

Universidade Federal de São Carlos

Centro de Ciências Biológicas e da Saúde

Programa de Pós-Graduação em Ecologia e
Recursos Naturais

Erick Mateus Barros

**Macroecologia Microbiana: dispersão bacteriana em lagos rasos
distribuídos pelo estado de São Paulo**

São Carlos
2018

Universidade Federal de São Carlos
Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Ecologia e
Recursos Naturais

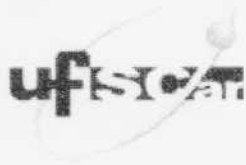
Erick Mateus Barros

**Macroecologia Microbiana: dispersão bacteriana em lagos rasos
distribuídos pelo estado de São Paulo**

Dissertação apresentada ao Programa de Pós-Graduação em Ecologia e Recursos Naturais da Universidade Federal de São Carlos, como parte dos requisitos para obtenção do título de MESTRE em ECOLOGIA, área de concentração: Ecologia e Recursos Naturais.

Orientador: Prof. Dr. Hugo Sarmiento

São Carlos
2018



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Ecologia e Recursos Naturais

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Erick Mateus Barros, realizada em 01/03/2018:

A handwritten signature in black ink, appearing to read 'Hugo Sarmento', is written above a horizontal line.

Prof. Dr. Hugo Miguel Preto de Moraes Sarmento
UFSCar

A handwritten signature in black ink, appearing to read 'Ricardo Koroiva', is written above a horizontal line.

Prof. Dr. Ricardo Koroiva
UFSCar

A handwritten signature in black ink, appearing to read 'Tadeu de Siqueira Barros', is written above a horizontal line.

Prof. Dr. Tadeu de Siqueira Barros
UNESP

à Sophia.

Agradecimentos

Aos meus pais (Selma e Messias), cujo empenho e dedicação inabaláveis em minha formação foram os principais responsáveis por tornar possível que eu chegasse até aqui.

Ao Prof. Dr. Hugo Sarmiento, pela orientação, a paciência em esperar por resultados, os ensinamentos e discussões e por acreditar no meu potencial.

Ao Prof. Dr. Adriano Caliman, meu co-orientador, por seu auxílio intelectual e por ter me ajudado a manter meu foco quando eu não sabia qual caminho seguir. Literalmente, minha tábua de salvação em um mar de dados.

À Profa. Dra. Inessa Bagatini por ter me auxiliado, tirado dúvidas e resolvido pessoalmente alguns dos mais cascudos problemas enfrentados durante todas as etapas do processo.

Aos membros das bancas de qualificação e defesa, por sua disponibilidade em participar do presente trabalho, contribuindo com dicas e sugestões para a sua melhoria.

À Gabrielle, minha companheira de todos os momentos, e entusiasta em nossas discussões filosófico-político-biológicas.

Ao Maurílio, pelas dezenas de discussões que permearam estes 8 anos e me permitiram expandir meus horizontes de conhecimento. Espero ter produzido o mesmo efeito em sua formação.

Ao Vinicius, grande amigo, companheiro de mestrado durante meu primeiro ano, e a pessoa que me convenceu a seguir o caminho da Ecologia Microbiana.

Ao Prof. Dr. Armando Vieira e à Profa. Dra. Inessa Bagatini, que abriram as portas de seu laboratório para a realização da maioria das etapas do meu mestrado, além de fornecerem os materiais coletados por eles e usados para a obtenção dos dados do presente trabalho; e aos demais membros do laboratório (em especial a Guilherme, Helena, Ingritt, Letícia, Naiara, Thaís e aos técnicos de campo (Luiz A. Joaquim) e laboratório (Antônio Luiz Sartori)) que auxiliaram tirando muitas dúvidas e realizando coletas e experimentos.

Ao PPGERN, capaz de tornar esta etapa ainda mais desafiadora.

Ao pessoal do Laboratório Multiusuário Centralizado para Sequenciamento de DNA em Larga Escala – LMSeq, localizado na UNESP – Jaboticabal, onde foi realizado o sequenciamento de minhas amostras. Em especial à Dra. Camila Fernandes e ao Dr. Luciano Kishi.

À FAPESP (processos 2014/14139-3 e 2011/50054-4) pelo financiamento.

À CAPES pela bolsa de estudos.

Aos companheiros do Laboratório de Processos Microbianos e Biodiversidade, Aylan, Davi, Mariana, Michaela, Pedro, Roberta e Vinícius, por todas as discussões de artigos, ideias adquiridas e sofrimento compartilhado durante essa fase corrida e cheia de pressão.

A todos os outros companheiros de caminhada que, apesar de não citados nominalmente, participaram ativamente de mais esta etapa da minha vida.

Aos meus pais por serem meus grandes fãs e cuja torcida é meu porto-seguro e minha motivação para seguir em frente com meus sonhos desde o início até o fim.

“O que um analfabeto em ecologia não percebe em relação a um ecossistema’, dizia Kynes, ‘é que se trata de um sistema. Um sistema! Um sistema mantém certa estabilidade fluida que pode ser destruída por um deslize em apenas um nicho. Um sistema tem ordem, uma correnteza que flui de um ponto a outro. Se algo represar a correnteza, a ordem desmoronará. Os inexperientes talvez só percebam esse desmoronamento quando já for tarde demais. É por isso que a função mais elevada da ecologia é a compreensão das consequências”’.

Trecho retirado do livro “Duna”, de Frank Herbert

Resumo

Estudos em ecologia microbiana vêm avançando rapidamente nas últimas décadas, em parte por influencia de avanços tecnológicos que nos permitiram perceber uma diversidade microbiana muito maior que se esperava. É cada vez mais evidente que padrões ecológicos inicialmente descritos para macro-organismos também podem ser observados em micro-organismos, fato que aproxima o estudo destes últimos à aplicação de teorias e modelos já consagrados pela ecologia clássica. Mesmo assim, a contribuição de microrganismos para o avanço de teorias ecológicas está ainda na sua infância. A presente dissertação tem como objetivo estudar os padrões de distribuição espacial de organismos procariotos de dezenas de pequenos lagos inseridos em uma região tropical com um perímetro superior a 3000 km. Esta dissertação está dividida em quatro capítulos. Primeiro foco no contexto histórico dos estudos ecológicos relacionados a atributos espaciais de distribuição de espécies e servindo de base para a criação de novos paradigmas como o das metapopulações e das metacomunidades. Além disso, discuti os avanços dos estudos microbiológicos em convergência com estes novos paradigmas e como tecnologias moleculares contribuíram para aproximar a ecologia microbiana da ecologia geral. O segundo capítulo visa apresentar de forma mais detalhada os métodos utilizados, além de problemas enfrentados e suas soluções, com o objetivo de facilitar futuros estudos onde se necessite aplicar as mesmas técnicas de extração, sequenciamento e processamento de dados (bioinformática). No terceiro capítulo foram aplicadas duas técnicas de extração bastante conhecidas para que pudéssemos avaliar sua influência em índices comumente utilizados em estudos ecológicos. O último capítulo demonstra a presença de uma distribuição bimodal de distribuição de frequências de ocupação através da análise de 60 lagoas próximas a nascentes distribuídas por uma macrorregião, abarcando todo o Estado de São Paulo, padrão já largamente demonstrado para macro-organismos, mas que ainda não havia sido testado em bactérias de água doce. Além disso, foram avaliados os grupos bacterianos mais abundantes dentre os organismos ‘*core*’ e os mais diversos dentre os ‘*satellite*’ e sua potencial relevância nas comunidades em que estão inseridas. Este estudo constitui uma contribuição da ecologia microbiana para a ecologia geral, visando uma plena compreensão dos processos que dirigem a composição das comunidades microbianas de determinado ambiente.

Palavras-chave: Macroecologia; Ecologia Microbiana; Influências do Método; Fenol:Clorofórmio; Kit de Extração Power Soil; Distribuição de Frequências de Ocupação; Hipótese da Bimodalidade.

Abstract

Studies in microbial ecology have had rapid advances in the last decades, in part due to the influence of technological innovations that allowed us to perceive a much greater microbial diversity than was previously found. There are increasing evidences that ecological patterns previously attributed only to macro-organisms can also be observed in microorganisms, a fact that brings the study of microorganisms closer to the application of theories and models already consecrated in classical ecology. Still, the contribution of microorganisms to theoretical ecology is still in its infancy. The present dissertation aims to study the patterns of spatial distribution in prokaryotic organisms of dozens of small lakes inserted in a tropical region in a perimeter of more than 3000 km. This thesis is divided in four chapters, first focusing on the historical context of ecological studies related to spatial attributes of species distribution, which was the basis for the creation of new paradigms such as metapopulations and metacommunities, and discussed the advances of microbiological studies in convergence with these new paradigms and how molecular technologies have contributed to bring microbial ecology closer to general ecology. The second chapter aims to present in a more detailed view on the methods used, problems faced and their solutions, in order to facilitate future studies where the same techniques of extraction, sequencing and data processing (bioinformatics) need to be applied. In the third chapter two well-known extraction methods were tested to evaluate their influence in indices commonly used in ecological studies. The last chapter demonstrates the presence of a bimodal distribution of occupancy-frequency distribution through the analysis of 60 shallow lakes near springs distributed in a large-scale matrix covering the entire state of São Paulo, a pattern already widely demonstrated for macro-organisms, but not yet tested in freshwater bacteria. In addition, the most abundant bacterial groups among the 'core' organisms and the most diverse among the 'satellite' and their potential relevance for the communities in which they were inserted were evaluated. This study is a contribution of microbial ecology to general ecology, aiming at a full understanding of the processes that drive the composition of microbial communities in a given environment.

Key-words: Macroecology; Microbial Ecology; Method Influences; Phenol:Chloroform; Power Soil extraction kit; Frequency-Occupancy Distribution; Bimodality Hypothesis.

Lista de Figuras

- Figura 1 - Modelo de Biogeografia de ilhas de MacArthur e Wilson. Em uma ilha vazia, o número inicial de imigração é alto e vai diminuindo com o tempo, a diminuição dos nichos disponíveis forçará uma queda na taxa de imigração e um aumento da extinção de espécies. O ponto em que as retas representando estas taxas de imigração e extinção se cruzam (a) é o ponto de equilíbrio do ambiente analisado. As ilhas podem ser grandes ou pequenas, e estarem próximas ou distantes do continente (ou a fonte de migrantes) (b) ilhas pequenas e distantes apresentam o menor ponto de equilíbrio. Ilhas grandes e distantes e pequenas e próximas apresentam equilíbrios com valores intermediários e ilhas grandes e próximas ao continente são as que têm o ponto de equilíbrio suportando as maiores taxas de diversidade local. (Fonte: MacArthur e Wilson, 1967). 27
- Figura 2 - Modelo de Metapopulações de Levins. Círculos preenchidos representam fragmentos ocupados pelo organismo em questão. Setas em linha reta representam eventos de colonização. Setas voltando para o mesmo fragmento representam extinção da população no fragmento (Fonte: Nee, 2007). 28
- Figura 3 - Representação do modelo de metapopulação de fonte-dreno. Áreas em cinza representam fragmentos “fonte”. Áreas brancas “drenos”. Setas representam migração. Note que a situação de fonte ou dreno pode não estar associada ao tamanho do fragmento, fragmentos pequenos podem atuar como fonte assim como grandes podem ser drenos. 29
- Figura 4** - Os quatro paradigmas das metacomunidades: Dinâmicas de fragmentos (Patch Dynamics) (a), Seleção de espécies (Species Sorting) (b), Efeito de massa (Mass Effect) (c) e Perspectiva neutra (Neutral perspective) (d). (Fonte: Leibold *et al.*, 2004). 31
- Figura 5 – Representação gráfica do Estado de São Paulo onde ocorreu a amostragem. Os 60 pontos escolhidos foram divididos em quatro quadrantes: Serra do Mar, Paranapanema, Tietê e Grande, nomes relacionados aos rios ou formação geográfica mais importante de cada região. As representações menores indicam a formação vegetal (A) e hidrológica (B) do Estado. 41
- Figure 6 – Location of the 15 shallow lakes samples used in this study. 65
- Figure 7** – A) Comparison of all OTU (Operational Taxonomic Units) found in 15 freshwater samples using two DNA extraction methods. The X-axis represents the kit extraction and Y-axis the organic extraction DNA . B) Boxplots represent the number of points overrepresented using the kit extraction (Bottom-right panel from Fig.1A) and the organic extraction method (Top-left panel from Fig.1A).The central full line indicates the median value, the dotted line indicates the arithmetic mean value, the boxes indicate the lower and upper quartiles, the vertical lines indicate the 10th and 90th percentiles, and the dots represent the 5th and 95th percentiles. 67

Figure 8 - Comparison of the sum of OTU (Operational Taxonomic Units) relative abundance grouped at class level, obtained by two DNA extraction methods. Highlighted points represent classes that were significantly different ($p < 0.05$, paired t-test) in their relative abundance between extraction methods.	67
Figure 9 - Comparison of commonly used indexes in community ecology: OTU (Operational Taxonomic Units) richness, Shannon-Weaver diversity and Pielou evenness for each DNA extraction method. No significant differences were observed by paired t-test comparisons. The central full line indicates the median value, the dotted line indicates the arithmetic mean value, the boxes indicate the lower and upper quartiles, the vertical lines indicate the 10th and 90th percentiles, and the dots represent the 5th and 95th percentiles.	69
Figure 10 – Location of the 60 sample sites, divided between 3 different sub-basins inserted into Paraná river basin (colored in Grey) and the adjacent coastal mountain formation. Samples were named T for Tietê river basin (green), P for Paranapanema river basin (blue), G for Grande river basin (Orange) and S for Serra do Mar formation (blue).	95
Figure 11 – Occupancy-frequency distributions for the total analyzed sites and for T (central green colored), P (southwest red colored), G (north orange colored) and S (Southwest blue colored) sub-regions. Significance: ‘****’ < 0.001 ; ‘***’ < 0.01 ; ‘*’ < 0.05	101
Figure 12 – Occupancy-frequency distributions concerning an increase of area range. Significance: ‘****’ < 0.001 ; ‘***’ < 0.01	102
Figure 13 - Venn’s plot of the core populations in the different sub-regions, with a list of the taxonomy of the OTUs present in all sites and those that occur exclusively in each sub-region. ‘*’ indicates the abundant core OTUs, ‘o’ indicates the rare core OTUs.	104
Figure 14 - Average core OTU abundance distribution in all samples (grey), in T (central green colored), in P (southwest red colored), in G (north orange colored) and in S (Southwest blue colored). A full detailed list of OTU’s taxonomy can be found in supplementary table S2.	109
Figure S 1 - Boxplot representing the percentage of bacterial classes overrepresentation which were significantly different ($p < 0.05$).	86

Lista de Tabelas

Table 1 - Mean, standard deviation (sd) and coefficient of Variation (cv) for community ecology commonly used indexes in samples extracted with the kit extraction and organic extraction.....	69
Table 2 – General features of the study area and environmental conditions in the shallow lakes studied (average values).	96
Table 3 – Sequencing dataset statistics and ecological indexes. ^(*) For Alpha, Gamma and Beta diversity, we used Simpson' index; ^(**) for evenness we used Hill' index. ^(***) Bimodality is tested against a Gaussian logarithmic curve and a significancy means that the occupancy-frequency distribution had not presented only one hump at one of these extremes.....	103
Table 4 – Distribution of core organisms' phyla among sub-basins.	108
Table S1 - Paired tests for bacterial phylum with p-values and α -values. Groups that shown normal distribution were performed by the Paired t-test, the non-normal data were performed by Wilcoxon Signed Rank Test.....	81
Table S2 - Paired tests for bacterial classes with p-values and α -values. Groups that shown normal distribution were performed by the Paired t-test, the non-normal data were performed by Wilcoxon Signed Rank Test. The significative values are represented in bold and marked with an “*”.....	82
Table S3 – Database containing basic information about the collection, geographic position and environmental analysis.....	120
Table S4 – The taxonomy of core organisms. The “X” indicates region where a given OTU was identified as core, if it is present in all regions, it is marked in the 4 columns representing each sub-region and in the column named as "Total Core". The taxonomy was separated by each level, some OTUs cannot be identified until species, in this case the spaces referring to a certain taxonomic level and the following ones remain empty.	120
Table S5 – Diversity ranking for the satellite microorganisms, by all the samples and by each sub-region.....	120
Table S6 – Taxonomic table and total abundance data considering only the Parcubacteria Phylum (OBS.: Esta tabela de dados pode ser acessada apenas em sua forma digital. Se você teve acesso a esta informação por via impressa da dissertação, e necessitar destes, envie um e-mail para erickmbarros@gmail.com).....	120

Lista de Siglas e Abreviações

Ambiente

UGRHi – Unidades de Gerenciamento de Recursos Hídricos

Amostras

T – Amostras coletadas de lagos localizados na Bacia do rio Tietê

P – Amostras coletadas de lagos localizados na Bacia do rio Paranapanema e outros rios menores

G – Amostras coletadas de lagos localizados nos rios ao Norte da Bacia do rio Tietê, que desaguam no rio Grande

S – Amostras coletadas de lagos localizados formação da Serra do Mar

O – ambiente Oligotrófico

M – ambiente Mesotrófico

E – ambiente Eutrófico

Metodologia

DGGE (Denaturing Gradient Gel Electrophoresis) - Eletroforese em Gel de Gradiente Desnaturante

NGS (Next-Generation Sequencing) – Sequenciamento da Próxima Geração

PCR – Polimerase Chain Reaction – Reação em Cadeia da Polimerase

qPCR (quantitative PCR) – PCR quantitativa

OTU (Operational Taxonomic Unit) – Unidade Taxonômica Operacional

CTAB (cetyl trimethyl ammonium bromide) – brometo de cetil trimetil amônio

SDS (Sodium dodecyl sulphate) – Dodecil sulfato de sódio

EBM (Earth Microbiome Project) – iniciativa que visa produzir protocolos padronizados para técnicas moleculares. Para mais detalhes, visite o site <http://www.earthmicrobiome.org/>

DNA (Deoxyribonucleic Acid) – Ácido Desoxirribonucleico

RNA (Ribonucleic Acid) – Ácido Ribonucleico

rRNA (Ribosomal Ribonucleic Acid) – Ácido Ribonucleico Ribossomal

Institucional

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo

BIOTA – Programa FAPESP de Pesquisas em Caracterização, Conservação, Restauração e Uso Sustentável da Biodiversidade

CCMA-UFSCar – Coleção de Culturas de Microalgas de Água Doce da Universidade Federal de São Carlos

Sumário

Prólogo.....	19
Introdução: Dinâmicas espaciais e Microrganismos.....	24
Biogeografia.....	25
<i>Origens da biogeografia</i>	25
<i>Biogeografia de ilhas e o impacto das ideias de MacArthur e Wilson</i>	26
Ambientes Interligados	27
<i>Metapopulações: População de populações</i>	27
<i>Metacomunidades: Comunidades locais interligadas por migração</i>	29
<i>Os quatro paradigmas do estudo de metacomunidades</i>	30
<i>Partição da Variância</i>	31
<i>E os microrganismos?</i>	32
<i>O ambiente seleciona</i>	32
<i>Diversidade microbiana dependente da técnica aplicada</i>	33
<i>OTU – As “espécies” de microrganismos</i>	34
<i>Novos paradigmas da diversidade microbiana</i>	35
Métodos	38
Coleta	39
<i>Geografia e escolha dos pontos</i>	39
<i>Clorofila a</i>	40
DNA.....	42
<i>Extração</i>	42
<i>Amplificação por PCR</i>	44
<i>Montagem e quantificação das bibliotecas e sequenciamento</i>	45
Bioinformática	46

Referências.....	48
Comparison of two widely used DNA extraction methods in aquatic microbial ecology	61
Abstract.....	61
Main Text.....	63
Acknowledgments	70
Literature Cited	71
Supplementary material	75
<i>Study sites and sampling</i>	75
DNA extraction methods	76
<i>Extraction by PowerSoil DNA isolation Kit</i>	76
<i>Extraction by Phenol:Chloroform Method</i>	77
Amplification and Sequencing.....	77
Bioinformatics and statistics	78
Literature Cited	79
Testing the occupancy-frequency distribution in freshwater bacterial communities.....	88
Abstract.....	88
Introduction.....	90
Methods	93
<i>Study sites and sampling</i>	93
<i>DNA extraction and amplification</i>	96
<i>Bioinformatics and statistics</i>	97
Results.....	99
<i>Exploring the dataset</i>	99
<i>Testing the occupancy-frequency distribution</i>	99
Discussion.....	104

Acknowledgments	112
References.....	113
Supplementary Material.....	120
Considerações Finais	134
Referências.....	136

Prólogo

Quando terminei o ensino médio e resolvi prosseguir meus estudos, minhas notas mais elevadas em geografia e história pareciam indicar que eu havia feito uma escolha equivocada ao resolver seguir a vereda da biologia, e encarar um primeiro ano repleto de matemática, física e química, meus nêmeses desde tenra idade serviram apenas para reforçar a sensação, mas eu resisti, conclui minha graduação no final de 2015 e iniciei minha pós-graduação sem que houvesse pausa, e meu esforço foi recompensado (logo o leitor perceberá) com um mestrado repleto de história da ciência e biogeografia.

É preciso ser justo, o Mestrado não é um mar de rosas em contrapartida de uma graduação cheia de dor e sofrimento, como eu fiz crer na frase acima. A adaptação a um novo ritmo de trabalho é cansativa e perigosa, já que o tempo é curto demais e a pressão por

resultados pode surgir a qualquer momento de várias fontes, prováveis e improváveis, enquanto é preciso adaptar-se a um ritmo de leitura gigante em uma língua não nativa e com temas que não foram aprendidos durante a graduação (mas que deveriam ter sido...).

Talvez o livro que mais impactou minha formação profissional e intelectual (o que nesta profissão não deixam de ser a mesma coisa) durante estes últimos dois anos foi o clássico de Thomas Kuhn, “A Estrutura das Revoluções Científicas”, pois mostrou que a área que resolvi estudar encontra-se bem no meio deste tornado que é o nascimento de um novo paradigma científico e apresentando todos os sintomas diagnosticados por Kuhn em seu livro (até mesmo a resistência por parte dos cientistas envolvidos, em aceitar que esta é uma revolução científica).

Foi somente ao entender que a Ecologia passa por uma revolução que está deixando de lado ideias que consideram os ambientes como locais isolados, para passar a considerar os efeitos da migração de indivíduos para as dinâmicas de comunidades, que eu consegui reunir serenidade suficiente pra atravessar o mestrado sem muitos traumas. Afinal, pude entender que o conhecimento que me era apresentado, completamente desconhecido, o era por ser um conjunto de informações que ainda não havia sido apresentado a mim e me permitiu buscar informações anteriores visando uma perspectiva histórica dos motivos que levaram à construção deste novo conhecimento.

Esta perspectiva me permitiu abrir a mente para ideias novas como as aqui apresentadas e outras mais que deverão ser aplicadas no futuro, além de me fazer repensar toda minha trajetória desde o início de minha educação formal até agora, na pós-graduação. Eu pretendia escrever um texto com este tema, utilizando como base principalmente as ideias que “*A Estrutura*” me fez enxergar, mas não teve tempo para tal. Então isso ficará para um futuro, espero não tão distante.

Foi também com imenso prazer que recebi, logo no começo, a missão de analisar mapas de localização para decidir qual seriam as melhores amostras a serem utilizadas neste projeto dentre mais de 320 coletadas pelo Laboratório de Ficologia. Foi como voltar à infância, quando eu rabiscava as bordas de um (hoje velho) atlas escolar com o único objetivo de transcrevê-las para uma folha sulfite. Esta etapa lembrada como uma feliz fase da infância é a chave para a produção dos textos aqui presentes, pois me permitiu selecionar um grupo de ambientes adequados a diversos tipos distintos de estudos, desde a Biogeografia e Macroecologia até análises Filogenéticas e Metodológicas.

Nos capítulos seguintes apresentarei meu projeto dividido em quatro seções, dois destes capítulos são introdutórios e escritos em Português, visando apresentar uma visão geral e o estado da arte no que diz respeito aos avanços da ecologia microbiana, os outros dois capítulos, redigidos em Inglês, apresentam artigos científicos desenvolvidos durante meu mestrado e encontram-se em processo de revisão para futura publicação.

No primeiro capítulo, intitulado “Dinâmicas espaciais e os Microrganismos” fiz uma apresentação teórica em um contexto histórico, repassando as principais contribuições ecológicas do último século que desencadearam a criação de hipóteses que consideram questões geográficas de dispersão e isolamento. Apresentei também, os motivos de a microbiologia não ter participado ativamente do desenvolvimento destas teorias, mas agora avançar largamente seu conhecimento e em sua contribuição para a ecologia geral. Este capítulo foi idealizado durante a disciplina “Ecologia Teórica”, ministrada pelo meu orientador Hugo Sarmiento e contou com imensa colaboração de Luis Felipe Ventura de Almeida.

No segundo capítulo, intitulado “Métodos” apresentei de forma detalhada todas as etapas necessárias para a obtenção dos dados necessários à etapa criativa que se segue nos capítulos seguintes. Neste capítulo, visei apresentar também os problemas aos quais fui obrigado a enfrentar durante minha aprendizagem e quais foram as soluções encontradas por mim e pelas pessoas que me auxiliaram (leia-se “salvaram”).

O terceiro capítulo, também o primeiro resultado efetivo destes dois anos de estudos, chama-se “Comparison of two widely used DNA extraction methods in aquatic microbial ecology” e foi produzido em colaboração com Aylan K. Meneghini, Inessa Lacativa Bagatini, Camila C. Fernandes, Luciano T. Kishi, Armando A. H. Vieira e claro, meu orientador Hugo Sarmiento. Neste capítulo, realizamos uma comparação de dois métodos de extração: Fenol-Clorofórmio e PowerSoil DNA isolation Kit, visando avaliar a real influência de cada um em análises ecológicas e, utilizando para tal o material coletado em 15 das 60 amostras utilizadas ao longo dos trabalhos que compõem esta dissertação. Este trabalho foi submetido no formato de nota para a revista **Aquatic Microbial Ecology** e, após a primeira revisão por pares, está agora sendo corrigida pelos autores.

O quarto capítulo, “Testing the occupancy-frequency distribution in freshwater bacterial communities”, mostrou pela primeira vez que bactérias de água doce apresentam uma distribuição bimodal no espaço, aproveitamos também para discutir a importância geral

de phyla bacterianos mais importantes dentre os organismos “core” e “satellite”. Pretendemos assim, realizar os últimos ajustes deste texto e submetê-lo à revista **Ecology**.

Sem mais delongas, partimos então ao resultado destes dois últimos anos de trabalho. Obrigado a você que se propôs a ler esta dissertação, desejo-lhe uma agradável leitura.

Erick Mateus-Barros

Introdução: Dinâmicas espaciais e Microrganismos

A observação de que as inúmeras espécies de animais e plantas presentes no planeta não se distribuem de forma homogênea no espaço tem sido um dos temas que mais se destaca na pesquisa ecológica. Variações espaciais e temporais na distribuição de recursos importantes bem como processos geológicos e ecológicos podem resultar em uma heterogeneidade do espaço e conseqüentemente trazer profundos impactos nas populações e suas interações dentro das comunidades ecológicas (Collinge, 2010). Esta heterogeneidade espacial é um dos principais fatores estudados no campo da ecologia espacial.

A ecologia espacial se dedica a compreender como a estrutura da paisagem e a distribuição de indivíduos e populações afetam as dinâmicas ecológicas. Esta é uma área em

expansão, onde novos paradigmas foram apresentados nas últimas décadas, tais como a teoria neutra (Hubbell, 2001), das metacomunidades (Leibold *et al.*, 2004) e do mosaico geográfico de coevolução (Thompson, 2005).

Estas novas teorias muitas vezes misturam-se aos pressupostos das antigas teorias de biogeografia, o que está dissolvendo os limites, antes bem claros, que separavam os estudos de biogeografia e da ecologia clássica (Jenkins e Ricklefs, 2011; Ricklefs e Jenkins, 2011).

E na esteira desta revolução científica, o estudo da ecologia microbiana vem se desenvolvendo. Passando simultaneamente por duas revoluções: esta conceitual que vem modificando a forma de todo o estudo ecológico e mais uma revolução tecnológica. Com novas técnicas de sequenciamento de DNA que pode ser retirado diretamente de amostras ambientais (Taberlet *et al.*, 2012) e de processamento bioinformático de dados massivos via computadores de última geração e softwares desenvolvidos especificamente para dados biológicos (e.g. Logares *et al.*, 2014; R. Core Team, 2016), a grande quantidade de dados alcançados em cada amostra coletada possibilitou que o estudo da ecologia microbiana se firmasse como uma das áreas da ecologia capazes de testar a robustez dessas novas teorias.

A seguir, farei um pequeno resumo, a tentativa de um levantamento histórico que explica em partes o caminho trilhado até o presente, como chegamos ao ponto de conseguirmos misturar ecologia e biogeografia novamente e lançar mão de modernas tecnologias para tentar, de forma mais integrada, explicar os padrões da vida terrestre.

Biogeografia

Origens da biogeografia

No século XVIII, alguns cientistas da época participaram de viagens intercontinentais, coletando espécies e observando diferentes dinâmicas em ambientes temperados e tropicais. Um de seus mais conhecidos expoentes, Alexander von Humboldt foi um dos primeiros a imaginar hipóteses envolvendo as ciências biológicas e da terra, consideradas as bases da biogeografia.

O naturalista alemão se tornou mundialmente conhecido após suas viagens pela América Latina (1799-1804) e Ásia central (1829), e publicou um texto intitulado “Ensaio sobre a biogeografia de plantas” (Humboldt e Bonpland, 1805), que descrevia a variação na

distribuição de plantas conforme um gradiente de altitude, no monte Chimborazo, Equador. O texto, produzido como consequência de sua viagem à América Latina é o primeiro registro sobre a distribuição geográfica de espécies, e marca o início do estudo de biogeografia.

Desde sua origem, esta área de estudo mantém como sua característica mais marcante a interdisciplinaridade, abrangendo áreas de estudo como geografia, geologia, paleontologia e ecologia que são integradas à procura de novas respostas para as questões ambientais.

Biogeografia de ilhas e o impacto das ideias de MacArthur e Wilson

Um dos modelos matemáticos de maior sucesso nas ciências biológicas, a teoria da biogeografia de ilhas foi concebida para explicar o número de espécies observadas em ilhas de diferentes tamanhos (MacArthur e Wilson, 1967). O modelo de MacArthur e Wilson considera que há uma forte ligação entre o tamanho de uma ilha e o número de espécies suportado e que o balanço entre as taxas de imigração e extinção levariam o ambiente ao equilíbrio (veja a Fig. 1 para mais detalhes).

Durante muitas décadas, este modelo foi considerado como ideal para a biogeografia de ilhas, outros autores o corroboraram em seus experimentos, e os trabalhos que diziam o contrário eram frequentemente ignorados. Após tantos anos, já sabemos que a ideia de que existe um equilíbrio entre as taxas de imigração e extinção não se aplica a todas as ilhas do mundo, mas o decorrer dos anos acabou por tornar esta teoria transcendente ao campo de estudo inicialmente proposto.

A hipótese de que ilhas maiores e menos isoladas apresentam mais espécies que ilhas menores e mais isoladas é hoje aplicada a qualquer tipo de ambiente fragmentado. De topos de montanhas (e.g. Endo *et al.*, 2015) a oceanos (e.g. Milici *et al.*, 2016), passando por cavernas (e.g. Fernandes *et al.*, 2013), lagos e lagoas (e.g. Schiaffino *et al.*, 2011), e até por fragmentos florestais contidos em uma matriz de áreas degradadas (e.g. Wolfe *et al.*, 2015), o estudo da biogeografia de ilhas está vinculado a diversos campos da ecologia e estes são muito importantes para estudos vinculados à conservação de biodiversidade.

Conceitos muito utilizados atualmente tais como metapopulação e metacomunidade, só puderam existir após a popularização de formas de enxergar os ambientes não como locais isolados, mas que trocam indivíduos frequentemente. Esta ideia já era imaginada

anteriormente a MacArthur e Wilson, mas não podia ser amplamente testada, principalmente pela dificuldade de se montar um modelo matemático simples e aplicável que levasse em consideração o fluxo de organismos entre ambientes.

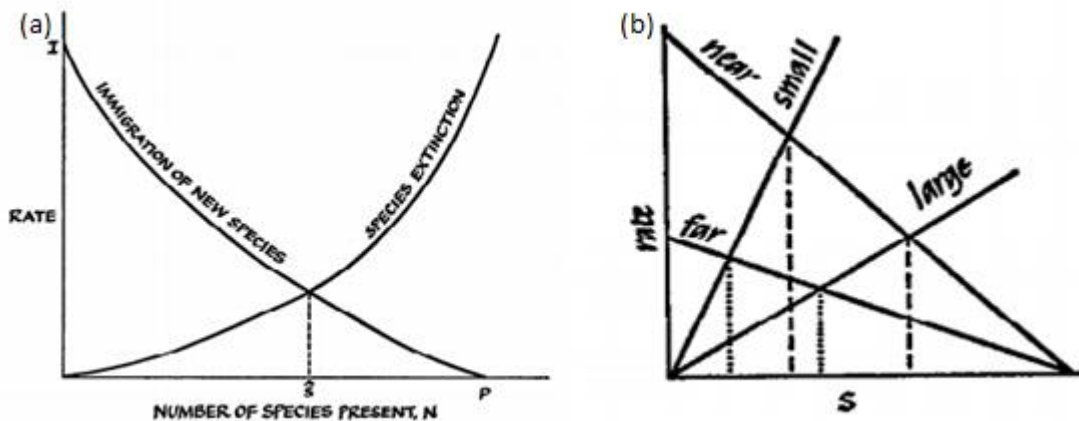


Figura 1 - Modelo de Biogeografia de ilhas de MacArthur e Wilson. Em uma ilha vazia, o número inicial de imigração é alto e vai diminuindo com o tempo, a diminuição dos nichos disponíveis forçará uma queda na taxa de imigração e um aumento da extinção de espécies. O ponto em que as retas representando estas taxas de imigração e extinção se cruzam (a) é o ponto de equilíbrio do ambiente analisado. As ilhas podem ser grandes ou pequenas, e estarem próximas ou distantes do continente (ou a fonte de migrantes) (b) ilhas pequenas e distantes apresentam o menor ponto de equilíbrio. Ilhas grandes e distantes e pequenas e próximas apresentam equilíbrios com valores intermediários e ilhas grandes e próximas ao continente são as que têm o ponto de equilíbrio suportando as maiores taxas de diversidade local. (Fonte: MacArthur e Wilson, 1967).

Ambientes Interligados

Metapopulações: População de populações

Segundo Levins (1970), metapopulações podem ser definidas como sendo uma “população de populações”, ou seja, um conjunto de populações locais que eventualmente se conectam através de migração, formando uma única população interligada por um fluxo de indivíduos. Este conceito pode ser de grande utilidade no estudo da dinâmica de organismos vivendo em habitats fragmentados, que estão assim sujeitos a processos de extinção local, migração entre populações locais e (re)colonização de fragmentos “vagos”, sendo que estes processos podem afetar diretamente a manutenção de uma espécie no longo prazo (Hanski e Thomas, 1994).

Diversos modelos têm sido criados de forma a tentar representar as dinâmicas espaciais em metapopulações, dentre estes, dois modelos de metapopulações podem ser apresentados como os mais relevantes, além de possuírem relativa simplicidade conceitual. São eles o modelo clássico (ou de Levins) e o de fonte-dreno (Nee, 2007).

O modelo clássico assume que todos os fragmentos possuem o mesmo tamanho, a mesma distância entre si, as mesmas taxas de extinção e colonização, além de possuírem dinâmicas independentes (Fig. 2). Apesar de este modelo ser pouco realista e improvável na natureza, ele pode ser utilizado como modelo nulo e base para a construção de modelos espacialmente mais realistas (Falke e Fausch, 2010).

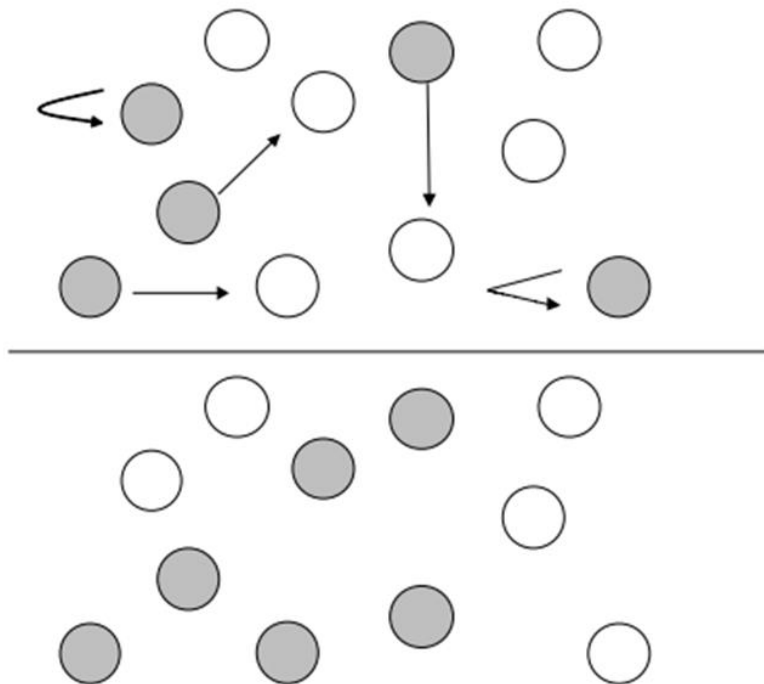


Figura 2 - Modelo de Metapopulações de Levins. Círculos preenchidos representam fragmentos ocupados pelo organismo em questão. Setas em linha reta representam eventos de colonização. Setas voltando para o mesmo fragmento representam extinção da população no fragmento (Fonte: Nee, 2007).

O modelo de fonte-dreno consiste em populações interligadas por dispersão e considera as diferenças de demografia (taxas de natalidade e mortalidade) entre as mesmas, porém sem levar em consideração a possibilidade de extinção local (Fig. 3). Assim, neste modelo populações em fragmentos de boa qualidade sustentam, por meio de migração, populações que se encontram em declínio (i.e. taxa de mortalidade maior que a natalidade) em habitats mais pobres (Nee, 2007).

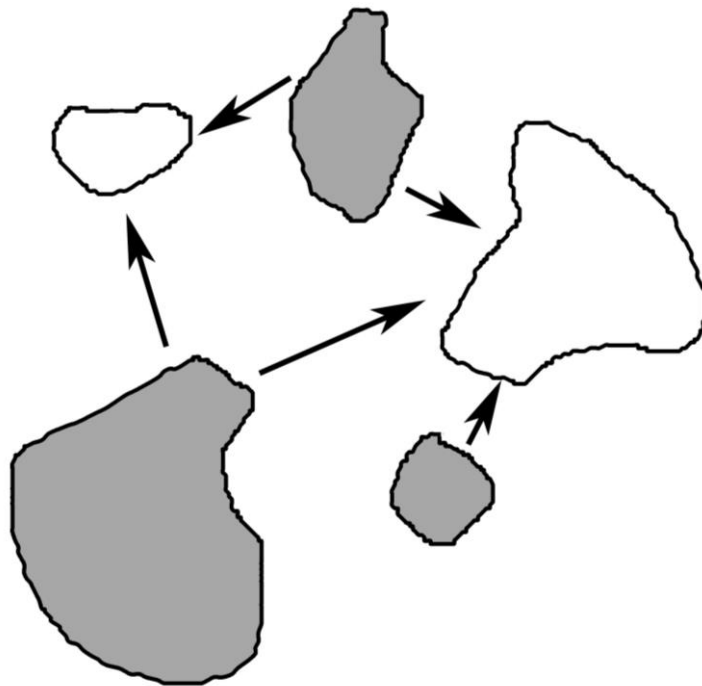


Figura 3 - Representação do modelo de metapopulação de fonte-dreno. Áreas em cinza representam fragmentos “fonte”. Áreas brancas “drenos”. Setas representam migração. Note que a situação de fonte ou dreno pode não estar associada ao tamanho do fragmento, fragmentos pequenos podem atuar como fonte assim como grandes podem ser drenos.

Metacomunidades: Comunidades locais interligadas por migração

Esse transplante das ideias da biogeografia de ilhas para outros ambientes, tentando explicar interações entre populações anteriormente consideradas isoladas foi mais tarde utilizada para o estudo das comunidades biológicas. Uma metacomunidade é definida como um grupo de comunidades locais interligadas por fluxo de indivíduos (Leibold *et al.*, 2004).

Em seu artigo de 1992, David Wilson produziu a primeira tentativa de criar modelos de distribuição para metacomunidades. Obviamente, o nível de complexidade do ambiente que é alcançado ao discutirmos comunidades torna mais difícil a criação de um modelo que explique satisfatoriamente todos os fenômenos da natureza, por isso outros modelos podem ser encontrados (e.g. Mouquet e Loreau, 2002).

De qualquer forma, experimentos em laboratório (Souffreau *et al.*, 2014; Fronhofer *et al.*, 2015) e em campo (Werner *et al.*, 2007; Ruiz-González *et al.*, 2015) para vários tipos distintos de organismos e ambientes, indicam que a ideia de que as comunidades são interligadas explica melhor o que é observado do que a perspectiva de que existam apenas comunidades isoladas.

Os quatro paradigmas do estudo de metacomunidades

São definidos quatro modelos que juntos explicam os processos que determinam os padrões de dispersão dos organismos em uma metacomunidade, explicando como ocorreriam os fluxos de indivíduos entre as comunidades locais. Estes serão brevemente explicados a seguir (Para mais detalhes ler: Leibold *et al.*, 2004; Hubert *et al.*, 2015).

A perspectiva de dinâmicas de fragmento (*Patch Dynamics*) leva em consideração que os fragmentos estudados são homogêneos, desta forma um ou poucos organismos são mais aptos a viver em todos os fragmentos e controlam os fluxos de migração. Os outros organismos, piores competidores, acabam migrando para os locais abandonados, o que gera um equilíbrio entre as taxas de competição e dispersão (Fig. 4a).

A perspectiva de seleção de espécies (*Species Sorting*) considera que os fragmentos são heterogêneos. Assim, cada organismo será melhor ou pior competidor dependendo do local onde está vivendo. Nestes casos, o resultado final é uma complexa dinâmica entre os fragmentos e indivíduos, com taxas de extinção e dispersão e níveis de diversidade, riqueza e dominância diferentes para cada fragmento que compõe a metacomunidade (Fig. 4b).

O efeito de massa (*Mass Effect*) ocorre quando as comunidades são muito abertas, e a ocorrência de altas taxas de dispersão impede que o ambiente selecione os organismos e o resultado final é uma diversidade regional homogênea, mesmo que a diversidade local das comunidades seja heterogênea (Fig. 4c).

A perspectiva neutra (*Neutral Perspective*) leva em consideração que as mudanças observadas no ambiente são estocásticas apenas. A tendência para este modelo é que as comunidades se tornem homogêneas, com tendência a grande dominância (Fig. 4d).

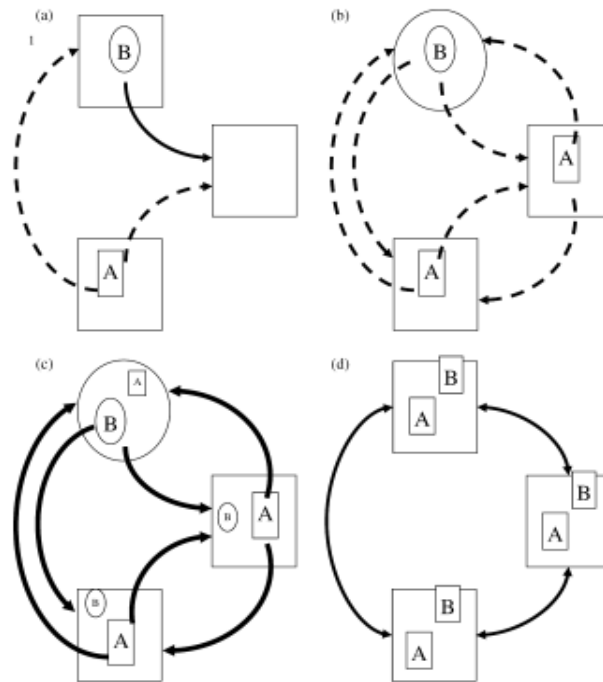


Figura 4 - Os quatro paradigmas das metacomunidades: Dinâmicas de fragmentos (Patch Dynamics) (a), Seleção de espécies (Species Sorting) (b), Efeito de massa (Mass Effect) (c) e Perspectiva neutra (Neutral perspective) (d). (Fonte: Leibold *et al.*, 2004).

Partição da Variância

A ideia de que seja possível analisar ambientes interconectados por fluxo de organismos despertou o interesse de pesquisadores do mundo inteiro, e o estudo das metacomunidades tende a crescer cada vez mais. No entanto, como era de se esperar, uma análise tão complexa acaba causando alguns problemas de entendimento (Cottenie, 2005; Brown *et al.*, 2016).

As quatro perspectivas acima listadas, também conhecidas como “big 4” (Brown *et al.*, 2016), possuem características que muitas vezes se repetem ou se misturam em outros caracteres que dificultam sua identificação e seu isolamento para estudos capazes de identificar as quatro perspectivas de forma plena em um único ambiente. Essa dificuldade

apresentada motivou que outros autores passassem a utilizar outra metodologia para determinar quais processos ecológicos estão sendo observados, metodologia esta conhecida como partição da variância.

A partição da variância tem em seu favor sua facilidade de aplicação. Nesta perspectiva, todos os caracteres que compõem o big 4 podem ser resumidos a apenas dois: espacial e ambiental, o que facilita sua aplicação em qualquer conjunto de dados (para mais informações sobre como fatores espaciais e ambientais se correlacionam com o "big 4" leia Cottenie, 2005).

Vale ressaltar que por espaço entende-se a distancia entre os diferentes ambientes analisados, assim, quanto mais distantes um do outro, maiores seriam as chances de os organismos não serem capazes de deslocar-se entre eles e que também aumentaria a dissimilaridade entre os diferentes ambientes e, portanto, a seleção destes organismos seria intensificada (Borcard *et al.*, 1992; Cottenie, 2005; Brown *et al.*, 2016).

Este método, graças a sua maior simplicidade de aplicação em relação ao *big 4* tornou-se bastante utilizado como um substituto. Tem, portanto sua importância histórica na formação das bases do conhecimento adquirido sobre as metacomunidades, mas começa a ser abandonada, em favor de análises que considerem uma separação não tão abrupta entre os componentes que explicam a distribuição das espécies.

E os microrganismos?

O ambiente seleciona

Uma das frases mais celebres do estudo de microrganismos talvez seja: “*Everything is everywhere, but the environment selects*” cunhada no original alemão por Baas Becking (1934), serviu de parâmetro aos trabalhos de microbiologia por muitas décadas. É que acreditava-se que, por sua elevada capacidade dispersiva, os microrganismos não seriam detidos por qualquer barreira geográfica e, portanto, diferenças encontradas entre ambientes distintos ocorreriam apenas por uma seleção de organismos que melhor se adaptassem a diferenças ambientais.

Mesmo que já questionada há muito tempo, esta ideia começou a cair por terra apenas nas duas últimas décadas, com o avanço de técnicas de sequenciamento mais avançadas, como o DGGE (*Denaturing Gradient Gel Electrophoresis*) e o NGS (*New Generation Sequencing*), que permitiram acessar a real diversidade microbiana em diferentes locais.

As novas técnicas permitiram também que pudéssemos acessar uma diversidade rara (Pedrós-Alió, 2012), antes não observável que demonstraram ter papéis importantes em seus ambientes, por exemplo, como bioindicadores ou responsáveis pela maior parte da diversidade local. Além disso, parece que na verdade, a dispersão pode influenciar mais a composição local de espécies microbianas que de macro-organismos (Farjalla *et al.*, 2012; Soininen *et al.*, 2013).

Diversidade microbiana dependente da técnica aplicada

Uma técnica clássica de identificação de microrganismos e ainda muito útil para análises ambientais (e.g. Mauroy *et al.*, 2017), bioquímicas (e.g. Eyer *et al.*, 2017) e biomédicas (e.g. Patel *et al.*, 2017) entre outros, a replicação *in vitro* promoveu uma revolução em sua época e ainda hoje é peça fundamental na identificação de espécies microbianas (Rosselló-Mora e Amann, 2001), mas seu poder de isolamento é limitado, muitas espécies não tem condições de sobreviver com as composições nutricionais conhecidas e também não temos condições de descobrir qual seria a composição ideal para todos os organismos.

Muitos organismos possuem necessidades específicas que desconhecemos, além disso, a rede de interações microbiana implica que alguns destes organismos dependem de metabólitos específicos liberados por outros microrganismos para sobreviver (Sarmiento *et al.*, 2015) e o isolamento destes organismos costuma significar sua morte. Portanto, somos incapazes de acessar toda a diversidade microbiana por meio deste método. É por este motivo que ideias como a de Bass Becking perduraram enquanto a replicação *in vitro* manteve-se como a principal forma de estudo da ecologia de microrganismos.

A técnica de DGGE foi a primeira a tornar possível a observação de uma variedade maior de organismos e, por sua vez, promoveu outra revolução na área que foi a primeira a permitir um sequenciamento direto de amostras ambientais e mostrou que a diversidade microbiana ambiental foi subestimada por muitos anos. É, por exemplo, o trabalho de Farjalla

e colaboradores (Farjalla *et al.*, 2012), que foi capaz de identificar os componentes que determinam a dispersão de espécies e que a distancia pode ser uma barreira para microrganismos maior que para macro-organismos.

A técnica de NGS talvez não tenha promovido uma mudança tão brusca na forma como se passou a obter dados sobre a diversidade microbiana quanto o DGGE, mas foi capaz de acessar um numero muito maior de organismos através de um método capaz de identificar características genéticas de indivíduos, permitindo, portanto um reconhecimento em nível de espécie. Esta técnica permitiu mostrar um numero maior de organismos raros e que há sim, isolamento geográfico para a maioria dos grupos estudados. No ultimo capítulo do presente trabalho, mostrarei de forma indireta este isolamento, por meio da identificação de organismos satélite em um conjunto de 60 amostras coletadas por todo o Estado de São Paulo e sequenciadas por meio da técnica NGS de sequenciamento.

Esta é uma técnica tão boa de sequenciamento direto de amostras ambientais que muitos cientistas estão usando-a para encontrar traços genéticos de animais e plantas! Neste caso o material obtido costuma ser chamado eDNA (Barnes e Turner, 2015) por encontrar características indiretas (uma amostra de solo pode ter traços genéticos deixados por um fio deixado para trás por uma onça-pintada, por exemplo) e não o DNA obtido diretamente da lise de uma célula microbiana.

E como a tecnologia não para de avançar, logo uma nova técnica deverá ser lançada e produzirá novo avanço na área. Talvez não para mostrar um número maior de organismos nos ambientes, até porque, discute-se muito se uma parte dessa variedade não seja produzida pela própria técnica de sequenciamento (He *et al.*, 2015), mas provavelmente para facilitar ainda mais o sequenciamento de amostras ambientais. É o caso da técnica NanoPore, capaz de realizar sequenciamentos de DNA e RNA em tempo real de DNA (Garalde *et al.*, 2018; Jenjaroenpun *et al.*, 2018).

OTU – As “espécies” de microrganismos

OTU é uma sigla em inglês para Unidades Taxonômicas Operacionais, muito utilizada para individualizar organismos por diferenças do DNA, mesmo que características morfológicas não sejam conhecidas. Estudos de ecologia microbiana normalmente lançam mão do uso das OTUs como ferramenta de identificação, que considera apenas o material

genético, em detrimento de conceitos tradicionais de definição de espécies, que utilizam caracteres morfológicos e/ou comportamentais, pois muitas vezes estes caracteres não são observáveis.

Para que haja uma separação efetiva das sequencias de aminoácidos disponíveis é preciso definir o nível de similaridade para que sejam provenientes de um mesmo organismo e, portanto, de mesma OTU. O padrão mais comum é uma similaridade de 97%, mas é possível encontrar trabalhos que utilizam uma similaridade de 99% entre as sequencias (e.g. Grice *et al.*, 2008; Caporaso *et al.*, 2012).

Apesar de promoverem uma individualização análoga, os conceitos de espécie e OTU distinguem-se por sua complexidade, como já exposto, OTU utiliza-se apenas do DNA para dar nomes aos microrganismos encontrados, já para uma definição de espécies conforme o conceito de espécie procariótica corrente, é necessário identificar similaridades nos caracteres genômicos e fenotípicos (Rosselló-Mora e Amann, 2001) e, para uma identificação adequada de caracteres morfológicos microbianos é preciso utilizar técnicas de isolamento *in vitro*.

Como já explicado anteriormente, técnicas de replicação de microrganismos *in vitro* apresentam limitações, e é por causa dessas limitações que utiliza-se o conceito de OTU. A diversidade microbiana observada é centenas de vezes superior ao número de espécies descritas e esta técnica de identificação oferece uma forma coerente de se realizar estudos ecológicos enquanto não conseguimos nomear todas as espécies observadas por análises genéticas. A técnica possui limitações, mas esforços para tornar a técnica sempre mais confiável são realizados (He *et al.*, 2015).

Novos paradigmas da diversidade microbiana

O conhecimento acerca da ecologia dos microrganismos vem se desenvolvendo rapidamente nas últimas décadas, e o principal responsável por esse avanço é o avanço da tecnologia, não apenas de técnicas moleculares, mas também de computadores capazes de processar e analisar a quantidade exagerada de dados obtidos, ou até mesmo em produzir modelos de previsão de padrões, através da meta-heurística. Sem dúvidas, conhecimento computacional torna-se cada vez mais importante para o estudante que deseja construir sua carreira na ecologia microbiana.

Estes estudos vêm demonstrando que, ao contrário do que se pensava até pouco tempo atrás, os padrões de distribuição dos microrganismos possuem muitas características encontradas para macro-organismos, assim toda a ecologia começa a se aproximar para um desenvolvimento comum (Barberan *et al.*, 2014).

Observa-se também que este estreitamento entre estas duas áreas funciona como uma via de mão dupla em que os ecólogos microbianos utilizam conhecimento já adquirido acerca da ecologia de macro-organismos como fonte inspiradora para a produção de trabalhos que promovem o avanço do conhecimento acerca dos microrganismos (e.g. Logue e Lindström, 2008; Lindh *et al.*, 2015; Ruiz-González *et al.*, 2015) enquanto as técnicas produzidas para análise molecular de microrganismos mostram-se ótimas ferramentas para estudos da ecologia de macro-organismos (e.g. Barnes e Turner, 2015; Shaw *et al.*, 2016).

Métodos

Ao decidir qual será a pergunta a ser respondida, dá-se início a uma etapa metodológica que consome a maior parte do tempo dedicado ao mestrado. Dos dois anos de estudos, pelo menos um ano resume-se a coletas, processamento e análise dos dados obtidos. A seguir apresento uma descrição detalhada de todas as etapas metodológicas enfrentadas durante meu mestrado, espero com isso apresentar as dificuldades comuns a qualquer pessoa que decida seguir a mesma linha de estudos que a minha e possíveis soluções para problemas comuns. A ideia é

tornar este capítulo um guia para futuros projetos de pesquisa, facilitando processos demorados e detalhados e mitigando fontes de desespero para futuros pesquisadores em ecologia microbiana.

Coleta

Geografia e escolha dos pontos

O primeiro desafio deste projeto foi a seleção das amostras. Partimos de coletas já realizadas entre 2012 e 2016, no âmbito do projeto BIOTA/FAPESP coordenado pelo Prof. Dr. Armando Augusto Henrique Vieira do laboratório de ficologia (departamento de Botânica da UFSCar). Foram visitados 320 lagos, durante várias campanhas realizadas por diferentes membros do laboratório. Foram coletados dados ambientais e amostras filtradas para sequenciamento de DNA e clorofila *a* (chl *a*), além de amostras frescas para identificação e isolamento de microalgas para ampliação da Coleção de Culturas de Microalgas de Água Doce (CCMA-UFSCar).

Como os pontos já haviam sido coletados, tivemos que adaptar as amostras disponíveis às nossas necessidades e definimos alguns parâmetros antes de iniciar a escolha dos pontos de coleta que utilizaríamos para este trabalho, assim separamos o Estado em quatro regiões, com base nos padrões geográficos que definem as Unidades de Gerenciamento de Recursos Hídricos (UGRHi). Estes foram nomeados: Quadrante Tietê, Quadrante Grande, Quadrante Paranapanema e Quadrante Serra do Mar.

Assim, o primeiro quadrante (Tietê) contém toda a bacia do rio Tietê, desde sua nascente próxima à serra do mar até sua foz no rio Paraná. Os rios próximos ao Estado de Minas, com sua foz principalmente no rio Grande, foram agrupados no segundo quadrante (Grande). A Bacia do Paranapanema e alguns rios menores formam o terceiro quadrante (Paranapanema) e, por fim, todos os rios que, nascendo na Serra do Mar, descem a encosta em direção ao mar fazem parte de nosso quarto quadrante (Serra do Mar). Para mais detalhes, veja a fig. 5.

Além disso, as amostras deveriam ser o mais equidistantes possível entre si, minimizando assim que os resultados encontrados sejam explicados pela falta de padrão da

escolha dos pontos (fig. 5). Decidimos também não utilizar coletas realizadas em grandes reservatórios de usinas hidrelétricas, pois entendemos que o tempo de residência muito superior ao de outros pontos coletados em lagos rasos poderia enviesar o nosso desenho amostral (Crump *et al.*, 2004; Lindström *et al.*, 2005).

Clorofila a

Definidos os pré-requisitos, escolhemos 60 amostras levando em consideração seu estado trófico. Para tal, separamos, em cada quadrante, 15 amostras baseando-nos no índice de estado trófico encontrado. Este índice foi estimado através dos valores de Chl *a* obtidos em cada um dos lagos escolhidos para este projeto.

Amostras de água foram filtradas em membranas de fibra de vidro (GF/C), e em congelados papel alumínio em freezer -20°C. No laboratório, o material foi extraído com álcool aquecido a 70°C e acondicionado em local escuro e refrigerado por 24 horas. A concentração de Chl *a* foi determinada em um espectrofotômetro.

O uso da chl *a* como estimativa do estado trófico foi proposta por Carlson (1977) e se mantém como uma das melhores formas de avaliar o nível de impacto de ambientes aquáticos, até hoje. No entanto, estes cálculos foram propostos utilizando como base, dados obtidos em altas latitudes. Nós ponderamos que esta diferença característica entre o ambiente aqui estudado e o ambiente estudado por Carlson poderia acabar influenciando em nosso resultado final, e decidimos utilizar como referencia outro trabalho mais recentemente produzido e com enfoque em ambiente tropical (Cunha *et al.*, 2013). Neste artigo, Cunha apoiou-se nos cálculos apresentados por Carlson e dados obtidos em campo, realizado no estado de São Paulo, para modelar um cálculo mais adequado ao ambiente aqui observado.

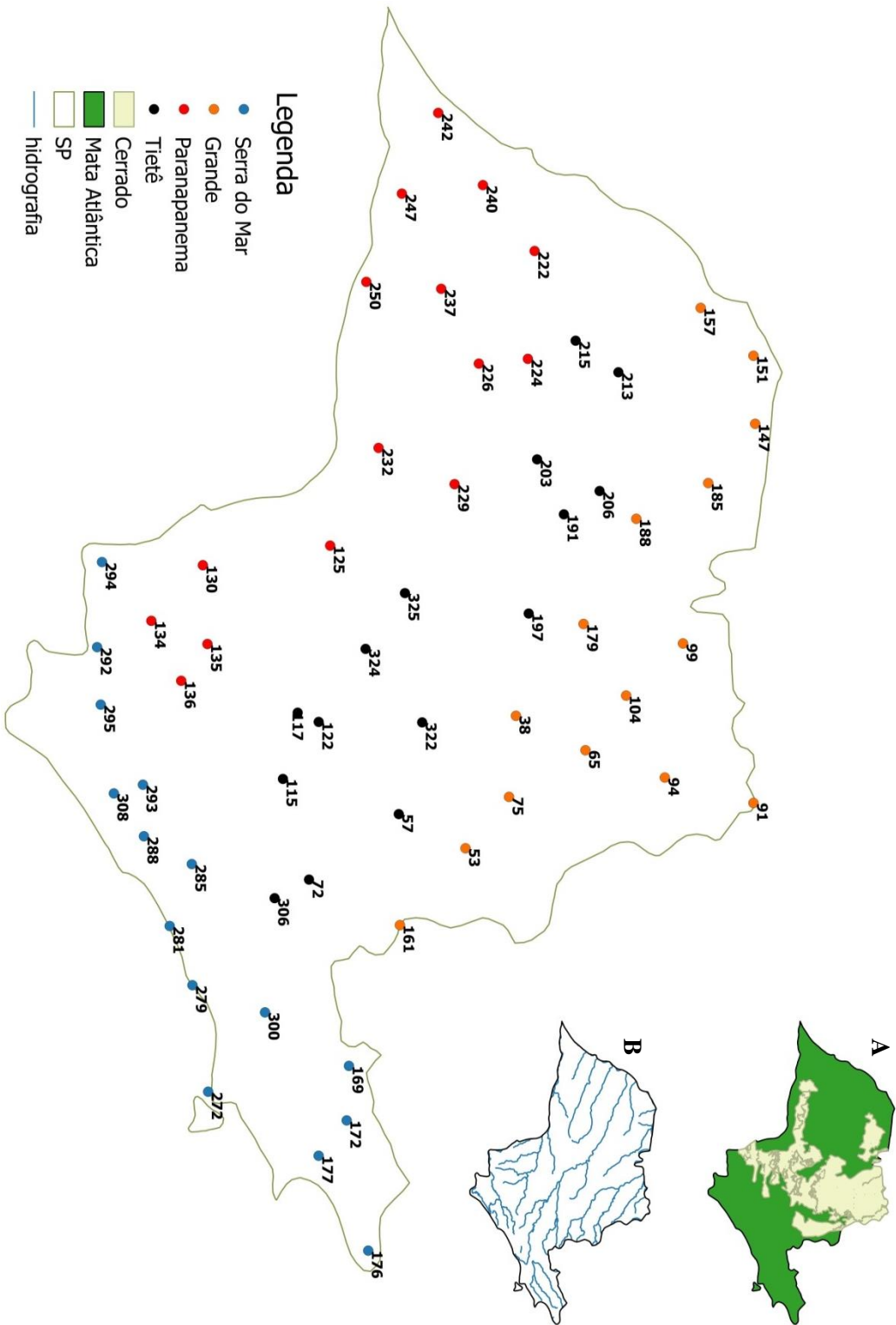


Figura 5 – Representação gráfica do Estado de São Paulo onde ocorreu a amostragem. Os 60 pontos escolhidos foram divididos em quatro quadrantes: Serra do Mar, Paranapanema, Tietê e Grande, nomes relacionados aos rios ou formação geográfica mais importante de cada região. As representações menores indicam a formação vegetal (A) e hidrológica (B) do Estado.

Ao final, separamos cinco amostras de lagos oligotróficos, cinco amostras de lagos mesotróficos e mais cinco amostras de lagos eutróficos para cada um dos quadrantes definidos anteriormente, com o objetivo de homogeneizar o efeito do aporte de nutrientes em nosso resultado final.

Vale a pena salientar que decidimos não utilizar todos os pontos com o mesmo índice de trofia porque entendemos que a observação da flutuação desta variável poderia revelar dados interessantes.

DNA

Extração

A etapa de extração tem por objetivo a obtenção de material genético de interesse, isolando-o do restante do material celular e de outras substâncias contidas nas amostras. Nesta etapa foi utilizada uma metodologia de extração baseada no isolamento por fenol:clorofórmio. É possível encontrar dezenas de protocolos com modificações sutis do método baseado em fenol:clorofórmio, neste caso, o protocolo utilizado baseia-se no utilizado pelo projeto Tara Oceans (para mais detalhes sobre o projeto, recomendo iniciar pesquisas à partir de Bork *et al.*, 2015).

Hoje, estão disponíveis no mercado, alguns kits de extração menos tóxicos (eg. PowerSoil DNA isolation Kit, MoBio®; Invisorb spin extraction Kit, Stratec Molecular®), mas, comparados à técnica do fenol:clorofórmio, ainda são materiais muito caros, e multiplicam-se ainda disputando mercado; este último ponto levantado é especialmente problemático, pois os vários kits ainda sendo testados e disputando mercado, acabam gerando trabalhos sem padrão rigoroso, o que dificulta a replicação de um resultado em outro laboratório. Para tentar sanar este problema, foi lançado Earth Microbiome Project (EMP), que visa padronizar protocolos para todas as etapas do sequenciamento (para mais detalhes: <http://www.earthmicrobiome.org/>).

Estes kits tem como vantagem a velocidade de aplicação, no máximo algumas horas são necessárias para que o DNA seja obtido; já o fenol:clorofórmio exige alguns dias inteiros

de dedicação exclusiva, em contrapartida, se aplicado com extremo rigor e cuidado extra nas lavagens que retiram o excesso de fenol:clorofórmio da amostra, o material obtido estará livre de agentes inibidores, e terá DNA suficiente para as etapas de amplificação e sequenciamento subsequentes. Esta é uma técnica mais rudimentar em que, basicamente, encubamos o material filtrado com detergentes e enzimas que dissolvem camadas protetoras externas liberadas pelas células além de quebrar as membranas celular e nucleica, deixando assim o DNA livre para ser isolado.

Resumidamente, os filtros foram cortados em pequenos pedaços e incubados (37 ° C; 45 minutos) com *lysis buffer* (3 mL) e lisozima (75 µL) e mais tarde, incubados (55 ° C; 60 minutos) com proteinase K (75 µL) e sódio dodecil-sulfato (SDS) 10% (300 µL), após a incubação, a extração foi realizada em dois passos de centrifugação utilizando isoamil fenol:clorofórmio:isoamil (2 mL, 10 minutos, 5000 rpm) e outra centrifugação com clorofórmio:isoamil (2 mL; 10 minutos; 5000 rpm) e, finalmente, purificamos o material extraído em colunas Amicon de 100KDa (Millipore® 100.000MWCO).

O material é posteriormente coletado em eppendorfs, as amostras tiveram também sua qualidade e quantidade avaliadas em Nanodrop® (ou Qubit®) e congelado em freezer -80°C.

Seguiu-se uma etapa extra de limpeza das amostras, ricas em substâncias húmicas. Nós realizamos um método de purificação que usa brometo de cetiltrimetilamônio (CTAB) 10% (Schneegurt, Mark A *et al.*, 2003). Este método consiste em uma etapa de incubação (12 µL; 65°C; 15 minutos) com CTAB pré-aquecido (65°C) dissolvido em NaCl (0.7M), seguida de uma etapa de centrifugação utilizando novamente clorofórmio:isoamil (128 µL; 5 minutos; 5000 rpm), segue-se uma etapa de incubação (-20°C; 15 minutos) e centrifugação (5 minutos; 5000 rpm) utilizando álcool 100% (256 µL), e em seguida uma ultima etapa de centrifugação (2 minutos, 5000 rpm), desta vez com álcool 70% (125 µL).

Aqui ficam duas dicas importantes para as etapas que utilizam álcool. Primeiro: a incubação que segue à adição do álcool 100% deve seguir no mínimo 15 minutos. Quanto mais tempo essa incubação ocorrer, maior será a precipitação de DNA, portanto, é possível deixar as amostras incubando por toda uma noite, em caso de baixa concentração de DNA. Segundo: A centrifugação visa a criação de um *pellet* que concentra o material genético em uma das bordas do frasco utilizado, a parte líquida é retirada, mas não é possível ver o *pellet*. É de suma importância, portanto, que se saiba para qual lado a centrífuga empurrou o

material. Recomenda-se orientar os tubos todos com suas tampas voltadas para a mesma direção.

O material é deixado em local protegido para que o álcool seque completamente (15 minutos) e finalmente, adiciona-se água ultra-pura para se obter o volume de amostra desejado.

Amplificação por PCR

Por termos escolhido amostras bastante heterogêneas, esta se mostrou a etapa mais desafiadora do projeto. Vários testes e adaptações foram necessários para que todos os dados fossem amplificados de forma a obtermos DNA de qualidade para as etapas seguintes.

Utilizamos aqui, o Mix KAPA HiFi HotStart ReadyMix (Kapa Biosystems®), com a reação indicada na bula do produto. Os *primers* utilizados, 341F (5'-CCTACGGGNGGCWGCAG-3') e 805R (5'-GACTACHVGGGTATCTAATCC-3') (Herlemann *et al.*, 2011), amplificam os genes das regiões V3-V4 da subunidade 16S do gene do RNA ribossomal (rRNA).

É importante realçar que temperatura de “*annealing*” do *primer* deve ser calculada especificamente para o *primer* utilizado, através de cálculos de temperatura de *melting*, e não a temperatura indicada pelo protocolo do mix. Isto irá facilitar a obtenção de material amplificado suficiente para os próximos passos.

Para a maioria das amostras, a amplificação do DNA microbiano ocorreu sem problemas. Estas foram, então, purificadas com beads magnéticas (utilizamos Beads Agentecourt® AMPure® XP, da empresa Beckman Coulter®) e acondicionadas em freezer -20°C, à espera da amplificação das amostras mais trabalhosas.

Algumas amostras possuíam elevada concentração de elementos inibidores de PCR, portanto, utilizar um volume menor de amostra, resultava em pouca concentração de material genético amplificado e, um volume maior da amostra, resultava em uma PCR inibida. Era possível perceber alguma reação ao observarmos o gel, mas nunca material suficiente era amplificado. Este problema foi resolvido ao fazermos uma diluição do material, para tal utilizamos 5µL de amostra e 45µL de água MiliQ. Após, realizamos algumas PCRs e

juntamos o material obtido para que pudéssemos obter DNA suficiente para o sequenciamento.

Montagem e quantificação das bibliotecas e sequenciamento

As amostras foram então indexadas com o kit de indexação Nextera® XT V.2. para 96 amostras da Illumina® e novamente purificadas com beads magnéticas.

O kit de indexação consiste em 20 frascos com tampas de duas cores distintas, que podem variar dependendo do lote e da marca do produto, e contendo fragmentos capazes de ligar-se às pontas do DNA amplificado na PCR. Estas duas cores distintas nos frascos são importantes pois identificam e separam os indexes capazes de ligar-se a cada extremidade 5' ou 3' do material amplificado.

Cada index contém um número de identificação que será fornecido ao computador que realizará o sequenciamento, este computador é capaz de reconhecer as diferentes sequencias que compõe o index e, comparando com os códigos fornecidos, reconhecer cada uma das amostras separadamente.

Cada amostra deve receber, portanto, alguns microlitros de um par dos indexes fornecidos, estes dois indexes devem vir de frascos com cores distintas, pois assim, cada um deles se ligará a cada extremidade do fragmento de DNA amplificado, além disso, essa combinação de dois indexes nunca poderá repetir-se para que o computador seja capaz de reconhecer todas as amostras fornecidas sem confundi-las. O produto possui indexes suficientes para criar uma rede lógica de até 96 amostras contendo um par distinto cada.

Talvez não seja preciso salientar que a contaminação de apenas um dos frascos com algum index vindo de outro frasco significa o comprometimento de todo o produto, já que o computador será incapaz de diferenciar o material proveniente deste tipo de contaminação daquele proveniente de outra amostra, mas que apresentar um mesmo par de códigos.

Por meio da técnica de eletroforese em gel das amostras indexadas e purificadas, em baixa voltagem de forma que as amostras se deslocassem vagarosamente pelo gel, constatamos que as amostras indexadas ainda não estavam puras, mas apresentavam anelamento inespecífico, com duas bandas além da banda de interesse, por isso, foi necessária uma nova purificação, desta vez em gel com um kit de purificação (DNA Clean &

Concentrator, Zymo Research®). O material purificado foi quantificado em qPCR, antes da preparação para o sequenciamento.

Após o preparo das bibliotecas, o material foi sequenciado utilizando o MiSeq® Reagent Kit V.3. Utilizamos para tal um sequenciador MiSeq® Illumina® (Sinclair *et al.*, 2015), informamos ao sistema quais os códigos dos índices utilizados na etapa anterior e a quais amostras estes índices estavam relacionados. O equipamento realiza o sequenciamento e disponibiliza os dados em uma plataforma online da Illumina para serem acessados com um login e baixados para o computador que irá realizar os tratamentos posteriores dos dados.

Bioinformática

Os dados do sequenciamento foram, então, processados através de um *pipeline* compilado por Logares e colaboradores (Logares *et al.*, 2014; Logares, 2017) e implementado no servidor do Laboratório de Processos e Biodiversidade Microbiana (DHb – UFSCar).

Em resumo, utilizamos uma rotina baseada em UPARSE, em uma sequência de programas com funções específicas. O primeiro tratamento utiliza o método de Bayes Hammer, uma aplicação do programa Spades para correção de erros. A etapa seguinte utiliza o PEAR para parear as sequências e eliminar qualquer dado em fita simples. A seguir, é acionado o usearch para eliminar sequências de baixa qualidade, aquelas com menos de 100 pares de bases e, após, o *hmm3*, outra aplicação Spades para alinhar as duas fitas na mesma direção, 5'-3'. A penúltima etapa utiliza o programa UPARSE, aplicação usearch, para eliminar quimeras e identificar as sequências em OTUs, delimitadas como sequências com pelo menos 97% de similaridade, uma base de dados deve ser fornecida como referência e, para tanto, utilizamos a base de dados SILVA versão 128 (Quast *et al.*, 2013; Yilmaz *et al.*, 2014). Finalmente, a etapa de classificação taxonômica compara as OTUs obtidas com o banco de dados fornecido para catalogar, com até 75% de similaridade, os grupos taxonômicos.

O resultado final é uma tabela de abundância de OTUs, cada uma destas com sua respectiva classificação taxonômica, identificando de seu filo até a espécie (quando possível). A classificação pode ser realizada utilizando qualquer base de dados disponível, basta que seja feito o download. Em outras ocasiões utilizamos, por exemplo, a base de dados FWOonly

(Newton *et al.*, 2011), que contém apenas *taxa* típicos de ambientes aquáticos. Cada base de dados tem a sua particularidade e pode provar-se útil para diversas análises, porém a base SILVA é a mais completa disponível no momento.

Referências

BAAS BECKING, L. G. M. **Geobiologie of inleiding tot de milieukunde**. Den Haag: Van Stockum, 1934.

BARBERAN, A.; CASAMAYOR, E. O.; FIERER, N. The microbial contribution to macroecology. **Front Microbiol**, v. 5, p. 203, 2014. ISSN 1664-302X (Print)

1664-302X (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24829564> >.

BARNES, M. A.; TURNER, C. R. The ecology of environmental DNA and implications for conservation genetics. **Conservation Genetics**, v. 17, n. 1, p. 1-17, 2015. ISSN 1566-0621 1572-9737.

BORCARD, D.; LEGENDRE, P.; DRAPEAU, P. Partialling out the Spatial Component of Ecological Variation. **Ecology**, v. 73, n. 3, p. 1045-1055, 1992. ISSN 1939-9170. Disponível em: < <http://dx.doi.org/10.2307/1940179> >.

BORK, P. et al. Tara Oceans studies plankton at planetary scale. **Science**, v. 348, n. 6237, p. 873, May 22, 2015 2015. Disponível em: < <http://www.sciencemag.org/content/348/6237/873.short> >.

BROWN, B. L. et al. Making sense of metacommunities: dispelling the mythology of a metacommunity typology. **Oecologia**, p. 1-10, 2016. ISSN 1432-1939. Disponível em: < <http://dx.doi.org/10.1007/s00442-016-3792-1> >.

CAPORASO, J. G. et al. The Western English Channel contains a persistent microbial seed bank. **The ISME journal**, v. 6, n. 6, p. 1089-1093, 2012. ISSN 1751-7362.

CARLSON, R. E. A trophic state index for lakes. **Limnology and Oceanography**, v. 22, n. 2, p. 361-369, 1977. ISSN 1939-5590. Disponível em: < <http://dx.doi.org/10.4319/lo.1977.22.2.0361> >.

COLLINGE, S. Spatial ecology and conservation. **Nature**, v. 1, p. 69, 2010.

COTTENIE, K. Integrating environmental and spatial processes in ecological community dynamics. **Ecology Letters**, v. 8, n. 11, p. 1175-1182, 2005. ISSN 1461-0248. Disponível em: < <http://dx.doi.org/10.1111/j.1461-0248.2005.00820.x> >.

CRUMP, B. C. et al. Microbial biogeography along an estuarine salinity gradient: combined influences of bacterial growth and residence time. **Applied and environmental microbiology**, v. 70, n. 3, p. 1494-1505, 2004. ISSN 0099-2240.

CUNHA, D. G. F.; CALIJURI, M. D. C.; LAMPARELLI, M. C. A trophic state index for tropical/subtropical reservoirs (TSIts_r). **Ecological Engineering**, v. 60, p. 126-134, 11// 2013.

ISSN 0925-8574. Disponível em: <

<http://www.sciencedirect.com/science/article/pii/S0925857413003091> >.

ENDO, Y. et al. Comparative phylogeography of alpine invertebrates indicates deep lineage diversification and historical refugia in the Australian Alps. **Journal of Biogeography**, v. 42, n. 1, p. 89-102, 2015. ISSN 1365-2699. Disponível em: < <http://dx.doi.org/10.1111/jbi.12387> >.

EYER, L. et al. Antiviral activity of the adenosine analogue BCX4430 against West Nile virus and tick-borne flaviviruses. **Antiviral Research**, v. 142, p. 63-67, 2017/06/01/ 2017. ISSN 0166-3542. Disponível em: <

<http://www.sciencedirect.com/science/article/pii/S0166354217300244> >.

FALKE, J. A.; FAUSCH, K. D. From metapopulations to metacommunities: linking theory with empirical observations of the spatial population dynamics of stream fishes. Community ecology of stream fishes: concepts, approaches, and techniques. American Fisheries Society, Symposium, 2010. p.207-233.

FARJALLA, V. F. et al. Ecological determinism increases with organism size. **Ecology**, v. 93, n. 7, p. 1752-1759, 2012. ISSN 1939-9170. Disponível em: < <http://dx.doi.org/10.1890/11-1144.1> >.

FERNANDES, C. S.; DE SIQUEIRA BUENO, S. L.; BICHUETTE, M. E. Distribution of Cave-Dwelling *Aegla* spp. (Decapoda: Anomura: Aeglidae) from the Alto Ribeira Karstic Area in Southeastern Brazil Based on Geomorphological Evidence. **Journal of Crustacean Biology**, v. 33, n. 4, p. 567-575, 2013. ISSN 0278-0372. Disponível em: < <http://dx.doi.org/10.1163/1937240X-00002159> >.

FRONHOFER, E. A. et al. Condition-dependent movement and dispersal in experimental metacommunities. **Ecol Lett**, v. 18, n. 9, p. 954-63, Sep 2015. ISSN 1461-0248 (Electronic) 1461-023X (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26206470> >.

GARALDE, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. **Nature Methods**, 01/15/online 2018. Disponível em: < <http://dx.doi.org/10.1038/nmeth.4577> >.

GRICE, E. A. et al. A diversity profile of the human skin microbiota. **Genome research**, v. 18, n. 7, p. 1043-1050, 2008. ISSN 1088-9051.

HANSKI, I.; THOMAS, C. D. Metapopulation dynamics and conservation: A spatially explicit model applied to butterflies. **Biological Conservation**, v. 68, n. 2, p. 167-180, 1994/01/01/ 1994. ISSN 0006-3207. Disponível em: < <http://www.sciencedirect.com/science/article/pii/0006320794903484> >.

HE, Y. et al. Stability of operational taxonomic units: an important but neglected property for analyzing microbial diversity. **Microbiome**, v. 3, p. 20, 2015. ISSN 2049-2618 (Print)

2049-2618 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25995836> >.

HERLEMANN, D. P. et al. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. **The ISME journal**, v. 5, n. 10, p. 1571-1579, 2011. ISSN 1751-7362.

HUBBELL, S. P. **The Unified Neutral Theory of Biodiversity and Biogeography**. 2001.

HUBERT, N. et al. Metacommunity speciation models and their implications for diversification theory. **Ecol Lett**, v. 18, n. 8, p. 864-81, Aug 2015. ISSN 1461-0248

(Electronic), 1461-023X (Linking). Disponível em: <

<http://www.ncbi.nlm.nih.gov/pubmed/26036711> >.

HUMBOLDT, A. V.; BONPLAND, A. **Essai sur la géographie des plantes : accompagné d'un tableau physique des régions équinoxiales, fondé sur des mesures exécutées, depuis le dixième degré de latitude boréale jusqu'au dixième degré de latitude australe,**

pendant les années 1799, 1800, 1801, 1802 et 1803. A Paris, :: Chez Levrault, Schoell et

compagnie, libraires, 1805. Disponível em: < <http://www.biodiversitylibrary.org/item/37872>

>.

JENJAROENPUN, P. et al. Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN.PK113-7D.

Nucleic Acids Research, p. gky014-gky014, 2018. ISSN 0305-1048. Disponível em: < <http://dx.doi.org/10.1093/nar/gky014> >.

JENKINS, D. G.; RICKLEFS, R. E. Biogeography and ecology: two views of one world.

Philosophical Transactions of the Royal Society B: Biological Sciences, v. 366, n. 1576, p. 2331-2335, 2011.

LEIBOLD, M. A. et al. The metacommunity concept: a framework for multi-scale community ecology. **Ecology Letters**, v. 7, n. 7, p. 601-613, 2004. ISSN 1461023X, 14610248.

LEVINS, R. A. Extinction. **Lect. Notes Math. Life Sci.**, v. 2, p. 75–107, 1970.

LINDH, M. V. et al. Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. **Environ Microbiol**, v. 17, n. 7, p. 2459-76, Jul 2015. ISSN 1462-2920 (Electronic)

1462-2912 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25403576> >.

LINDSTRÖM, E. S.; KAMST-VAN AGTERVELD, M. P.; ZWART, G. Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water

retention time. **Applied and environmental microbiology**, v. 71, n. 12, p. 8201-8206, 2005. ISSN 0099-2240.

LOGARES, R. ramalok/amplicon_processing: Workflow for Analysing MiSeq Amplicons based on Uparse. 2017. Disponível em: < <https://doi.org/10.5281/zenodo.259579> >.

LOGARES, R. et al. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. **Environ Microbiol**, v. 16, n. 9, p. 2659-71, Sep 2014. ISSN 1462-2920 (Electronic), 1462-2912 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24102695> >.

LOGUE, J. B.; LINDSTRÖM, E. S. Biogeography of Bacterioplankton in Inland Waters. **Freshwater Reviews**, v. 1, n. 1, p. 99-114, 2008. ISSN 1755-084X.

MACARTHUR, R. H.; WILSON, E. O. **The Theory of Island Biogeography**. Princeton University Press, 1967. ISBN 9780691088365. Disponível em: < <https://books.google.com.br/books?id=a10cdkywhVgC> >.

MAUROY, A. et al. High-throughput sequencing analysis reveals the genetic diversity of different regions of the murine norovirus genome during in vitro replication. **Archives of Virology**, v. 162, n. 4, p. 1019-1023, April 01 2017. ISSN 1432-8798. Disponível em: < <https://doi.org/10.1007/s00705-016-3179-y> >.

MILICI, M. et al. Bacterioplankton Biogeography of the Atlantic Ocean: A Case Study of the Distance-Decay Relationship. **Front Microbiol**, v. 7, p. 590, 2016. ISSN 1664-302X (Electronic), 1664-302X (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/27199923> >.

MOUQUET, N.; LOREAU, M. Coexistence in Metacommunities: The Regional Similarity Hypothesis. **The American Naturalist**, v. 159, n. 4, p. 420-426, 2002. ISSN 00030147, 15375323. Disponível em: < <http://www.jstor.org/stable/10.1086/338996> >.

NEE, S. **Metapopulations and their spatial dynamics**. Oxford University Press: Oxford, UK, 2007.

NEWTON, R. J. et al. A guide to the natural history of freshwater lake bacteria. **Microbiol Mol Biol Rev**, v. 75, n. 1, p. 14-49, Mar 2011. ISSN 1098-5557 (Electronic) 1092-2172 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21372319> >.

PATEL, S. et al. HIV specific t cells expressing an X5-GPI artificial receptor can suppress hiv replication in vitro—implications for a cure strategy for HIV+ individuals with hematologic malignancies. **Cytotherapy**, v. 19, n. 5, p. S7-S8, 2017. ISSN 1465-3249. Disponível em: < <http://dx.doi.org/10.1016/j.jcyt.2017.02.005> >. Acesso em: 2018/01/16.

PEDRÓS-ALIÓ, C. The rare bacterial biosphere. **Ann Rev Mar Sci**, v. 4, p. 449-66, 2012. ISSN 1941-1405 (Print)

1941-0611 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22457983> >.

QUAST, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. **Nucleic Acids Research**, v. 41, p. D590-D596, 2013.

R. CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing 2016.

RICKLEFS, R. E.; JENKINS, D. G. Biogeography and ecology: towards the integration of two disciplines. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 366, n. 1576, p. 2438-2448, 2011.

ROSSELLÓ-MORA, R.; AMANN, R. The species concept for prokaryotes. **FEMS Microbiology Reviews**, v. 25, n. 1, p. 39-67, 2001. ISSN 0168-6445. Disponível em: < <http://dx.doi.org/10.1111/j.1574-6976.2001.tb00571.x> >.

RUIZ-GONZÁLEZ, C.; NIÑO-GARCÍA, J. P.; DEL GIORGIO, P. A. Terrestrial origin of bacterial communities in complex boreal freshwater networks. **Ecology Letters**, v. 18, n. 11, p. 1198-1206, 2015. ISSN 1461-0248. Disponível em: < <http://dx.doi.org/10.1111/ele.12499> >.

SARMENTO, H. et al. Microbial food web components, bulk metabolism, and single-cell physiology of piconeuston in surface microlayers of high-altitude lakes. **Front Microbiol**, v. 6, p. 361, 2015. ISSN 1664-302X (Electronic)

1664-302X (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25999921> >.

SCHIAFFINO, R., M. et al. Bacterial community structure in a latitudinal gradient of lakes: the roles of spatial versus environmental factors. **Freshwater Biology**, v. 56, n. 10, p. 1973-1991, 2011. ISSN 00465070.

SCHNEEGURT, M. A.; DORE, S. Y.; KULPA JR, C. F. Direct extraction of DNA from soils for studies in microbial ecology. **Current issues in molecular biology**, v. 5, n. 1, p. 1-8, 2003. ISSN 1467-3037.

SHAW, J. L. A.; WEYRICH, L.; COOPER, A. Using environmental (e)DNA sequencing for aquatic biodiversity surveys: a beginner's guide. **Marine and Freshwater Research**, 2016. ISSN 1323-1650.

SINCLAIR, L. et al. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the illumina platform. **PLoS One**, v. 10, n. 2, p. e0116955, 2015. ISSN 1932-6203 (Electronic), 1932-6203 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25647581> >.

SOININEN, J.; KORHONEN, J. J.; LUOTO, M. Stochastic species distributions are driven by organism size. **Ecology**, v. 94, n. 3, p. 660-670, 2013. ISSN 00129658, 19399170.

Disponível em: < <http://www.jstor.org/stable/23436269> >.

SOUFFREAU, C. et al. An experimental analysis of species sorting and mass effects in freshwater bacterioplankton. **Freshwater Biology**, v. 59, n. 10, p. 2081-2095, 2014. ISSN 00465070.

TABERLET, P. et al. Environmental DNA. **Molecular Ecology**, v. 21, n. 8, p. 1789-1793, 2012. ISSN 1365-294X. Disponível em: < <http://dx.doi.org/10.1111/j.1365-294X.2012.05542.x> >.

THOMPSON, J. N. **The geographic mosaic of coevolution**. University of Chicago Press, 2005. ISBN 022611869X.

WERNER, E. E. et al. Turnover in an amphibian metacommunity: the role of local and regional factors. **Oikos**, v. 116, n. 10, p. 1713-1725, 2007. ISSN 1600-0706.

WILSON, D. S. Complex Interactions in Metacommunities, with Implications for Biodiversity and Higher Levels of Selection. **Ecology**, v. 73, n. 6, p. 1984-2000, 1992. ISSN 1939-9170. Disponível em: < <http://dx.doi.org/10.2307/1941449> >.

WOLFE, J. D. et al. Island vs. countryside biogeography: an examination of how Amazonian birds respond to forest clearing and fragmentation. **Ecosphere**, v. 6, n. 12, p. 1-14, 2015.

ISSN 2150-8925. Disponível em: < <http://dx.doi.org/10.1890/ES15-00322.1> >.

YILMAZ, P. et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. **Nucleic Acids Research**, v. 42, p. D643-D648, 2014.

Comparison of two widely used DNA extraction methods in aquatic microbial ecology

Manuscript submitted as Note to Aquatic Microbial Ecology

Abstract

In recent years, the rapid advances of culture-independent methods and new molecular tools have revolutionized our understanding of microbial biodiversity and ecological functions. The extraction of environmental DNA is a critical step in this process and several methods have been proposed and used. However, there is a lack of information on how these different methods may influence the final outcome and ultimately

the ecological inferences from those results. Our aim was to compare two of the most commonly used extraction methods in aquatic microbial ecology and to determine if they both methods yield comparable results for community ecology analyses purposes. We extracted DNA from 15 different shallow lakes with Phenol:Chloroform, (a classical and widely used extraction method), and the PowerSoil DNA isolation Kit. We found that despite some observable differences between methods (notably, the improved efficiency by Phenol:Chlorofom method in extract groups like Planctomycetes and Verrucomicrobia, known to have atypical cell wall composition), there were no significant differences in diversity, abundance and evenness indexes, widely used in community ecology. Therefore, samples extracted with any of these different methods may be mixed in downstream community ecology analyses with no major biases.

Keywords: Phenol:Chloroform extraction, PowerSoil extraction kit, next generation sequencing, high throughput sequencing, aquatic microbial ecology.

Main Text

High throughput sequencing has become a molecular tool widely used in a daily basis by microbial ecologists (Nielsen *et al.*, 2014; Newton & Mclellan, 2015; Granja-Salcedo *et al.*, 2017). This tool has the potential to reveal all the genes in a given environmental sample as well as the taxonomy of the microorganisms, providing insights to the dynamics of their biodiversity and functions at different temporal and spatial scales (Pylro *et al.*, 2014; Woodhouse *et al.*, 2016). Nevertheless, in order to achieve good sequencing results, the microbial community DNA obtained must have high quality and sufficient quantity for downstream analyses of microbial communities such as amplicon sequencing, metagenomics and metatranscriptomics.

In aquatic samples, the microbial biomass is usually obtained by filtering a certain volume of water through membranes of different porosities (e.g. Mestre *et al.*, 2017). The particle attached bacteria represent a relevant fraction of microbial biodiversity (Zhang *et al.*, 2016), but sometimes the amount of particles or the high content in humic substances may turn into a methodological limitation, selecting certain organisms (Mestre *et al.*, 2017) or inhibiting the subsequent analyses (Huggett *et al.*, 2008; Sidstedt *et al.*, 2015; Wang *et al.*, 2017). Therefore, it is particularly important to perform an appropriate and rigorous methodological design, in order to minimize the negative effects of these issues on downstream analysis (Olson & Morrow, 2012; Taberlet *et al.*, 2012).

Indeed, some studies have already explored methodological approaches through different purification procedures both on metagenomic or amplicon DNA aiming to maximize the quality of the material obtained (e.g. Arbeli & Fuentes, 2007; Dai *et al.*, 2016; Wang *et al.*, 2017), and others have compared extraction methods in order to maximize DNA integrity (e.g. Seumahu *et al.*, 2012; Li *et al.*, 2015). It is clear that the quality of the extracted material may influence our ability to detect some organisms or genes (Deiner *et al.*, 2015), however

there is a lack of information regarding how different DNA extraction methods may influence the overall results obtained after sequencing, and ultimately our ecological interpretations of those results.

Here, we evaluated two standard DNA extraction methods widely used in aquatic microbial ecology: 1) the PowerSoil DNA isolation kit (MoBio Inc. Carlsbad CA, USA) following the manufacturer's protocol with adaptations (Supplementary material for further details) referred here as kit extraction, and 2) the Phenol:Chloroform extraction, followed by purification in Amicon columns (Millipore® 100KDa/100.000MWCO), with an additional purification step with cetyl trimethyl ammonium bromide (CTAB) 10% (Schneegurt, M. A. *et al.*, 2003), referred here as organic extraction.

To achieve comparable results, approximately 500 mL of water samples were pre-filtered through a glass fiber filter with 1.2 µm mesh, and then through polycarbonate 0.22 µm membranes, each half part of this last filter was used to perform each of the extraction method above mentioned. After that, the DNA total from the 15 samples collected (Fig. 6) were analyzed by the exact same PCR protocol, sequenced in the same Illumina MiSeq run and processed in an Uparse based pipeline (Logares *et al.*, 2014; Logares, 2017) using the SILVA database (Quast *et al.*, 2013; Yilmaz *et al.*, 2014) to determine operational taxonomical units (OTUs) at 97% similarity threshold (Supplementary material for further details). All freshwater samples tested here were collected from 15 shallow lakes in Sao Paulo State (Brazil). In order to compare the two DNA extraction methods, community ecology and statistical analyses were performed through 16S rRNA gene amplicon sequencing. Also, a set of community ecology indexes were calculated with BiodiversityR package (Kindt & Coe, 2005). Finally, paired t-test and Wilcoxon signed-rank test with an alpha of 0.05% of significance and graphs were performed.

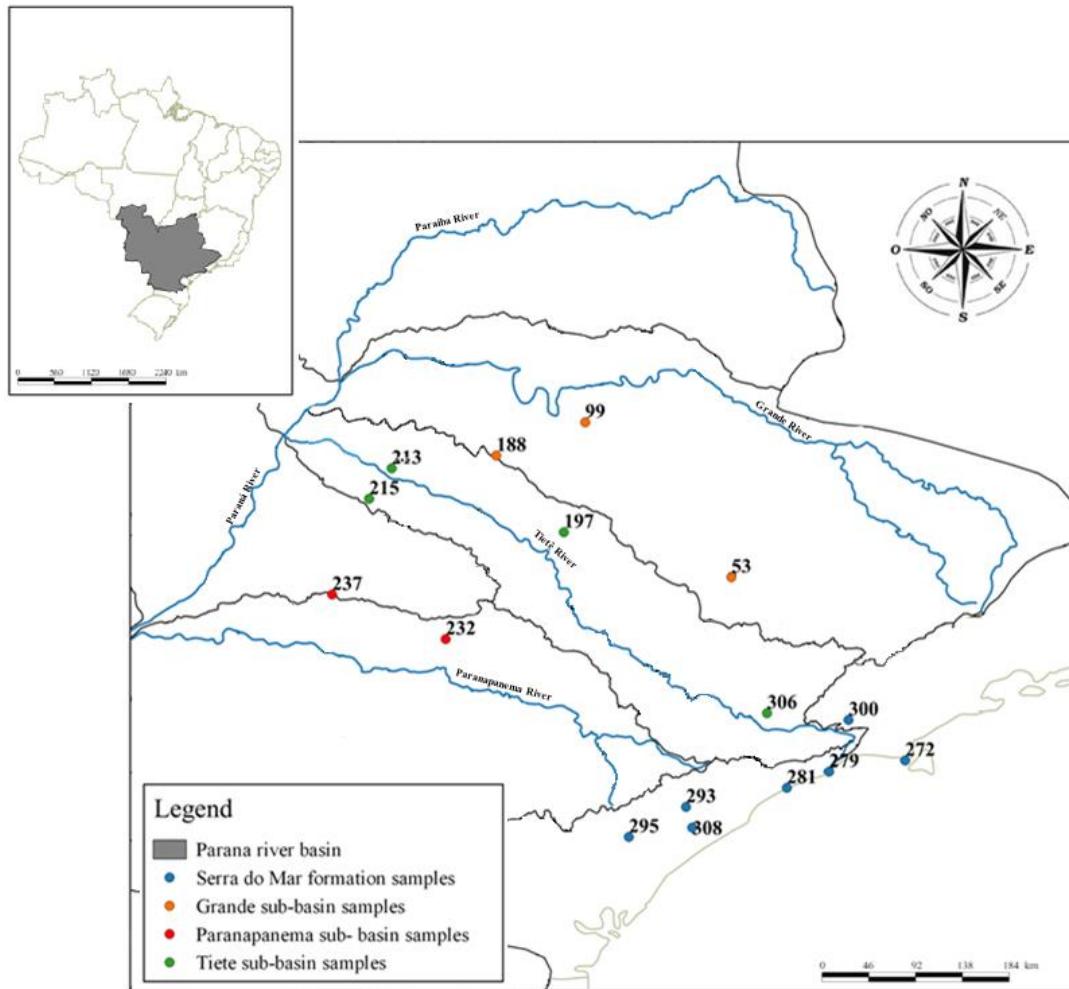


Figure 6 – Location of the 15 shallow lakes samples used in this study.

We obtained an average of 184,114 reads per sample. In order to make these results comparable, total reads per sample were rarefied using the 'rrarefy' function from the 'vegan' package on R software (Oksanen *et al.*, 2016; R. Core Team, 2016). The sample with the lowest number of OTUs had 39,599 reads. The table achieved was a matrix of 3,224 OTUs with a mean of 1,115 OTUs per sample (the minimum and maximum OTUs/samples were 827 and 1303, respectively).

An increased number of OTUs was observed in the organic extraction, while a decreased number was observed in the kit extraction (Fig. 7A). The OTUs observed in the first DNA extraction method were more aggregated and closer to the 1:1 line, while OTUs representing the kit extraction had a more heterogeneous distribution with points more scattered in the bottom-right half of the plot. However, despite there were an increased number of OTUs obtained by the organic extraction, no significant differences could be found between the number of OTUs represented in each of the two extraction methods (Fig. 7B, see also supplementary material Tab. S1 and Tab. S2).

Regarding the question "which taxonomical groups were overrepresented with each extraction method?", about 8% of total classes had statistically significant differences in the paired t-test. Most significant differences were observed in groups which were overrepresented with the organic extraction (Tab. S1 and S2, Fig. 8). Some Classes like Bacteroidia (phylum Bacteroidetes), Melainabacteria (phylum Cyanobacteria), Verrucomicrobia Incertae Sedis (Phylum Verrucomicrobia) and Planctomycetacia (Phylum Planctomycetes) had their relative abundance increased in approximately 75% in the organic extraction comparing to the Soil extraction Kit. The exception was the class Chloroflexia, a rare group of Chloroflexi phylum, which was 28.1% overrepresented with the kit extraction.

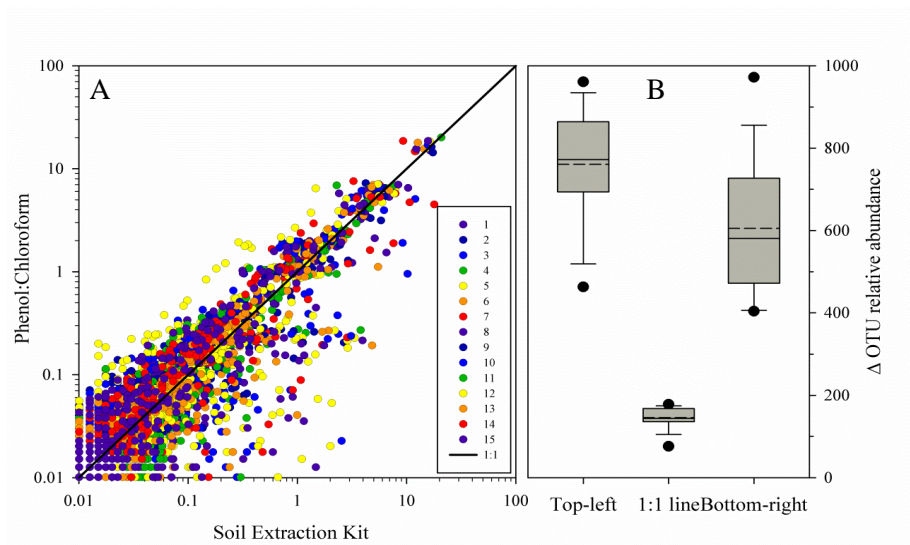


Figure 7 – A) Comparison of all OTU (Operational Taxonomic Units) found in 15 freshwater samples using two DNA extraction methods. The X-axis represents the kit extraction and Y-axis the organic extraction DNA . B) Boxplots represent the number of points overrepresented using the kit extraction (Bottom-right panel from Fig.1A) and the organic extraction method (Top-left panel from Fig.1A).The central full line indicates the median value, the dotted line indicates the arithmetic mean value, the boxes indicate the lower and upper quartiles, the vertical lines indicate the 10th and 90th percentiles, and the dots represent the 5th and 95th percentiles.

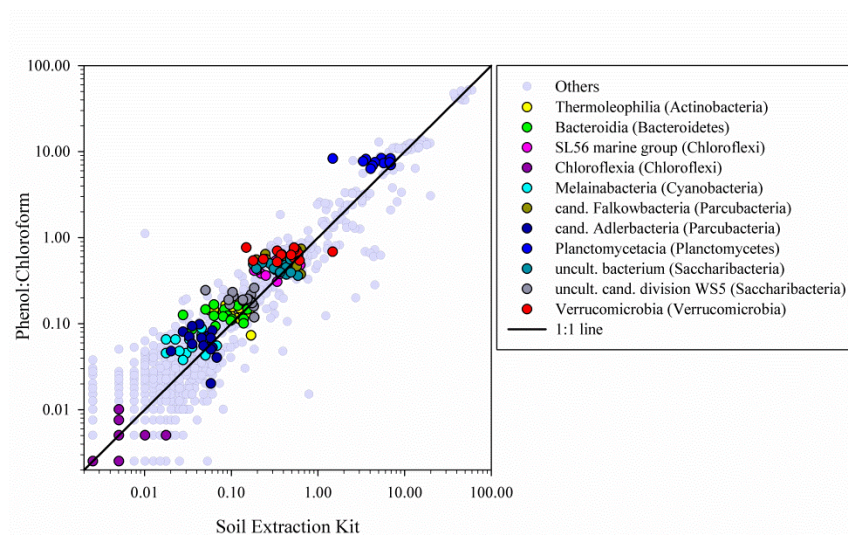


Figure 8 - Comparison of the sum of OTU (Operational Taxonomic Units) relative abundance grouped at class level, obtained by two DNA extraction methods. Highlighted points represent classes that were significantly different ($p < 0.05$, paired t-test) in their relative abundance between extraction methods.

Therefore, the organic extraction was more efficient extracting Gram-negative bacteria (Jeske *et al.*, 2015; Nelson & Stegen, 2015). The use of SDS, lysozyme and other organic compounds in the organic extraction method denature more efficiently the complex carbohydrates and proteins present in the membrane of Gram-negative bacteria. However, some groups which were better extracted by the Phenol:Chloroform method, such as Planctomycetes and Verrucomicrobia, have an unusual membrane composition that surrounds the bacterial DNA that protects genetic material, which cannot yet be called nuclear membrane (Jeske *et al.*, 2015), but might provide an extra barrier to extraction. Our results indicate that the organic extraction method is more efficient in breaking that type of cell wall than the kit extraction, this may be important in cases that the detection of specific groups represent a distinct interpretation of a result (e.g. Janabi *et al.*, 2016).

We also compared some commonly used indexes in community ecology (Tab. 1) and no significant differences in richness, diversity or evenness were observed between the two extraction methods. Nevertheless, a wider variation in all indexes was observed for the extraction kit (Fig. 9, Tab. 1). These results demonstrate that there is no difference between the DNA extraction methods evaluated and that samples using any of those methods could be mixed for community ecology analysis purposes.

This study demonstrates that for purposes of community ecology analysis it is possible to mix samples extracted with organic extraction and kit extraction, provided the DNA is of good quality. Furthermore, our results also suggest that mixing DNA samples extracted from either methods for functional studies might be reasonable as well, even if those minor differences in extraction efficiency might be amplified in metagenomic studies.

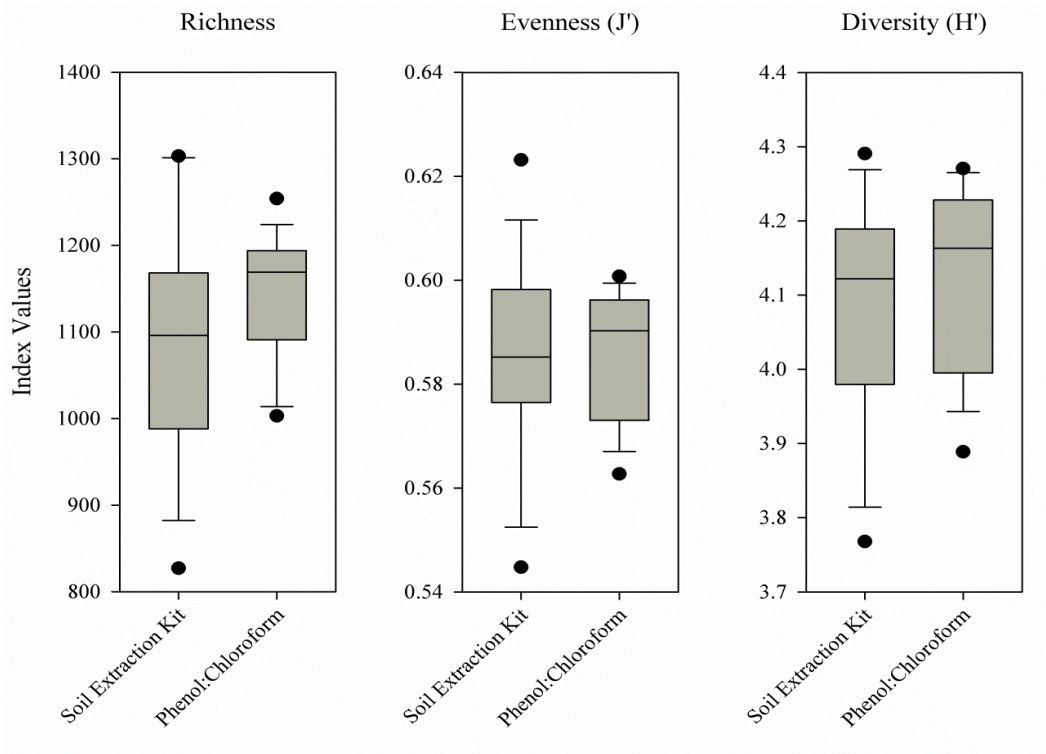


Figure 9 - Comparison of commonly used indexes in community ecology: OTU (Operational Taxonomic Units) richness, Shannon-Weaver diversity and Pielou evenness for each DNA extraction method. No significant differences were observed by paired t-test comparisons. The central full line indicates the median value, the dotted line indicates the arithmetic mean value, the boxes indicate the lower and upper quartiles, the vertical lines indicate the 10th and 90th percentiles, and the dots represent the 5th and 95th percentiles.

Table 1 - Mean, standard deviation (sd) and coefficient of Variation (cv) for community ecology commonly used indexes in samples extracted with the kit extraction and organic extraction.

	SoilExtraction Kit			Phenol:Chloroform		
	<i>mean</i>	<i>sd</i>	<i>cv</i>	<i>mean</i>	<i>sd</i>	<i>cv</i>
Richness (S)	1085	134	0.12	1145	71	0.06
Shannon-Waever (H')	4.09	0.15	0.04	4.12	0.12	0.03
Simpson (D)	0.95	0.01	0.01	0.95	0.01	0.01
Pielou (J')	0.59	0.02	0.03	0.59	0.01	0.02
Hill (E)	0.06	0.01	0.15	0.05	0.00	0.06

Acknowledgments

This research was founded by the São Paulo Research Foundation (FAPESP grants: 2014/14139-3 and 2011/50054-4). We thank Universidade Estadual Paulista – UNESP, Departamento de Tecnologia, Laboratório Multiusuário Centralizado para Sequenciamento de DNA em Larga Escala e Análise de Expressão Gênica – LMSeq for sequencing (grants: 2009/53984-2, 2016/07679-7). We also would like to thank CAPES for the scholarship provided for the first author (EMB) and FAPESP (grant: 2013/18083-0) for the post-doctoral scholarship provided for the third author (ILB). Finally, we are thankful to Cilene C. Mori, Fabrício S. Maeccheri, Guilherme P. de Moraes, Helena H. Vieira, Ingritt C. Moreira, Letícia P. Tessarolli, Lucas S. Tiberti, Luiz A. Joaquim, Naiara C. P. Archanjo, Rodrigo V. de Mello and Thaís G. da Silva who has participated in the field sampling.

Literature Cited

- Arbeli Z, Fuentes CL (2007) Improved purification and PCR amplification of DNA from environmental samples. *FEMS microbiology letters* 272:269-275
- Dai X, Chen S, Li N, Yan H (2016) Quantitative and qualitative validations of a sonication-based DNA extraction approach for PCR-based molecular biological analyses. *Anal Biochem* 501:44-46
- Deiner K, Walser J-C, Mächler E, Altermatt F (2015) Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation* 183:53-63
- Granja-Salcedo YT, Ramirez-Uscategui RA, Machado EG, Duarte Messana J, Takeshi Kishi L, Lino Dias AV, Berchielli TT (2017) Studies on bacterial community composition are affected by the time and storage method of the rumen content. *PloS one* 12:e0176701
- Huggett JF, Novak T, Garson JA, Green C, Morris-Jones SD, Miller RF, Zumla A (2008) Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon. *BMC Res Notes* 1:70
- Janabi AHD, Kerkhof LJ, McGuinness LR, Biddle AS, McKeever KH (2016) Comparison of a modified phenol/chloroform and commercial-kit methods for extracting DNA from horse fecal material. *Journal of Microbiological Methods* 129:14-19
- Jeske O, Schuler M, Schumann P, Schneider A, Boedeker C, Jogler M, Bollschweiler D, Rohde M, Mayer C, Engelhardt H, Spring S, Jogler C (2015) Planctomycetes do possess a peptidoglycan cell wall. *Nature communications* 6:7116
- Kindt R, Coe R (2005) Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies, Vol. World Agroforestry Centre

- Li P, Yang SF, Lv BB, Zhao K, Lin MF, Zhou S, Song X, Tang XM (2015) Comparison of extraction methods of total microbial DNA from freshwater. *Genetics and molecular research : GMR* 14:730-738
- Logares R (2017) ramalok/amplicon_processing: Workflow for Analysing MiSeq Amplicons based on Uparse.
- Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmiento H, Hingamp P, Ogata H, de Vargas C, Lima-Mendez G, Raes J, Poulain J, Jaillon O, Wincker P, Kandels-Lewis S, Karsenti E, Bork P, Acinas SG (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental microbiology* 16:2659-2671
- Mestre M, Borrull E, Sala M, Gasol JM (2017) Patterns of bacterial diversity in the marine planktonic particulate matter continuum. *The ISME journal* 11:999-1010
- Nelson WC, Stegen JC (2015) The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Frontiers in microbiology* 6:713
- Newton RJ, McLellan SL (2015) A unique assemblage of cosmopolitan freshwater bacteria and higher community diversity differentiate an urbanized estuary from oligotrophic Lake Michigan. *Frontiers in microbiology* 6:1028
- Nielsen S, Minchin T, Kimber S, van Zwieten L, Gilbert J, Munroe P, Joseph S, Thomas T (2014) Comparative analysis of the microbial communities in agricultural soil amended with enhanced biochars or traditional fertilisers. *Agriculture, Ecosystems & Environment* 191:73-82
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2016) Vegan: community ecology package.

- Olson ND, Morrow JB (2012) DNA extract characterization process for microbial detection methods development and validation. *BMC Research Notes* 5:668-668
- Pylro VS, Roesch LFW, Morais DK, Clark IM, Hirsch PR, Tótola MR (2014) Data analysis for 16S microbial profiling from different benchtop sequencing platforms. *Journal of Microbiological Methods* 107:30-37
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner F (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41:D590-D596
- R. Core Team (2016) R: A Language and Environment for Statistical Computing. In. R Foundation for Statistical Computing, Vienna, Austria
- Schneegurt MA, Dore SY, Kulpa CF, Jr. (2003) Direct extraction of DNA from soils for studies in microbial ecology. *Curr Issues Mol Biol* 5:1-8
- Seumahu CA, Suwanto A, Rusmana I, Solihin DD (2012) Comparison of DNA Extraction Methods for Microbial Community Analysis in Indonesian Tempe Employing Amplified Ribosomal Intergenic Spacer Analysis. *HAYATI Journal of Biosciences* 19:93-98
- Sidstedt M, Jansson L, Nilsson E, Noppa L, Forsman M, Rådström P, Hedman J (2015) Humic substances cause fluorescence inhibition in real-time polymerase chain reaction. *Anal Biochem* 487:30-37
- Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Environmental DNA. *Molecular ecology* 21:1789-1793
- Wang H, Qi J, Xiao D, Wang Z, Tian K (2017) A re-evaluation of dilution for eliminating PCR inhibition in soil DNA samples. *Soil Biology and Biochemistry* 106:109-118

- 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* 10:1337-1351
- Yilmaz P, Parfrey L, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glockner F (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research* 42:D643-D648
- Zhang Y, Xiao W, Jiao N (2016) Linking biochemical properties of particles to particle-attached and free-living bacterial community structure along the particle density gradient from freshwater to open ocean. *J Geophys Res G Biogeosci* 121:2261-2274

Supplementary material

Study sites and sampling

In this study, we used 15 samples collected from shallow lakes distributed in São Paulo state (Brazil), located in a tropical region, a transition area between the Cerrado (Brazilian savannah) and the Atlantic Forest (a semi-deciduous seasonal forest), all within the coastal mountain formation and in the Paraná river basin, the most important river in this region. Most of this area was replaced by pastures and agricultural monocultures (mainly sugar cane). These shallow lakes were mostly dammed creeks build with the purpose of fulfill human's needs.

To determine the trophic state index, we used the chlorophyll *a* with reference values adapted for tropical regions (Cunha *et al.*, 2013), obtaining a mix of oligotrophic, mesotrophic and eutrophic samples and, thus, ensuring representativeness of different types of freshwater environments and decreasing chances of observed variances being caused by trophic state.

In each shallow lake, several environmental variables were recorded *in situ*, such as temperature, pH, dissolved oxygen and water conductivity, and water samples were taken to determine chlorophyll *a* and DNA extraction. These lakes had a reduced depth, ranging from several centimeters to a few meters and, because this, secchi disk depth was not always possible to determine.

For chlorophyll *a*, 100 to 500 ml of water was passed through a glass fiber filter (Macherey-Nagel® GF-6). Pigments were extracted with ethanol (90% v/v at 80°C) in the dark (Marker, 1980; Nush, 1980), quantified by spectrophotometry (Lorenzen, 1967), and used as a proxy of the trophic state (Cunha *et al.*, 2013).

For molecular analyses, 500 ml were pre-filtered through a glass fiber filter with 1.2 μm mesh (BOECO® MGC), to retain eukaryotes, large particles and attached prokaryotes, and then through polycarbonate 0.22 μm membranes (Millipore® Isopore™ 0.2 μm GTBP) to retain the free-living prokaryotes. Filters were cut in half, immediately frozen and, once arrived to the lab kept at -80°C until DNA extraction.

From the purposes for this work, we used the free-living prokaryotes retention filters and, then, we conducted the extraction step by the two different methods using each half filter.

DNA extraction methods

Extraction by PowerSoil DNA isolation Kit

The Earth Microbiome Project (EMP) proposes a unified protocol to proceed the molecular procedures. For the extraction, EMP proposed the use of the PowerSoil DNA isolation kit as a common method from all environments. As this is a kit for extraction of soil samples, some modifications to the manufacturer's protocol were necessary to optimize the procedure and were proposed by EMP.

Below we describe the adaptation steps of the protocol used for DNA extraction. The non-cited steps are the same as manufacturer's protocol.

The white filter membrane containing filtered water was inserted into the "Power Bead Tubes" and then, manually mixed. After adding the C1 solution, the samples were vortexed for 5 minutes and centrifuged for 1 minute, 10000 RPM. Following the protocol, then we added the C2 solution, and the samples were incubated at -20°C for 5 minutes.

Extraction by Phenol:Chloroform Method

The total DNA was extracted with a Phenol:Chloroform protocol. Briefly, filters were cut in small pieces and incubated with lysis buffer and lysozyme and later with proteinase K and sodium dodecyl sulphate (SDS) 10% and purified in 100KDa Amicon columns (Millipore® 100.000MWCO).

In order to clear the samples that were rich in humic substances, we performed a purification method after the extraction using cetyl trimethyl ammonium bromide (CTAB) 10% (Schneegurt, Mark A *et al.*, 2003). This method consists in one incubation step with NaCl and CTAB followed by a centrifugation with chloroform-isoamyl and a sequence of incubations at -20°C and centrifugations with alcohol 100% and 70%. Finally, the material was dried to eliminate the remaining alcohol and resuspended in pure water. Despite these procedures, we still had to dilute the samples in ultrapure water in a 1:5 proportion in order to avoid PCR inhibition by humic substances.

Amplification and Sequencing

The V3-V4 regions of the 16S rRNA gene were amplified with the 341F (5'-CCTACGGGNGGCWGCAG-3') and 805R (5'-GACT ACHVGGGTATCTAATCC-3') primers (Herlemann *et al.*, 2011), using the KAPA HiFi HotStart ReadyMix PCR kit (Kapa Biosystems®) and the thermocycling protocol proposed by the PCR kit manufacturer. After a beads purification step and multiplex indexation with Nextera XT kit V2 (Illumina®), samples were sequenced in an Illumina MiSeq platform (2 x 250 bp) installed at the Universidade Estadual Paulista – UNESP, Departamento de Tecnologia, Laboratório Multiusuário Centralizado para Sequenciamento de DNA em Larga Escala e Análise de Expressão Gênica – LMSeq.

Bioinformatics and statistics

Sequence data was processed using an UPARSE based pipeline internally implemented (Logares *et al.*, 2014; Logares, 2017). This pipeline uses UPARSE routines (Edgar, 2013) to recognize the paired sequences, clear from singletons and chimera, and, comparing with database SILVA version 128 (Quast *et al.*, 2013; Yilmaz *et al.*, 2014), determine OTUs at 97% similarity threshold.

The statistical analyses were carried in R Software (R. Core Team, 2016). Total reads were obtained after rarefaction by the lowest sample richness using the “*rrarefy*” function from “*vegan*” package (Oksanen *et al.*, 2016). Richness, diversity and evenness indexes were calculated with the “*BiodiversityR*” package (Kindt & Coe, 2005), using Simpson and Shannon-Waever diversity index and Hill’s and Pielou’s index for evenness.

Literature Cited

- Cunha DGF, Calijuri MdC, Lamparelli MC (2013) A trophic state index for tropical/subtropical reservoirs (TSI_{tr}). *Ecological Engineering* 60:126-134
- Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods* 10:996-998
- Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *The ISME journal* 5:1571-1579
- Kindt R, Coe R (2005) Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies, Vol. World Agroforestry Centre
- Logares R (2017) ramalok/amplicon_processing: Workflow for Analysing MiSeq Amplicons based on Uparse.
- Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmiento H, Hingamp P, Ogata H, de Vargas C, Lima-Mendez G, Raes J, Poulain J, Jaillon O, Wincker P, Kandels-Lewis S, Karsenti E, Bork P, Acinas SG (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental microbiology* 16:2659-2671
- Lorenzen CJ (1967) Determination of chlorophyll and phaeo-pigments: spectrophotometric equations. *Limnology and oceanography* 12:343-346
- Marker A (1980) The measurement of photosynthetic pigments in freshwaters and standardization of methods: conclusions and recommendations. *Arch Hydrobiol Ergebn Limnol* 14:91-106
- Nush E (1980) Comparison of different methods for chlorophyll and phaeopigment determination. *Arch Hydrobiol Beih* 14:14-36

- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2016) Vegan: community ecology package.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner F (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41:D590-D596
- R. Core Team (2016) R: A Language and Environment for Statistical Computing. In. R Foundation for Statistical Computing, Vienna, Austria
- Schneegurt MA, Dore SY, Kulpa Jr CF (2003) Direct extraction of DNA from soils for studies in microbial ecology. *Current issues in molecular biology* 5:1-8
- Yilmaz P, Parfrey L, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glockner F (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research* 42:D643-D648

Table S1 - Paired tests for bacterial phylum with p-values and α -values. Groups that shown normal distribution were performed by the Paired t-test, the non-normal data were performed by Wilcoxon Signed Rank Test..

Phylum	Test	P (*<0.05)	Alfa (*>0.8)
Actinobacteria	Paired t Test	0.295	0.065
Proteobacteria	Paired t Test	0.597	0.05
Cyanobacteria	Paired t Test	0.3	0.063
Planctomycetes	Paired t Test	<0.001*	0.998*
Parcubacteria	Paired t Test	0.08	0.308
Verrucomicrobia	Paired t Test	0.042*	0.454
Bacteroidetes	Wilcoxon Signed Rank Test	0.421	-
Saccharibacteria	Wilcoxon Signed Rank Test	0.151	-
Chloroflexi	Paired t Test	0.006*	0.836*
Chlamydiae	Wilcoxon Signed Rank Test	0.762	-
Candidatus Berkelbacteria	Paired t Test	0.051	0.41
TM6 (Dependentiae)	Paired t Test	0.056	0.386
Firmicutes	Wilcoxon Signed Rank Test	0.296	-
Gracilibacteria	Paired t Test	0.121	0.22
CPR2	Paired t Test	0.436	0.05
Spirochaetae	Wilcoxon Signed Rank Test	0.679	-
Omnitrophica	Paired t Test	0.677	0.05
Acidobacteria	Paired t Test	0.331	0.05
WWE3	Paired t Test	0.464	0.05
Microgenomates	Paired t Test	0.14	0.191
Peregrinibacteria	Paired t Test	0.241	0.096
Gemmatimonadetes	Wilcoxon Signed Rank Test	0.426	-
SR1 (Absconditabacteria)	Paired t Test	0.086	0.292
Chlorobi	Wilcoxon Signed Rank Test	0.639	-
Fusobacteria	Paired t Test	0.162	0.164
Elusimicrobia	Wilcoxon Signed Rank Test	0.301	-
Latescibacteria	Paired t Test	0.105	0.25
Lentisphaerae	Wilcoxon Signed Rank Test	0.359	-
WS2	Paired t Test	0.277	0.075
WS6	Wilcoxon Signed Rank Test	0.938	-
Nitrospirae	Wilcoxon Signed Rank Test	0.054	-
Deinococcus-Thermus	Wilcoxon Signed Rank Test	0.844	-
Fibrobacteres	Wilcoxon Signed Rank Test	0.375	-
Ignavibacteriae	Wilcoxon Signed Rank Test	0.75	-
Armatimonadetes	Wilcoxon Signed Rank Test	1	-
Tenericutes	Wilcoxon Signed Rank Test	0.75	-

Table S2 - Paired tests for bacterial classes with p-values and α -values. Groups that shown normal distribution were performed by the Paired t-test, the non-normal data were performed by Wilcoxon Signed Rank Test. The significative values are represented in bold and marked with an “**”.

Phylum	Class	Test	P (*<0.05)	Alfa (*>0.8)
Acidobacteria	Subgroup 7	Paired t Test	0.971	0.05
Acidobacteria	Holophagae	Paired t Test	0.166	0.159
Acidobacteria	Blastocatellia	Paired t Test	0.415	0.05
Acidobacteria	Acidobacteria	Wilcoxon Signed Rank Test	0.813	-
Acidobacteria	Solibacteres	Wilcoxon Signed Rank Test	0.5	-
Acidobacteria	Subgroup 2	Wilcoxon Signed Rank Test	1	-
Actinobacteria	Actinobacteria	Paired t Test	0.861	0.05
Actinobacteria	Acidimicrobiia	Wilcoxon Signed Rank Test	0.33	-
Actinobacteria	Thermoleophilia	Paired t Test	0.027*	0.549
Actinobacteria	MB-A2-108	Paired t Test	0.223	0.108
Actinobacteria	Coriobacteriia	Wilcoxon Signed Rank Test	1	-
Actinobacteria	OPB41	Wilcoxon Signed Rank Test	0.375	-
Armatimonadetes	Fimbriimonadia	Wilcoxon Signed Rank Test	1	-
Bacteroidetes	Sphingobacteriia	Wilcoxon Signed Rank Test	0.252	-
Bacteroidetes	Cytophagia	Paired t Test	0.645	0.05
Bacteroidetes	Bacteroidia	Paired t Test	0.004*	0.87*
Bacteroidetes	Flavobacteriia	Paired t Test	0.552	0.005
Candidatus Berkelbacteria	uncultured bacterium	Paired t Test	0.058	0.381
Candidatus Berkelbacteria	uncultured soil bacterium	Wilcoxon Signed Rank Test	0.08	-
Candidatus Berkelbacteria	Berkelbacteria bacterium GW2011 GWE1 39 13	Wilcoxon Signed Rank Test	0.461	-
Candidatus Berkelbacteria	uncultured candidate division WS6 bacterium	Wilcoxon Signed Rank Test	0.563	-
Candidatus Berkelbacteria	uncultured organism	Wilcoxon Signed Rank Test	1	-
Chlamydiae	Chlamydiae	Wilcoxon Signed Rank Test	0.762	-
Chlamydiae	LD1-PA33	Wilcoxon Signed Rank Test	0.844	-
Chlorobi	Chlorobia	Wilcoxon Signed Rank Test	0.639	-
Chloroflexi	SL56 marine group	Paired t Test	0.005*	0.859*
Chloroflexi	Chloroflexia	Wilcoxon Signed Rank Test	0.042*	-
Chloroflexi	Thermomicrobia	Paired t Test	0.31	0.059
Chloroflexi	JG30-KF-CM66	Paired t Test	0.253	0.089
Chloroflexi	Dehalococcoidia	Wilcoxon Signed Rank Test	0.563	-
Chloroflexi	SAR202 clade	Wilcoxon Signed Rank Test	0.125	-
Chloroflexi	KD4-96	Wilcoxon Signed Rank Test	1	-
Chloroflexi	Ktedonobacteria	Wilcoxon Signed Rank Test	0.5	-
CPR2	uncultured bacterium	Paired t Test	0.552	0.05
CPR2	uncultured Clostridium sp.	Paired t Test	0.299	0.064
CPR2	candidate division CPR2 bacterium GW2011	Wilcoxon Signed Rank Test	0.188	-

	GWD2 39 7			
Cyanobacteria	Cyanobacteria	Paired t Test	0.312	0.058
Cyanobacteria	Melainabacteria	Paired t Test	0.001*	0.966*
Cyanobacteria	ML635J-21	Wilcoxon Signed Rank Test	0.622	-
Deinococcus-Thermus	Deinococci	Wilcoxon Signed Rank Test	0.844	-
Elusimicrobia	Elusimicrobia	Wilcoxon Signed Rank Test	0.301	-
Elusimicrobia	Fibrobacteria	Wilcoxon Signed Rank Test	0.375	-
Firmicutes	Bacilli	Wilcoxon Signed Rank Test	0.454	-
Firmicutes	Clostridia	Wilcoxon Signed Rank Test	0.119	-
Firmicutes	Negativicutes	Paired t Test	0.363	0.05
Firmicutes	Erysipelotrichia	Wilcoxon Signed Rank Test	1	-
Fusobacteria	Fusobacteriia	Paired t Test	0.162	0.164
GAL15	uncultured bacterium	Wilcoxon Signed Rank Test	0.426	-
Gemmatimonadetes	Gemmatimonadetes	Paired t Test	0.242	0.095
Gracilibacteria	uncultured bacterium	Paired t Test	0.415	0.05
Gracilibacteria	mixed culture isolate koll6	Paired t Test	0.073	0.327
Gracilibacteria	uncultured soil bacterium	Paired t Test	0.167	0.158
Gracilibacteria	Candidatus Peregrinibacteria bacterium GW2011 GWA2 33 10	Wilcoxon Signed Rank Test	0.078	-
Gracilibacteria	Candidatus Peregrinibacteria bacterium GW2011 GWF2 43 17	Wilcoxon Signed Rank Test	0.438	-
Gracilibacteria	wastewater metagenome	Wilcoxon Signed Rank Test	1	-
Gracilibacteria	uncultured Microgenomates group bacterium	Wilcoxon Signed Rank Test	1	-
Ignavibacteriae	Ignavibacteria	Wilcoxon Signed Rank Test	0.75	-
Latescibacteria	uncultured bacterium	Paired t Test	0.228	0.105
Latescibacteria	uncultured Latescibacteria bacterium	Paired t Test	0.07	0.338
Lentisphaerae	Lentisphaeria	Wilcoxon Signed Rank Test	0.359	-
Microgenomates	Candidatus Woesebacteria	Paired t Test	0.285	0.07
Microgenomates	Candidatus Levybacteria	Paired t Test	0.363	0.05
Microgenomates	Candidatus Pacebacteria	Paired t Test	0.104	0.25
Microgenomates	Candidatus Beckwithbacteria	Paired t Test	0.271	0.078
Microgenomates	uncultured bacterium	Paired t Test	0.683	0.05
Microgenomates	candidate division CPR3 bacterium GW2011 GWF2 35 18	Wilcoxon Signed Rank Test	0.305	-
Microgenomates	Candidatus Shapirobacteria	Paired t Test	1	0.05
Microgenomates	Candidatus Collierbacteria	Wilcoxon Signed Rank Test	0.685	-
Nitrospirae	Nitrospira	Wilcoxon Signed Rank Test	0.054	-
Omnitrophica	uncultured bacterium	Paired t Test	0.71	0.05
Omnitrophica	Omnitrophica Incertae Sedis	Paired t Test	0.386	0.05

Omnitrophica	uncultured planctomycete	Paired t Test	0.562	0.05
Omnitrophica	uncultured organism	Paired t Test	0.894	0.05
Omnitrophica	uncultured Verrucomicrobia bacterium	Paired t Test	0.595	0.05
Omnitrophica	uncultured Banisveld landfill bacterium BVC57	Wilcoxon Signed Rank Test	0.297	-
Omnitrophica	uncultured Omnitrophica bacterium	Wilcoxon Signed Rank Test	0.5	-
Omnitrophica	uncultured Firmicutes bacterium	Wilcoxon Signed Rank Test	1	-
Parcubacteria	Candidatus Nomurabacteria	Paired t Test	0.149	0.179
Parcubacteria	uncultured bacterium	Paired t Test	0.029	0.533
Parcubacteria	Candidatus Moranbacteria	Paired t Test	0.172	0.152
Parcubacteria	Candidatus Azambacteria	Paired t Test	0.137	0.195
Parcubacteria	Candidatus Falkowbacteria	Paired t Test	0.028*	0.541
Parcubacteria	Candidatus Magasanikbacteria	Paired t Test	0.071	0.334
Parcubacteria	Parcubacteria bacterium GW2011	Paired t Test	0.093	0.275
Parcubacteria	Candidatus Uhrbacteria	Wilcoxon Signed Rank Test	0.188	-
Parcubacteria	Candidatus Adlerbacteria	Paired t Test	0.037*	0.479
Parcubacteria	wastewater metagenome	Paired t Test	0.873	0.05
Parcubacteria	Candidatus Kuenenbacteria	Paired t Test	0.022	0.597
Parcubacteria	Candidatus Jorgensenbacteria	Paired t Test	0.13	0.205
Parcubacteria	uncultured Parcubacteria group bacterium	Paired t Test	0.056	0.386
Parcubacteria	Candidatus Wolfebacteria	Paired t Test	0.524	0.05
Parcubacteria	Candidatus Campbellbacteria	Wilcoxon Signed Rank Test	0.639	-
Parcubacteria	Parcubacteria Incertae Sedis	Paired t Test	0.571	0.05
Parcubacteria	uncultured anaerobic bacterium	Paired t Test	0.351	0.05
Parcubacteria	uncultured planctomycete	Paired t Test	0.413	0.05
Parcubacteria	marine metagenome	Wilcoxon Signed Rank Test	0.438	-
Parcubacteria	Parcubacteria bacterium OLB19	Wilcoxon Signed Rank Test	0.438	-
Parcubacteria	groundwater metagenome	Wilcoxon Signed Rank Test	0.875	-
Parcubacteria	Candidatus Yanofskybacteria	Wilcoxon Signed Rank Test	0.125	-
Parcubacteria	Parcubacteria bacterium DG 74 3	Wilcoxon Signed Rank Test	1	-
Parcubacteria	uncultured deep-sea bacterium	Wilcoxon Signed Rank Test	0.5	-
Parcubacteria	Candidatus Giovannonibacteria	Wilcoxon Signed Rank Test	0.75	-
Parcubacteria	uncultured organism	Wilcoxon Signed Rank Test	0.75	-
Peregrinibacteria	uncultured organism	Paired t Test	0.324	0.052

Peregrinibacteria	Candidatus Peribacteria	Paired t Test	0.355	0.05
Planctomycetes	Planctomycetacia	Paired t Test	<0.001*	0.998*
Planctomycetes	Phycisphaerae	Paired t Test	0.221	0.11
Planctomycetes	vadinHA49	Wilcoxon Signed Rank Test	0.578	-
Proteobacteria	Betaproteobacteria	Paired t Test	0.362	0.05
Proteobacteria	Alphaproteobacteria	Paired t Test	0.115	0.231
Proteobacteria	Gammaproteobacteria	Wilcoxon Signed Rank Test	0.847	-
Proteobacteria	Deltaproteobacteria	Wilcoxon Signed Rank Test	0.454	-
Proteobacteria	Epsilonproteobacteria	Paired t Test	0.792	0.05
Proteobacteria	SPOTSOCT00m84	Wilcoxon Signed Rank Test	0.844	-
Saccharibacteria	uncultured bacterium	Paired t Test	0.042*	0.45
Saccharibacteria	uncultured Candidatus Saccharibacteria bacterium	Wilcoxon Signed Rank Test	0.121	-
Saccharibacteria	uncultured candidate division WS5 bacterium	Paired t Test	0.005*	0.846*
Saccharibacteria	uncultured soil bacterium	Paired t Test	0.869	0.05
Saccharibacteria	bacterium WWH38	Wilcoxon Signed Rank Test	0.083	-
Saccharibacteria	Unknown Class	Wilcoxon Signed Rank Test	1	-
Spirochaetae	Spirochaetes	Wilcoxon Signed Rank Test	0.679	-
SR1 (Absconditabacteria)	uncultured candidate division SR1 bacterium	Paired t Test	0.203	0.124
SR1 (Absconditabacteria)	uncultured bacterium	Paired t Test	0.28	0.073
SR1 (Absconditabacteria)	uncultured organism	Wilcoxon Signed Rank Test	0.813	-
Tenericutes	Mollicutes	Wilcoxon Signed Rank Test	0.75	-
TM6 (Dependentiae)	uncultured bacterium	Paired t Test	0.06	0.372
TM6 (Dependentiae)	uncultured soil bacterium	Paired t Test	0.61	0.05
TM6 (Dependentiae)	uncultured prokaryote	Wilcoxon Signed Rank Test	1	-
TM6 (Dependentiae)	candidate division TM6 bacterium GW2011 GWF2 30 66	Wilcoxon Signed Rank Test	0.375	-
TM6 (Dependentiae)	candidate division TM6 bacterium JCVI TM6SC1	Wilcoxon Signed Rank Test	0.375	-
Verrucomicrobia	Spartobacteria	Wilcoxon Signed Rank Test	0.252	-
Verrucomicrobia	Opitutae	Wilcoxon Signed Rank Test	0.561	-
Verrucomicrobia	Verrucomicrobia Incertae Sedis	Wilcoxon Signed Rank Test	0.015*	-
Verrucomicrobia	OPB35 soil group	Paired t Test	0.383	0.05
Verrucomicrobia	Verrucomicrobiae	Wilcoxon Signed Rank Test	0.123	-
WS3	uncultured bacterium	Paired t Test	0.277	0.075
WS7	uncultured bacterium	Wilcoxon Signed Rank Test	0.938	-
WWE3	uncultured candidate division WWE3 bacterium	Paired t Test	0.494	0.05
WWE3	uncultured bacterium	Paired t Test	0.425	0.05
WWE3	uncultured anaerobic bacterium	Wilcoxon Signed Rank Test	0.502	-

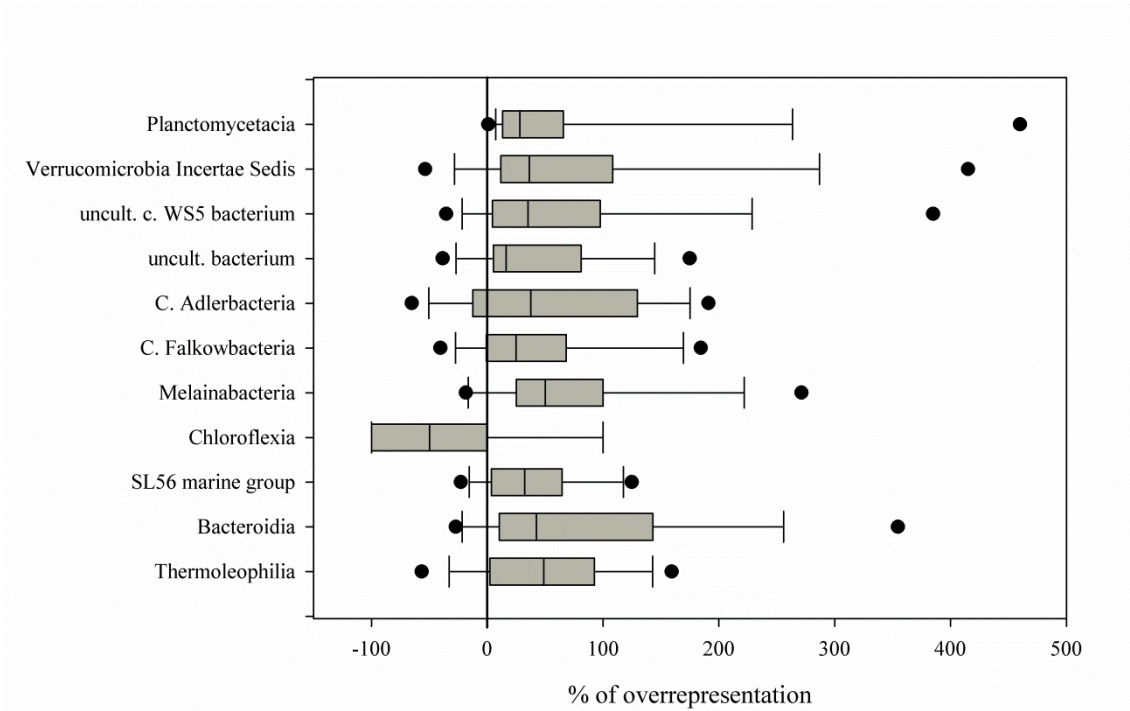


Figure S 1 - Boxplot representing the percentage of bacterial classes overrepresentation which were significantly different ($p < 0.05$).

Testing the occupancy-frequency distribution in freshwater bacterial communities

Manuscript under preparation to be submitted to Ecology

Abstract

Understanding how populations are distributed at different spatial scales is a fundamental question in ecology. The Raunkiaer's rule predicts a bimodal occupancy-frequency distribution with some species found in all locations (core), and other found in only one location (satellite). Here, we aimed to test if freshwater bacterial communities had a bimodal occupancy-frequency distribution, and to explore its consequences for their regional populations distribution, analyzing high throughput sequencing data from samples collected in 60 shallow lakes evenly distributed in a 3.000 kilometers perimeter. We divided the dataset in four sub-regions by their geological configuration, with the aim to characterize the core community of tropical

shallow lakes and to relate organisms' ubiquity to their abundance. We found that, despite the low connectivity between sites and the large area covered by our sampling, bacterial communities followed a significant bimodal occupancy-frequency distribution. The core populations were generally, but not always, composed by the most abundant organisms, and presented two consistent patterns: the "abundant core" populations with a global distribution, and the "rare core" populations, in most cases encountered in only one sub-region (and never in all sites). Furthermore, there was a geographical coherence in regional bimodal distributions in detriment of a random effect of increasing area. These results demonstrate the robustness of bimodality even in an extreme situation such as highly unconnected water bodies combined with high dispersive capacity, and highlight the role of biogeographical processes (well known for macro-organisms) on microbial communities.

Introduction

Raunkiaer's rule states that species dispersion display a bimodal frequency distribution in a set of local samplings (Gleason, 1929; Papp and Izsák, 1997), where most organisms are encountered only in one or few sites, while some others can be found in all (or most) sites. When species sampling is ideal, bimodality should be observed, where the major number of species would be found in the first (0-20%) and the last (81-100%) frequency classes (Papp and Izsák, 1997), at least for terrestrial macro-organisms.

This rule gained popularity when Hanski (Hanski, 1982) used its assumptions to explain the distribution of species in the context of metapopulations studies (Papp and Izsák, 1997), and became one of the most prominent models that opposed to the unimodality (Levins, 1969).

While Levins (1969) considered in his mathematical models that there was a relationship between immigration and extinction rates, which eventually created a distribution model with only one hump (or an unimodal distribution of species), Hanski (1982) argued that local species extinctions were independent of immigration rates, but both depend on regional species distributions. In this mathematical model it was possible to observe two humps on frequency-occupancy distribution plots, one at each end of a distribution graph, Hanski called organisms that fall on each extreme of: "core" and "satellite" organisms (Levins, 1969; Hanski, 1982; Gotelli, 1991).

The bimodal hypothesis, also known as the "core-satellite hypothesis" or the "occupancy-frequency distribution", advocates that there is a clear relationship between the local abundance and the regional distributions of species (Hanski, 1982). Therefore, the most abundant organisms tend to be found in all sampling sites (Preston, 1948; Brown, 1984), probably by a mix of deterministic and stochastic spatiotemporal processes that select a set of

organisms (Brown, 1984; Magurran and Henderson, 2003), allied to the dispersal capability intrinsic to each species (Brown, 1984; Mehranvar and Jackson, 2001) providing a “rescue effect” that facilitates the recolonization of core species when they are locally extinct (Hanski and Gyllenberg, 1993; Gaston *et al.*, 2000). In this sense, satellite species can also be understood as a potential pool of rare species (reactive seed bank) that can become core if the environmental conditions fluctuate enough to allow better-adapted species to settle (Magurran and Henderson, 2003; Ruiz-González *et al.*, 2017).

There is also an inverse relationship between the scale of the studied area and the strength of the bimodal distribution (Van Rensburg *et al.*, 2000). The two more probable causes of this pattern is a range area effect, in which the increase of area implies a decrease in chances to detect core organisms, whereas at lower scales the dispersal capacity of the organisms allied to specific rescue effects increase the chances to individual recolonizations (Van Rensburg *et al.*, 2000).

This hypothesis has been key to understand biogeographical distributions patterns of key species (Lindh *et al.*, 2017), and has been tested in a variety of environments, mainly for macro-organisms (e.g. Van Rensburg *et al.*, 2000; Mehranvar and Jackson, 2001; Heino, 2015; Jokimäki *et al.*, 2016; Tonkin *et al.*, 2016), but also begins to be explored for microorganisms (Soininen and Heino, 2005; Lindh *et al.*, 2017). Despite the bimodality has not been observed in freshwater diatoms (Soininen and Heino, 2005), a study using a marine bacterial dataset has demonstrated the presence of a bimodal distribution pattern (Lindh *et al.*, 2017).

Our understanding of microbial biogeographical patterns has increased (Martiny *et al.*, 2006; Logue and Lindström, 2008) and we are now beginning to acknowledge the importance of microorganisms as a model to test mathematical models and to evaluate dispersal

mechanisms and community structure in space and time (Barberan *et al.*, 2014). Still, the occupancy-frequency hypothesis may be used to understand freshwater bacteria spatial distribution patterns.

In order to test the limits of bacterial dispersal capacity, we tested the bimodal hypothesis in bacterial communities of headwater shallow lakes with no water flow connection between them, using a mathematical model that reflect the degree of connectivity versus isolation of these communities.

In this way, we tested the core-satellite hypothesis in a set of 60 shallow lakes, with the aim to analyze if bimodal patterns in occupancy-frequency distributions of tropical freshwater bacterial communities could be observed, and their consequences for the regional distribution of microbial populations. We also divided the study area in four coherent geographical regions (that we called sub-regions) and in three random subsets with increasing area, to test if there is geographical coherence in OTU distribution or if it is simply a neutral effect of the area covered by our sampling. We hypothesized that: 1) freshwater bacterial populations have a bimodal occupancy-frequency distribution; 2) core-satellite distributions reflect local abundance, with most abundant organisms being present in the core community, and rare organisms in the satellite community; 3) this bimodal distribution of occupancy-frequency would continue to be observed at smaller scales, with a stronger bimodality pattern within each sub-region compared to the total area; and 4) there is geographical coherence in regional bimodal distributions, with stronger bimodality in geographic sub-regions than in random distributed samples in increasing areas. Moreover, we wanted to characterize freshwater's core community, in a scale over than 3.000 kilometers perimeter of a tropical region and, in ultimately, to analyze the limits of the microbial dispersive capacity and also the limits of the bimodal hypothesis.

Methods

Study sites and sampling

We collected bacterial samples in 60 shallow lakes scattered in all São Paulo state (Brazil), located in a tropical region, in a transition area between the Cerrado (Brazilian savannah biome), and the tropical Atlantic Forest biome, a semi-deciduous seasonal forest. The geography includes part of the Brazilian coastal formation and the Paraná river basin, the most important hydrological formation of this region (Fig. 10, Tab. 2). However, most of the area was replaced by pastures and agricultural monocultures (mainly sugar cane). Most of these shallow lakes are dammed creeks build with the purpose of fulfill human's needs.

We used chlorophyll *a* (chl *a*) to determine the trophic state index using reference values adapted for tropical regions (Cunha *et al.*, 2013). Based on the trophic state we selected 60 samples (from a total of 326 samples taken) forming a sample set with an equal number of oligotrophic, mesotrophic and eutrophic lakes in order to reach an homogeneous distributed set of samples in order to avoid variability linked to trophic state.

The 60 samples were equally distributed in the four sub-regions that were named after their most important geographical features: 1) 15 samples located within the Tietê river basin, called “Tietê” (T); 2) 15 samples located in the southwest part of the state, and comprehends the major part of Paranapanema river basin and other small tributaries, here called “Paranapanema” (P); 3) 15 samples, located at the north part of the state, that contains a set of micro-basins called according to the largest river of the region, the “Grande” river (G); 4) Coastal area that comprehends a mountain formation in the east part of the state, starting contiguously with the others sub-regions described in an altitude around of 800 meters and ends up in the ocean, and here called “Serra do Mar” (S) (Fig. 10, Fig. 11). Aiming to

facilitate the following discussion, we named the core organisms encountered in all samples as “global” and these encountered in only one region as “regional” ones.

To evaluate the basin-independent range area effect on the occupancy-frequency distribution, we defined three sample subsets containing 20 random samples within three concentric circles (increasing in area), regardless of the draining basin or geological formation (Fig. 12), named as: small, medium and large.

In each site, several environmental variables were recorded *in situ*, such as temperature, pH, dissolved oxygen and water conductivity (Tab. S3), and therefore water samples were taken to determine chlorophyll *a* and collect material for DNA extraction.

For chlorophyll *a*, 100 to 500 ml of water was passed through a glass fiber filter (Macherey-Nagel® GF-6). Pigments were extracted with ethanol (90% v/v at 80°C) in the dark (Marker, 1980; Nush, 1980), quantified by spectrophotometry (Lorenzen, 1967), and used as a proxy of the trophic state (Cunha *et al.*, 2013).

For molecular analyses, 400 to 500 ml were pre-filtered through a glass fiber filter with 1.2 µm mesh (BOECO® MGC), to retain eukaryotes, large particles and attached prokaryotes, and then through polycarbonate 0.22 µm membranes (Millipore® Isopore™ 0.2 µm GTBP) to retain the free-living prokaryotes. The filters were cut in a half, immediately frozen and, once arrived to the lab kept at -80°C until DNA extraction.

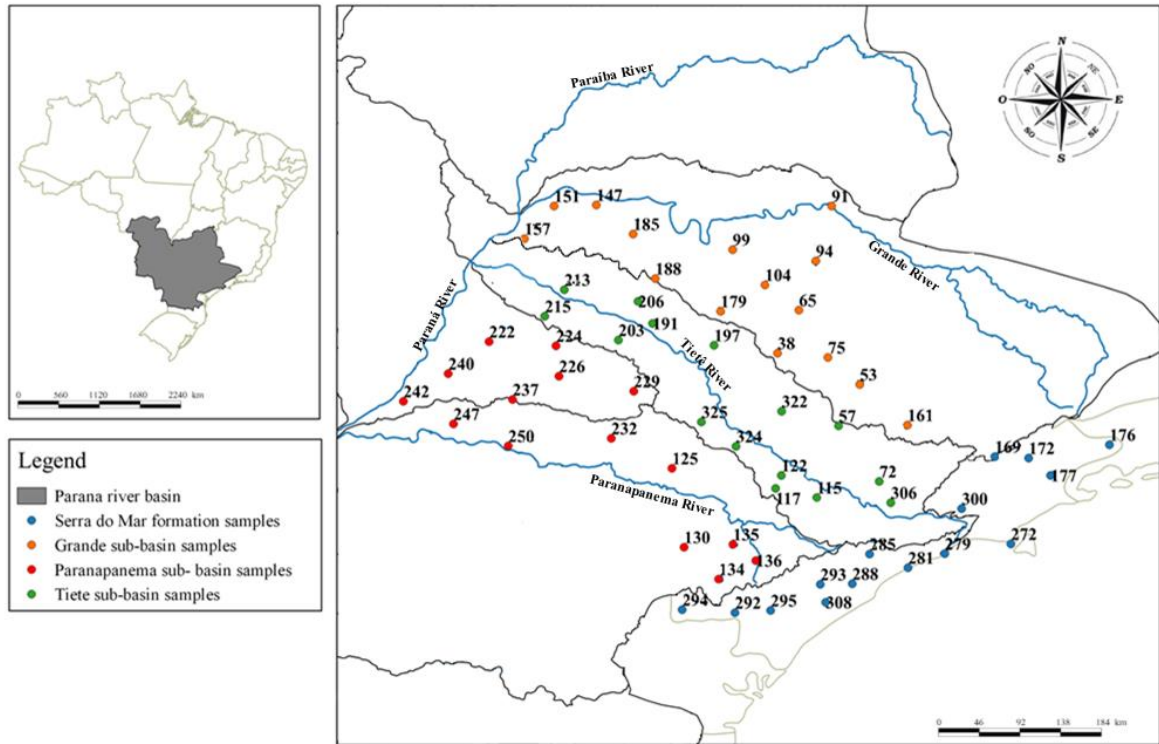


Figure 10 – Location of the 60 sample sites, divided between 3 different sub-basins inserted into Paraná river basin (colored in Grey) and the adjacent coastal mountain formation. Samples were named T for Tietê river basin (green), P for Paranapanema river basin (blue), G for Grande river basin (Orange) and S for Serra do Mar formation (blue).

Table 2 – General features of the study area and environmental conditions in the shallow lakes studied (average values).

	Tietê Basin	Parapanema Basin	Grande Basin	Serra do Mar Formation
Area (Km ²)	1910	2083	1669	991
Perimeter (Km)	1706	1647	1575	1607
River length (Km)	1120	929	1286	-
Maximal length (km)	658	572	414	146
Maximal width (km)	134	230	178	540
Vegetation	mostly Atlantic Forest	mostly Atlantic Forest	mostly Cerrado-Savannah	Atlantic Forest
pH	6.54	7.21	6.66	6.65
Temperature (°C)	23.60	21.49	24.99	22.55
chlorophyll <i>a</i> (mg/L)	23.42	15.39	9.20	18.46
Dissolved Organic Carbon (mg/L)	9.41	8.91	5.67	11.01
Nitrate (mg/L)	0.11	3.60	0.21	0.16
Phosphate (mg/L)	0.00	0.01	0.05	0.01

DNA extraction and amplification

The DNA was extracted with a Phenol:Chloroform protocol. Briefly, filters were cut in small pieces and incubated (37°C; 45 minutes) with lysis buffer (3 mL) and lysozyme (75µL) and later, incubated (55°C; 60 minutes) with proteinase K (75µL) and sodium dodecyl sulphate (SDS) 10% (300µL), after the incubation, the extraction were performed in two centrifugation steps using isoamil phenol-chloroform (2mL; 10 minutes; 5000 rpm) and another centrifugation with isoamil chloroform (2mL; 10 minutes; 5000 rpm), and, finally, we purified the extracted material in 100KDa Amicon columns (Millipore® 100.000MWCO).

In order to clear the samples that were rich in humic substances, we performed a purification method after the extraction using cetyl trimethyl ammonium bromide (CTAB)

10% (Schneegurt *et al.*, 2003). This method consists in one incubation step (65°C; 15 minutes) with NaCl 5M (16.25µL) and CTAB (hot; 12µL) followed by a centrifugation with chloroform-isoamyl (128µL), an incubation with alcohol 100% (freeze; 256µL) at -20°C (overnight) followed by a centrifugation (5 minutes; 5000 rpm), and a last centrifugation with alcohol 70% (freeze; 125µL; 5 minutes; 5000 rpm). Finally, the material was dried to eliminate the remaining alcohol and resuspended in pure water. Despite these procedures, we still had to dilute the samples in ultrapure water in a 1:9 proportion in order to avoid PCR inhibition by humic substances.

The V3-V4 regions of the 16S rRNA gene were amplified with the 341F (5'-CCTACGGGNGGCWGCAG-3') and 805R (5'-GACT ACHVGGGTATCTAATCC-3') primers (Herlemann *et al.*, 2011), using the KAPA HiFi HotStart ReadyMix PCR kit (Kapa Biosystems®) and the thermocycling protocol proposed by the PCR kit manufacturer. After, a purification step using magnetic beads (Agencourt® AMPure® XP) and multiplex indexation with Nextera XT kit V2 (Illumina®), samples were sequenced in an Illumina MiSeq platform installed at the Universidade Estadual Paulista – UNESP, Departamento de Tecnologia, Laboratório Multiusuário Centralizado para Sequenciamento de DNA em Larga Escala e Análise de Expressão Gênica – LMSeq.

Bioinformatics and statistics

Sequence data was processed using an UPARSE based pipeline internally implemented (Logares *et al.*, 2014; Logares, 2017). This pipeline uses UPARSE routines (Edgar, 2013) to recognize the paired sequences, clear from singletons and chimera, and, comparing with database SSU-SILVA version 128 (Quast *et al.*, 2013; Yilmaz *et al.*, 2014), determine the Operational Taxonomic Units (OTUs) at 97% similarity threshold. Since the vast majority of

microorganisms is yet uncultured, the current prokaryotic species definition is based in similarities in genomic characters (Rosselló-Mora and Amann, 2001). Most microbial ecology studies use the term “OTU”, considering similarity of marker gene sequence (16S rRNA) with a 97% similarity threshold (Stackebrandt and Goebel, 1994) as a proxy for prokaryotic species.

The statistical analyses were carried in R Software (R. Core Team, 2016), using the following packages: “stats” (R. Core Team, 2016), “RColorBrewer” (Neuwirth, 2014), “reshape” (Wickham, 2007) and “ggplot2” (Wickham, 2009). Total reads were obtained after rarefaction by the lowest sample richness using the “rrarefy” function from “vegan” package (Oksanen *et al.*, 2016).

Bimodality was tested using the “MOSstest” function from “vegan” package (Oksanen *et al.*, 2016). The “MOSstest” function uses the Mitchell-Olds & Shaw’s test of quadratic extremes (Mitchell-Olds and Shaw, 1987) to determine if the data analyzed have a Gaussian occupancy-frequency distribution, so the pattern will be considered bimodal if the h_0 is denied and form an inverse-Gaussian shape (Ter Braak and Looman, 1986; Oksanen *et al.*, 2016). For the Venn’s plot, we used the browser application Venny 2.0 (Oliveros, 2007-2015).

We separated the core and satellite OTUs by their abundance: the “abundant” group include organisms that represented more than 1% of the total abundance in each site, while the “rare” group includes organisms that represented less than 0.1% of local abundance according to the definition proposed in Pedrós-Alió (2006). Here, we defined as “abundant core” OTUs those that were abundant at least in half of the samples (thus, more than 30 samples and more than 7 considering each sub-region), and as “rare core” those that were rare in more than half of the samples. It is important to emphasize that we are dealing with core

organisms, that were present in all samples and this definition only serves to relate occupancy and abundance.

For satellite organisms we defined the “abundant satellite” as those that were abundant in the only site where they were found and the “rare satellite” as those that were rare (<0.1%) in their site

Results

Exploring the dataset

We found a total of 882120 reads (average of 14702 sequences per sample) that were clustered in 3946 OTUs (Tab. 3). The most diverse phylum was Parcubacteria containing 1737 OTUs, followed by Proteobacteria (853), Actinobacteria (149), Chlamydiae (138), Saccharibacteria (138), Omnitrophica (132), Bacteroidetes (89), Planctomycetes (85), Firmicutes (62) and Verrucomicrobia (56).

On the other hand, the most abundant phylum was Actinobacteria (477413 reads), followed by Proteobacteria (209815), Planctomycetes (47971), Verrucomicrobia (47095), Cyanobacteria (38698), Parcubacteria (32991), Bacteroidetes (10976), Saccharibacteria (4895), Chloroflexi (3882) and finally Chlamydiae (1933).

Testing the occupancy-frequency distribution

We observed a significant inverse-Gaussian shape ($P < 0.05$), confirming the bimodal distribution of OTUs' of occupancy-frequency (Fig. 11a). In all four sub-regions we found a significant bimodal distribution, with different shape strengths within sub-regions, being less

pronounced in T and G ($P < 0.05$), intermediate in S ($P < 0.01$), and stronger in P ($P < 0.001$) (Fig. 11b, Tab. 3).

Concerning the core communities, 15 OTUs were found in all samples (Tab. 3) and, on a regional scale, T presented 34 core organisms, 10 of them encountered only in this region (Fig. 10, Tab. 3), S presented 33 core organisms, 8 exclusives (Fig. 10, Tab. 3), for P, 24 core and 4 exclusives (Fig. 10, Tab. 3) and, finally, in G, 29 core and 3 unique. For satellite OTUs, we found 1196 OTUs considering all samples and, in a regional scale, T had the higher number of satellite OTUs, followed by P, G and S (Tab. 3)

Concerning the random effect of the area on the occupancy-frequency distribution, we found a significant bimodal distribution for the three subsets, with a greater significance in the medium area ($P > 0.001$), and lower significance for small and large areas ($P > 0.01$) (Fig. 12).

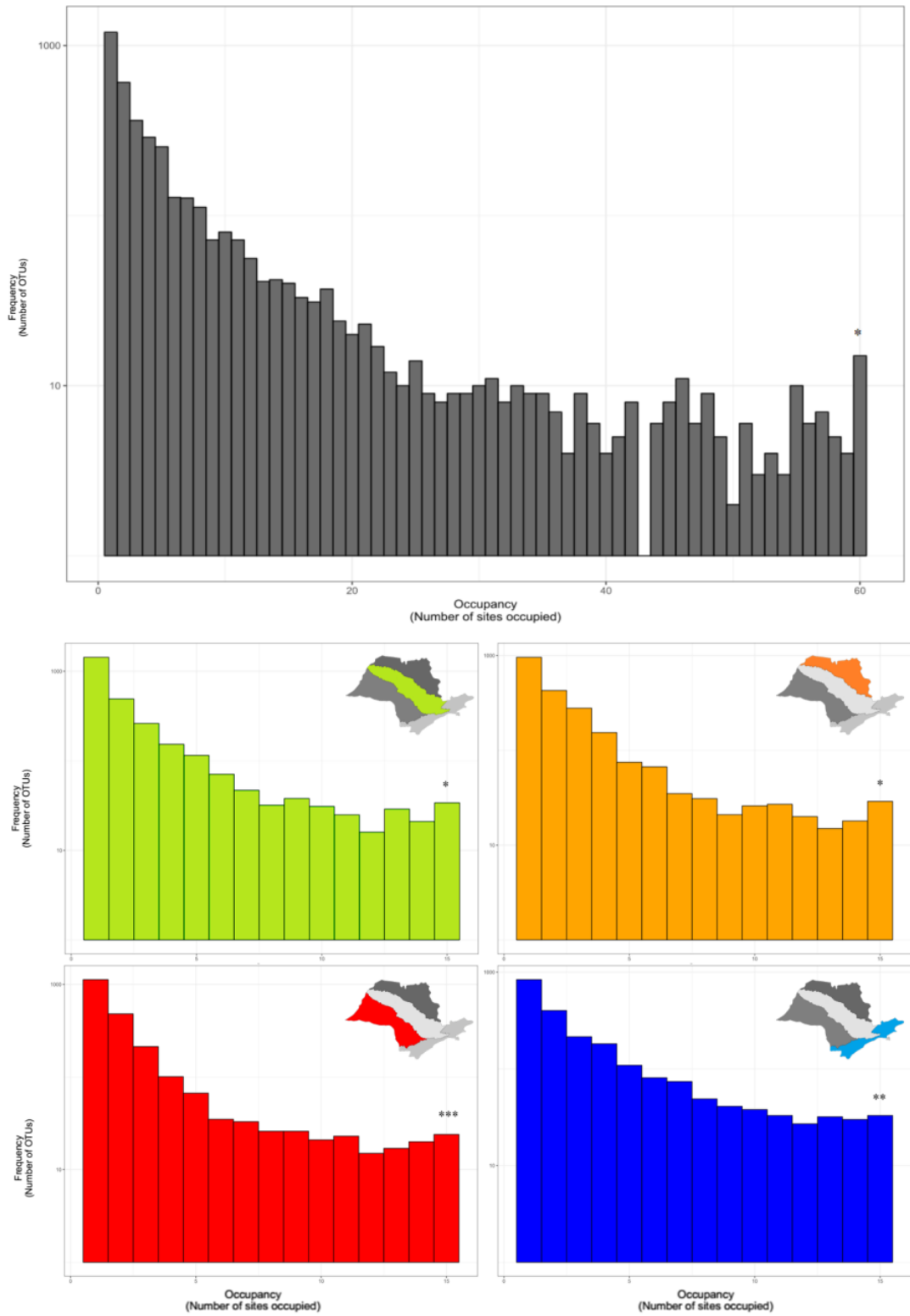


Figure 11 – Occupancy-frequency distributions for the total analyzed sites and for T (central green colored), P (southwest red colored), G (north orange colored) and S (Southwest blue colored) sub-regions. Significance: ‘***’ < 0.001; ‘**’ < 0.01; ‘*’ < 0.05.

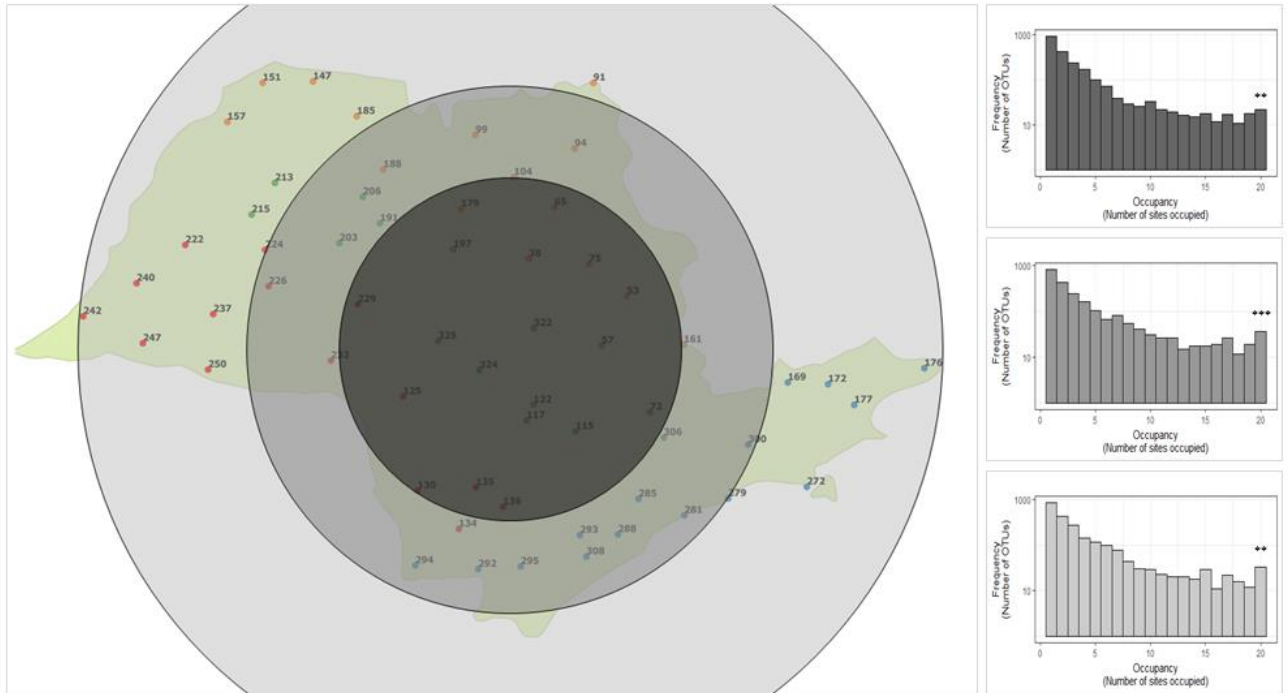


Figure 12 – Occupancy-frequency distributions concerning an increase of area range. Significance: ‘***’ < 0.001; ‘**’ < 0.01.

Actinobacteria was the phylum with the greatest number of representatives in core, mainly by *hgcl_clade* (Order: Frankiales; Family: Sporichthyaceae). Another prominent phylum was Proteobacteria, notably by *Rhizobiales* (Alphaproteobacteria) and *Burkholderiales* (Betaproteobacteria). Other phyla with representatives in core were Bacteroidetes, Cyanobacteria, Planctomycetes, and Verrucomicrobia (Tab. 4). The taxonomical composition of core populations is depicted in Fig. 13 and Tab. S4.

Parcubacteria was the most diverse phylum among satellite populations in all regions, followed by Proteobacteria (Tab. S5). About the Proteobacteria group, it is interesting to note that the most diverse representatives of the satellite group are *Oligoflexales* and *Bdellovibrionales* (Deltaproteobacteria) and *Legionellales* (Gammaproteobacteria).

Abundant core populations were represented by 9 OTUs distributed among the Phyla Actinobacteria, Proteobacteria and Cyanobacteria, notably by the *hgcl_clade*, a genus of Actinobacteria (Tab. 3, Fig. 13). Analyzing the sub-regions separately, we found that T and S had 12 abundant core OTUs each one, followed by G (9) and P (8). Rare core populations were not encountered for the totality of sites, but they are present in each sub-region, mostly in P (6), after G (5), S (4) and T (3) (Tab. 3). All satellite bacteria were rare (contributed less than 0.1% of relative abundance).

Table 3 – Sequencing dataset statistics and ecological indexes. (*) For Alpha, Gamma and Beta diversity, we used Simpson' index; (**) for evenness we used Hill' index. (***) Bimodality is tested against a Gaussian logarithmic curve and a significance means that the occupancy-frequency distribution had not presented only one hump at one of these extremes.

	All Samples	Tietê Basin	Parapanema Basin	Grande Basin	Serra do Mar Basin
Total nb. reads	882120	220530	220530	220530	220530
Mean nb. reads	14702	14702	14702	14702	14702
Total nb. OTUs	3946	2798	2228	2184	2182
Total Richness	3946	2798	2228	2184	2182
Mean Richness	461	513	390	427	515
Mean Alpha Diversity*	0.913	0.938	0.907	0.877	0.930
Gamma Diversity*	0.967	0.967	0.961	0.962	0.962
Beta Diversity*	0.956	0.876	0.881	0.872	0.849
Total Evenness**	0.023	0.035	0.035	0.031	0.036
Mean Evenness**	0.093	0.111	0.095	0.080	0.084
Bimodality significance***	< 0.05	< 0.05	< 0.001	< 0.05	< 0.01
Stress	0.0125	0.0237	0.0213	0.0302	0.0394
Core	15	34	24	29	33
“Abundant core”	9	12	8	9	12
“Rare core”	0	3	6	5	4
Satellite	1196	1433	1126	959	837

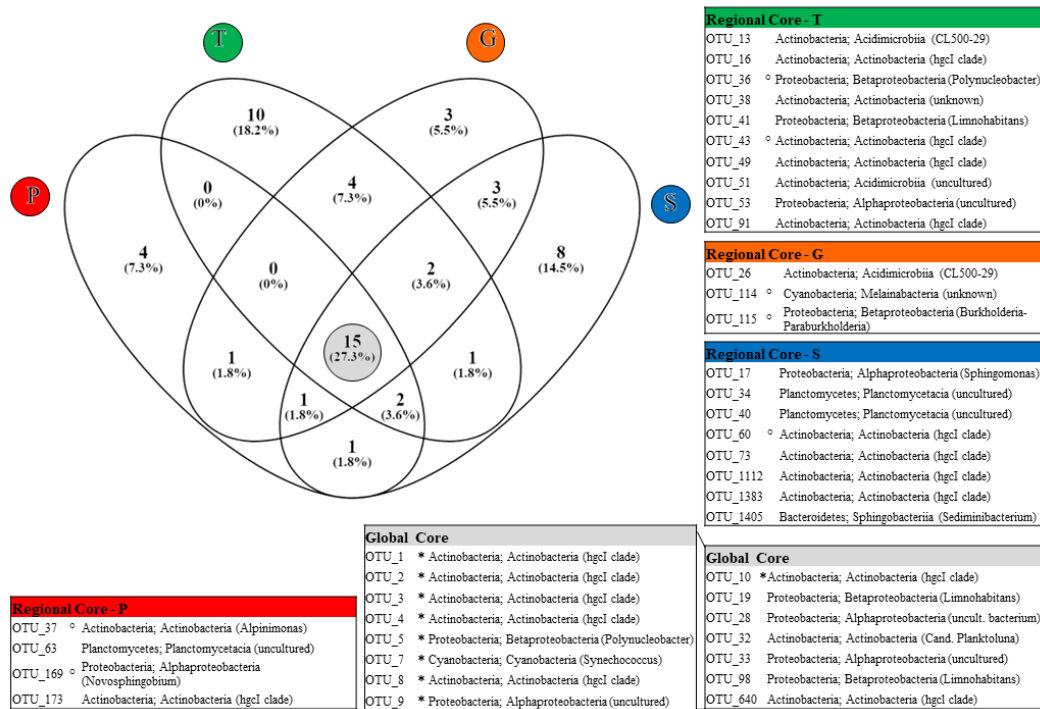


Figure 13 - Venn's plot of the core populations in the different sub-regions, with a list of the taxonomy of the OTUs present in all sites and those that occur exclusively in each sub-region. ‘*’ indicates the abundant core OTUs, ‘o’ indicates the rare core OTUs.

Discussion

We found a significant bimodal distribution in occupancy-frequency of freshwater bacteria in 60 shallow lakes distributed in the order of thousands of kilometers scale. These results are in accordance with previous report that analyzed these patterns in bacterial communities in the ocean (Lindh *et al.*, 2017), but not with those from freshwater diatom (Soininen and Heino, 2005), which makes this work the first, as far as we know, to identify a bimodal occupancy-frequency pattern for unicellular freshwater organisms.

In their study, Soininen and Heino could not find a bimodal distribution of occupancy-frequency in their dataset. They argued that these results were probably caused by the lack of

connectivity between sampling sites (from distinct river basins), which prevented the flux of individuals between sites. In our case, bacterial dispersal capacity has probably overcome geographic isolation or historical factors. Similar arguments were used to explain why birds have core populations while other such herpetiles or mammals do not (Mehranvar and Jackson, 2001). Another possible explanation for this pattern is the terrestrial seed bank that constantly feeds aquatic environments with microorganisms, a prominent process in headwater environments that promotes recolonization and avoid local extinctions (Ruiz-González *et al.*, 2015).

Another hypothesis suggested by Soininen and Heino to explain the absence of bimodal pattern in their dataset was a sampling effect. This artifact was cited by Gaston *et al.* (1997) as one of the most frequently used arguments to explain distribution patterns in macro-organisms. Performing sub-sampling of their dataset, Lindh *et al.* (2017) demonstrated that a minimum number of sequences per sample of 3000 to 5000 was required to observe bimodality, value easily reached with high throughput sequencing techniques used nowadays in microbial biodiversity studies (e.g. Ruiz-González *et al.*, 2015; Mentés *et al.*, 2017; Niño-García *et al.*, 2017). This indicates that microorganisms emerge as an excellent model to study macroecological patterns of organism's distribution.

We also observed that, the bimodal distribution of occupancy-frequency was more prominent in sub-regions rather than in the whole region. In order to elucidate if this was a simple effect of increasing area, we did a random subsampling of our data, but these results were not conclusive. We expected to observe a decrease in the bimodal effect with increasing area, already demonstrated in macro-organisms (Van Rensburg *et al.*, 2000), essentially as a correlation between local abundances and area range, since the increase in local abundance represents an increase in the population capacities to occupy other sites by immigration (Gaston *et al.*, 1997), so the larger the analyzed area, smaller would be the number of

organisms with sufficient local abundance to be able to occupy all sites, but this was not the case. Instead, we observed maximal bimodal signal in the intermediate scale (medium circle), indicating that it is not simply an effect of the area covered by sampling, instead, these patterns of distribution seem to be more related to a geographical coherence.

Similarly to our results, Lindh *et al.* (2017) had also found a coherence in the bimodal distribution of microorganisms when they separated their dataset in two regions according to salinity (the Baltic Sea and the Kattegat Sea). In their work, Soininen and Heino (2005) argued that the absence of bimodality may have occurred due to a lack of connectivity between sampling sites located in different river basins, based on Riis and Sand-Jensen (2002) report for stream plants, that found a more prominent bimodal distribution within watersheds than among basins, and argued that this occur because the connection via water flow would increase the dispersive capacity of these organisms.

In our case, however, we took samples near springs, and therefore, not connected via a water flow. The high bacterial dispersive capacity, may explain why we encountered a bimodal distribution in unconnected streams, but the argument that geographical coherence would be related only to connectivity between sites via water flow is not sustained in our case. Instead, we suggest that bacterial colonization success that ultimately determine which bacteria will be able to become core in each geographical region would depend more of basal environmental features (geological features that affect the type of soil, pH, ions concentration, etc.) than dispersal limitations (but see Farjalla *et al.*, 2012; Soininen *et al.*, 2013). This solution is compatible with the lack of bimodality in river diatoms (Soininen and Heino 2005), and the bimodality found in marine bacteria correlated with a strong gradient of one environmental factor geographically distributed (in this case salinity), in a much more connected environment (Lindh *et al.* 2017).

Besides that, the correlation between abundance and bimodality was less prominent for microorganisms. The core populations mostly composed by the most abundant OTUs, especially at the larger scale, but other organisms, including some of the rarest, were also part of this ubiquitous group, mainly on smaller scales. This ubiquity of rare organisms may be a successful strategy, in which remaining rare these organisms avoid predation and viral lysis (Galand *et al.*, 2009; Lindh *et al.*, 2017), or may be part of a seed bank for further increase in abundance, forming a reactive group capable of significantly increase their abundance when ideal environmental conditions are reached (Ruiz-González *et al.*, 2017). We had not performed time series which would help us determine which of these is the dominant strategy, but these “rare core” organisms have a majority of organisms that are never abundant and others that achieve this in only one or a few environments, suggesting that this division by “rare core” ends up grouping representatives of these two strategies into a single group.

Regarding which groups were among the regional core, the hgcl_clade stands out as the most frequent group. This is a still unclassified genus of Actinobacteria, strongly present in the abundant global core, but also present in the rare regional scale in T and P. Actinobacteria is an ubiquitous phylum, abundant in terrestrial, freshwater and marine environments (Newton *et al.*, 2011). The ubiquity of these groups has already been analyzed and seems to be related with their structural characteristics (Newton *et al.*, 2011) and a potential capability to generate energy driven by light (Sharma *et al.*, 2008). In fact, the hgcl_clade was already found as an abundant group in other freshwater environments (Lliros *et al.*, 2014; Hosen *et al.*, 2017; Montes *et al.*, 2017) and also in the estuarine environments (Golebiewski *et al.*, 2017), and its abundance seems to be related to an increase in the nitrogen supply caused by human impacts (Hosen *et al.*, 2017), which may explain their high abundance in our dataset, collected in a region deeply affected by human activities, mainly agriculture.

Table 4 – Distribution of core organisms' phyla among sub-basins.

Core per Phylum	All Samples	Tietê Basin	Paranapanema Basin	Grande Basin	Serra do Mar Formation
Actinobacteria	8	18	10	13	14
Bacteroidetes	-	-	1	1	3
Cyanobacteria	1	1	2	3	1
Planctomycetes	-	1	2	-	3
Alphaproteobacteria	3	5	4	5	6
Betaproteobacteria	3	8	3	6	4
Gammaproteobacteria	-	-	1	1	1
Verrucomicrobia	-	1	1	-	1
Total Core	15	34	24	29	33

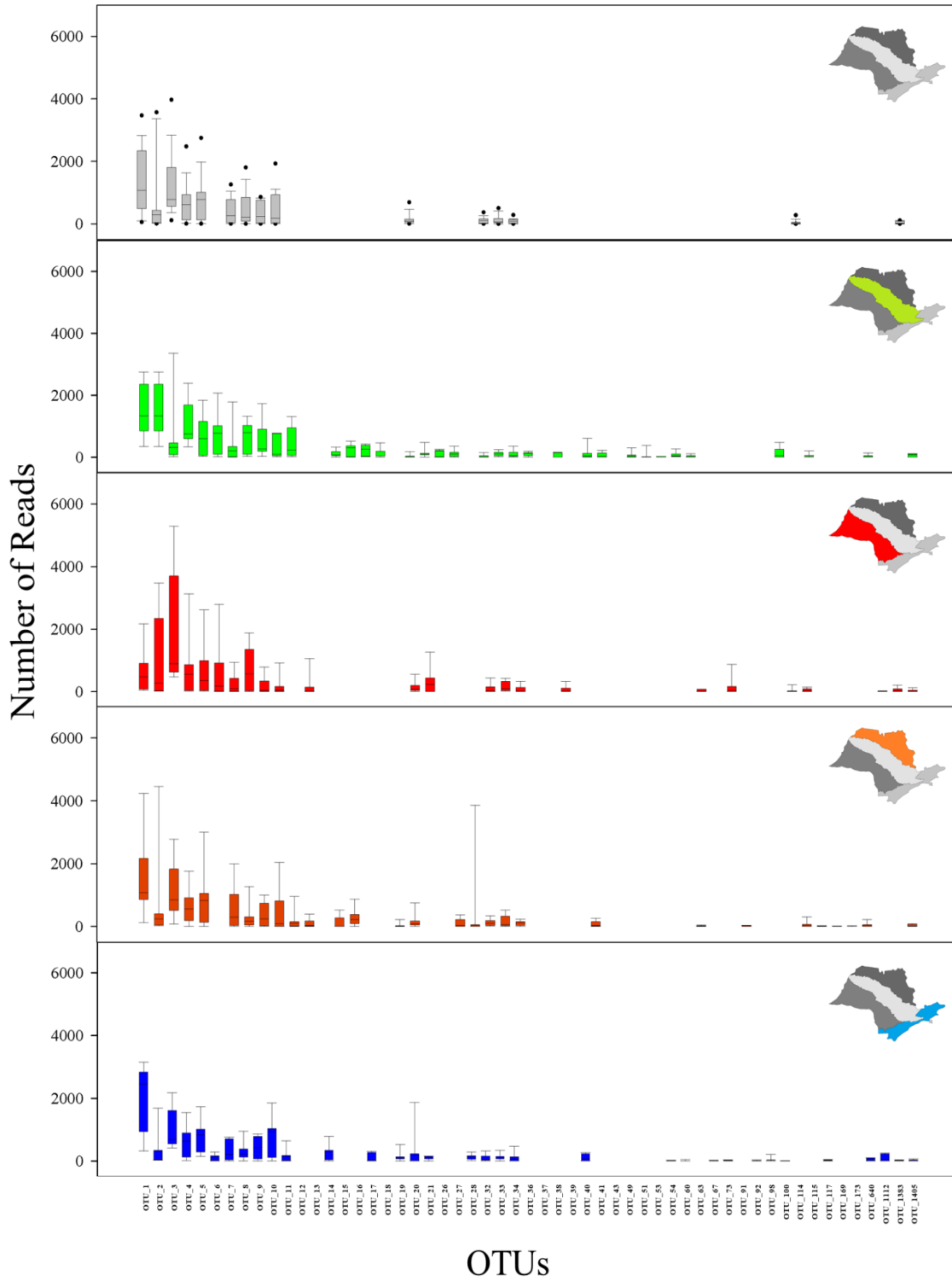


Figure 14 - Average core OTU abundance distribution in all samples (grey), in T (central green colored), in P (southwest red colored), in G (north orange colored) and in S (Southwest blue colored). A full detailed list of OTU's taxonomy can be found in supplementary table S2.

Parcubacteria was the most diverse group among satellite organisms. There are evidences that this is a group entirely composed by organisms specialized in symbiotic life (Nelson and Stegen, 2015). Therefore, it is not unexpected that these were found among satellite organisms, since its presence in specific sites is dependent to certain environmental conditions that enable the presence of those bacteria and the associated organism. It is interesting that this group presented a high diversity within satellite organisms, having representative OTUs in almost all environments analyzed, and even presenting more than 300 distinct OTUs in a single environment.

These data point to an intricate process of speciation, dependent not only on the environment, or on geographic isolation processes, but also related to specific needs and speciation processes related to the organisms that they are associated. These results may also be indicating a presence of a technical artifact in which the division of OTUs by 97% threshold does not match the speciation rates of this phylum. The restrictions of the OTU technique have been discussed and other identification technique are already proposed (Callahan *et al.*, 2017). The main problem of the OTU is their arbitrary detection system of taxonomic units that do not considers speciation characteristics intrinsic to each group, limiting all the organisms to the same value to separate the sequences and this may create misunderstandings that we cannot predict.

Unfortunately, this is one of the least known prokaryotic groups, hampering the discussion about satellite organisms, since Parcubacteria were the most diverse satellite organism in this study. Furthermore, it is also one of the most poorly classified organisms, most known only at the Phylum level and still having its classification at class level as *Candidatus*. Its relevance to diversity in tropical inland environments demonstrated here calls attention to the need for future efforts to expand knowledge about the biology of this group. (This can be also observed in Tab. S5 and Tab. S6 of supplementary material).

The occupancy-frequency distribution was already tested in sets of macro-organisms over the years (e.g. Mehranvar and Jackson, 2001; Heino, 2015; Jokimäki *et al.*, 2016), but many times the distributions were not bimodal. Many authors credited this to an insufficient sampling effort effect (Mehranvar and Jackson, 2001; Mcgeoch and Gaston, 2002). In this sense, microbial datasets processed with high throughput sequencing tools provide an unique opportunity to test the robustness of these theories, as compiling comprehensive data from thousands of plants or animals could prove to be a Herculean work, taking into account that the ideal sampling to test bimodality seems to be between 3000 to 5000 individuals for each site (Lindh *et al.*, 2017).

In addition, this work highlights the applicability of this theory to provide information on the distribution of microorganisms along freshwater habitats facilitating the recognition of groups of importance to local processes and global functions. This is the case, for example, of the Parcubacteria Phylum which, although still little known, proved to be a key *phylum* in shallow lake environments. Finally, we emphasize the need to carry out studies in larger areas so that's possible to best understand the effect of area increase on the prokaryotic theory of bimodality.

Acknowledgments

This research was funded by the São Paulo Research Foundation (FAPESP process: 2014/14139-3 and 2011/50054-4). We thank Universidade Estadual Paulista – UNESP, Departamento de Tecnologia, Laboratório Multiusuário Centralizado para Sequenciamento de DNA em Larga Escala e Análise de Expressão Gênica – LMSeq for sequencing. We also would like to thank CAPES for the scholarship provided to the first author (EMB) and FAPESP (grant: 2013/18083-0) for the post-doctoral scholarship provided for the third author (ILB). Finally, we are thankful to Prof. Armando A. H. Vieira for coordinating the project that funded the sampling, and Cilene C. Mori, Fabrício S. Meccheri, Guilherme P. de Moraes, Helena H. Vieira, Ingritt C. Moreira, Letícia P. Tessarolli, Lucas S. Tiberti, Luiz A. Joaquim, Naiara C. P. Archanjo, Rodrigo V. de Mello and Thaís G. da Silva who participated in the field sampling.

References

- Barberan, A., E. O. Casamayor, and N. Fierer. 2014. The microbial contribution to macroecology. *Front Microbiol* **5**:203.
- Brown, J. H. 1984. On the relationship between abundance and distribution of species. *The American Naturalist* **124**:255-279.
- Callahan, B. J., P. J. McMurdie, and S. P. Holmes. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* **11**:2639-2643.
- Cunha, D. G. F., M. d. C. Calijuri, and M. C. Lamparelli. 2013. A trophic state index for tropical/subtropical reservoirs (TSIts_r). *Ecological Engineering* **60**:126-134.
- Edgar, R. C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**:996-998.
- Farjalla, V. F., D. S. Srivastava, N. A. C. Marino, F. D. Azevedo, V. Dib, P. M. Lopes, A. S. Rosado, R. L. Bozelli, and F. A. Esteves. 2012. Ecological determinism increases with organism size. *Ecology* **93**:1752-1759.
- Galand, P. E., E. O. Casamayor, D. L. Kirchman, and C. Lovejoy. 2009. Ecology of the rare microbial biosphere of the Arctic Ocean. *Proceedings of the National Academy of Sciences* **106**:22427-22432.
- Gaston, K. J., T. M. Blackburn, J. J. Greenwood, R. D. Gregory, R. M. Quinn, and J. H. Lawton. 2000. Abundance–occupancy relationships. *Journal of Applied Ecology* **37**:39-59.
- Gaston, K. J., T. M. Blackburn, and J. H. Lawton. 1997. Interspecific Abundance-Range Size Relationships: An Appraisal of Mechanisms. *Journal of Animal Ecology* **66**:579-601.
- Gleason, H. 1929. The significance of Raunkiaer's law of frequency. *Ecology* **10**:406-408.

- Golebiewski, M., J. Calkiewicz, S. Creer, and K. Piwosz. 2017. Tideless estuaries in brackish seas as possible freshwater-marine transition zones for bacteria: the case study of the Vistula river estuary. *Environ Microbiol Rep* **9**:129-143.
- Gotelli, N. J. 1991. Metapopulation models: the rescue effect, the propagule rain, and the core-satellite hypothesis. *The American Naturalist* **138**:768-776.
- Hanski, I. 1982. Dynamics of Regional Distribution: The Core and Satellite Species Hypothesis. *Oikos* **38**:210-221.
- Hanski, I., and M. Gyllenberg. 1993. Two general metapopulation models and the core-satellite species hypothesis. *The American Naturalist* **142**:17-41.
- Heino, J. 2015. Deconstructing occupancy frequency distributions in stream insects: effects of body size and niche characteristics in different geographical regions. *Ecological Entomology* **40**:491-499.
- Herlemann, D. P., M. Labrenz, K. Jürgens, S. Bertilsson, J. J. Waniek, and A. F. Andersson. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J* **5**:1571-1579.
- Hosen, J. D., C. M. Febria, B. C. Crump, and M. A. Palmer. 2017. Watershed Urbanization Linked to Differences in Stream Bacterial Community Composition. *Front Microbiol* **8**:1452.
- Jokimäki, J., J. Suhonen, and M.-L. Kaisanlahti-Jokimäki. 2016. Urbanization and species occupancy frequency distribution patterns in core zone areas of European towns. *European Journal of Ecology* **2**.
- Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological society of America* **15**:237-240.

- Lindh, M. V., J. Sjostedt, B. Ekstam, M. Casini, D. Lundin, L. W. Hugerth, Y. O. Hu, A. F. Andersson, A. Andersson, C. Legrand, and J. Pinhassi. 2017. Metapopulation theory identifies biogeographical patterns among core and satellite marine bacteria scaling from tens to thousands of kilometers. *Environ Microbiol* **19**:1222-1236.
- Lliros, M., O. Inceoglu, T. Garcia-Armisen, A. Anzil, B. Leporcq, L. M. Pigneur, L. Viroux, F. Darchambeau, J. P. Descy, and P. Servais. 2014. Bacterial community composition in three freshwater reservoirs of different alkalinity and trophic status. *PLoS One* **9**:e116145.
- Logares, R. 2017. ramalok/amplicon_processing: Workflow for Analysing MiSeq Amplicons based on Uparse.
- Logares, R., S. Sunagawa, G. Salazar, F. M. Cornejo-Castillo, I. Ferrera, H. Sarmiento, P. Hingamp, H. Ogata, C. de Vargas, G. Lima-Mendez, J. Raes, J. Poulain, O. Jaillon, P. Wincker, S. Kandels-Lewis, E. Karsenti, P. Bork, and S. G. Acinas. 2014. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol* **16**:2659-2671.
- Logue, J. B., and E. S. Lindström. 2008. Biogeography of Bacterioplankton in Inland Waters. *Freshwater Reviews* **1**:99-114.
- Lorenzen, C. J. 1967. Determination of chlorophyll and pheo-pigments: spectrophotometric equations. *Limnology and Oceanography* **12**:343-346.
- Magurran, A. E., and P. A. Henderson. 2003. Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**:714-716.
- Marker, A. 1980. The measurement of photosynthetic pigments in freshwaters and standardization of methods: conclusions and recommendations. *Arch. Hydrobiol. Ergebn. Limnol.* **14**:91-106.

- Martiny, J. B. H., B. J. M. Bohannan, J. H. Brown, R. K. Colwell, J. A. Fuhrman, J. L. Green, M. C. Horner-Devine, M. Kane, J. A. Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Ovreas, A.-L. Reysenbach, V. H. Smith, and J. T. Staley. 2006. Microbial biogeography: putting microorganisms on the map. *Nat Rev Micro* **4**:102-112.
- McGeoch, M. A., and K. J. Gaston. 2002. Occupancy frequency distributions: patterns, artefacts and mechanisms. *Biol Rev Camb Philos Soc* **77**:311-331.
- Mehranvar, L., and D. A. Jackson. 2001. History and taxonomy: their roles in the core-satellite hypothesis. *Oecologia* **127**:131-142.
- Mentes, A., A. Szabo, B. Somogyi, B. Vajna, N. Tugyi, B. Csitari, L. Voros, and T. Felfoldi. 2017. Differences in planktonic microbial communities associated with three types of macrophyte stands in a shallow lake. *FEMS Microbiol Ecol*.
- Mitchell-Olds, T., and R. G. Shaw. 1987. REGRESSION ANALYSIS OF NATURAL SELECTION: STATISTICAL INFERENCE AND BIOLOGICAL INTERPRETATION. *Evolution* **41**:1149-1161.
- Nelson, W. C., and J. C. Stegen. 2015. The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Front Microbiol* **6**:713.
- Neuwirth, E. 2014. RColorBrewer: ColorBrewer Palettes.
- Newton, R. J., S. E. Jones, A. Eiler, K. D. McMahon, and S. Bertilsson. 2011. A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* **75**:14-49.
- Niño-García, J. P., C. Ruiz-González, and P. A. Del Giorgio. 2017. Exploring the Ecological Coherence between the Spatial and Temporal Patterns of Bacterioplankton in Boreal Lakes. *Front Microbiol* **8**:636.
- Nush, E. 1980. Comparison of different methods for chlorophyll and phaeopigment determination. *Arch. Hydrobiol. Beih* **14**:14-36.

- Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2016. *Vegan: community ecology package*.
- Oliveros, J. C. 2007-2015. *Venny*. An interactive tool for comparing lists with Venn's diagrams.
- Papp, L., and J. Izsák. 1997. Bimodality in occurrence classes: a direct consequence of lognormal or logarithmic series distribution of abundances: a numerical experimentation. *Oikos* **79**:191-194.
- Pedrós-Alió, C. 2006. Marine microbial diversity: can it be determined? *Trends Microbiol* **14**:257-263.
- Preston, F. W. 1948. The commonness, and rarity, of species. *Ecology* **29**:254-283.
- Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. Glockner. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**:D590-D596.
- R. Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Riis, T., and K. Sand-Jensen. 2002. Abundance-range size relationships in stream vegetation in Denmark. *Plant Ecology* **161**:175-183.
- Rosselló-Mora, R., and R. Amann. 2001. The species concept for prokaryotes. *FEMS Microbiol Rev* **25**:39-67.
- Ruiz-González, C., J. P. Niño-García, and P. A. del Giorgio. 2015. Terrestrial origin of bacterial communities in complex boreal freshwater networks. *Ecology Letters* **18**:1198-1206.
- Ruiz-González, C., J. P. Niño-García, S. W. Kembel, and P. A. Del Giorgio. 2017. Identifying the core seed bank of a complex boreal bacterial metacommunity. *ISME J*.

- Schneegurt, M. A., S. Y. Dore, and C. F. Kulpa Jr. 2003. Direct extraction of DNA from soils for studies in microbial ecology. *Current issues in molecular biology* **5**:1-8.
- Sharma, A. K., O. Zhaxybayeva, R. T. Papke, and W. F. Doolittle. 2008. Actinorhodopsins: proteorhodopsin-like gene sequences found predominantly in non-marine environments. *Environ Microbiol* **10**:1039-1056.
- Soininen, J., and J. Heino. 2005. Relationships between local population persistence, local abundance and regional occupancy of species: distribution patterns of diatoms in boreal streams. *Journal of Biogeography* **32**:1971-1978.
- Soininen, J., J. J. Korhonen, and M. Luoto. 2013. Stochastic species distributions are driven by organism size. *Ecology* **94**:660-670.
- Stackebrandt, E., and B. Goebel. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Evol Microbiol* **44**:846-849.
- Ter Braak, C. J., and C. W. Looman. 1986. Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* **65**:3-11.
- Tonkin, J. D., F. O. Arimoro, and P. Haase. 2016. Exploring stream communities in a tropical biodiversity hotspot: biodiversity, regional occupancy, niche characteristics and environmental correlates. *Biodiversity and Conservation* **25**:975-993.
- van Rensburg, B. J., M. A. McGeoch, W. Matthews, S. L. Chown, and A. S. van Jaarsveld. 2000. Testing generalities in the shape of patch occupancy frequency distributions. *Ecology* **81**:3163-3177.
- Wickham, H. 2007. Reshaping data with the reshape package. *Journal of Statistical Software* **21**.
- Wickham, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Yilmaz, P., L. Parfrey, P. Yarza, J. Gerken, E. Pruesse, C. Quast, T. Schweer, J. Peplies, W. Ludwig, and F. Glockner. 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research* **42**:D643-D648.

Supplementary Material

Table S3 – Database containing basic information about the collection, geographic position and environmental analysis.

Table S4 – The taxonomy of core organisms. The “X” indicates region where a given OTU was identified as core, if it is present in all regions, it is marked in the 4 columns representing each sub-region and in the column named as "Total Core". The taxonomy was separated by each level, some OTUs cannot be identified until species, in this case the spaces referring to a certain taxonomic level and the following ones remain empty.

Table S5 – Diversity ranking for the satellite microorganisms, by all the samples and by each sub-region.

Table S6 – Taxonomic table and total abundance data considering only the Parcubacteria Phylum (**OBS.:** Esta tabela de dados pode ser acessada apenas em sua forma digital. Se você teve acesso a esta informação por via impressa da dissertação, e necessitar destes, envie um e-mail para erickmbarros@gmail.com).

Sites	GE065	GE075	GE104	GE151	GE188	GM038	GM094	GM157	GM161
Collect date	18-Jan-2013	9-Apr-2013	26-Jul-2013	6-Nov-2013	22-Jan-2014	14-Dec-2012	23-Jul-2013	8-Nov-2013	2-Dec-2013
Hour	10h50min	8h40min	10h20min	13h40min	11h42min	8h30min	-	8h40min	10h50min
Location (GPS coordinates)	21.170283 S 47.855717 W	22.628067 S 47.062333 W	20.884950 S 48.239083 W	19.992283 S 50.622133 W	20.813750 S 49.478733 W	21.657833 S 48.097267 W	20.613900 S 47.664400 W	20.362217 S 50.957633 W	22.471683 S 46.629350 W
Altitude (ft)	1806	2191	1609	1328	1782	1857	2203	1121	2838
Altitude (m)	550	668	490	405	543	566	671	342	865
pH	6.80	6.40	7.05	8.79	6.98	6.01	8.65	7.02	6.71
Temperature (°C)	24.70	24.34	17.68	27.48	29.96	25.62	21.03	24.07	25.70
chlorophyll a (mg/L)	20.930	9.770	12.000	29.222	24.275	7.530	4.460	4.350	5.024
Trophic State Index	Eutrophic	Eutrophic	Eutrophic	Eutrophic	Eutrophic	Mesotrophic	Mesotrophic	Mesotrophic	Mesotrophic
Dissolved Organic Carbon (mg/L)	3.468	5.376	4.748	10.000	8.204	2.769	2.726	4.555	5.273
Total Carbon (mg/L)	8.888	9.338	8.433	13.440	13.800	6.772	5.339	8.758	10.050
Inorganic Carbon (mg/L)	5.420	3.962	3.685	3.435	5.597	4.004	2.613	4.203	4.775
Total Nitrogen (mg/L)	0.182	0.342	0.240	0.223	0.725	0.206	0.045	0.249	0.546
Chloride (mg/L)	1.245	1.986	0.349	0.920	1.292	0.347	0.082	5.611	2.474
Nitrite (mg/L)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Nitrate (mg/L)	0.000	0.000	0.000	0.000	0.047	0.018	0.047	0.000	0.851
Phosphate (mg/L)	0.000	0.000	0.000	0.000	0.000	0.000	0.756	0.000	0.000
Sulfate (mg/L)	0.104	0.388	0.042	0.092	1.209	0.034	0.084	0.000	1.185
Lithium (mg/L)	0.000	0.022	0.023	0.012	0.000	0.000	0.021	0.012	0.000
Sodium (mg/L)	2.722	1.763	2.952	6.598	3.565	1.307	2.650	5.637	6.319
Ammonium (mg/L)	0.124	0.223	0.233	0.215	0.160	0.160	0.261	0.239	0.184
Potassium (mg/L)	1.438	1.828	0.638	4.135	3.078	1.190	0.514	2.986	1.095
Magnesium (mg/L)	2.341	1.559	0.000	0.986	2.082	1.654	0.212	4.076	1.520
Calcium (mg/L)	5.851	3.055	2.211	1.900	6.644	3.484	0.942	11.164	5.512
Fluorescence (Em. 455nm, Ab. 355nm)	0.798	2.031	1.210	3.149	1.446	0.799	0.378	2.900	1.443

Sites	GM179	GO053	GO091	GO099	GO147	GO185	PE134	PE222	PE224
Collect date	20-Jan-2014	11-Dec-2012	23-Jul-2013	25-Jul-2013	5-Nov-2013	21-Jan-2014	12-Sep-2013	3-Jun-2014	4-Jun-2014
Hour	9h26min	8h30min	9h00min	8h50min	10h41min	14h05min	14h25min	16h00min	11h15min
Location (GPS coordinates)	21.183883 S 48.741150 W	22.011700 S 47.181700 W	19.992250 S 47.486033 W	20.486350 S 48.604500 W	19.980117 S 50.144900 N	20.309389 S 49.729222 W	24.318250 S 49.146250 W	21.527000 S 51.356567 W	21.574233 S 50.600417 W
Altitude (ft)	1969	1945	1748	1427	1376	1588	2755	1335	1283
Altitude (m)	600	593	533	435	419	484	840	407	391
pH	6.01	5.07	5.30	7.15	5.47	6.55	9.10	10.10	6.90
Temperature (°C)	25.77	29.92	22.30	19.15	26.53	30.60	26.29	23.40	19.55
chlorophyll a (mg/L)	6.045	3.260	2.790	3.630	3.649	1.082	13.250	105.193	9.341
Trophic State Index	Mesotrophic	Oligotrophic	Oligotrophic	Oligotrophic	Oligotrophic	Oligotrophic	Eutrophic	Eutrophic	Eutrophic
Dissolved Organic Carbon (mg/L)	5.193	5.445	3.130	4.144	11.980	7.980	8.103	26.320	5.709
Total Carbon (mg/L)	10.420	6.988	10.370	9.147	13.210	11.630	17.690	29.530	18.990
Inorganic Carbon (mg/L)	5.223	1.543	7.244	5.003	1.236	3.653	9.591	3.201	13.280
Total Nitrogen (mg/L)	0.359	0.169	0.476	0.221	2.629	0.524	0.646	1.620	0.329
Chloride (mg/L)	2.201	0.251	0.183	0.675	7.391	1.100	1.980	3.499	10.183
Nitrite (mg/L)	0.000	0.000	0.000	0.000	0.093	0.000	0.000	0.145	0.000
Nitrate (mg/L)	0.051	0.000	0.921	0.233	0.902	0.077	0.000	0.000	0.000
Phosphate (mg/L)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sulfate (mg/L)	0.100	0.000	0.039	0.168	3.655	0.059	0.377	0.156	0.315
Lithium (mg/L)	0.013	0.000	0.003	0.021	0.000	0.000	0.021	0.000	0.024
Sodium (mg/L)	5.235	0.671	3.056	3.957	2.941	2.032	4.276	2.651	19.112
Ammonium (mg/L)	0.240	0.047	0.113	0.313	2.741	0.178	0.310	0.248	0.585
Potassium (mg/L)	9.238	0.245	3.742	1.287	6.620	3.338	3.301	9.262	4.748
Magnesium (mg/L)	2.750	0.148	1.312	0.000	0.172	1.487	0.000	0.481	3.073
Calcium (mg/L)	5.295	0.449	3.859	2.975	0.585	2.730	6.823	1.343	7.801
Fluorescence (Em. 455nm, Ab. 355nm)	0.956	0.893	0.676	1.145	5.377	2.436	1.318	3.511	2.034

Sites	PE237	PE240	PM125	PM130	PM135	PM226	PM232	PO136	PO229
Collect date	29-Jul-2014	29-Jul-2014	10-Sep-2013	11-Sep-2013	12-Sep-2013	4-Jun-2014	6-Jun-2014	13-Sep-2013	5-Jun-2014
Hour	8h06min	14h50min	13h30min	14h35min	17h40min	15h50min	11h00min	8h20min	14h40min
Location (GPS coordinates)	22.182067 S 51.091517 W	21.889517 S 51.818833 W	22.960783 S 49.290567 W	23.853683 S 49.153483 W	24.216583 S 48.769167 W	21.918183 S 50.566917 W	22.621383 S 49.975800 W	24.005400 S 48.342614 W	22.087977 S 49.722099 W
Altitude (ft)	1478	1260	2054	2227	2730	1532	1675	2191	2020
Altitude (m)	450	384	626	679	832	467	511	668	616
pH	6.70	8.00	6.80	6.80	-	7.00	5.60	7.40	6.30
Temperature (°C)	16.11	20.16	21.69	23.88	22.65	21.66	20.70	21.56	22.12
chlorophyll a (mg/L)	51.550	13.899	5.930	4.530	7.670	4.630	7.483	4.190	0.804
Trophic State Index	Eutrophic	Eutrophic	Mesotrophic	Mesotrophic	Mesotrophic	Mesotrophic	Mesotrophic	Oligotrophic	Oligotrophic
Dissolved Organic Carbon (mg/L)	21.820	7.050	3.262	7.454	2.840	6.218	3.784	11.290	3.866
Total Carbon (mg/L)	23.940	11.740	6.892	11.140	5.764	12.270	5.006	20.700	7.626
Inorganic Carbon (mg/L)	2.123	4.695	3.631	3.687	2.925	6.052	1.222	9.413	3.759
Total Nitrogen (mg/L)	1.044	0.514	0.175	0.306	0.314	0.564	0.247	0.929	11.120
Chloride (mg/L)	1.417	2.317	0.546	0.602	0.228	2.008	0.749	3.825	7.359
Nitrite (mg/L)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.029	0.000
Nitrate (mg/L)	0.016	0.252	0.431	0.020	0.569	0.155	0.189	0.563	45.292
Phosphate (mg/L)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sulfate (mg/L)	0.236	0.290	0.000	0.000	0.000	0.141	0.018	0.000	0.926
Lithium (mg/L)	0.000	0.011	0.068	0.000	0.022	0.057	0.014	0.023	0.016
Sodium (mg/L)	2.064	2.901	1.591	0.960	2.187	2.810	1.614	6.982	4.788
Ammonium (mg/L)	0.159	0.273	0.242	0.300	0.258	0.330	0.000	0.448	0.404
Potassium (mg/L)	3.322	4.714	1.045	0.874	0.717	3.163	0.321	1.475	6.449
Magnesium (mg/L)	0.230	1.464	0.000	0.000	0.000	2.493	0.225	0.000	9.293
Calcium (mg/L)	0.573	2.074	2.597	2.795	1.758	5.069	0.498	7.380	0.000
Fluorescence (Em. 455nm, Ab. 355nm)	2.293	2.108	0.684	0.868	0.616	1.550	0.929	1.427	0.397

Sites	PO242	PO247	PO250	SE172	SE272	SE288	SE294	SE295	SM177
Collect date	30-Jul-2014	31-Jul-2014	1-Aug-2014	5-Dec-2013	24-Jun-2015	25-Aug-2015	27-Aug-2015	27-Aug-2015	6-Dec-2013
Hour	10h45min	13h50min	10h09min	12h25min	10h04min	12h44min	8h40min	16h50min	16h20min
Location (GPS coordinates)	22.202933 S 52.325100 W	22.457983 S 51.758983 W	22.707367 S 51.140267 W	22.845000 S 45.260050 W	23.586667 S 45.348667 W	24.267666 S 47.265693 W	24.569967 S 49.175350 W	24.569833 S 48.175467 W	23.041633 S 45.012100 W
Altitude (ft)	1054	1263	1064	1970	-247	1938	1790	522	2693
Altitude (m)	321	385	324	600	-75	591	546	159	821
pH	6.50	6.30	7.40	5.59	6.50	6.07	6.89	6.70	8.10
Temperature (°C)	20.94	21.78	19.79	26.12	19.80	19.75	18.71	21.95	31.41
chlorophyll a (mg/L)	0.628	0.736	1.033	93.577	93.577	18.560	20.025	11.775	5.582
Trophic State Index	Oligotrophic	Oligotrophic	Oligotrophic	Eutrophic	Eutrophic	Eutrophic	Eutrophic	Eutrophic	Mesotrophic
Dissolved Organic Carbon (mg/L)	2.901	19.050	3.967	14.730	7.305	5.260	14.750	17.740	8.035
Total Carbon (mg/L)	5.187	21.220	11.180	19.610	12.600	10.270	21.880	28.170	10.500
Inorganic Carbon (mg/L)	2.286	2.172	7.216	4.877	5.300	5.012	7.123	10.430	2.463
Total Nitrogen (mg/L)	1.320	3.174	0.626	1.090	0.409	0.227	0.917	1.137	0.922
Chloride (mg/L)	1.464	0.703	1.916	1.409	26.251	4.670	3.730	1.087	0.865
Nitrite (mg/L)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Nitrate (mg/L)	4.775	0.160	1.552	0.032	0.274	0.012	0.000	0.102	0.000
Phosphate (mg/L)	0.159	0.000	0.000	0.000	0.000	0.005	0.013	0.014	0.000
Sulfate (mg/L)	0.284	0.090	1.610	0.325	2.735	0.586	0.254	0.348	0.193
Lithium (mg/L)	0.022	0.022	0.000	0.000	0.000	0.000	0.001	0.000	0.000
Sodium (mg/L)	1.957	1.983	4.097	4.959	18.190	5.778	8.930	4.557	4.127
Ammonium (mg/L)	0.261	0.284	0.000	0.318	0.000	0.000	0.013	0.001	0.209
Potassium (mg/L)	2.349	0.734	1.091	2.226	1.944	2.008	5.822	3.341	1.314
Magnesium (mg/L)	1.140	0.149	2.110	1.799	0.000	0.000	0.000	0.000	0.597
Calcium (mg/L)	0.751	0.349	5.213	4.473	1.938	1.916	4.246	1.884	1.844
Fluorescence (Em. 455nm, Ab. 355nm)	0.796	2.457	1.499	6.317	1.604	1.807	3.170	3.250	1.636

Sites	SM279	SM292	SM300	SM308	SO169	SO176	SO281	SO285	SO293
Collect date	25-Jun-2015	26-Aug-2015	16-Nov-2015	19-Nov-2015	4-Dec-2013	6-Dec-2013	26-Jun-2015	24-Aug-2015	26-Aug-2015
Hour	15h16min	11h36min	15h15min	15h13min	11h30min	11h29min	11h56min	14h00min	14h30min
Location (GPS coordinates)	23.928796 S 46.209002 W	24.595250 S 48.578683 W	23.422217 S 46.022000 W	24.477772 S 47.552501 W	22.829000 S 45.642139 W	22.693933 S 44.346833 W	24.086111 S 46.623139 W	23.931225 S 47.057456 W	24.536472 S 48.726889 W
Altitude (ft)	-4	142	1936	303	3622	1787	-265	2283	2702
Altitude (m)	-1	43	590	92	1104	545	-81	696	824
pH	6.15	6.90	8.15	5.97	6.04	6.75	6.28	6.65	7.05
Temperature (°C)	20.30	21.32	27.98	26.22	19.66	28.02	18.10	20.66	18.30
chlorophyll a (mg/L)	5.878	5.482	6.821	7.005	0.092	1.800	2.035	2.836	1.805
Trophic State Index	Mesotrophic	Mesotrophic	Mesotrophic	Mesotrophic	Oligotrophic	Oligotrophic	Oligotrophic	Oligotrophic	Oligotrophic
Dissolved Organic Carbon (mg/L)	5.756	3.443	7.427	13.630	3.620	6.588	3.802	42.060	-
Total Carbon (mg/L)	8.916	15.000	17.670	17.350	5.065	8.266	5.660	45.180	-
Inorganic Carbon (mg/L)	3.160	11.560	10.240	3.726	1.445	1.678	1.858	3.118	-
Total Nitrogen (mg/L)	0.354	0.347	0.924	0.796	0.233	0.432	0.273	0.412	-
Chloride (mg/L)	74.587	1.988	5.308	5.299	0.733	0.709	10.422	2.515	1.970
Nitrite (mg/L)	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000
Nitrate (mg/L)	0.565	0.012	0.442	0.025	0.287	0.035	0.538	0.067	0.067
Phosphate (mg/L)	0.003	0.019	0.005	0.006	0.000	0.000	0.019	0.008	0.010
Sulfate (mg/L)	8.252	0.982	3.857	0.656	0.395	0.415	1.489	0.493	0.231
Lithium (mg/L)	0.000	0.003	0.000	0.000	0.013	0.000	0.000	0.008	0.000
Sodium (mg/L)	9.147	4.318	7.771	5.235	3.138	3.201	8.658	3.613	0.000
Ammonium (mg/L)	0.000	0.000	0.006	0.190	0.314	0.031	0.000	0.004	0.000
Potassium (mg/L)	0.627	2.153	5.543	1.383	0.825	0.352	0.291	1.472	0.000
Magnesium (mg/L)	0.000	0.000	0.000	0.000	0.478	0.440	0.000	0.000	0.000
Calcium (mg/L)	1.489	5.507	12.537	3.290	1.187	0.716	1.156	1.911	0.000
Fluorescence (Em. 455nm, Ab. 355nm)	1.331	1.385	1.849	3.695	1.204	1.447	0.861	1.673	-

Sites	TE057	TE191	TE206	TE306	TE322	TM115	TM117	TM122	TM197
Collect date	8-Jan-2013	23-Jan-2014	19-Mar-2014	18-Nov-2015	14-Jul-2016	22-Aug-2013	23-Aug-2013	24-Aug-2013	24-Jan-2014
Hour	9h00min	10h15min	10h25min	10h00min	10h44min	11h15min	8h20min	10h00min	14h02min
Location (GPS coordinates)	22.479188 S 47.407954 W	21.321783 S 49.509950 W	21.071467 S 49.673600 W	23.352850 S 46.825867 W	22.315650 S 48.051167 W	23.292017 S 47.654533 W	23.188633 S 48.119283 W	23.041333 S 48.054417 W	21.568650 S 48.814617 W
Altitude (ft)	2235	1493	1410	2504	2143	1827	1726	1629	1616
Altitude (m)	681	455	430	763	653	557	526	497	493
pH	5.94	7.04	8.50	6.71	8.31	6.90	7.10	7.10	5.39
Temperature (°C)	27.14	28.94	29.20	25.62	19.61	19.22	18.77	21.19	27.65
chlorophyll a (mg/L)	17.440	17.807	86.437	32.750	61.281	8.840	4.460	4.880	4.377
Trophic State Index	Eutrophic	Eutrophic	Eutrophic	Eutrophic	Eutrophic	Mesotrophic	Mesotrophic	Mesotrophic	Mesotrophic
Dissolved Organic Carbon (mg/L)	6.582	8.900	16.660	5.431	40.790	2.036	10.170	13.910	3.983
Total Carbon (mg/L)	13.950	12.670	17.870	17.120	41.580	7.192	13.470	24.940	6.119
Inorganic Carbon (mg/L)	7.370	3.766	1.216	11.690	0.791	5.156	3.297	11.030	2.137
Total Nitrogen (mg/L)	0.522	0.726	1.238	0.349	1.907	0.269	0.423	1.148	0.202
Chloride (mg/L)	1.729	6.628	2.336	4.142	-	1.621	1.478	28.039	0.924
Nitrite (mg/L)	0.000	0.000	0.000	0.000	-	0.421	0.000	0.000	0.000
Nitrate (mg/L)	0.036	0.000	0.021	0.026	-	0.496	0.000	0.060	0.000
Phosphate (mg/L)	0.000	0.000	0.000	0.004	-	0.000	0.000	0.000	0.000
Sulfate (mg/L)	0.222	0.470	0.097	1.694	-	0.603	0.048	0.603	0.000
Lithium (mg/L)	0.000	0.062	0.000	0.000	-	0.021	0.000	0.017	0.019
Sodium (mg/L)	3.162	4.290	2.541	5.628	-	3.701	2.165	7.405	2.483
Ammonium (mg/L)	0.309	0.277	0.238	0.182	-	0.373	0.249	0.324	0.316
Potassium (mg/L)	2.984	5.315	6.479	2.473	-	0.994	2.574	16.957	2.520
Magnesium (mg/L)	1.876	1.482	0.222	0.000	-	0.000	0.872	4.421	0.403
Calcium (mg/L)	6.334	4.513	0.607	11.219	-	3.651	1.189	12.567	0.809
Fluorescence (Em. 455nm, Ab. 355nm)	1.910	2.094	5.093	1.610	0.691	0.960	2.669	3.178	1.762

Sites	TM213	TO072	TO203	TO215	TO324	TO325
Collect date	21-Mar-2014	8-Apr-2013	18-Mar-2014	21-Mar-2014	21-Jul-2016	21-Jul-2016
Hour	9h15min	9h00min	10h40min	14h30min	10h07min	13h45min
Location (GPS coordinates)	20.939617 S 50.506483 W	23.164950 S 47.134117 W	21.509617 S 49.895500 W	21.239517 S 50.727733 W	22.713917 S 48.565139 W	22.436483 S 48.907417 W
Altitude (ft)	1118	2251	1405	1312	2024	1844
Altitude (m)	341	686	428	400	617	562
pH	5.77	4.79	6.74	5.03	6.74	6.04
Temperature (°C)	27.33	20.09	29.90	27.03	15.04	17.29
chlrophyll α (mg/L)	6.405	0.280	1.144	1.105	51.933	52.137
Trophic State Index	Mesotrophic	Oligotrophic	Oligotrophic	Oligotrophic	Oligotrophic	Oligotrophic
Dissolved Organic Carbon (mg/L)	10.960	4.318	6.387	4.547	1.274	5.138
Total Carbon (mg/L)	14.820	15.630	10.210	7.190	5.372	9.727
Inorganic Carbon (mg/L)	3.855	11.310	3.819	2.643	4.098	4.589
Total Nitrogen (mg/L)	1.279	0.441	0.509	0.185	1.004	0.181
Chloride (mg/L)	0.244	1.071	1.967	3.091	-	-
Nitrite (mg/L)	0.000	0.048	0.000	0.000	-	-
Nitrate (mg/L)	0.026	0.563	0.033	0.038	-	-
Phosphate (mg/L)	0.000	0.000	0.000	0.000	-	-
Sulfate (mg/L)	0.040	0.397	0.057	0.167	-	-
Lithium (mg/L)	0.000	0.020	0.018	0.013	-	-
Sodium (mg/L)	0.228	7.218	4.197	8.484	-	-
Ammonium (mg/L)	0.042	0.489	0.212	0.000	-	-
Potassium (mg/L)	0.417	3.374	3.430	10.425	-	-
Magnesium (mg/L)	0.086	1.891	1.236	4.692	-	-
Calcium (mg/L)	0.085	6.866	2.391	12.734	-	-
Fluorescence (Em. 455nm, Ab. 355nm)	5.229	1.690	1.206	2.150	0.492	2.237

OTU_id	Total Core	Tiete basin	Parapananema basin	Grande basin	Serra do Mar/formation	Phylum	Class	Order	Family	Genus	Species
OTU_11	-	-	-	X	X	Actinobacteria	Acidimicrobia	Acidimicrobiales	Acidimicrobiaceae	CL500-29_marine_group	uncultured_bacterium
OTU_13	-	X	-	-	-	Actinobacteria	Acidimicrobia	Acidimicrobiales	Acidimicrobiaceae	CL500-29_marine_group	uncultured_bacterium
OTU_26	-	X	-	-	-	Actinobacteria	Acidimicrobia	Acidimicrobiales	Acidimicrobiaceae	CL500-29_marine_group	uncultured_bacterium
OTU_51	-	X	-	-	-	Actinobacteria	Acidimicrobia	Acidimicrobiales	uncultured_bacterium	uncultured_bacterium	uncultured_bacterium
OTU_1	X	X	X	X	X	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_2	X	X	X	X	X	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_3	X	X	X	X	X	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_4	X	X	X	X	X	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_8	X	X	X	X	X	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_10	X	X	X	X	X	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_14	-	X	-	X	X	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_15	-	X	-	X	X	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_16	-	X	-	-	-	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_43	-	X	-	-	-	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_49	-	X	-	-	-	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_60	-	-	-	-	-	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_73	-	-	-	-	-	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_91	-	X	-	-	-	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_173	-	-	X	-	-	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_640	X	X	X	X	X	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_1112	-	-	-	-	-	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_1383	-	-	-	-	-	Actinobacteria	Actinobacteria	Frankiales	Sporichthyaceae	hgcl_clade	uncultured_bacterium
OTU_37	-	-	X	-	-	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Alphimonas	uncultured_bacterium
OTU_32	X	X	X	X	X	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	Candidatus_planktoluna	uncultured_bacterium
OTU_39	-	X	-	-	-	Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	uncultured_bacterium	uncultured_bacterium
OTU_38	-	X	-	-	-	Actinobacteria	Actinobacteria	PEM15	uncultured_bacterium	uncultured_bacterium	uncultured_bacterium
OTU_100	-	-	-	X	X	Bacteroides	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	Bacteroidaceae_bacterium_MS4
OTU_92	-	-	X	-	-	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	Sediminibacterium	uncultured_bacterium
OTU_1405	-	-	-	X	X	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	Sediminibacterium	uncultured_bacterium
OTU_114	-	-	-	X	-	Cyanobacteria	Melainabacteria	Obscuribacteriales	uncultured_bacterium	uncultured_bacterium	uncultured_bacterium
OTU_7	X	X	X	X	X	Cyanobacteria	Cyanobacteria	SubsectionI	FamilyI	Synechococcus	uncultured_bacterium
OTU_12	-	-	X	X	X	Planctomycetes	Planctomycetada	Planctomycetales	Planctomycetaceae	Synechococcus	uncultured_bacterium
OTU_20	-	X	-	-	-	Planctomycetes	Planctomycetada	Planctomycetales	Planctomycetaceae	uncultured_bacterium	uncultured_bacterium
OTU_34	-	-	-	X	X	Planctomycetes	Planctomycetada	Planctomycetales	Planctomycetaceae	uncultured_bacterium	uncultured_bacterium
OTU_40	-	-	-	-	-	Planctomycetes	Planctomycetada	Planctomycetales	Planctomycetaceae	uncultured_bacterium	uncultured_bacterium
OTU_63	-	-	X	-	-	Planctomycetes	Planctomycetada	Planctomycetales	Planctomycetaceae	uncultured_bacterium	uncultured_bacterium
OTU_115	-	-	-	X	-	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Burkholderia-Paraburkholderia	uncultured_bacterium
OTU_18	-	X	-	X	-	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Limnobabians	uncultured_bacterium
OTU_19	X	X	X	X	X	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Limnobabians	uncultured_bacterium
OTU_41	-	X	-	-	-	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Limnobabians	uncultured_bacterium
OTU_98	X	X	X	X	X	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Limnobabians	uncultured_bacterium
OTU_5	X	X	X	X	X	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Polynucleobacter	uncultured_bacterium
OTU_36	-	X	-	-	-	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Polynucleobacter	uncultured_bacterium
OTU_27	-	X	-	X	-	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	uncultured_bacterium	uncultured_bacterium
OTU_67	-	X	-	X	-	Proteobacteria	Betaproteobacteria	Caulobacteriales	Caulobacteraceae	uncultured_bacterium	uncultured_bacterium
OTU_21	-	X	-	X	-	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	Candidatus_Methylophilus	uncultured_bacterium
OTU_9	X	X	X	X	X	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylophilaceae	uncultured_bacterium	uncultured_bacterium
OTU_28	X	X	X	X	X	Proteobacteria	Alphaproteobacteria	Rhizobiales	alphal_cluster	uncultured_bacterium	uncultured_bacterium
OTU_33	X	X	X	X	X	Proteobacteria	Alphaproteobacteria	Rhizobiales	FukuN57	uncultured_bacterium	uncultured_bacterium
OTU_53	-	X	-	-	-	Proteobacteria	Alphaproteobacteria	Rhizobiales	FukuN57	uncultured_bacterium	uncultured_bacterium
OTU_117	-	X	-	X	-	Proteobacteria	Alphaproteobacteria	Rhizobiales	alphal_cluster	uncultured_bacterium_FukuS110	uncultured_bacterium
OTU_169	-	-	-	X	-	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Novosphingobium	uncultured_bacterium
OTU_17	-	-	-	X	-	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	Sphingomonas
OTU_54	-	-	X	X	X	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Nevskiaceae	Nevskia	uncultured_bacterium
OTU_6	-	X	-	-	X	Verrucomicrobia	Spartobacteria	Chthoniobacteriales	FukuN18_freshwater_group	uncultured_bacterium	uncultured_bacterium

Phylum	All Samples	Phylum	Tiete	Phylum	Parapanama	Phylum	Grande	Phylum	Serra do Mar
Parcubacteria	586	Parcubacteria	778	Parcubacteria	509	Parcubacteria	439	Parcubacteria	377
Proteobacteria	248	Proteobacteria	258	Proteobacteria	265	Proteobacteria	199	Proteobacteria	180
Omnitrophica	57	Omnitrophica	73	Saccharibacteria	47	Chlamydiae	35	Saccharibacteria	33
Chlamydiae	47	Chlamydiae	45	Chlamydiae	36	Saccharibacteria	33	Chlamydiae	28
Saccharibacteria	26	Saccharibacteria	41	Actinobacteria	28	Bacteroidetes	28	Actinobacteria	27
Firmicutes	21	Bacteroidetes	22	Omnitrophica	27	Actinobacteria	26	Planctomycetes	20
Candidatus_Berkelbacte	17	Actinobacteria	20	Bacteroidetes	27	Firmicutes	22	Omnitrophica	17
Planctomycetes	16	Firmicutes	19	Planctomycetes	21	Firmicutes	17	Firmicutes	16
WWE3	15	Candidatus_Berkelt	18	Firmicutes	19	Candidatus_Berkelb	16	WWE3	13
Cyanobacteria	15	WWE3	18	Cyanobacteria	16	Cyanobacteria	16	Candidatus_Berkelt	12
Acidobacteria	15	Planctomycetes	16	Verrucomicrobia	13	WWE3	15	Bacteroidetes	12
Bacteroidetes	13	Cyanobacteria	13	WWE3	11	Planctomycetes	14	Acidobacteria	10
Actinobacteria	12	TM6_(Dependentia)	13	CPR2	10	TM6_(Dependentia)	14	Verrucomicrobia	10
Woesearchaeota_(DHVE	12	Acidobacteria	12	Gracilibacteria	10	Acidobacteria	13	Cyanobacteria	10
Elusimicrobia	12	Elusimicrobia	10	Chloroflexi	9	Verrucomicrobia	13	TM6_(Dependentia)	8
Verrucomicrobia	11	Verrucomicrobia	9	Acidobacteria	9	Woesearchaeota_(I	8	Elusimicrobia	7
Gracilibacteria	11	Gracilibacteria	8	TM6_(Dependentia	9	CPR2	7	Gracilibacteria	7
TM6_(Dependentiae)	9	Nitrospirae	7	Candidatus_Berkelt	8	Microgenomates	6	Microgenomates	6
Peregrinibacteria	9	Peregrinibacteria	7	Elusimicrobia	7	Peregrinibacteria	4	Peregrinibacteria	6
WS2	6	Chloroflexi	6	Peregrinibacteria	6	Spirochaetae	4	Spirochaetae	4
Chloroflexi	6	WS2	6	Microgenomates	4	Gracilibacteria	4	Spirochaetae	4
Nitrospirae	4	Microgenomates	6	WS2	4	Elusimicrobia	4	CPR2	4
Spirochaetae	4	Woesearchaeota_(I	5	Fusobacteria	4	Gemmatimonadete	4	SR1_(Absconditaba	4
Gemmatimonadetes	3	Spirochaetae	5	Gemmatimonadete	4	Gemmatimonadete	4	Woesearchaeota_(3
Microgenomates	3	CPR2	4	SR1_(Absconditaba	3	Nitrospirae	4	Fusobacteria	3
Deinococcus-Thermus	2	Gemmatimonadete	4	Woesearchaeota_(I	3	Chloroflexi	2	Euryarchaeota	3
CPR2	2	SR1_(Absconditaba	2	Lentisphaerae	3	Euryarchaeota	2	Gemmatimonadete	2
WS6	2	Fusobacteria	2	Euryarchaeota	2	SR1_(Absconditaba	2	Chloroflexi	2
Ignavibacteriae	2	Lentisphaerae	2	Ignavibacteriae	2	Tenericutes	1	Nitrospirae	2
SR1_(Absconditabacteri	2	Fibrobrates	1	Nitrospirae	2	Armatimonadetes	1	WS6	2
Parvarchaeota	1	Euryarchaeota	1	Deinococcus-Therr	2	Armatimonadetes	1	WS2	2
Fusobacteria	1	Parvarchaeota	1	Aenigmarchaeota	1	Deinococcus-Therm	1	Diapherotrites	1
Aenigmarchaeota	1	Ignavibacteriae	1	Spirochaetae	1	Ignavibacteriae	1	Lentisphaerae	1
Lentisphaerae	1	Fibrobrates	1	Nitrospirae	1	Lentisphaerae	1	Chlorobi	1
Diapherotrites	1	Nitrospirae	1	Lentisphaerae	1	Fibrobrates	1	Fibrobrates	1
Nitrospirae	1	Latescibacteria	1	Tenericutes	1	Ignavibacteriae	1	Ignavibacteriae	1
Fibrobrates	1	Tenericutes	1	Thaumarchaeota	1	Latescibacteria	1	Latescibacteria	1
Thaumarchaeota	1	Thaumarchaeota	1	Thaumarchaeota	1	Latescibacteria	1	Latescibacteria	1
Total	1196	Total	1433	Total	1126	Total	959	Total	837

Considerações Finais

Análises envolvendo ecologia em ambientes interligados ganham força nas últimas décadas (Ricklefs, 2008). Estes são estudos que explicam de maneira muito mais satisfatória os processos que contribuem para a distribuição de espécies em cada ambiente. No entanto, são modelos relativamente novos, paradigmas ainda em processo de adaptação, digladiando-se para determinar qual é a melhor forma de se analisar os ambientes e absorvendo informações uns dos outros e de modelos anteriores para que assim, crie-se uma teoria que explique de forma concisa e coerente os fenômenos observados na natureza.

Um ótimo exemplo é a teoria da neutralidade conhecida principalmente após Hubbell (2001), apoia-se na teoria de Biogeografia de ilhas (MacArthur e Wilson, 1967) para propor

sua teoria, e foi alvo de intensos debates (i.e. Hubbell, 2003; Ricklefs, 2003). Até que sua aplicabilidade como modelo complementar ao clássico modelo de nichos começou a ser reconhecida (Gaston e Chown, 2005; Alonso *et al.*, 2006) e a teoria neutra foi finalmente absorvida por outra grande teoria de distribuição de espécies, a das metacomunidades (Leibold *et al.*, 2004).

Da mesma forma, a ecologia microbiana se desenvolve rapidamente tendo como alicerce o avanço tecnológico, superando e criando novos paradigmas a cada nova técnica aplicada à ecologia microbiana. As teorias de Baas Becking (1934) foram parcialmente superadas e hoje se sabe que há sim limites para a dispersão de microrganismos, apesar de o ambiente também selecionar quais poderão estar presentes em determinados locais (Martiny *et al.*, 2006).

Estas descobertas colocam estudos ecológicos da microbiota mais próximos de teorias largamente utilizadas e exaustivamente testadas em macro-organismos. E essa aproximação trás consequências muito positivas para o estudo ecológico. Primeiro o mais óbvio é o de entender de forma efetiva quais são os padrões que influenciam na distribuição das espécies microbianas (Martiny *et al.*, 2006; Logue e Lindström, 2008) e segundo é a contribuição que o estudo de microrganismos trás para os estudos de ecologia geral, fornecendo dados suficientes para se analisar a robustez de certas teorias científicas (Barberan *et al.*, 2014).

Na presente dissertação, procurei apresentar os dois aspectos mais importantes desta revolução nos estudos microbianos. Como fator tecnológico procurei mostrar as consequências do uso de diferentes técnicas em uma mesma etapa do processo metodológico para o resultado final das análises ecológicas e como fator ecológico, demonstrei a aplicabilidade de um modelo largamente utilizado em ecologia geral para a ecologia microbiana.

Desta forma, espero ter contribuído através dessa dissertação para lançar luz sobre a importância do uso de microrganismos em estudos de ecologia geral (Barberan *et al.*, 2014), em um momento crucial de desenvolvimento de novos modelos ecológicos de distribuição espacial de organismos é essencial aproximar as duas áreas de conhecimento e, entendendo seus pontos convergentes e divergentes, para que seja possível fundi-las, sempre que possível, em estudos amplos que visem uma avaliação mais plena de impactos ambientais ou da conservação de ambientes.

Referências

ALONSO, D.; ETIENNE, R. S.; MCKANE, A. J. The merits of neutral theory. **Trends Ecol Evol**, v. 21, n. 8, p. 451-7, Aug 2006. ISSN 0169-5347 (Print)

0169-5347 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16766082> >.

BAAS BECKING, L. G. M. **Geobiologie of inleiding tot de milieukunde**. Den Haag: Van Stockum, 1934.

BARBERAN, A.; CASAMAYOR, E. O.; FIERER, N. The microbial contribution to macroecology. **Front Microbiol**, v. 5, p. 203, 2014. ISSN 1664-302X (Print)

1664-302X (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24829564> >.

GASTON, K. J.; CHOWN, S. L. Neutrality and the niche. **Functional Ecology**, v. 19, n. 1, p. 1-6, 2005. ISSN 1365-2435. Disponível em: < [http://dx.doi.org/10.1111/j.0269-](http://dx.doi.org/10.1111/j.0269-8463.2005.00948.x)

[8463.2005.00948.x](http://dx.doi.org/10.1111/j.0269-8463.2005.00948.x) >.

HUBBELL, S. P. **The Unified Neutral Theory of Biodiversity and Biogeography**. 2001.

HUBBELL, S. P. Modes of speciation and the lifespans of species under neutrality: a response to the comment of Robert E. Ricklefs. **Oikos**, v. 100, n. 1, p. 193-199, 2003. ISSN 1600-0706. Disponível em: < <http://dx.doi.org/10.1034/j.1600-0706.2003.12450.x> >.

LEIBOLD, M. A. et al. The metacommunity concept: a framework for multi-scale community ecology. **Ecology Letters**, v. 7, n. 7, p. 601-613, 2004. ISSN 1461023X, 14610248.

LOGUE, J. B.; LINDSTRÖM, E. S. Biogeography of Bacterioplankton in Inland Waters. **Freshwater Reviews**, v. 1, n. 1, p. 99-114, 2008. ISSN 1755-084X.

MACARTHUR, R. H.; WILSON, E. O. **The Theory of Island Biogeography**. Princeton University Press, 1967. ISBN 9780691088365. Disponível em: < <https://books.google.com.br/books?id=a10cdkywhVgC> >.

MARTINY, J. B. H. et al. Microbial biogeography: putting microorganisms on the map. **Nat Rev Micro**, v. 4, n. 2, p. 102-112, 02//print 2006. ISSN 1740-1526. Disponível em: < <http://dx.doi.org/10.1038/nrmicro1341> >.

RICKLEFS, R. E. A comment on Hubbell's zero-sum ecological drift model. **Oikos**, v. 100, n. 1, p. 185-192, 2003. ISSN 1600-0706.

RICKLEFS, R. E. Disintegration of the ecological community. **Am Nat**, v. 172, n. 6, p. 741-50, Dec 2008. ISSN 1537-5323 (Electronic) 0003-0147 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18954264> >.