Microbial Ecology
https://doi.org/10.1007/s00248-020-01560-3

MICROBIOLOGY OF AQUATIC SYSTEMS

Local and Geographic Factors Shape the Occupancy-Frequency Distribution of Freshwater Bacteria

Erick Mateus-Barros[1,2] · Michaela L. de Melo[1,3] · Inessa L. Bagatini[4] · Adriano Caliman[5] · Hugo Sarmento[1]

Received: 14 April 2020 / Accepted: 14 July 2020

Abstract

Species prevalence across the landscape is related to their local abundance, which is a result of deterministic and stochastic processes that select organisms capable of recolonizing sites where they were once extinct, a process known as the rescue effect. The occupancy-frequency distribution (OFD) describes these patterns and has been extensively used to understand organism's distribution but has been poorly tested on microorganisms. In order to test OFD on freshwater bacteria, we collected data from 60 shallow lakes distributed across a wide area in southeastern Brazil, to determine the bacterial operational taxonomic units (OTUs) that were present in all sites (core) and at only one site (satellite). Then, we analyzed the spatial abundance distributions of individual OTUs to understand the influence of local abundances on regional occupancy patterns. Finally, we tested the environmental factors that influenced occupancy and abundance. We found a significant bimodal OFD for freshwater bacteria using both OTUs (97% clustering) and amplicon sequence variants (ASVs, unique sequences), with 13 core OTUs and 1169 satellite OTUs, but only three core ASVs. Core organisms had a bimodal or gamma abundance distribution. The main driver of the core community was pH, while nutrients were key when the core community was excluded and the rest of the community (mild and satellite *taxa*) was considered. This study demonstrates the close relationship between local environmental conditions and the abundance and dispersion of microorganisms, which shapes their distribution across the landscape.

Keywords Bimodal distribution · Core and satellite organisms · Diversity patterns · Microbial metacommunity

Introduction

The occupancy-frequency distribution, or OFD, describes species prevalence at different sites, which is helpful for understanding metapopulations and species spatial distribution. Raunkiaer's rule states that, if sampling is optimal, a bimodal distribution should be observed, with most species at the extremes of an OFD plot, meaning that most organisms will be found in one/few sites or in all/most sites [1, 2]. This concept has been extensively used to understand the geographical distribution of organisms, but has hardly been observed in natural communities [e.g., 3, 4] and has been poorly tested in microbial communities [but see 5, 6].

This rule gained popularity when Hanski [7] used its assumptions to explain species distribution in a metapopulation context [2], becoming one of the most prominent models to oppose the unimodal model [8]. In his mathematical models, Levins [8] considered a relationship between immigration and extinction rates, which eventually created a distribution with only one hump (unimodal distribution). On the other hand, Hanski [7] argued that local species extinctions were independent from immigration

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00248-020-01560-3) contains supplementary material, which is available to authorized users.

✉ Erick Mateus-Barros
erickmbarros@ufscar.br

[1] Department of Hydrobiology, Laboratory of Microbial Processes and Biodiversity, Federal University of São Carlos (UFSCar), São Carlos, SP 13565-905, Brazil

[2] Post Graduate Program in Ecology and Natural Resources (PPG-ERN), UFSCar, São Carlos, SP 13565-905, Brazil

[3] Département Des Sciences Biologiques, Université du Québec à Montréal, Montreal, Canada

[4] Department of Botany, Laboratory of Phycology, Federal University of São Carlos (UFSCar), São Carlos, SP 13565-905, Brazil

[5] Departamento de Ecologia, Universidade Federal do Rio Grande do Norte, Natal, RN 59078-900, Brazil

Published online: 23 July 2020

rates, and that both depend on regional species distributions; thus, there should be two humps in OFD across plots [7, 9]. Hanski labeled these two humps along the distribution as "core" (more frequent) and "satellite" (less frequent) organisms, resulting in the "core-satellite hypothesis."

Another important statement is that the OFD is indirectly related to local abundances, since the most abundant organisms are more likely to be part of the core, while the satellite should be composed of organisms with low abundance [7, 10]. In their classic study, Brown [10] argued that the geographic abundance distribution of a species should present a pattern with a center showing greater abundances towards the periphery, where it is increasingly rarer until eventually becoming extinct. Furthermore, the presence of a species in a certain location depends on its ability to adapt to the locations' specific conditions. Therefore, ubiquitous organisms are those that can settle in new areas after migrating from places where they are most abundant, increasing their abundance and migrating to new locations at further distances from the starting point. In this way, organisms recolonize sites where they are less abundant or extinct, originating from locations where they are more abundant, and the species that spread to more sites tend to be those that reach higher abundances in certain sites, a process known as the rescue effect [11]. This process probably occurs due to a mixture of deterministic and stochastic processes that select a set of organisms [10, 12, 13] and can be consistent in space [13] and time [14].

Historically, the dominant idea about the ecology of microorganisms (bacteria and protists) was that species could potentially be found in any site around the world [15]. Since they have extreme dispersal capacity and rapid evolution rates, only local environmental processes would be able to select organisms in each location [16–19]. However, the advance of culture-independent molecular tools that enabled researchers to access the entire microbial diversity changed that view, and microorganisms have become useful models in research fields from applied research and human health to ecology [16, 17]. Nowadays, we know that not all microorganisms are ubiquitous and that geographical barriers can also affect microbial dispersion [17, 18, 20]. In the last decades, concepts and theories about dispersion of macroscopic animals have been used to build a framework explaining microbial spatial distribution [18, 19]. Moreover, microorganisms are useful for stretching the assumptions of theoretical models and mathematical hypotheses to their limits, as they are present in high numbers everywhere and large datasets are freely available [16].

Therefore, OFD analysis is extremely useful for understanding the macroecological distribution of key species [5], as well as the influence of species abundance and local factors on regional dispersion patterns. Although the bimodal hypothesis has been tested in a wide range of environments, it has mainly been tested on macroorganisms [4, 21]. In the meantime, our understanding of microbial macroecological patterns has notably increased [17, 18, 20], and microorganisms have proven to be valuable models to test ecological hypotheses and evaluate dispersal mechanisms and community structure across space and time [16]. However, the OFD has just started to be tested on microbial communities [5, 6]. In this sense, our aim was to test the OFD on freshwater bacteria and provide insights about the influence of local environmental (i.e., pH, nutrients) and geographical (i.e., distance between sites) factors on OTU abundance and distribution across a tropical metacommunity—a set of local communities connected by species flow [22]. We hypothesized that (1) the OFD should be bimodal due to the high dispersal capacity of bacteria; (2) the most abundant organisms should be very abundant in some sites but rarer in others, so an abundance distribution analysis should show a dominance of bimodal arrangement; and (3) these patterns of occupancy and abundance must be related to a mix of environmental (pH, temperature) and spatial (distance, geographical barriers) factors.

Material and Methods

Study Sites and Sampling Design

Microbial communities were sampled from 60 shallow, headwater lentic systems scattered over an area of approximately 250,000 km^2 in São Paulo state, southeastern Brazil. The sampling area presents a tropical climate and encompasses a transition zone between Cerrado (Brazilian savannah biome) and tropical Atlantic Forest biomes. The geography includes four distinct areas that are nearly equivalent in area and perimeter (Tab. S1). The first area comprises part of the Brazilian coastal formation (named Serra do Mar), and the other three are watersheds (the Tietê River basin, the Paranapanema River basin, the Grande River basin), which all originate in the Paraná River basin, the most important watershed in this region (Fig. 1). Most of the sampled aquatic systems are small artificial reservoirs formed by creeks that are dammed to meet the needs of local humans. Natural vegetation in catchments of some of these reservoirs has been converted to pasture and cropland, mainly sugar cane. Samplings were carried out between June 2012 and July 2016. To avoid any bias, we defined two a priori criteria to select the spatial distribution of sampling design: (1) an equal sample size and approximate spatial distribution of sampling sites across the four watersheds and (2) within each watershed an equal sample size across trophic state categories (oligotrophic, mesotrophic, and eutrophic).

Environmental Variables, Sampling Procedures, and Laboratory Analysis

Water temperature, conductivity, and pH were obtained in situ, with a multi-parameter probe (YSI, Yellow Springs,

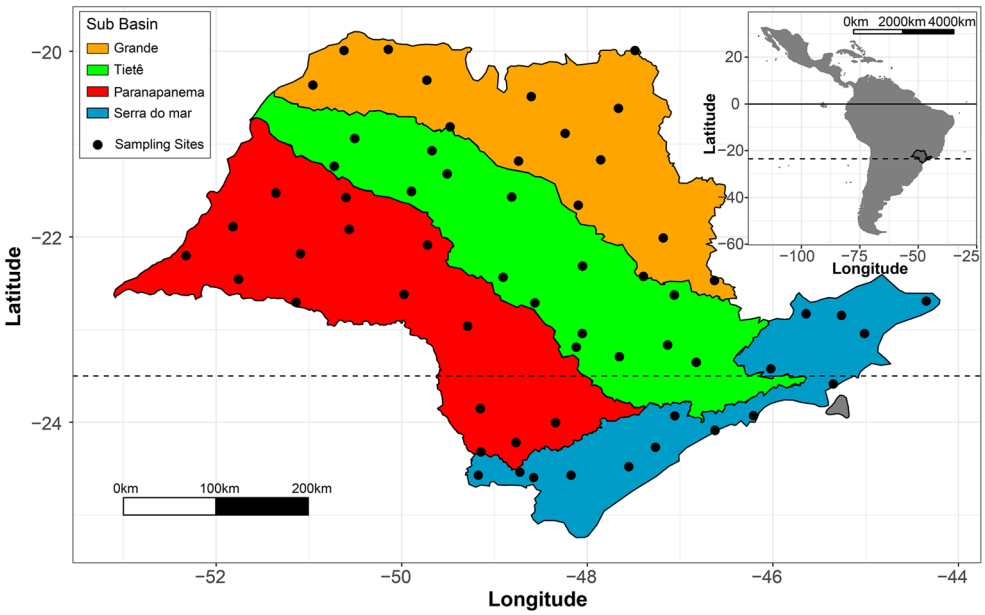

Fig. 1 Geographic location of 60 sampled shallow lakes in the study region within São Paulo State. All the sites were located inside a sub-region of the Parana river basin (gray) and selected to equally represent the four hydrological formations in this region: coastal formation (Blue), Grande river (orange), Tietê river sub-basin (green), and Paranapanema river sub-basin (red)

USA) and altitude was obtained with a GPS. The surface water samples for all analyses were collected with a bucket on the shore of the environments. For nutrient analysis, water samples were filtered through a polycarbonate membrane with 0.45-μm mesh. At each site, the membrane was washed with ultrapure water and rinsed with lake water, thus preventing carbon implement caused by the filter. Afterwards, 20 mL of water was slowly filtered, and samples were frozen and stored in amber bottles. In the laboratory, the samples were stored in a freezer at − 20 °C until posterior analysis. Dissolved organic carbon (DOC), inorganic carbon (IC), total carbon (TC), and total nitrogen (TN) were obtained using a TOC-V (Shimadzu®, Kyoto, Japan). The concentration of dissolved nutrients was obtained in an Ion Chromatography System (Thermo Scientific®, Waltham, Massachusetts, USA) and after, dissolved inorganic nitrogen (DIN) was obtained by adding the observed values of nitrite, nitrate, and ammonium. Tryptophan-like fluorescent dissolved organic matter (T-fDOM), known as a biological activity indicator [23, 24], was measured in a FS5 Spectrofluorometer (Edinburgh Instruments®, Livingston, UK), calculating the ratio between dissolved organic matter fluorescence and quinine sulfate (0.001 mg/L dissolved in 0.1 M H_2SO_4) at 455 nm excitation and 355 nm emission. The concentration of phytoplanktonic chlorophyll *a* (chl-*a*) was used to determine the trophic state of each sampling site according to Cunha and collaborators [25]. To measure chl-*a* concentration, 100 to 500 ml of water was passed through a glass fiber filter (Macherey-Nagel® GF-6), extracted in ethanol (90% v/v at 80 °C) in the dark [26, 27], and quantified by spectrophotometry [28]. To test the importance of environmental factors on spatial distribution, we decided to use altitude, pH, temperature, conductivity, DOC, DIC, DIN, T-fDOM, and chl-*a* as variables because they are widely known as essential factors that shape microbial communities, except for phosphate, which was removed from analyses due values that were very low or close to the detection limit of the method (see Tab. S1).

Sampling Procedures, Molecular Analysis, and Bioinformatics for Microbial Communities

For molecular analyses, 400 to 500 ml of water was collected as described above and pre-filtered through a glass fiber filter with 1.2-μm mesh (BOECO® MGC) to retain eukaryotes, large particles, and attached prokaryotes. Then, it was filtered through polycarbonate 0.22-μm membranes (Millipore® Isopore™ 0.2 μm GTBP) to retain the free-living prokaryotes. At each site, one filter for eukaryotes and particle attached bacteria and another for free-living bacteria were obtained. In this study, we only used the free-living bacterial community. The filters with microbial samples were divided in two, frozen in liquid nitrogen, and then stored in a − 80 °C ultrafreezer until DNA extraction.

A detailed description of the molecular procedure and bioinformatics were provided by Mateus-Barros and collaborators [29]. Briefly, we used one half of a filter to extract the bacterial DNA from each sample with phenol-chloroform. The amplification of V3-V4 regions of 16S rRNA was performed using the 341F (5′-CCTACGGGNGGCWGCAG-3′) and 805R (5′-GACT ACHVGGGTATCTAATCC-3′) primers [30] and a KAPA HiFi HotStart ReadyMix PCR Kit (Kapa Biosystems®). Finally, the fragments were sequenced in an Illumina MiSeq platform.

The sequences obtained were processed to determine operational taxonomic units (OTUs) with 97% similarity. We obtained the OTUs table using an internally implemented

pipeline [31, 32] based on UPARSE routines [33] to recognize paired sequences, eliminate singletons and chimeras, and, finally, compare the obtained sequences with the SILVA version 128 database [34, 35] to determine OTUs taxonomy. For further analyses, all chloroplast, mitochondria, and Archaea sequences were removed. The OTUs table was rarefied by the lowest sample richness (14699 sequences) and transformed into relative abundances.

As a complement to the analysis described above, we also processed the data to obtain a ASVs table. This newest approach is obtained by applying the DADA2 pipeline [36] to the software R [37], using the SILVA version 128 database [34, 35] for taxonomy assignment. These exact sequences were then blasted against the NCBI database (https://www.ncbi.nlm.nih.gov) and manually recorded the geographic coordinates of each sequence (> 300 bp, with 100% nucleotide match). Our aim was to test the OFD and observe if the approach would strongly influence the ecological pattern.

Statistical Analysis

Aiming to observe the distribution pattern throughout the metacommunity, we performed the OFD analysis based on the method showed by Lindh and collaborators [5], using the *MOStest* function from vegan package [38] and a presence/absence transformed OTUs table. The *MOStest* function is based on Mitchell-Olds and Shaw [39] to determine if analyzed data has only one hump at one of the extremes of its occupancy-frequency distribution; however, if the occupancy distribution shape has two humps, one at each extreme of the graph, the distribution is considered bimodal [38, 40]. After the OFD analysis, three distinct groups were observed. We called the bacteria found in all sites the "core," and those restricted to only one site as the "satellite." The bacteria that did not fit in the core nor the satellite criteria were called "mild."

Then, to understand the relationship between occupancy and abundance, we used a $\log_{10}(x + 1)$-transformed OTUs table to perform the SpaDs analysis. To determine the pattern of variation, we subsampled our dataset to create three tables containing the core, mild, and satellite communities separately. This analysis was implemented here as proposed by Niño-García and collaborators [13]. It is an iterative procedure that determines the frequency of abundances of each OTU and categorizes these frequencies in graphical curves to visualize the abundance tendency of these organisms in a metacommunity. The analysis first used the *Hartigan's dip test* implemented in the dip test package [41] to detect the non-unimodal abundance distributions and classify them as bimodal. After, the *distfit* function in the *distrplus* package [42] was used to carry out a maximum likelihood estimation and classify the unimodal abundance distributions as Cauchy, exponential, gamma, logistic, lognormal, normal, or Weibull categories. In their study, Niño-García and collaborators [13] decided to include some groups with unimodal distribution under the "normal like" category. Since we did not observe OTUs with normal distribution, we preferred to keep the groups separated to avoid artificial grouping.

Finally, in order to evaluate whether and how the sampled environmental factors affect bacterial communities, we used a $\log_{10}(x + 1)$-transformed OTUs table to perform a Mantel test and a distance-based redundancy analysis (dbRDA) with the *mantel* and *capscale* functions of *vegan* package [38], respectively. Before these analyses, the environmental dataset was standardized using the *decostand* function (method: standardize) in the vegan package [38]. The dbRDA was performed using the complete dataset and the dissimilarity matrices were constructed based on the Bray-Curtis distance analysis. We used this approach instead of the classical RDA because it permits the use of Bray-Curtis distance matrices, which is more appropriate for our OTU table that presents many zeros, as expected from a biological data table for microorganisms, obtained from high-throughput techniques. To perform the Mantel test, the OTU table was divided to obtain sub-tables considering the results of OFD (core, mild, and satellite) and SpaDs (bimodal, gamma, and other) and then, each dataset (the sub-tables and the complete table) was transformed into Bray-Curtis distance matrices. These matrices were then compared to the environmental data and geographic distances. A Euclidean distance approach was applied to the environmental data, while the geographic data was transformed into distance in a way that considers the distance between each site measured by the longitude/latitude coordinates in decimal degrees. For the Mantel comparisons between distance matrices, we used a Spearman method. The datasets were transformed into distance matrices using the *vegdist* vegan function for environmental and biological datasets, and *rdist.earth* function from the fields package [43] for geographical matrix. All the statistical analyses and plots were performed in the R software [37, 44].

Results

After rarefaction, 3738 OTUs were obtained from 881,940 reads (14,699 sequences per sample) and classified within 38 different phyla (Tab. S2). The *phyla* with the highest number of OTUs were Parcubacteria (1682 registered OTUs), Proteobacteria (799), and Actinobacteria (166), while the most abundant *phyla* were Actinobacteria (54.24% of the reads), Proteobacteria (23.64%), and Planctomycetes (5.41%). For more details about OTUs count and relative abundance per phyla, see Tables S2 and S3, and Figure S1.

The OFD had an inverse-Gaussian shape, with one hump at each extreme of the frequency distribution. Furthermore, the MOStest was significant for both ends of the distribution

curve ($p < 0.05$), which means that freshwater bacteria had a bimodal occupancy-frequency distribution in our study sites (Fig. 2a). We performed the MOStest in the ASVs table and the results were also significant, despite the fact that only three ASV could be classified as core (Fig. S2).

Core organisms (present in all sites) were represented by 13 OTUs, which represent only 0.35% of the total OTUs number. On the other hand, satellite (present in only one site) were represented by 1169 OTUs (i.e., around 31% of all OTUs) distributed in over 38 *phyla* (Tab. S2). On average, core OTUs represented 50% of relative abundances, while satellite were less than 1% of relative abundances (Tab. S2, Fig. 2b). The most prominent bacteria among the core were from the hgcI clade (phylum Actinobacteria), while the Parcubacteria was the most prominent group among satellite organisms, accounting for more than half of the total identified OTUs (Tab. S2, S3).

Concerning the SpaDs results, we found three different categories of abundance distribution (Fig. 3): (1) the gamma category, observed in a large number of OTUs, which is characteristic of organisms that are mostly absent (in the case of satellite) or not abundant (in the case of core) in the studied sites (Fig. 3a); (2) the bimodal category includes OTUs with high abundances in many sites, but are less abundant in many others (Fig. 3a); and (3) other abundance categories (Cauchy, log-normal, and logistic) that have a more heterogeneous form, mostly grouping OTUs that are less abundant than others that can be very abundant in few sites (Fig. 3a and Figure S3). The gamma was represented by core, mild, and satellite OTUs although was most represented by mild and satellite. On the contrary, the bimodal was more represented by mild and core OTUs. Lastly, the other category presented some representatives of mild and core (Fig. 3b).

Finally, the Mantel test showed that five variables were significantly correlated to bacterial community dissimilarity matrices ($p < 0.05$). The pH had the strongest correlation with aquatic bacterial community composition, which was also highlighted by the dbRDA (see Fig. S4). The bacteria found in this study were also influenced by chemical factors (pH and conductivity), nutrients and organic matter composition (DIN and T-fDOM), and geographic distances (Fig. 4). The relative importance of each significant environmental factor varied when compared to the OFD groups: pH was also the strongest factor when compared to the core OTUs, while the mild OTUs distribution were more related to nutrients (DIN) and organic matter composition (T-fDOM), and the satellite OTUs were not associated with any environmental variable (Fig. 4a, Tab. S5). The categorized SpaDs analysis also showed variation of OTU correlations with the environmental variables. The category of OTUs with bimodal distribution was more related to pH, while Gamma category had similar correlation with pH, DIN, and T-fDOM. The other category seemed to be homogeneously influenced by pH, T-fDOM, DIN, and conductivity (Fig. 4b, Tab. S6).

Discussion

The Occupancy-Frequency Distribution of Freshwater Bacteria Is Bimodal

We found a significant bimodal distribution for the OFD of freshwater bacteria, which agrees with a previous report on marine bacterial communities [5] and freshwater bacteria in a large lake [6], but disagrees with another report on stream diatoms [3]. Soininen and Heino [3] argued that the lack of bimodality in their dataset for eukaryotic microalgae was probably caused by a lack of connectivity between sampling sites in distinct river basins, which prevented the flow of individuals between sites. Herein, bacterial dispersal capacity probably overcame geographic isolation or historical factors. Similar arguments were used to explain why birds have core populations, while other vertebrates as herptiles or mammals do not [45]. Another possible explanation for these results is

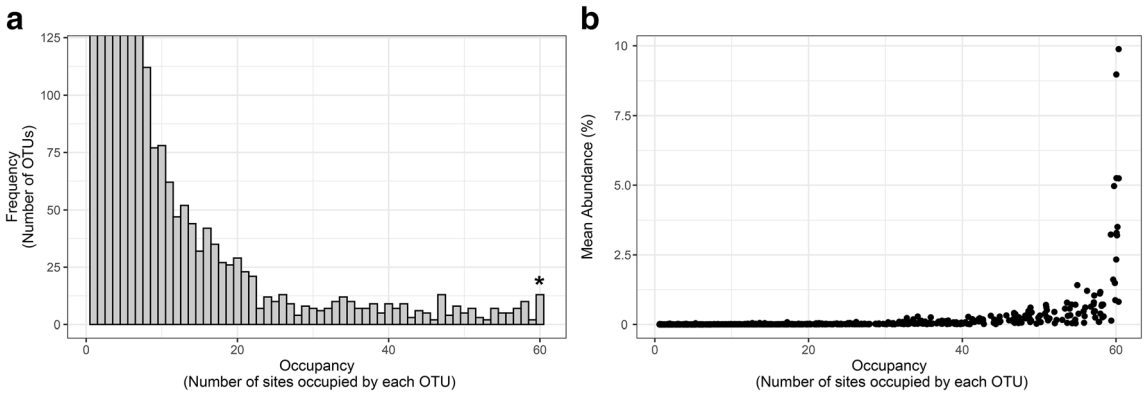

Fig. 2 **a** Occupancy-frequency distribution from 60 shallow lakes scattered throughout São Paulo State. The asterisk (*) represents a significant result in the MOStest ($p < 0.05$), which means that the distribution presents a consistent hump in each extreme. **b** Occupancy-abundance distribution plot. The dots represent the mean abundance of each OTU

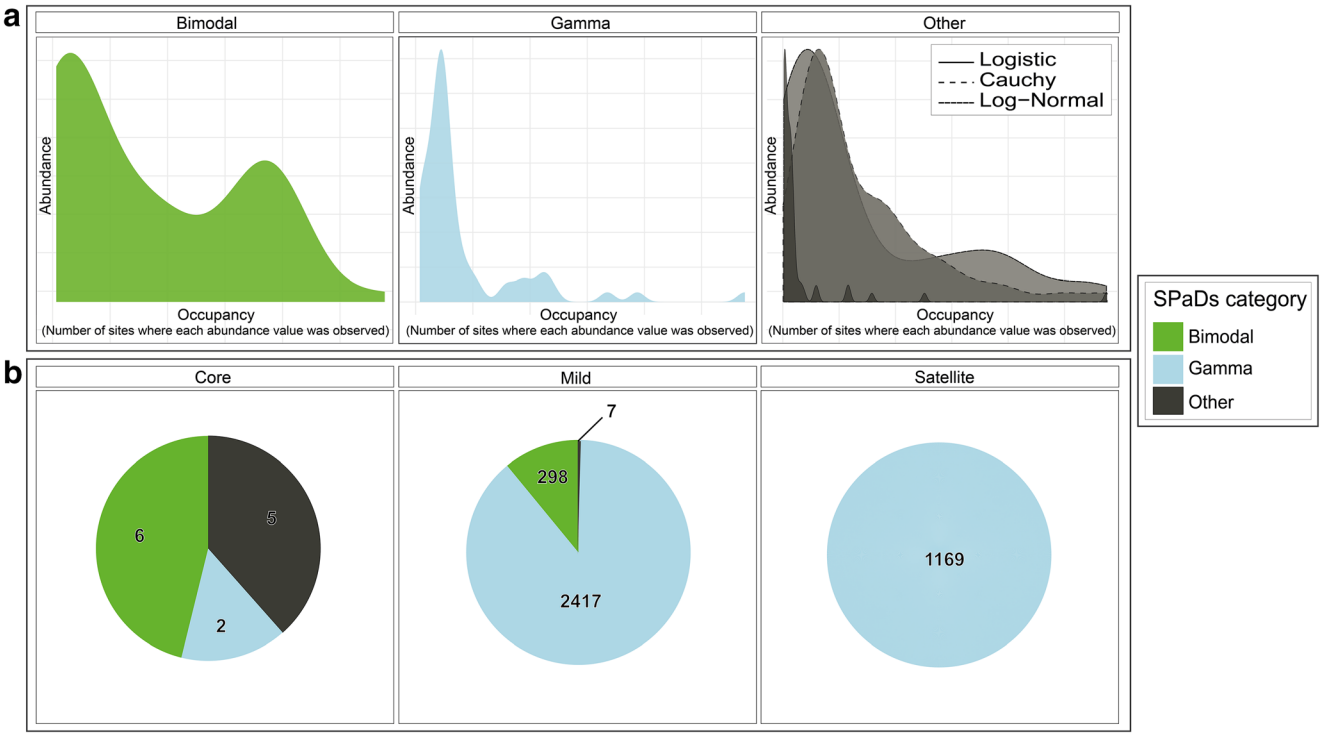

Fig. 3 **a** Examples for each SpaDs category recovered in this study (bimodal, gamma, Cauchy, log-normal, and logistic). Each curve represents the abundance distribution for one OTU. **b** The absolute number of OTUs of each OFD type that falls into each SpaDs category

the existence of a terrestrial seed bank that constantly feeds aquatic environments with bacteria, a process that occurs in headwater environments that frequently promotes recolonization and prevents local extinctions [46].

The same analysis using a higher clustering resolution (ASVs) yielded slightly different results: the bimodality remained significant, while the bimodal hump was weaker and only three ASVs were core. The inaccuracy of molecular methods used to access the microbial biodiversity, from DNA extraction [29], amplification [47], and sequencing errors [48], could bias the interpretation of ecological processes at a high clustering resolution level, such as ASVs. Actually, both methods used to determine taxonomic units are prone to criticism, as they define taxonomy based on an arbitrary threshold value (97% similarity for OTUs; 100% similarity for ASVs), OTU does not consider different diversification rates for each bacterial phylogenetic group, and ASV may artificially overestimate diversity [48].

Core Group Is Dominated by Native Freshwater Bacteria

The hgcI clade stands out as the most frequent and abundant group in the core. This Actinobacteria clade (also known as acI lineage) seems to be closely related to freshwater environments and was abundant in other studies of freshwater [49–52] and estuarine systems [53]. Members belonging to the hgcI clade are small (< 0.1 μm^3), free-living, and have

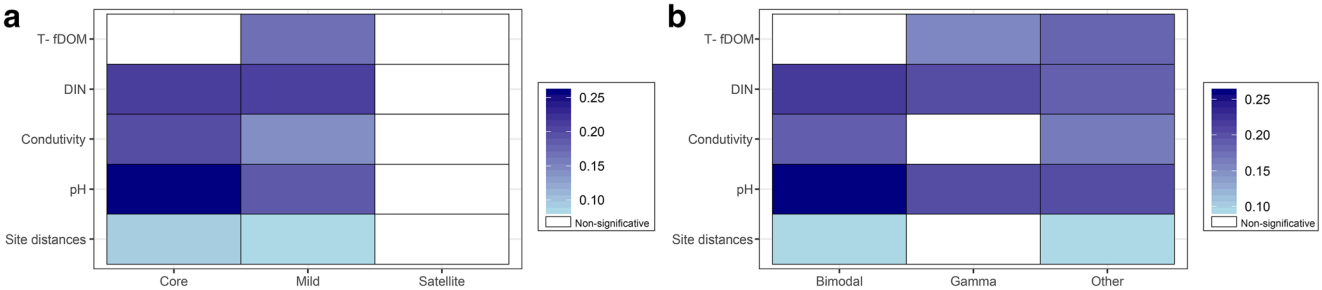

Fig. 4 Mantel test results showing significant correlations ($p < 0.05$) between environmental variables and OTUs relative abundances for the OFD groups: core, mild, and satellite bacteria (**a**) and abundance distribution categories: bimodal, gamma, and others (**b**). The correlation intensity was expressed in blue. Darker colors represent stronger correlations between variables and OTUs; white boxes represent non-significant correlation

been highlighted as the most abundant and cosmopolitan bacterial group in freshwater ecosystems [54, 55]. They can potentially take up carbohydrate and N-rich organic compounds, but can also utilize sunlight via actinorhodopsin, and perform anaplerotic carbon fixation [56]. Despite the accumulation of genomic data from this group through culture-independent methods, the first strains have only recently been isolated from a natural lake and maintained in pure cultures using catalase supplementation [55].

Polynucleobacter (Pnecc) is a well-studied genus of ultramicrobacteria, which also represents a ubiquitous and abundant component of freshwater bacterioplankton [57, 58]. In contrast to the hgcI clade, strains of Pnecc have been isolated and studied from a wide variety of freshwater ecosystems, differing in environmental conditions, which suggests high diversification within the group [59, 60]. This diversification could be due to adaptations to specific ecological conditions that likely provide competitive advantages that make them globally distributed [58]. Moreover, they assimilate subproducts from the photodegradation of humic substances [61], which are organic substrates that are very abundant in freshwaters worldwide. Another well-represented phylum in the core was Proteobacteria, represented by some typical freshwater organisms like *Polynucleobacter* [62, 63]. Additionally, the bacterial genus *Synechococcus* was included in the core community. This genus has been well-studied and are freshwater Cyanobacteria [57] commonly found in tropical waters [51, 52] where they contribute greatly to primary production. Other bacteria that are typically found in nutrient-rich freshwaters (i.e., *Nevskia* and Methylophilaceae) and in both terrestrial and freshwater environments (i.e., Microbacteriaceae and Rhizobiales) were also expected to dominate the shallow lakes due to the massive influx of terrestrial organisms from soil to headwaters [46]. Despite the fact that some of these bacteria were not included in the core, they presented abundant representatives, reinforcing the previous argument that the soil seed bank is probably a key factor for maintaining core bacteria in the environments analyzed herein.

Parcubacteria Is the Most Represented Phyla Among Satellite OTUs

There is evidence that this group is entirely composed of symbiotic bacteria [64], which can explain its presence between the satellite in almost all environments studied herein. Unfortunately, Parcubacteria is one of the least known prokaryotic groups and was the most present organism within the satellite recorded herein, thus hampering the discussion about satellite organism. Furthermore, it is also one of the most poorly classified organisms, mostly known at the Phylum level only. Herein, its relevance to diversity in tropical inland environments highlights the need for future studies about the biology of this group.

The Relationship Between Occupancy and Abundance

The SpaDs showed that the most common abundance distribution categories were bimodal and gamma, with higher representativeness of the bimodal category among core organisms, indicating that the core may in fact be locally rare or even extinct, but recolonization is possible from other sites where they are more abundant. The same categories were dominant in temperate regions [13], and a similar description was used by Lindh and collaborators [5] to discriminate the satellite and core from other occupancy-frequency distribution categories. They called the organisms with high relative abundances and occupancy the "core" and the organisms with low abundances and occupancy the "satellite." Since a bimodal category implies more frequent high local abundances than expected for the other categories, core organisms present more sites in which they are a source population capable of recolonizing other sites. If this assumption is correct, core organisms would present this abundance arrangement in space [12] and time [14].

The presence of OFD is historically related to local abundances [7, 10] and is mainly explained by a process known as rescue effect [11], which predicts that locally abundant organisms tend to easily colonize the landscape. The importance of core for local abundances (on average, more than 50% relative abundances) compared to the satellite (less than 1% relative abundances) is a good indication that a rescue effect allowed more abundant organisms to frequently recolonize any site where they were less abundant and prevented an eventual local extinction risk. However, a study including the temporal analysis of abundance fluctuations must be conducted to confirm this.

Environmental Factors That Shape Occupancy and Abundance Distribution

Finally, the compartmentalized Mantel test showed that, in general, the main factor related to OFD and SpaDs is pH, as reported in other inland aquatic environments [12, 46, 65, 66]. Indeed, environmental factors seem to modulate the bacterial OFD, pH in soil and freshwater environments [12, 46, 65, 66], and salinity in marine environments [5]. Interestingly, the removal of core bacteria showed that other factors might also be relevant. These results seem to indicate that pH is more strongly related to the composition of the most abundant bacteria, while the mild bacteria are related to pH, along with nutrients, and the satellite bacteria with stochasticity. The increased importance of T-fDOM, an indirect indicator of autochthonous carbon supply [23, 24], highlights the importance of an autochthonous metabolism at these sites [67] and is a potential indicator of phytoplankton-bacteria interaction, which is presumably high in tropical regions due to high phytoplankton excretion rates at high temperatures and light [68].

Niño-García and collaborators [13] argued that the presence of a large portion of bacteria in lakes was linked to downstream transport from headwaters. Herein, only the satellite bacteria were not related to any environmental condition, indicating that sorting may be more important in lotic headwater environments. After, these selected organisms are taken by the water flow and colonize other locations. Another possible explanation is that the high degree of isolation between our samples made the sorting action more evident than in other sites strongly connected by water flow. The mechanisms that select this larger portion of rare microbial organisms have been extensively addressed over the past few years. The rare biosphere (i.e., microbial taxa that are found in low abundance in a location that can be rare or dormant [69, 70]) can have diverse persistence strategies and be sorted by a set of variables [69]. Herein, the large portion of rare bacteria may be explained by autochthonous nutrient inputs. Besides that, the significant correlation between bacterial community composition and geographic distance, although weaker than environmental sorting, indicates that geographic components such as dispersion and isolation can influence the bacterial freshwater distribution and abundance. However, it is necessary to sample on a larger scale to discover from which distances a bimodal OFD can no longer be observed, and then investigate if it was caused by geographic barriers or environmental selection and bacterial occupancy.

In this study, we demonstrated a bimodal pattern of the OFD for freshwater bacteria across a tropical landscape, with the core being dominated by typical freshwater bacteria. Our dataset also includes the presence of bacteria typically found in nutrient-rich freshwaters and other bacteria found in both terrestrial and freshwater environments, indicating an important contribution from soil seed bank to OFD. The abundance distribution also showed an arrangement with almost half of the core OTUs fitting in the bimodal abundance category, which indicates that, despite the fact that core bacteria was found in low abundances in many sites, their high abundances in other sites may compensate and assure their presence across the landscape. On the other hand, the gamma category was more related to less abundant organisms and indicated that they are consistently rare across the landscape. Furthermore, we found that pH was the major driver of bacterial community composition at our study sites, but a compartmentalized analysis indicates that other nutritional (i.e., DIN and T-fDOM) and stochastic factors (i.e., geographic distance and randomness) are also relevant for the bacterial community composition.

In the last decades, we have overcome the traditional idea that microorganisms are not influenced by geographical barriers and only by environmental selection [16–19]. Meanwhile, microbes have been emerging as ideal model organisms to stress ecological theory and mathematical hypotheses [16]. Our observations support the view of a close relationship between bacterial abundance and their ability to spread across a landscape, with environmental factors playing an important role in selecting their capacity to become ubiquitous or rare in natural systems. In addition, we also demonstrated the applicability of OFD allied with SpaDs to provide information about the macroecology of microorganisms along freshwater habitats, enabling the recognition of *taxa* that are important for local processes and global functions.

Acknowledgments We are grateful to Prof. Armando Vieira for providing the samples from the Biota-FAPESP project and to his lab team 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 participated in the field sampling and performed physio-chemical analyses. We also thank the LMSeq Laboratory of UNESP for DNA sequencing. Adriano Caliman gratefully acknowledges continuous funding through Research Productivity Grants provided by CNPq.

Funding Information This research was funded by the São Paulo Research Foundation - FAPESP (projects 2014/14139-3 and 2011/50054-4). EMB was funded by a CAPES scholarship (grant 001); HS was supported by CNPq (grant 309514/2017-7).

Data Availability The 16S rRNA amplicon results have been deposited in the NCBI repository under accession number PRJNA411849.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Code Availability The R scripts used in these analyses can be found at: https://github.com/LMPB/Occupancy-Frequency-Distribution

References

1. Gleason H (1929) The significance of Raunkiaer's law of frequency. Ecology 10:406–408
2. Papp L, Izsák J (1997) Bimodality in occurrence classes: a direct consequence of lognormal or logarithmic series distribution of abundances: a numerical experimentation. Oikos 79:191–194
3. Soininen J, Heino J (2005) Relationships between local population persistence, local abundance and regional occupancy of species: distribution patterns of diatoms in boreal streams. J Biogeogr 32: 1971–1978
4. Tonkin JD, Arimoro FO, Haase P (2016) Exploring stream communities in a tropical biodiversity hotspot: biodiversity, regional occupancy, niche characteristics and environmental correlates. Biodivers Conserv 25:975–993
5. Lindh MV, Sjostedt J, Ekstam B, Casini M, Lundin D, Hugerth LW, Hu YO, Andersson AF, Andersson A, Legrand C, Pinhassi J

(2017) Metapopulation theory identifies biogeographical patterns among core and satellite marine bacteria scaling from tens to thousands of kilometers. Environ Microbiol 19:1222–1236
6. Jeong SY, Choi JY, Kim TG (2020) Coordinated metacommunity assembly and spatial distribution of multiple microbial kingdoms within a lake. Microb Ecol 79:801–814
7. Hanski I (1982) Dynamics of regional distribution: the core and satellite species hypothesis. Oikos 38:210–221
8. Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. Bull Entomol Soc Am 15:237–240
9. Gotelli NJ (1991) Metapopulation models: the rescue effect, the propagule rain, and the core-satellite hypothesis. Am Nat 138: 768–776
10. Brown JH (1984) On the relationship between abundance and distribution of species. Am Nat 124:255–279
11. Gaston KJ, Blackburn TM, Greenwood JJ, Gregory RD, Quinn RM, Lawton JH (2000) Abundance–occupancy relationships. J Appl Ecol 37:39–59
12. Niño-García JP, Ruiz-González C, del Giorgio PA (2016) Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. ISME J 10:1755–1766
13. Niño-García JP, Ruiz-González C, Del Giorgio PA (2016) Landscape-scale spatial abundance distributions discriminate core from random components of boreal lake bacterioplankton. Ecol Lett 19:1506–1515
14. Niño-García JP, Ruiz-González C, Del Giorgio PA (2017) Exploring the ecological coherence between the spatial and temporal patterns of bacterioplankton in boreal lakes. Front Microbiol 8: 636
15. Baas Becking LGM (1934) Geobiologie of inleiding tot de milieukunde. Van Stockum, Den Haag
16. Barberán A, Casamayor EO, Fierer N (2014) The microbial contribution to macroecology. Front Microbiol 5:203
17. Green JL, Bohannan BJ, Whitaker RJ (2008) Microbial biogeography: from taxonomy to traits. Science 320:1039–1043
18. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JB (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol 10:497–506
19. Horner-Devine MC, Bohannan BJ (2006) Unifying ecology to include all creatures great and small. Trends Ecol Evol 21:473
20. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreas L, Reysenbach A-L, Smith VH, Staley JT (2006) Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 4:102–112
21. Heino J (2015) Deconstructing occupancy frequency distributions in stream insects: effects of body size and niche characteristics in different geographical regions. Ecol Entomol 40:491–499
22. Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman D, Loreau M, Gonzalez A (2004) The metacommunity concept: a framework for multi-scale community ecology. Ecol Lett 7:601–613
23. Cammack WKL, Kalff J, Prairie YT, Smith EM (2004) Fluorescent dissolved organic matter in lakes: relationships with heterotrophic metabolism. Limnol Oceanogr 49:2034–2045
24. Bertilsson S, Tranvik LJ (1998) Photochemically produced carboxylic acids as substrates for freshwater bacterioplankton. Limnol Oceanogr 43:885–895
25. Cunha DGF, Calijuri MC, Lamparelli MC (2013) A trophic state index for tropical/subtropical reservoirs (TSItsr). Ecol Eng 60:126–134
26. Marker A (1980) The measurement of photosynthetic pigments in freshwaters and standardization of methods: conclusions and recommendations. Arch Hydrobiol Ergebn Limnol 14:91–106
27. Mush E (1980) Comparison of different methods for chlorophyll and phaeopigment determination. Arch Hydrobiol Beih 14:14–36
28. Lorenzen CJ (1967) Determination of chlorophyll and pheo-pigments: spectrophotometric equations. Limnol Oceanogr 12:343–346
29. Mateus-Barros E, Meneghine AK, Bagatini IL, Fernandes CC, Kishi LT, Vieira AAH, Sarmento H (2019) Comparison of two DNA extraction methods widely used in aquatic microbial ecology. J Microbiol Methods 159:12–17
30. 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. ISME J 5:1571–1579
31. Logares R (2017) Ramalok/amplicon_processing: workflow for analysing MiSeq amplicons based on Uparse
32. Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmento 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. Environ Microbiol 16:2659–2671
33. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996–998
34. 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 Res 41:D590–D596
35. 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 Res 42:D643–D648
36. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP (2016) DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583
37. R. Core Team (2016) R: A Language and Environment for Statistical Computing. In: Editor (ed.)^(eds.) Book R: A Language and Environment for Statistical Computing, vol. R Foundation for Statistical Computing, City
38. 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
39. Mitchell-Olds T, Shaw RG (1987) Regression analysis of natural selection: statistical inference and biological interpretation. Evolution 41:1149–1161
40. Ter Braak CJ, Looman CW (1986) Weighted averaging, logistic regression and the Gaussian response model. Vegetatio 65:3–11
41. Maechler M (2016) Diptest: Hartigańs Dip Test Statistic for Unimodality - corrected. In: editor (ed.)^(eds.) Book diptest: Hartigańs dip test statistic for unimodality - corrected, vol., City
42. Delignette-Muller ML, Dutang C (2015) fitdistrplus: an R package for fitting distributions. J Stat Softw 64:1–34
43. Nychka D, Furrer R, Paige J, Sain S (2017) Fields: tools for spatial data (R package), version 8.3-6. In: editor (ed.)^(eds.) Book fields: tools for spatial data (R package), version 8.3-6, vol., City
44. Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York
45. Mehranvar L, Jackson DA (2001) History and taxonomy: their roles in the core-satellite hypothesis. Oecologia 127:131–142
46. Ruiz-González C, Niño-García JP, del Giorgio PA (2015) Terrestrial origin of bacterial communities in complex boreal freshwater networks. Ecol Lett 18:1198–1206
47. Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL, Polz MF (2004) Fine-scale phylogenetic architecture of a complex bacterial community. Nature 430:551–554

48. Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res 43:e37
49. Hosen JD, Febria CM, Crump BC, Palmer MA (2017) Watershed urbanization linked to differences in stream bacterial community composition. Front Microbiol 8:1452
50. Mentes A, Szabo A, Somogyi B, Vajna B, Tugyi N, Csitari B, Voros L, Felfoldi T (2017) Differences in planktonic microbial communities associated with three types of macrophyte stands in a shallow lake. FEMS Microbiol Ecol 94:fix164
51. de Melo ML, Bertilsson S, Amaral JHF, Barbosa PM, Forsberg BR, Sarmento H (2019) Flood pulse regulation of bacterioplankton community composition in an Amazonian floodplain lake. Freshw Biol 64:108–120
52. Camara Dos Reis M, Lacativa Bagatini I, de Oliveira VL, Bonnet MP, da Motta MD, Sarmento H (2019) Spatial heterogeneity and hydrological fluctuations drive bacterioplankton community composition in an Amazon floodplain system. PLoS One 14:e0220695
53. Golebiewski M, Calkiewicz J, Creer S, Piwosz K (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
54. Newton RJ, Jones SE, Helmus MR, McMahon KD (2007) Phylogenetic ecology of the freshwater Actinobacteria acI lineage. Appl Environ Microbiol 73:7169–7176
55. Kim S, Kang I, Seo J-H, Cho J-C (2019) Culturing the ubiquitous freshwater actinobacterial acI lineage by supplying a biochemical 'helper' catalase. ISME J 13:2252–2263
56. Ghylin TW, Garcia SL, Moya F, Oyserman BO, Schwientek P, Forest KT, Mutschler J, Dwulit-Smith J, Chan L-K, Martinez-Garcia M (2014) Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. ISME J 8:2503–2516
57. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011) A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol Rev : MMBR 75:14–49
58. Hahn MW, Jezberová J, Koll U, Saueressig-Beck T, Schmidt J (2016) Complete ecological isolation and cryptic diversity in Polynucleobacter bacteria not resolved by 16S rRNA gene sequences. ISME J 10:1642–1655
59. Jezbera J, Jezberová J, Brandt U, Hahn MW (2011) Ubiquity of Polynucleobacter necessarius subspecies asymbioticus results from ecological diversification. Environ Microbiol 13:922–931
60. Salcher MM (2014) Same same but different: ecological niche partitioning of planktonic freshwater prokaryotes. J Limnol 73: 74–87
61. Buck U, Grossart HP, Amann R, Pernthaler J (2009) Substrate incorporation patterns of bacterioplankton populations in stratified and mixed waters of a humic lake. Environ Microbiol 11:1854–1865
62. Kasalický V, Jezbera J, Hahn MW, Šimek K (2013) The diversity of the Limnohabitans genus, an important group of freshwater bacterioplankton, by characterization of 35 isolated strains. PLoS One 8:e58209
63. Hahn MW (2003) Isolation of strains belonging to the cosmopolitan Polynucleobacter necessarius cluster from freshwater habitats located in three climatic zones. Appl Environ Microbiol 69:5248–5254
64. Nelson WC, Stegen JC (2015) The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. Front Microbiol 6:713
65. Lindström ES, Kamst-Van Agterveld MP, Zwart G (2005) Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. Appl Environ Microbiol 71:8201–8206
66. Tripathi BM, Stegen JC, Kim M, Dong K, Adams JM, Lee YK (2018) Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. ISME J 12:1072–1083
67. de Melo ML, Kothawala DN, Bertilsson S, Amaral JH, Forsberg B, Sarmento H (2020) Linking dissolved organic matter composition and bacterioplankton communities in an Amazon floodplain system. Limnol Oceanogr 65:63–76
68. Morana C, Sarmento H, Descy J-P, Gasol JM, Borges AV, Bouillon S, Darchambeau F (2014) Production of dissolved organic matter by phytoplankton and its uptake by heterotrophic prokaryotes in large tropical lakes. Limnol Oceanogr 59:1364–1375
69. Lynch MD, Neufeld JD (2015) Ecology and exploration of the rare biosphere. Nat Rev Microbiol 13:217–229
70. Pedrós-Alió C (2012) The rare bacterial biosphere. Annu Rev Mar Sci 4:449–466