red-tide algae, heterotrophic nanoflagellates, and ciliates. *Marine Ecology Progress Series* 322:85–97
- Sevrin-Reyssac J, Pletikosic M (1990) Cyanobacteria in fish ponds. *Aquaculture* 88(1):1–20
- Seymour JR, Amin SA, Raina JB, Stocker R (2017) Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nature microbiology* 2(7):1–12
- Shade A, Kent AD, Jones SE, Newton RJ, Triplett EW, McMahon KD (2007) Interannual dynamics and phenology of bacterial communities in a eutrophic lake. *Limnology and Oceanography* 52(2):487–494
- Shade A, Jones SE, McMahon KD (2008) The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics. *Environmental microbiology* 10(4):1057–1067
- Shade A, Peter H, Allison SD, Baho D, Berga M, Bürgmann H, Huber DH, Langenheder S, Lennon JT, Martiny JB, et al. (2012) Fundamentals of microbial community resistance and resilience. *Frontiers in microbiology* 3:417
- Shade A, Caporaso JG, Handelsman J, Knight R, Fierer N (2013) A meta-analysis of changes in bacterial and archaeal communities with time. *The ISME journal* 7(8):1493–1506
- Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N, Gilbert JA (2014) Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *MBio* 5(4)
- Sherr E, Sherr B (1994) Bacterivory and herbivory: key roles of phagotrophic protists in pelagic food webs. *Microbial Ecology* 28(2):223–235
- Sherr E, Sherr BF (1988) Role of microbes in pelagic food webs: a revised concept. *Limnology and Oceanography* 33:1225–1227
- Sherr EB, Sherr BF (2002) Significance of predation by protists in aquatic microbial food webs. *Antonie van Leeuwenhoek* 81(1-4):293–308
- Sherstyankin P, Alekseev S, Abramov A, Stavrov K, De Batist M, Hus R, Canals M, Casamor J (2006) Computer-based bathymetric map of lake baikal. *Doklady earth sciences* 408(1):564
- Shi X, Li S, Zhang M, Liu C, Wu Q (2020) Temperature mainly determines the temporal succession of the photosynthetic picoeukaryote community in lake chaohu, a highly eutrophic shallow lake. *Science of The Total Environment* 702:134,803

- Shimaraev MN, Domyshcheva VM (2013) Trends in hydrological and hydrochemical processes in lake baikal under conditions of modern climate change. *Climatic change and global warming of inland waters* pp 43–66
- Shiratori T, Suzuki S, Kakizawa Y, Ishida Ki (2019) Phagocytosis-like cell engulfment by a planctomycete bacterium. *Nature communications* 10(1):1–11
- Shoemaker WR, Lennon JT (2018) Evolution with a seed bank: the population genetic consequences of microbial dormancy. *Evolutionary applications* 11(1):60–75
- Sigman DM, Hain MP (2012) The biological productivity of the ocean. *Nature Education Knowledge* 3(6):1–16
- Šimek K, Kasalický V, Jezbera J, Horňák K, Nedoma J, Hahn MW, Bass D, Jost S, Boenigk J (2013) Differential freshwater flagellate community response to bacterial food quality with a focus on limnhabitans bacteria. *The ISME journal* 7(8):1519–1530
- Simek K, Gruzic V, Nedoma J, Jezberova J, Sorf M, Matous A, Pechar L, Posch T, Bruni EP, Vrba J (2019) Microbial food webs in hypertrophic fishponds: Omnivorous ciliate taxa are major protistan bacterivores. *Limnology and Oceanography* 64(5):2295–2309
- Simon M, López-García P, Moreira D, Jardillier L (2013) New haptophyte lineages and multiple independent colonizations of freshwater ecosystems. *Environmental Microbiology Reports* 5(2):322–332
- Simon M, Jardillier L, Deschamps P, Moreira D, Restoux G, Bertolino P, López-García P (2014) Complex communities of small protists and unexpected occurrence of typical marine lineages in shallow freshwater systems. *Environmental microbiology* 17(10):3610–3627
- Simon M, López-García P, Deschamps P, Moreira D, Restoux G, Bertolino P, Jardillier L (2015) Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. *The ISME journal* 9(9):1941–1953
- Simon M, López-García P, Deschamps P, Restoux G, Bertolino P, Moreira D, Jardillier L (2016) Resilience of freshwater communities of small microbial eukaryotes undergoing severe drought events. *Frontiers in Microbiology* 7:812
- Simpson EH (1949) Measurement of diversity. *nature* 163(4148):688–688
- Smith VH (2003) Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environmental Science and Pollution Research* 10(2):126–139



- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proceedings of the National Academy of Sciences* 103(32):12,115–12,120
- Solden L, Lloyd K, Wrighton K (2016) The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Current opinion in microbiology* 31:217–226
- Sommer U, Adrian R, De Senerpont Domis L, Elser JJ, Gaedke U, Ibelings B, Jeppesen E, Lüring M, Molinero JC, Mooij WM, et al. (2012) Beyond the plankton ecology group (peg) model: mechanisms driving plankton succession. *Annual review of ecology, evolution, and systematics* 43:429–448
- Søndergaard M, Jeppesen E, Jensen JP (2005) Pond or lake: does it make any difference? *Archiv für Hydrobiologie* 162(2):143–165
- Spencer SJ, Tamminen MV, Preheim SP, Guo MT, Briggs AW, Brito IL, Weitz DA, Pitkänen LK, Vigneault F, Virta MP, et al. (2016) Massively parallel sequencing of single cells by epicpcc links functional genes with phylogenetic markers. *The ISME journal* 10(2):427–436
- Stackebrandt E, Goebel BM (1994) Taxonomic note: a place for dna-dna reassociation and 16s rRNA sequence analysis in the present species definition in bacteriology. *International journal of systematic and evolutionary microbiology* 44(4):846–849
- Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ (2015) Species sorting and seasonal dynamics primarily shape bacterial communities in the upper mississippi river. *Science of the Total Environment* 505:435–445
- Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, Chow CET, Sachdeva R, Jones AC, Schwalbach MS, et al. (2011) Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *The ISME journal* 5(9):1414–1425
- Stegen JC, Lin X, Fredrickson JK, Chen X, Kennedy DW, Murray CJ, Rockhold ML, Konopka A (2013) Quantifying community assembly processes and identifying features that impose them. *The ISME journal* 7(11):2069–2079
- Stewart RI, Andersson GK, Brönmark C, Klatt BK, Hansson LA, Zülsdorff V, Smith HG (2017) Ecosystem services across the aquatic–terrestrial boundary: Linking ponds to pollination. *Basic and applied ecology* 18:13–20
- Stockner JG (1988) Phototrophic picoplankton: an overview from marine and freshwater ecosystems. *Limnology and Oceanography* 33(4part2):765–775

- Stoeck T, Zuendorf A, Breiner HW, Behnke A (2007) A molecular approach to identify active microbes in environmental eukaryote clone libraries. *Microbial Ecology* 53(2):328–339
- Strahler AN (1957) Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union* 38(6):913–920
- Su X, Steinman AD, Xue Q, Zhao Y, Tang X, Xie L (2017) Temporal patterns of phyto- and bacterioplankton and their relationships with environmental factors in lake taihu, china. *Chemosphere* 184:299–308
- Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, et al. (2015) Structure and function of the global ocean microbiome. *Science* 348(6237)
- Takai K, Horikoshi K (2000) Rapid detection and quantification of members of the archaeal community by quantitative pcr using fluorogenic probes. *Applied and environmental microbiology* 66(11):5066–5072
- Tammert H, Tšertova N, Kiprovskaia J, Baty F, Nõges T, Kisand V (2015) Contrasting seasonal and interannual environmental drivers in bacterial communities within a large shallow lake: evidence from a seven year survey. *Aquatic Microbial Ecology* 75(1):43–54
- Teachey ME, McDonald JM, Ottesen EA (2019) Rapid and stable microbial community assembly in the headwaters of a third-order stream. *Applied and environmental microbiology* 85(11)
- Teittinen A, Soininen J (2015) Testing the theory of island biogeography for microorganisms patterns for spring diatoms. *Aquatic microbial ecology* 75(3):239–250
- Thébault E, Fontaine C (2010) Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* 329(5993):853–856
- Thingstad TF (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography* 45(6):1320–1328
- Tiessen H (2008) Phosphorus in the global environment. In: *The ecophysiology of plant-phosphorus interactions*, Springer, pp 1–7
- Tillmann U, Elbrächter M, Krock B, John U, Cembella A (2009) *Azadinium spinosum* gen. et sp. nov.(dinophyceae) identified as a primary producer of azaspiracid toxins. *European Journal of Phycology* 44(1):63–79

- Tobias-Hünefeldt SP, Wing SR, Espinel-Velasco N, Baltar F, Morales SE (2019) Depth and location influence prokaryotic and eukaryotic microbial community structure in new zealand fjords. *Science of the Total Environment* 693:133,507
- Torruella G, Grau-Bové X, Moreira D, Karpov SA, Burns JA, Sebé-Pedrós A, Völcker E, López-García P (2018) Global transcriptome analysis of the aphelid paraphelidium *tribonemae* supports the phagotrophic origin of fungi. *Communications biology* 1(1):1–11
- Trench-Fiol S, Fink P (2020) Metatranscriptomics from a small aquatic system: Microeukaryotic community functions through the diurnal cycle. *Frontiers in Microbiology* 11
- Treusch AH, Vergin KL, Finlay LA, Donatz MG, Burton RM, Carlson CA, Giovannoni SJ (2009) Seasonality and vertical structure of microbial communities in an ocean gyre. *The ISME journal* 3(10):1148–1163
- Troitskaya E, Blinov V, Ivanov V, Zhdanov A, Gnatovsky R, Sutyryna E, Shimaraev M (2015) Cyclonic circulation and upwelling in lake baikal. *Aquatic Sciences* 77(2):171–182
- Troussellier M, Escalas A, Bouvier T, Mouillot D (2017) Sustaining rare marine microorganisms: macroorganisms as repositories and dispersal agents of microbial diversity. *Frontiers in microbiology* 8:947
- Tsilimigras MC, Fodor AA (2016) Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Annals of epidemiology* 26(5):330–335
- Tyner S, Briatte F, Hofmann H (2017) Network visualization with ggplot2
- UNDP-GEF (2015) The ecological atlas of the baikal basin. United Nations Office for Project Services (UNOPS) p 145
- Vadstein O (2000) Heterotrophic, planktonic bacteria and cycling of phosphorus. In: *Advances in microbial ecology*, Springer, pp 115–167
- Van Der Heijden MG, Bardgett RD, Van Straalen NM (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology letters* 11(3):296–310
- Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE (1980) The river continuum concept. *Canadian journal of fisheries and aquatic sciences* 37(1):130–137
- Vellend M (2010) Conceptual synthesis in community ecology. *The Quarterly review of biology* 85(2):183–206

- Vigneron A, Cruaud P, Langlois V, Lovejoy C, Culley AI, Vincent WF (2020) Ultra-small and abundant: Candidate phyla radiation bacteria are potential catalysts of carbon transformation in a thermokarst lake ecosystem. *Limnology and Oceanography Letters* 5(2):212–220
- Ward CS, Yung CM, Davis KM, Blinebry SK, Williams TC, Johnson ZI, Hunt DE (2017) Annual community patterns are driven by seasonal switching between closely related marine bacteria. *The ISME journal* 11(6):1412–1422
- Ward JG, Blomberg P, Hoffman N, Yao MC (1997) The intranuclear organization of normal, hemizygous and excision-deficient *rrna* genes during developmental amplification in *tetrahymena thermophila*. *Chromosoma* 106(4):233–242
- Warnes MGR, Bolker B, Bonebakker L, Gentleman R, Huber W, et al. (2016) Package ‘gplots’. Various R Programming Tools for Plotting Data
- Weinbauer MG, Christaki U, Nedoma J, Simek K (2003) Comparing the effects of resource enrichment and grazing on viral production in a meso-eutrophic reservoir. *Aquatic Microbial Ecology* 31(2):137–144
- Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, Xia LC, Xu ZZ, Ursell L, Alm EJ, et al. (2016) Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *The ISME journal* 10(7):1669–1681
- Wellborn GA, Skelly DK, Werner EE (1996) Mechanisms creating community structure across a freshwater habitat gradient. *Annual review of ecology and systematics* 27(1):337–363
- Wheeler B, Torchiano M, Torchiano MM (2016) Package ‘Imperm’. R package version pp 1–1
- Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, Cordero OX, Brown SP, Momeni B, Shou W, et al. (2016) Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME journal* 10(11):2557–2568
- Wilburn P, Shchapov K, Theriot EC, Litchman E (2019) Environmental drivers define contrasting microbial habitats, diversity and functional redundancy in lake baikal, siberia. *bioRxiv* p 605899
- Wilhelm SW, LeCleir GR, Bullerjahn GS, McKay RM, Saxton MA, Twiss MR, Bourbonniere RA (2014) Seasonal changes in microbial community structure and activity imply winter production is linked to summer hypoxia in a large lake. *FEMS microbiology ecology* 87(2):475–485
- Williams RJ, Howe A, Hofmockel KS (2014) Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Frontiers in microbiology* 5:358

- Williamson CE, Dodds W, Kratz TK, Palmer MA (2008) Lakes and streams as sentinels of environmental change in terrestrial and atmospheric processes. *Frontiers in Ecology and the Environment* 6(5):247–254
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. *Proceedings of the National Academy of Sciences* 87(12):4576–4579
- Wood SA, Kelly LT, Bouma-Gregson K, Humbert JF, Laughinghouse HD, Lazorchak J, McAllister TG, McQueen A, Pokrzywinski K, Puddick J, et al. (2020) Toxic benthic freshwater cyanobacterial proliferations: Challenges and solutions for enhancing knowledge and improving monitoring and mitigation. *Freshwater Biology*
- Woodhouse JN, Kinsela AS, Collins RN, Bowling LC, Honeyman GL, Holliday JK, Neilan BA (2016) Microbial communities reflect temporal changes in cyanobacterial composition in a shallow ephemeral freshwater lake. *The ISME journal* 10(6):1337–1351
- Woodward G, Perkins DM, Brown LE (2010) Climate change and freshwater ecosystems: impacts across multiple levels of organization. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1549):2093–2106
- Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ (2015) Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* 347(6223)
- Xiao X, Sogge H, Lagesen K, Tooming-Klunderud A, Jakobsen KS, Rohrlack T (2014) Use of high throughput sequencing and light microscopy show contrasting results in a study of phytoplankton occurrence in a freshwater environment. *PloS one* 9(8):e106,510
- Yan Q, Stegen JC, Yu Y, Deng Y, Li X, Wu S, Dai L, Zhang X, Li J, Wang C, et al. (2017) Nearly a decade-long repeatable seasonal diversity patterns of bacterioplankton communities in the eutrophic lake donghu (wuhan, china). *Molecular ecology* 26(14):3839–3850
- Yannarell A, Kent A, Lauster G, Kratz T, Triplett E (2003) Temporal patterns in bacterial communities in three temperate lakes of different trophic status. *Microbial ecology* 46(4):391–405
- Yannarell AC, Triplett EW (2004) Within-and between-lake variability in the composition of bacterioplankton communities: investigations using multiple spatial scales. *Applied and environmental microbiology* 70(1):214–223
- Yannarell AC, Triplett EW (2005) Geographic and environmental sources of variation in lake bacterial community composition. *Applied and environmental microbiology* 71(1):227–239

- Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16s rRNA gene sequences. *Nature Reviews Microbiology* 12(9):635–645
- Yi Z, Berney C, Hartikainen H, Mahamdallie S, Gardner M, Boenigk J, Cavalier-Smith T, Bass D (2017) High-throughput sequencing of microbial eukaryotes in lake baikal reveals ecologically differentiated communities and novel evolutionary radiations. *FEMS Microbiology Ecology* 93(8):fix073
- Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH, Yang EC, Duffy S, Bhattacharya D (2011) Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* 332(6030):714–717
- Yvon-Durocher G, Hulatt CJ, Woodward G, Trimmer M (2017) Long-term warming amplifies shifts in the carbon cycle of experimental ponds. *Nature Climate Change* 7(3):209–213
- Zakharenko AS, Galachyants YP, Morozov IV, Shubenkova OV, Morozov AA, Ivanov VG, Pimenov NV, Krasnopeev AY, Zemskaya TI (2019) Bacterial communities in areas of oil and methane seeps in pelagic of lake baikal. *Microbial ecology* 78(2):269–285
- Zakharova YR, Galachyants YP, Kurilkina MI, Likhoshvay AV, Petrova DP, Shishlyannikov SM, Ravin NV, Mardanov AV, Beletsky AV, Likhoshway YV (2013) The structure of microbial community and degradation of diatoms in the deep near-bottom layer of lake baikal. *PLoS one* 8(4):e59,977
- Zancarini A, Echenique-Subiabre I, Debross D, Taïb N, Quiblier C, Humbert JF (2017) Deciphering biodiversity and interactions between bacteria and microeukaryotes within epilithic biofilms from the Loue river, France. *Scientific reports* 7(1):1–13
- Zehr JP, Shilova IN, Farnelid HM, del Carmen Muñoz-Marín M, Turk-Kubo KA (2016) Unusual marine unicellular symbiosis with the nitrogen-fixing cyanobacterium UCYN-A. *Nature Microbiology* 2(1):1–11
- Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR (2015) Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences* 112(20):6449–6454
- Zeng J, Jiao C, Zhao D, Xu H, Huang R, Cao X, Yu Z, Wu QL (2019) Patterns and assembly processes of planktonic and sedimentary bacterial community differ along a trophic gradient in freshwater lakes. *Ecological Indicators* 106:105,491

- Zhao D, Shen F, Zeng J, Huang R, Yu Z, Wu QL (2016) Network analysis reveals seasonal variation of co-occurrence correlations between cyanobacteria and other bacterioplankton. *Science of the Total Environment* 573:817–825
- Zhong ZP, Liu Y, Miao LL, Wang F, Chu LM, Wang JL, Liu ZP (2016) Prokaryotic community structure driven by salinity and ionic concentrations in plateau lakes of the tibetan plateau. *Applied and Environmental Microbiology* 82(6):1846–1858
- Zhu C, Zhang J, Nawaz MZ, Mahboob S, Al-Ghanim KA, Khan IA, Lu Z, Chen T (2019) Seasonal succession and spatial distribution of bacterial community structure in a eutrophic freshwater lake, lake taihu. *Science of the Total Environment* 669:29–40
- Zubkov MV, Tarran GA (2008) High bacterivory by the smallest phytoplankton in the north atlantic ocean. *Nature* 455(7210):224–226
- Zwart G, Crump BC, Kamst-van Agterveld MP, Hagen F, Han SK (2002) Typical freshwater bacteria: an analysis of available 16s rRNA gene sequences from plankton of lakes and rivers. *Aquatic microbial ecology* 28(2):141–155
- Zwirgmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaultot D, Not F, Massana R, Ulloa O, Scanlan DJ (2008) Global phylogeography of marine *synechococcus* and *prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environmental microbiology* 10(1):147–161

## Résumé en français

Les écosystèmes d'eau douce représentent moins de 3% de l'eau terrestre, mais ils sont très divers, aussi bien en terme de fonctionnement que de taille. En effet, ils peuvent être divisés entre écosystèmes lenticques et lotiques, ainsi qu'entre grands et petits écosystèmes. Les grands écosystèmes d'eau douce contiennent un volume important d'eau et sont considérés comme plus stables dans le temps. En effet, leur grand volume d'eau induit un fort effet tampon ce qui permet de réguler les changements environnementaux. Cependant, ces grands écosystèmes sont plus sujets à des variations spatiales au sein de l'écosystème, via des gradients horizontaux et verticaux. Les petits écosystèmes d'eau douce sont eux très largement distribués et sont très nombreux. Leur superficie cumulée en zone tempérée est quasiment équivalente à celle des grands écosystèmes. Les petits écosystèmes d'eau douce sont caractérisés par une profondeur et un volume restreints, ce qui correspond à un faible effet tampon. Ces écosystèmes ont donc un fonctionnement particulier, et sont en principe plus sensibles aux perturbations, aux variations saisonnières et aux changements climatiques globaux. De plus, ils contribuent significativement au cycle du carbone, notamment dans les émissions de méthane et de dioxyde de carbone des eaux continentales.

Les microorganismes sont omniprésents dans une grande partie des écosystèmes terrestres. Ils sont des acteurs majeurs du fonctionnement de ces écosystèmes, en particulier de celui des écosystèmes aquatiques. Les microorganismes représentent la majeure partie de la biodiversité terrestre et peuvent être répartis dans trois grands domaines : les bactéries, les archées et les eucaryotes. Indépendamment



---

des domaines, les microorganismes présentent des caractéristiques communes, par exemple la petite taille, la diversité dans les populations et l'abondance. Cependant, selon les groupes phylogénétiques, les microorganismes peuvent présenter des caractères très divers, notamment en ce qui concerne leur physiologie, leur capacité métabolique ou encore leur mode trophique. Les microorganismes des trois domaines du vivant sont en constante interaction. De ce fait, il est important de s'intéresser aux trois domaines du vivants lorsque l'on veut étudier un écosystème et les communautés qu'il héberge.

L'un des grands objectifs de l'écologie microbienne est de mieux comprendre le fonctionnement des écosystèmes. Pour chercher à apporter des éléments de réponse à cette question globale dans un écosystème, plusieurs étapes peuvent être réalisées. Tout d'abord, il est nécessaire de décrire la composition et la structuration des communautés dans l'écosystème. Ensuite, il faut identifier les facteurs abiotiques qui peuvent influencer la structuration de la communauté. En effet, les communautés microbiennes peuvent être influencées par une large gamme de conditions physico-chimiques. Il est donc important de comprendre comment l'environnement influence les communautés. Enfin, il est important de s'intéresser aux interactions biotiques entre microorganismes, car elles ont également un grand rôle dans la structuration des communautés. L'ensemble de ces études doit être intégré dans un cadre spatio-temporel. L'étude d'un écosystème à un moment ou dans un endroit donné ne donne pas une vision représentative de la diversité et de la complexité des communautés.

L'objectif de cette thèse de doctorat a été d'étudier le fonctionnement des écosystèmes d'eau douce en se focalisant sur leurs communautés microbiennes. Pour cela, deux types d'écosystèmes ont été explorés : un grand lac, le lac Baïkal (Russie), qui est le plus ancien et le plus profond et le plus profond de la planète ainsi que cinq petits écosystèmes d'eau douce localisés dans la région Ile-de-France (France) : quatre étangs présentant différentes caractéristiques environnementales (Etang des Vallées, La Claye, Mare Gabard et Saint Robert) et un ruisseau (Ru

---

Sainte Anne). L'étude des différents écosystèmes d'eau douce s'est faite via (i) l'exploration de la diversité et de l'autoécologie des communautés qui habitent les écosystèmes, (ii) l'identification de l'influence des facteurs top-down et bottom-up sur les communautés, (iii) la conduite d'analyses multivariées selon des échelles spatio-temporelles cohérentes, de façon à avoir une vue d'ensemble des différents mécanismes en jeu dans les écosystèmes.

## **Structuration latitudinale et verticale des communautés de protistes dans le Lac Baïkal**

Le lac Baïkal est situé en Sibérie (Russie). Parmi ses différentes caractéristiques, ce lac est le plus vieux lac terrestre, il contient le plus grand volume d'eau douce et c'est également le lac le plus profond. Ses différentes caractéristiques font que le lac Baïkal peut être visualisé comme une "mer d'eau douce", ce qui le différencie des autres lacs et peut conduire à des variations dans son fonctionnement et dans les communautés microbiennes qui y résident. De ce fait, le lac Baïkal fait l'objet de recherches intensives. Cependant, ses communautés microbiennes restent peu étudiées, notamment selon les gradients horizontaux et verticaux. Dans cette étude, nous avons cherché à déterminer l'impact de ces gradients sur les eucaryotes planctoniques du lac. Pour cela, nous avons étudié des échantillons collectés pendant l'été 2017, le long d'un gradient latitudinal Nord-Sud (sur environ 600 km) recouvrant les trois bassins du lac (Nord, Central et Sud). Les prélèvements ont été conduits dans les eaux côtières et pélagiques, de la surface aux zones les plus profondes (5 à 1400 m, avec une répartition des échantillons dans trois zones de profondeur, épi-, méso- et bathypélagique). Les données sur les communautés microbiennes ont été obtenues par une approche de metabarcoding, avec un séquençage Illumina MiSeq de l'ARNr 18S.

Tout d'abord, j'ai caractérisé les communautés microbiennes eucaryotes planc-

toniques (taille des cellules de 0,2 à 30  $\mu\text{m}$ ) dans le lac. Mes résultats montrent que les communautés de protistes dans le lac sont complexes et diverses. Elles sont dominées par des alvéolés (ciliés et dinoflagellés), des ochrophytes et des lignées d'holomycètes, avec des cryptophytes, des haptophytes ou encore des katablepharides en abondance modérée. J'ai également détecté de nombreuses lignées présentes en faible abondance, dont notamment plusieurs lignées typiquement marines, telles que les diplonemides, les syndiniales et les radiolaires. Ces observations suggèrent que la barrière de salinité est relativement facile à franchir et que les déterminants "marins" pourraient être davantage liés à la nature oligotrophe du système et la présence d'une colonne d'eau profonde plutôt qu'à la salinité elle-même.

Ensuite, l'étude des communautés selon le gradient latitudinal a révélé que celui-ci jouait un faible rôle sur la structuration des communautés. En effet, il n'y a pas de différence significative dans la composition des communautés selon le bassin de prélèvement, malgré la grande échelle de distance. De même, les communautés sont similaires entre les échantillons prélevés en zones côtières et en zones pélagiques. En revanche, j'ai pu observer que la profondeur a un effet significatif fort sur la stratification des communautés protistes. En effet, les communautés de protistes planctoniques sont fortement et significativement stratifiées dans le lac Baïkal. Cela suggère que la profondeur, qui peut être considérée comme un proxy de la lumière, de la température, du pH, de l'oxygène ou encore de la conductivité, est un déterminant majeur de la structuration des communautés microbiennes.

Enfin, j'ai étudié les interactions entre les microorganismes au sein des différentes classes de profondeur. Pour cela, j'ai construit un réseau de co-occurrence par classe de profondeur, en prenant en compte uniquement les OTUs (Unité Taxonomique Opérationnelle) les plus abondantes. L'analyse des réseaux de cooccurrence a montré que les communautés microbiennes de surface (épipélagiques) sont beaucoup plus interconnectées que les communautés des zones méso- et bathypélagiques. Les réseaux de co-occurrence obtenus suggèrent également que

---

les interactions biotiques entre les microorganismes autotrophes, hétérotrophes et parasitaires influencent la structure des communautés protistes.

## **Similarité des patrons temporels au cours d'un suivi mensuel de deux ans dans de petits écosystèmes d'eau douce malgré des communautés microbiennes différentes**

Malgré leur importance écologique, les petits écosystèmes d'eau douce sont négligés dans les études scientifiques. De ce fait, peu de connaissances ont été acquises sur la diversité et la composition des communautés microbiennes qui les habitent. Un précédent travail, conduit par M. Simon, a décrit la composition de petits eucaryotes microbiens dans cinq petits écosystèmes d'eau douce situés au sud-ouest de Paris (France). L'échantillonnage a été effectué mensuellement pendant deux ans, sur la fraction de taille de 0,2 à 5  $\mu\text{m}$ . Les données ont été obtenues par metabarcoding, grâce au pyroséquençage 454 FLX de gènes ARNr 18S. Cette étude a révélé que ces petits écosystèmes d'eau douce abritaient une grande variété d'eucaryotes microbiens. En outre, les communautés se sont révélées être très dynamiques, avec des patrons saisonniers très marqués. Dans ce travail, j'ai analysé les communautés procaryotes du même échantillonnage, en m'intéressant aux gènes d'ARNr 16S. J'ai également considéré l'ensemble des données collectées sur les deux années pour étudier simultanément l'ensemble de la communauté microbienne (bactéries, archées et eucaryotes microbiens) afin de mieux comprendre la dynamique des communautés au cours du temps.

Tout d'abord, j'ai décrit la diversité et la composition des communautés bactériennes et archéennes dans ces petits écosystèmes. J'ai pu observer une très grande diversité de procaryotes, avec l'ensemble des grands groupes connus représentés. De façon inattendue pour les environnements oxiques, j'ai détecté de nombreux membres de la Candidate Phyla Radiation (CPR), qui étaient localement

---

très abondants. La comparaison des communautés microbiennes entre les différents écosystèmes a montré qu'ils présentaient des communautés très différentes. Par exemple, une faible proportion d'OTUs (<1%) était commune aux cinq écosystèmes malgré leur proximité géographique. Ces communautés microbiennes presque uniques à chaque écosystème mettent en évidence la forte influence de sélection des conditions environnementales locales.

Ensuite, j'ai exploré les variations temporelles de l'ensemble de la communauté microbienne. J'ai pu observer des patrons de saisonnalité dans les cinq écosystèmes, pour les procaryotes et les eucaryotes microbiens. Cette tendance est marquée, malgré la présence de fort changements au cours du temps dans la composition des communautés. En parallèle de ces dynamiques saisonnières, j'ai pu remarquer un changement global de communautés au cours des deux années.

Enfin, j'ai cherché à identifier les paramètres abiotiques impliqués dans la composition de la communauté microbienne et à déterminer leur importance relative. L'analyse des variations temporelles des paramètres physico-chimiques mesurés a révélé qu'ils présentaient également une saisonnalité marquée. En revanche, contrairement aux communautés microbiennes, ces paramètres sont caractérisés par une relative stabilité au cours des deux années. L'analyse de l'effet global des paramètres abiotiques mesurés a montré qu'ils n'expliquaient qu'une faible partie de la variation des communautés microbiennes. Cette observation souligne l'importance probable des interactions biotiques dans la composition et la structure des communautés microbiennes

## Suivi saisonnier des communautés microbienne eucaryotes et procaryotes pendant 8 ans dans de petits écosystèmes d'eau douce : patrons intra- et inter- annuels

L'étude des communautés microbiennes sur un suivi mensuel de deux ans a permis de mettre en évidence l'importance des petits écosystèmes d'eau douce en terme de diversité microbienne (intra et inter écosystèmes) et en terme de complexité dans leurs dynamiques temporelles. Cependant, une durée d'étude de deux ans n'est pas suffisante pour valider les dynamiques temporelles observées. En effet, des études sur le plus long terme sont essentielles pour décrire et prédire les dynamiques temporelles des communautés microbiennes. Elles sont d'autant plus importantes pour des petits écosystèmes d'eau douce, qui réagissent rapidement aux perturbations de l'environnement. Un long suivi est donc nécessaire pour couvrir un large éventail des conditions environnementales potentielles. Nous avons donc effectué un échantillonnage saisonnier pendant huit ans, ce qui représente environ 30 échantillons pour chaque écosystème. Les données ont été obtenues grâce à Illumina Séquençage MiSeq des amplicons d'ADNr 16S et 18S. Dans cette nouvelle approche, j'ai ciblé une plus grande fraction de taille de communautés microbiennes, à savoir de 0,22 à 30  $\mu\text{m}$  au lieu de 0,22 à 5  $\mu\text{m}$ .

Tout d'abord, cette étude m'a permis de conduire une description plus poussée de la composition microbienne dans les cinq petits écosystèmes d'eau douce, car aucun plateau de saturation de la diversité n'avait été atteint dans les précédentes études portant sur les deux années de suivi. Même si le plateau de saturation n'a toujours pas été atteint, ce suivi saisonnier de huit années a permis d'identifier respectivement 22 764 et 25 536 OTUs à 95% d'identité pour les procaryotes et les eucaryotes. Les OTUs appartiennent à tous les grands groupes microbiens actuellement connus. En outre, comme l'a montré le suivi mensuel sur deux années, les communautés microbiennes sont spécifiques à chaque écosystème malgré leur proximité géographique. Ces observations confirment la grande diversité microbi-

---

enne abritée par les petits écosystèmes d'eau douce, et donc l'importance de les étudier.

Ensuite, j'ai pu confirmer les tendances temporelles observées à l'échelle des deux ans grâce aux données du suivi sur huit ans. Les communautés microbiennes des petits écosystèmes d'eau douce présentent une forte saisonnalité, c'est-à-dire un patron intra-annuel, les plus grandes différences étant observées entre les communautés d'hiver et d'été. Le patron inter-annuel d'augmentation des dissimilarités au cours du temps entre les communautés a également été observé dans tous les écosystèmes, à l'exception des eucaryotes microbiens de l'écosystème Etang des Vallées. Cette absence de changements globaux au cours du temps dans cet écosystème peut s'expliquer par sa taille. C'est en effet le plus grand écosystème échantillonné dans cette étude, et on peut raisonnablement faire l'hypothèse qu'il présente un effet tampon plus fort que les autres, donc des communautés microbiennes plus stables. A l'échelle des OTUs, j'ai pu observer des patrons de récurrence saisonnière pour moins de 2% des OTUs. Les OTUs, ainsi que les groupes auxquels elles sont affiliées et leur abondance étaient variables selon les écosystèmes.

Enfin, j'ai pu observer que les paramètres physico-chimiques mesurés n'avaient de nouveau qu'une faible influence sur la variance de la communauté microbienne. Cela suggère que les facteurs biotiques, tels que la prédation, les mutualismes ou le parasitisme jouent un rôle majeur dans la composition des communautés. J'ai donc construit un réseau de co-occurrence basé sur les OTUs les plus abondantes détectées dans les cinq écosystèmes. Les interactions inférées montrent que les OTUs sont regroupées par même groupes taxonomiques, ce qui suggère qu'il existe un partage des OTUs par niche dans les écosystèmes.

## Etude de la topologie des réseaux de co-occurrence saisonniers dans des petits écosystèmes d'eau douce

L'étude des communautés microbiennes dans cinq petits écosystèmes d'eau douce à différentes fréquences et durées m'ont permis de découvrir une grande diversité de microorganismes présentant des patrons intra-annuels marqués. Sachant que les paramètres physico-chimiques n'étaient pas suffisants pour expliquer la variabilité des communautés microbiennes, j'ai décidé de me concentrer sur les paramètres biotiques, c'est-à-dire sur les interactions entre les microorganismes. J'ai donc inféré des réseaux de co-occurrence pour chaque écosystème à toutes les saisons, basés sur les microorganismes les plus abondants (c'est-à-dire ceux dont l'abondance relative est supérieure à 0,05% et qui sont détectés dans plus de 70% des échantillons). J'ai suivi l'approche développée par Chiquet et al., 2019, PLNmodels. Afin de déterminer la structure des réseaux de co-occurrence dans chaque écosystème pendant les quatre saisons, je me suis intéressée aux indices topologiques caractérisant les réseaux (nombre de noeuds, nombre de liens, connectance, etc.). La comparaison des indices entre les différentes saisons et les différents écosystèmes a révélé des similitudes dans la structuration des communautés microbiennes. Notamment, les réseaux obtenus pour l'été et le printemps sont majoritairement caractérisés par une plus grande taille, avec plus de noeuds, et avec une proportion plus élevée d'interactions positives. A l'inverse, comparés aux autres saisons, les réseaux d'hiver présentent plus d'interactions impliquant des parasites potentiels. Ces différences ont pu être testés grâce à des modèles statistiques. J'ai également pu observer de nombreuses interactions entre producteurs primaires et ciliés (prédateurs) dans les réseaux, ce qui valide écologiquement notre méthode. Il existe dans dans les écosystèmes des similitudes dans la structuration des communautés au cours de l'année, et ce malgré d'importantes différences en terme de composition des communautés microbiennes. Cette étude exploratoire donne par conséquent un premier aperçu de la structuration saisonnière des communautés microbiennes dans de petits écosystèmes d'eau douce. Elle



introduit également une méthodologie de comparaison de réseaux d'interactions et souligne l'importance de certains groupes taxonomiques dans le fonctionnement des écosystèmes étudiés.

## Conclusions

L'analyse menée dans le lac Baïkal a permis d'établir une image de la diversité et de la composition des communautés microbiennes eucaryotes sur l'ensemble du lac, ce qui n'avait jamais été fait auparavant. Cependant, cette image est uniquement celle des micro-organismes de la période estivale. Il faudrait pouvoir réaliser une analyse comparative dans le temps, aux différentes saisons et sur plusieurs années pour avoir une représentation complète des microorganismes eucaryotes dans le lac Baïkal. La caractérisation réalisée ici constitue un référent unique pour le suivi temporel de cet écosystème dans le contexte du changement climatique qui affecte particulièrement la Sibérie et offre des pistes pour savoir quels facteurs environnementaux vont impacter de manière plus sévère les microorganismes planctoniques qui s'y développent. Les analyses menées sur les petits écosystèmes d'eau douce ont permis de confirmer l'importance écologique de ces systèmes le plus souvent négligés dans les études. Ils hébergent des communautés microbiennes très diverses et très dynamiques dans le temps. Les microorganismes sont des acteurs clés de ces écosystèmes, et il est essentiel de mieux comprendre leur composition et leur structuration dans le temps. Par conséquent, ce travail peut contribuer à améliorer la gestion de ces écosystèmes communs afin de mieux préserver la biodiversité.

# Curriculum Vitae

## Publications

### Published peer-reviewed articles

**David G.M.**, Staentzel C., Schlumberger O., Perrot-Minnot M-J., Beisel J-N., Hardion L. A minimalist macroparasite diversity in the round goby of the Upper Rhine reduced to an exotic acanthocephalan lineage. *Parasitology*, 2017; <https://doi.org/10.1017/S0031182017002177>

### Submitted articles

**David G.M.**, Moreira D., Annenkova N.V. ,Reboul G., Deschamps P., Bertolino P., Jardillier L., López-García P. Deciphering eukaryotic microbial diversity along a horizontal and vertical spatial survey in Lake Baikal. *Submitted in Environmental Microbiology, published on bioRxiv* <https://doi.org/10.1101/2020.09.26.308536>

**David G.M.**, López-García P., Moreira D., Alric B., Deschamps P., Bertolino P., Restoux G., Rochelle-Newall E., Thebault E., Simon M., Jardillier L. Small freshwater ecosystems with dissimilar microbial communities exhibit similar temporal patterns. *Submitted in Molecular Ecology*

### Articles in preparation

**David G.M.**, Jardillier L., Moreira D., Deschamps, P. Restoux G., Rochelle-Newall E., Bertolino P., Simon M., López-García P. Dissimilar protists and prokaryotes communities display similar

---

intra- and inter-annual temporal patterns in small freshwater ecosystems

**David G.M.**, López-García P., Jardillier L., Moreira D., Thébault E. Topology of seasonal co-occurrence networks reveals similar community structuration in different small freshwater ecosystems

## Conferences

### Oral presentations

**David G.M.**, López-García P., Deschamps P., Moreira D., Simon M., Restoux G., Alric B., Bertolino P., Jardillier L. (Janvier 2019). Fairly common but mostly unexplored: small freshwater ecosystems harbour microbes with seasonal patterns combined to global community changes. IDEEV Days. Gif-sur-Yvette (France)

David G.M., López-García P., Deschamps P., Moreira D., Simon M., Restoux G., Alric B., Bertolino P., **Jardillier L.** (Janvier 2019). Disentangling microbial diversity and ecology in small freshwater ecosystems. Colloque RESOMAR. Wimereux (France)

**David G.M.**, López-García P., Deschamps P., Moreira D., Simon M., Restoux G., Alric B., Bertolino P., Jardillier L. (Octobre 2017). Communautés de procaryotes de petits écosystèmes d'eau douce : structure et dynamique temporelle complexes. VIIIe Colloque de l'Association Francophone d'Ecologie Microbienne. Camaret-sur-mer (France)

### Posters

**David G.M.**, López-García P., Alric B., Restoux G., Deschamps P., Moreira D., Simon M., Bertolino P., Jardillier L. (September 2019). Similar temporal patterns of microbial communities at the intra-annual and inter-annual scales in five small freshwater ecosystems. SAME16. Potsdam (Germany)

**David G.M.**, Jardillier L., Deschamps P., Moreira D., Alric B., Restoux G., Simon M., Bertolino P., López-García P. (Mars 2019). Fairly common but mostly unexplored: small freshwater ecosystems harbor microbes with seasonal patterns combined to a global community change. Paris-Saclay MICROBES day. Gif-sur-Yvette (France)

**David G.M.**, López-García P., Deschamps P., Moreira D., Simon M., Restoux G., Alric B., Bertolino P., Jardillier L. (Août 2018). Temporal dynamics and spatial distribution of prokaryotes and microbial eukaryotes in shallow temperate freshwater systems. ISME17. Leipzig (Germany)

## Teaching

### Monitorat

Total number of hours teaching at Paris-Saclay University: 213 hours

#### **2016 to 2017, 21 hours in total**

Practical work for third year of Licence students. Title: Biotic indices (21 hours, in French)

#### **2017 to 2019, 192 hours in total**

Practical work for third year Licence students. Title: Earth's energy budget (27 hours per year, in French)

Practical work for third year Licence students. Title: Hydrology (25 hours per year, in French)

In-class tutorials for second year Licence students. Title: Statistics applied to ecology (12 hours per year, in French)

## Internship

**2019** Supervision of a volunteer internship (third year of Licence undergraduate student, 4 months) - *Découverte des techniques de biologie moléculaire et des méthodes d'analyses de données biologiques*

## Sampling

**October 2016 to April 2019** Seasonal sampling of five small freshwater ecosystems of the “Parc Naturel Régional de la Haute Vallée de Chevreuse”. The studied ecosystems and the sampling procedure are described in Chapter 2).

**August 2019, 3 weeks** Sampling of ponds and lakes around Syrdakh (Yakoutia, Russia). The Central Yakutia region is an area where a strong thawing of the permafrost is observed, resulting from climatic and anthropogenic drivers. A group of researchers, conducted by Dr. A. Séjourné, monitor this area since several years, in order to investigate the regional changes in terms of hydrological, thermal, and geomorphological characteristics. In particular, they investigate the release of the dissolved carbon from ponds and lakes that are created by the thawing. These freshwater ecosystems present various ages of formations, that can be determined with the help of satellite images. As microorganisms are key players in the functioning of aquatic ecosystems and in particular in the carbon cycle, knowing the microbial community composition of these particular freshwater ecosystems is important. The aim of this sampling campaign was to collect for the first time samples to decipher the microbial community composition in several freshwater ecosystems following different ages of formation.

## Others

**January 2017 to March 2018** Non-permanent representative of the laboratory ESE (Ecologie, Systématique et Evolution)



**Titre:** Structuration spatio-temporelle des communautés microbiennes dans les écosystèmes d'eau douce  
**Mots clés:** Microorganismes, Ecologie, Diversité, Co-occurrence, Fonctionnement écosystèmes, Statistiques

**Résumé:** Les microorganismes constituent la forme de vie la plus abondante et diverse sur Terre et ils présentent une grande diversité phylogénétique et métabolique. Ils sont donc impliqués dans les cycles biogéochimiques et les réseaux trophiques, ce qui en fait des acteurs clés du fonctionnement des écosystèmes. Pour décrypter l'écologie des microorganismes, il est essentiel de prendre en compte les échelles spatiales, temporelles et taxonomiques. Bien que des paramètres abiotiques et biotiques aient été identifiés comme influençant la composition des communautés microbiennes dans les écosystèmes aquatiques (e.g. la température, la prédation), des études complémentaires sont nécessaires pour mieux comprendre la structure des communautés microbiennes. Cependant, l'étude des interactions biotiques entre microorganismes est difficile en raison de leur petite taille, de la grande diversité et du peu d'individus cultivés. Cette thèse de doctorat vise à décrire la diversité microbienne au sein de deux types d'écosystèmes d'eau douce encore peu étudiés, et à identifier les facteurs qui déterminent la composition de leurs communautés microbiennes. La première partie de cette thèse vise à décrire la distribution spatiale (horizontale et verticale) des protistes planctoniques du lac Baïkal (Sibérie, Russie). Nous nous sommes intéressés à des échantillons collectés en été 2017 le long d'un transect de ~600 km couvrant les trois bassins du lac, de la surface aux profondeurs (~1500 m) et du littoral au pélagique. Les trois autres parties présentent une étude de huit ans de la composition et de la dynamique temporelle des communautés microbiennes des trois domaines du vivant, à la surface de cinq petits écosystèmes d'eau douce (sud-ouest de Paris, France). Les échantillons ont été collectés à deux fréquences, mensuelle (2011-2013) et saisonnière (2011-2019). Les communautés planctoniques ont été caractérisées par le séquençage des gènes ARNr 16S

et 18S. Dans tous les écosystèmes, les communautés microbiennes sont très diverses, couvrant tous les super-groupes eucaryotes et procaryotes connus. Elles incluent des lignées typiquement marines (e.g. diploméide, MAST), ce qui suggère que la frontière entre le marin et l'eau douce pourrait être plus fine que prévu. Des taxons encore peu connus ont aussi été détectés, tels que des bactéries du Candidate Phyla Radiation. Des analyses multivariées ont montré que seule une faible fraction de la variance des communautés peut être expliquée par les paramètres abiotiques étudiés. Pour les variations spatiales, nous avons constaté une faible variabilité des communautés du lac Baïkal dans les différents bassins, mais avec une forte stratification le long de la colonne d'eau. La profondeur, qui traduit les variations environnementales (e.g. la lumière) dans la colonne d'eau, semble influencer significativement les communautés. Les petits écosystèmes abritent différentes communautés microbiennes malgré leur proximité géographique. Pour les variations temporelles, deux dynamiques ont été identifiées. À l'échelle intra-annuelle, les communautés sont caractérisées par une forte saisonnalité. Cependant, moins de 2% des unités taxonomiques opérationnelles présentent une récurrence saisonnière. Cela suggère que les écosystèmes ont un fonctionnement saisonnier, malgré des dynamiques individuelles imprévisibles. À l'échelle interannuelle, les communautés microbiennes sont de plus en plus différentes au cours des huit années, indiquant des changements continus dans leur composition. Enfin, l'inférence des interactions microbiennes grâce aux réseaux de cooccurrence reflète les variations spatio-temporelles précédemment observées. En effet, les communautés sont plus complexes à la surface du lac Baïkal qu'en profondeur. De plus, les petits écosystèmes partagent des topologies similaires pour chaque saison. Cela souligne l'importance des interactions écologiques chez les communautés microbiennes, dans l'espace et le temps.

**Title:** Spatio-temporal structuring of microbial communities in freshwater ecosystems  
**Keywords:** Microorganisms, Ecology, Diversity, Co-occurrence, Ecosystem functioning, Statistics

**Abstract:** Microorganisms are the most abundant and diverse forms of life on Earth and are characterized by high phylogenetic and metabolic diversities. They are thus involved in biogeochemical cycles and trophic webs, which make them key players in ecosystem functioning. To decipher the ecology of microorganisms, it is crucial to include spatial, temporal and taxonomic scales. Although several abiotic and biotic parameters have been identified as drivers of microbial community composition in aquatic ecosystems (e.g. temperature, orthophosphate concentration, predation, symbiosis), more investigations are needed to better understand how microbial community structure is shaped. However, investigating biotic interactions involving microbes is challenging because of microbial features (e.g. small size, high diversity, low cultivation efficiency). This PhD thesis aims at describing the microbial diversity inside two overlooked types of freshwater ecosystems and at identifying the factors driving microbial community composition. The first section of this thesis aims at comprehensively describing the spatial distribution (horizontal and vertical) of planktonic microbial eukaryotes in Lake Baikal (Siberia, Russia). We focus on samples collected in summer 2017 along a transect of ~600 km across the three basins of the lake, from the surface to the deepest areas (~1500 m) and from littoral to open waters. The three other sections present an eight-year investigation of the composition and temporal dynamics of microbial communities belonging to the three domains of life at the surface of five small freshwater ecosystems (located in the South West of Paris, France). Samples were collected at two different frequencies, monthly (2011-2013) and seasonally (2011-2019). The composition of planktonic communities was assessed by the sequencing of the phylogenetic marker genes 16S and 18S rRNA. In all the ecosystems studied, the microbial communi-

ties were diverse, covering all eukaryotic and prokaryotic supergroups. Moreover, they included typically marine lineages, especially in Lake Baikal, (e.g. diplomid, MAST) which suggested that the frontiers between marine and freshwater systems may be thinner than previously thought. They also included taxa that remain enigmatic, such as bacteria of the Candidate Phyla Radiation. Multivariate analysis showed that only a low fraction of the variance can be explained by the measured physico-chemical parameters. In terms of spatial variations, there was a weak variability of communities in Lake Baikal in summer across sampling basins, but a strong stratification along the water column. Depth, which is a proxy and a summary of the variations of the environmental conditions (e.g. light) along the water column, appeared to be a major driver of community composition. The small freshwater ecosystems harbored different microbial communities despite their geographic proximity. In terms of temporal variations, two types of patterns were detected. At the intra-annual scale, global communities were characterized by a strong seasonality. However, at the Operational Taxonomic Unit level, less than 2% of the community were characterized by recurrent seasonal patterns. This suggests that ecosystems have a yearly seasonal functioning, despite the presence of some unpredictable microbial dynamics. At the inter-annual scale, microbial communities experienced an increase of dissimilarities over the eight years, indicating turnovers in community composition. Finally, the structure of the communities studied through co-occurrence network inference reflected the spatio-temporal variations previously observed. Indeed, communities were more connected at the surface of Lake Baikal compared to the bottom. Moreover, ecosystems shared similar structural properties at each season. This underlines the importance of ecological interactions in the composition of microbial community over space and time.