Received: 11 February 2021 | Revised: 10 May 2021 | Accepted: 3 June 2021

DOI: 10.1111/1755-0998.13444

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Opening a next-generation black box: Ecological trends for hundreds of species-like taxa uncovered within a single bacterial >99% 16S rRNA operational taxonomic unit

Martin W. Hahn | Andrea Huemer | Alexandra Pitt | Matthias Hoetzinger

Research Department for Limnology, University of Innsbruck, Mondsee, Austria

Correspondence
Martin W. Hahn, Research Department for Limnology, University of Innsbruck, Mondsee, Austria.
Email: martin.hahn@uibk.ac.at

Present address
Matthias Hoetzinger, Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, SE-75651, Sweden

Funding information
This study was funded in whole by the Austrian Science Fund (FWF) [project 27160-B22]

Abstract

Current knowledge on environmental distribution and taxon richness of free-living bacteria is mainly based on cultivation-independent investigations employing 16S rRNA gene sequencing methods. Yet, 16S rRNA genes are evolutionarily rather conserved, resulting in limited taxonomic and ecological resolutions provided by this marker. The faster evolving protein-encoding gene priB was used to reveal ecological patterns hidden within a single operational taxonomic unit (OTU) defined by >99% 16S rRNA sequence similarity. The studied subcluster PnecC of the genus *Polynucleobacter* represents a ubiquitous group of abundant freshwater bacteria with cosmopolitan distribution, which is very frequently detected by diversity surveys of freshwater systems. Based on genome taxonomy and a large set of genome sequences, a sequence similarity threshold for delineation of species-like taxa could be established. In total, 600 species-like taxa were detected in 99 freshwater habitats scattered across three regions representing a latitudinal range of 3,400 km (42°N to 71°N) and a pH gradient of 4.2 to 8.6. In addition to the unexpectedly high richness, the increased taxonomic resolution revealed structuring of *Polynucleobacter* communities by a couple of macroecological trends, which was previously only demonstrated for phylogenetically much broader groups of bacteria. An unexpected pattern was the almost complete compositional separation of *Polynucleobacter* communities of Ca^{2+}-rich and Ca^{2+}-poor habitats. This compositional pattern strongly resembled the vicariance of plant species on silicate and limestone soils. The new cultivation-independent approach presented opened a window to an incredible, previously unseen diversity, and enables investigations aiming on deeper understanding of how environmental conditions shape bacterial communities and drive evolution of free-living bacteria.

KEYWORDS
amplicon sequencing, bacterioplankton, barcoding, biogeography, Cryptic bacterial diversity, freshwater, latitudinal gradient, Polynucleobacter, priB, 16S rRNA

1 | INTRODUCTION

Prokaryotes are the most numerous organisms on earth. Our current knowledge on diversity of prokaryotes and composition of prokaryotic communities is mainly based on cultivation-independent investigations employing 16S rRNA gene sequencing methods. The insights obtained by such investigations are, however, limited by the taxonomic resolution of the 16S rRNA gene sequences (Stackebrandt & Ebers, 2006). It has repeatedly been shown that prokaryotes affiliated with distinct but related species cannot be discriminated by using their full-length 16S rRNA gene sequences (Hahn et al., 2016; Jaspers & Overmann, 2004; Stackebrandt & Ebers, 2006). In addition to the limited taxonomic resolution of the gene, the majority of the current studies using high-throughput amplicon sequencing of 16S rRNA genes, do not use the entire gene length and thus not the full information content of 16S rRNA gene sequences. Some studies have even been based on <150 bp sequence fragments (García-García et al., 2019) equivalent to only about 10% of the entire gene. Furthermore, the taxonomic ranks of the operational taxonomic units (OTUs) typically established in 16S rRNA-sequence-based investigations are unknown. Despite being frequently stated in publications, and independent of the applied sequence similarity threshold used for OTU demarcation, the established OTUs do not represent species-like taxa. Due to all these limitations, prokaryotic communities remained at the species level black boxes. A few alternative markers and primers have been developed for profiling the composition and structure of bacterial communities with higher resolution than that provided by the 16S rRNA gene (Hill et al., 2002; Ogier et al., 2019; Sánchez et al., 2014). Such approaches have to deal with the trade-off of either relying on strongly degenerated primers ensuring a taxonomically broad match of the primers (Hill et al., 2002; Ogier et al., 2019), or using less degenerated primers, which only match to genes of much more narrow taxonomic groups (Sánchez et al., 2014).

Here, we present a method targeting only a phylogenetically narrow (>99% 16S rRNA sequence similarity, $OTU_{99\%}$) but important group of freshwater bacteria. The method presented here is not intended to be a replacement of standard 16S rRNA amplicon sequencing but to be a supplemental method enabling a species level taxonomic resolution for an important fraction of freshwater bacterioplankton. For the targeted protein-encoding single-copy housekeeping gene (primosomal replication protein N, priB), involved in DNA replication, a sequence similarity threshold could be established, which largely resolves communities of the investigated $OTU_{99\%}$ at the species level. The establishment of this threshold is based on large sets of priB gene sequences and corresponding genome sequences obtained from a large culture collection. These two corresponding data sets enabled the search for a priB gene sequence similarity threshold equal to the 95% genome-wide average nucleotide identity (ANI) threshold used in genome taxonomy and in standard taxonomy of prokaryotes for genome-based species demarcation (Jain et al., 2018; Konstantinidis et al., 2006).

The genus *Polynucleobacter* can be phylogenetically subdivided into four subclusters: PnecA, PnecB, PnecC and PnecD (Hahn, 2003). In contrast to the ecological diversification at the subcluster level (Jezbera et al., 2012; Newton & McLellan, 2015; Nuy et al., 2020), little is known about ecological diversification within these subclusters (Hahn et al., 2016; Jezbera et al., 2011). The method presented here targeted only subcluster PnecC (Hahn, 2003), which consists of many described and undescribed species all sharing 16S rRNA sequence similarity values ≥99%, and thus the whole subcluster PnecC represents a single $OTU_{99\%}$. This subcluster harbours strains of two distinct lifestyles. The evolutionary primary lifestyle is a free-living and planktonic one, while a derived lifestyle is found in strains dwelling as obligate endosymbionts of benthic ciliates (Boscaro et al., 2013; Heckmann & Schmidt, 1987). Strains of the planktonic lifestyle have been frequently detected in diversity surveys, which together cover an ecologically broad range of freshwater systems (Allen & Cavicchioli, 2017; Bahr et al., 1996; Percent et al., 2008; Zwart et al., 2002). They represent abundant freshwater bacteria (Jezbera et al., 2012) ubiquitously appearing in the water column of freshwater habitats worldwide (Comte et al., 2016; Hahn et al., 2015; Jezberova et al., 2010; Peixoto et al., 2011). Conversely to the planktonic strains, endosymbiotic strains have not previously been reported to appear with significant abundance in the water column of freshwater systems.

Comparative genomic analyses suggested that species of the genus *Polynucleobacter* are biologically maintained by intensive intraspecific recombination, which is opposed by interspecific recombination barriers that separate core gene pools of related species (Hoetzinger & Hahn, 2017). The coherence of the intraspecific gene pool of at least one species is even maintained across populations separated by geographic distances of up to 3,000 km, indicative of a high dispersal potential of the species (Hoetzinger et al., 2021). On the other hand, microdiversification of *Polynucleobacter* species is influenced by horizontal acquisition of accessory genomic islands that can be transferred among different species (Hoetzinger et al., 2017).

Here, we present a largely exploratory study designed to reveal the breadth of diversity within the targeted $OTU_{99\%}$. We investigated and compared the PnecC diversity of 99 freshwater habitats representing a broad limnological range and scattered across a latitudinal range of 3,400 km. This resulted in the detection of an astonishingly high total species richness across the investigated habitats. We tested if the revealed diversity is structured by environmental factors and known macroevolutionary trends. Since the structuring of bacterial communities and evolutionary processes takes place at the species level and not primarily at the level of 16S rRNA OTUs, the developed method provides a supplementary tool which partially increases the resolution of diversity surveys. In addition, the method is suitable to reveal drivers structuring the community composition and the selective forces of evolutionary adaptation of a model group of planktonic freshwater bacteria.

2 | MATERIALS AND METHODS

2.1 | Investigated habitats and sampling

In total, 117 water samples from 102 freshwater habitats were obtained (Table S1 and Figure S1); however, only the results of 114 samples from 99 habitats could be analysed (see below). This included 11 habitats sampled 2–3 times. Surface water (0.1–0.5 m depths) samples were taken from the shoreline or from piers, if available. Biomasses were collected by filtration onto 0.2 µm nuclepore membrane filters, preserved by storage in absolute ethanol, and transported in a mobile refrigerator. Water temperature, pH, conductivity and oxygen were measured on location. Water samples were filtered through GF/F filters (Whatman) for determination of concentrations of major ions and measured by ion chromatography (Thermo Scientific DIONEX ICS-1100).

2.2 | Reference priB sequences of cultured strains

In total, a set of 377 priB (primosomal replication protein N) sequences of cultured *Polynucleobacter* strains were established, representing 254 unique sequences (Table S2). Of those priB sequences, 102 originated from genome sequences. The remaining 275 sequences were obtained by Sanger sequencing of PCR products. The sequenced amplicons were generated by using primers priBausF 5′-CGTCARATGGCTTACATGATC-3′ and priBausR 5′-CAATAACGYTTACGCTTGAAC-3′. The sequenced fragments obtained with this primer pair included the binding sites of the priBinnFd/priBinnRd primer pair (see below), which enabled checking of the binding sites in strains without an available genome sequence.

2.3 | Amplicon sequencing and processing of reads

Genomic DNA was extracted from environmental samples as previously described (Jezbera et al., 2011), and purified using the Wizard DNA clean-up kit (Promega). The priB gene of *Polynucleobacter* bacteria (subcluster PnecC) was amplified by using primers priBinnFd 5′-YGGCGTTGAATCATTTMAC-3′ and priBinnRd 5′-TTCCAAACGCCATGRTGATT-3′ (annealing temperature 62°C, 30 cycles, Q5 polymerase [New England Biolabs]). The primers were tagged with Illumina adaptors and sample-specific tags (fusion primers). Amplicons of 117 environmental samples, one technical replicate and two controls consisting of amplicons from one and four cultured strains, respectively, were paired-end sequenced (300 bp) by Illumina MiSeq. Reads were processed using QIIME2 (Bolyen et al., 2019). This included demultiplexing, trimming of adapter and primer sequences, quality filtering, joining of paired reads, exclusion of too long (>288 bp) and too short (<285 bp) sequences (all reference sequences of PnecC strains were in the length range of 285–288 bp), removal of reads with copy number <10 present only in single samples, and rarefying to 25,230 reads per sample. Due to too small read numbers, three environmental samples had to be excluded, therefore, only the results from 114 water samples obtained from 99 habitats are presented (Supporting Information S1). Reads were clustered into OTUs by employing a 98% sequence similarity threshold (see results). OTUs were taxonomically classified by employing a reference set of priB sequences obtained from *Polynucleobacter* strains (Table S2). The two groups of $OTU_{98\%}$ sharing ≥98% or <98% sequence similarity with reference taxa are termed reference operational taxonomic units (refOTUs) and environmental OTUs (eOTUs), respectively.

2.4 | Data analyses

The OTU table exported from Qiime2 and the environmental data were analysed using R version 3.6.1 (R Core Team, 2019). The vegan package (Oksanen et al., 2019) was used for most of the analyses performed. Geographic distances between sampled habitats were determined by calculation of great-circle-distance using the haversine method, which assumes a spherical earth, from the R package "geosphere" (Hijmans, 2019). Site-specific climate data were obtained from the WorldClim data set (Fick & Hijmans, 2017) using the DIVA-GIS software (Hijmans et al., 2001). Furthermore, the R packages ggplot2 (Wickham, 2016) and maps (Minka & Deckmyn, 2018) were used.

Bray-Curtis dissimilarities were calculated without prior transformation of the OTU table. Occupancy of OTU was defined and calculated as follows. Because *Polynucleobacter* species are basically unable to exist across the whole studied environmental pH range of more than four units, occupancy of each OTU was assessed within a specific pH range of two units, that is, only samples within a pH range of two units around (+1 and −1) the relative-abundance-weighted average pH of samples with detection of the respective OTU were considered. The weighted average pH indicates the pH optimum of the respective OTU. For instance, the relative-abundance-weighted average pH of detections of *P. paneuropaeus* was 5.9, therefore occupancy refers to the samples of the pH range 4.9–6.9. Forty-one of the investigated samples belonged to this pH range and *P. paneuropaeus* was detected in 31 (with >25 reads, i.e., >0.1% of reads per sample), and thus showed a pH-specific occupancy of 73%.

3 | RESULTS

3.1 | Development of PCR primers targeting a protein-encoding gene

We aimed to develop a primer pair suitable for specific amplification of a protein-encoding gene present in all *Polynucleobacter* bacteria. The strategy employed for primer development and the faced limitations and results are described in Supporting Information S1. In brief, a primer pair for amplification of the primosomal replication protein N (priB) gene of *Polynucleobacter* bacteria affiliated with

subcluster PnecC was developed. Detailed analyses suggested a sequence similarity threshold of 98% for discrimination of species-like OTUs (Figure 1; Supporting Information S1). Discrimination based on this threshold agreed with average nucleotide identity (ANI) based species discrimination (95% identity threshold) for 99.2% of the pairwise comparisons among the 102 strains with available genome sequences (Figure 1). Species that could not be discriminated by priB similarities of <98% were combined together to species complexes.

3.2 | Amplicon sequencing of environmental samples

In total, results from 99 freshwater habitats (114 samples) including small ponds, lakes, streams and rivers (Table S1) located in three regions (Lapland, Central Europe, Corsica; Figure S1) along a European South-North cross-section (42°N to 71°N) were obtained. The selection of habitats aimed for maximizing the covered habitat diversity in order to maximize the insights into diversity of *Polynucleobacter* taxa. Details on the results of the amplicon sequencing are given in Supporting Information S1. Rarefaction analyses suggested that sequencing depth after rarefication (25,230 reads) was large enough to completely cover the amplicon sequence variant (ASV) numbers in the respective samples (Figure S2A). In total, 600 $OTUs_{98\%}$ were detected in the samples. Rarefaction analyses of the $OTU_{98\%}$ data suggested that the number of investigated samples was not high enough to completely cover the total $OTU_{98\%}$ richness in the investigated area (Figure S2B). The established $OTU_{98\%}$ were taxonomically classified by using reference sequences and a threshold of 98% sequence similarity (Figure 1). The reference sequences represented 108 species-like taxa, seven species complexes, and two taxa representing the same species but separated by priB sequence similarities <98% (Supporting Information S1 and Table S2). Environmental $OTU_{98\%}$ classified in one of the reference taxa are called refOTU, while all remaining $OTU_{98\%}$ sharing <98% similarity with reference sequences are called eOTU. Only 13% of the 600 detected OTUs were represented by refOTUs, that is, by cultivated reference strains (Figure 2), but this minor fraction of the total number of detected OTUs recruited 59% of the total amount of priB reads. The rank-abundance plot (Figure 2) sorting the detected taxa according to their relative read abundance shows that only a few OTUs recruited most of the obtained reads. The top ranked OTU (*P. paneuropaeus*) recruited 11.3% of all reads. The top-seven-ranked OTUs (including two species complexes) recruited together almost half of all reads, while the vast majority of the detected OTUs represented rare taxa. The percentage of eOTUs, that is, environmental OTUs sharing <98% sequence similarity with all reference strains, increased with decreasing read numbers recruited by the respective OTUs (Figure 2). While only 30% of the top-ten-ranked OTUs were eOTUs, the first quarter of the ranking contained 70% and the last quarter 95% eOTUs.

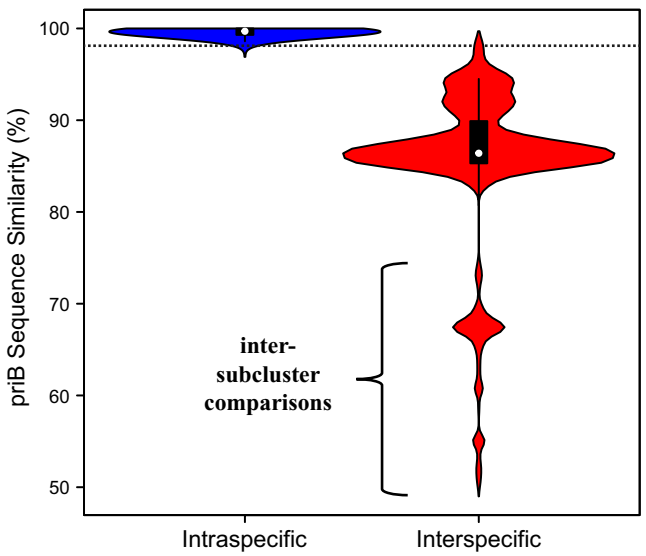


FIGURE 1 Violin plot showing frequencies of priB sequence similarity values for intra- and interspecific (95% ANI threshold) pairwise comparisons of genome-sequenced strains. The dotted horizontal line indicates 98% priB sequence similarity. Sequence similarities of priB genes below 80% represent comparisons of strains affiliated with different *Polynucleobacter* subclusters (PnecA, PnecB, PnecC and PnecD)

3.3 | Structuring of *Polynucleobacter* communities by environmental factors

The investigated samples and habitats represent broad ranges of environmental parameters (Table S1), for instance, including a pH range of 4.2 to 8.6. The composition of the PnecC communities in the investigated samples was strongly changed along the environmental gradients defined by the respective habitats. A constrained gradient analysis by canonical correspondence analysis (CCA) suggested a discontinuous variation in composition mainly corresponding to the concentration of dissolved Ca^{2+} ions (Figure 3a). Further analyses of community compositions along the Ca^{2+} gradient of the investigated samples suggested a sharp breakpoint in community composition at calcium concentrations of about 12 mg Ca^{2+} L^{-1} (Figure 4b). An ANOSIM analysis confirmed significant differences in the composition of communities in samples below and above this concentration (9,999 permutations, R = 0.6499, p = .0001). Communities from low Ca^{2+} habitats were rather diverse and showed continuous changes along environmental gradients (mainly pH), while communities from high Ca^{2+} environments appeared in the NMDS ordination as a much smaller cluster (Figure 3b and Figure S3). A variation partitioning analysis indicated that a set of 15 explanatory variables including physicochemical, geographic and climatic parameters as well as habitat characteristics (Figure S4) explained together about 25% of variance in community composition across the 114 samples (Figure 5c). Not much less than half of the explained variance in community composition was explained by the set of physicochemical parameters (Env, 9.7%) alone (Figure S4), while the habitat characteristics (habitat type and size) explained the variance of composition only marginally. By contrast,

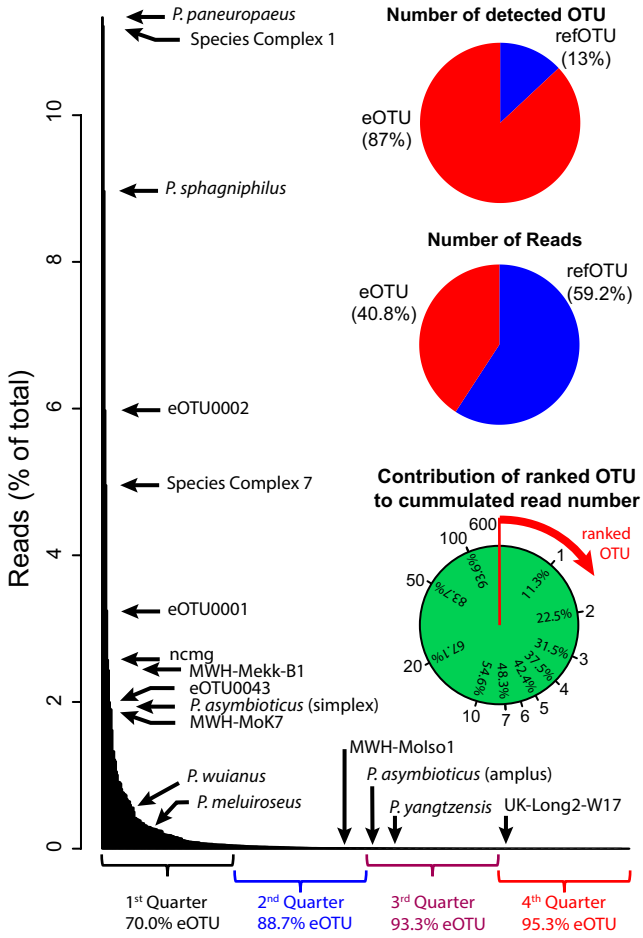


FIGURE 2 Frequencies of the detected $OTUs_{98\%}$. The bar plot shows the rank-abundance distribution of the 600 operational taxonomic units (OTUs) detected in the 99 investigated habitats. Detections of repeatedly sampled habitats were downweighted in order to give detections from all habitats the same weight. Individual rank-abundance curves of each investigated sample are shown in Figure S2C. The pie charts depict shares of refOTUs (representing cultured strains) and eOTUs (sharing <98% sequence similarity with priB sequences of cultured strains). Top pie chart, shares of refOTUs and eOTUs of the total number of detected $OTUs_{98\%}$. Middle, share of reads assigned to refOTUs and eOTUs. Bottom, cumulative contribution of detected OTUs sorted by increasing rank to the total number of reads. For instance, the top-ranked $OTU_{98\%}$ (*P. paneuropaeus*) recruited 11.3% of the total number of reads (habitats weighted equally) and the top-seven ranked $OTUs_{98\%}$ recruited in total 48.3% of reads

variance in OTU richness of samples was much better explained by environmental conditions. About 80% of variation of richness could be explained by the set of environmental variables (Figure 5c). The largest explained fraction of variance was explained by habitat characteristics (habitat size and type, 31%). The second largest fraction was explained by physicochemical variables (including pH) together with habitat characteristics (21% of variability in OTU richness). The rest of the variance was mainly explained by various combinations of parameters (Figure S4).

Communities of the most acidic samples tended to be dominated by only a single OTU, that is, *P. sphagniphilus*. In the four samples with lowest pH (<4.6) this species represented 99.7%–99.8% of reads (Figure 6b). When excluding rare OTUs defined by relative abundances in the samples of <1% (i.e., <1% of reads), these four samples showed an OTU richness of 1.0. By contrast, samples from streams and rivers were characterized by 66–89 OTUs when rare OTUs were excluded. In general, both the OTU richness (Figure 6b) and the size of habitats (Figure S5) tended to increase with pH values of the habitats. Interestingly, similar z values describing the relationship of habitat size and species (OTU) richness were observed for *Polynucleobacter* OTU (Figure S5) and whole bacterioplankton communities of freshwater lakes characterized by using 16S rRNA sequence data (Reche et al., 2005).

3.4 | Occupancy of particular OTUs along the pH gradient

Only eight (including two species complexes) of the 600 detected OTUs represented common taxa with pH-specific occupancy >50%. These 1.5% of all detected OTUs frequently occurred with high average relative abundances and tended to appear with more even relative abundance across the occupied habitats. In contrast to these common taxa, 15.5% of the detected OTUs showed occupancies between 10% and 50% and appeared by average with maximum relative abundances of 15.9% of the reads (range 0.3% to 87%). Interestingly, locally abundant taxa showed occupancy values <10% but rather high relative abundances in a few habitats or samples. Examples for locally abundant taxa are *P. wuianus* and *P. meluiroseus* (Figure S1). The former species was discovered in October 2002 in a small slightly acidic pond designated Pond-1 (Hahn et al., 2005). The priB amplicon sequencing included three samples of that pond, which were taken in August and October 2009, and in June 2010. *P. wuianus* was detected in all three samples of its type locality with relative abundances ranging from 15% to 86% of the reads. Across the other 111 investigated samples, this species was only detected in six samples from two ponds located 200 m and 40 km away from Pond-1. Both ponds were sampled three times but *P. wuianus* was detected only with low relative abundances of 0.004% to 0.135% of the reads. The occupancy of this species was only 3.7% and the average relative abundance was despite the local abundance peak in Pond-1 only 0.56%, which was twenty-times smaller than the relative abundance of the common species *P. paneuropaeus*. Similarly, *P. meluiroseus* showed an occupancy of 6.1% with detection in nine samples representing seven habitats but this species only appeared in three samples with relative abundances of >1%. Interestingly, the two highest values of 41% and 8% relative abundance were observed for the lake from which the type strain of the species was isolated (Pitt et al., 2018). In contrast to *P. wuianus*, the detections of *P. meluiroseus* were geographically broader scattered (Figure S1). Obviously, both species are characterized by high local relative abundances, low pH-specific occupancy and high local persistence.

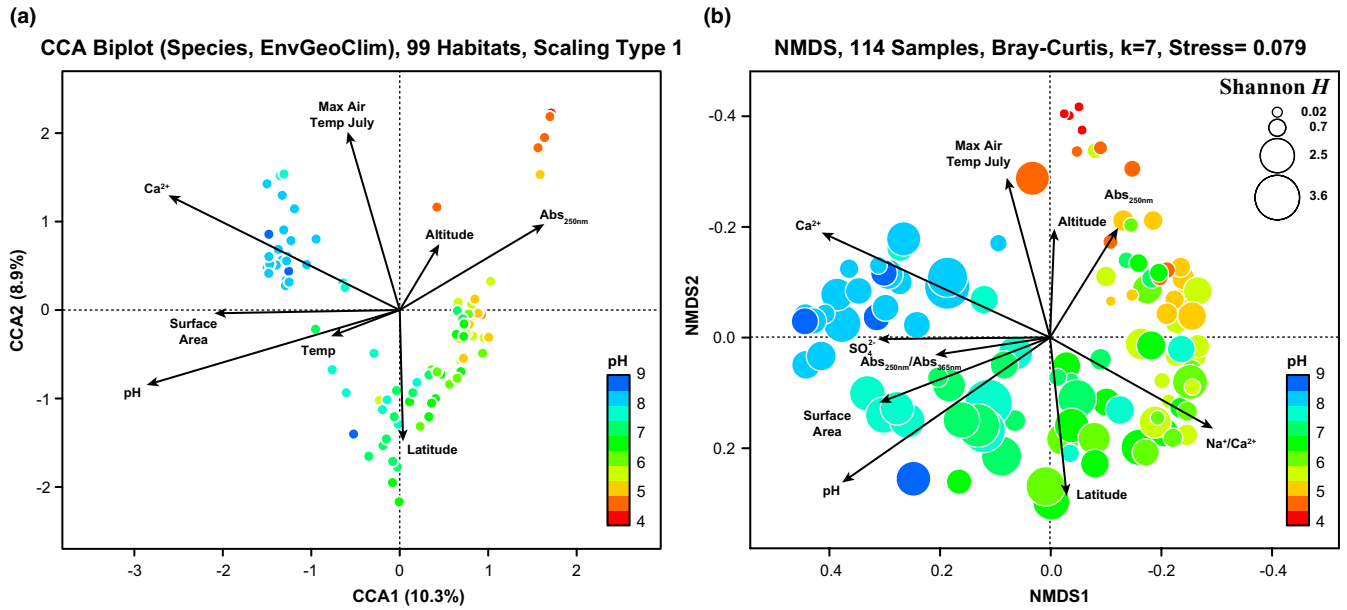


FIGURE 3 (a) Direct gradient analysis by canonical correspondence analysis (CCA) of Bray-Curtis community dissimilarity values and environmental variables (permutation test for the whole CCA model, $p = .001$). In this constrained ordination, dots represent the 99 habitats colour-coded by the pH of the respective sample. (b) Indirect gradient analysis by nonmetric multidimensional scaling (NMDS) exclusively based on Bray-Curtis dissimilarities of the 114 environmental samples. Each bubble represents a sample colour-coded by pH. The diameter of the bubbles is nonlinearly scaled by the Shannon index H' of the respective sample. Both ordinations show environmental variables significantly ($p < .05$) correlated with the ordination models

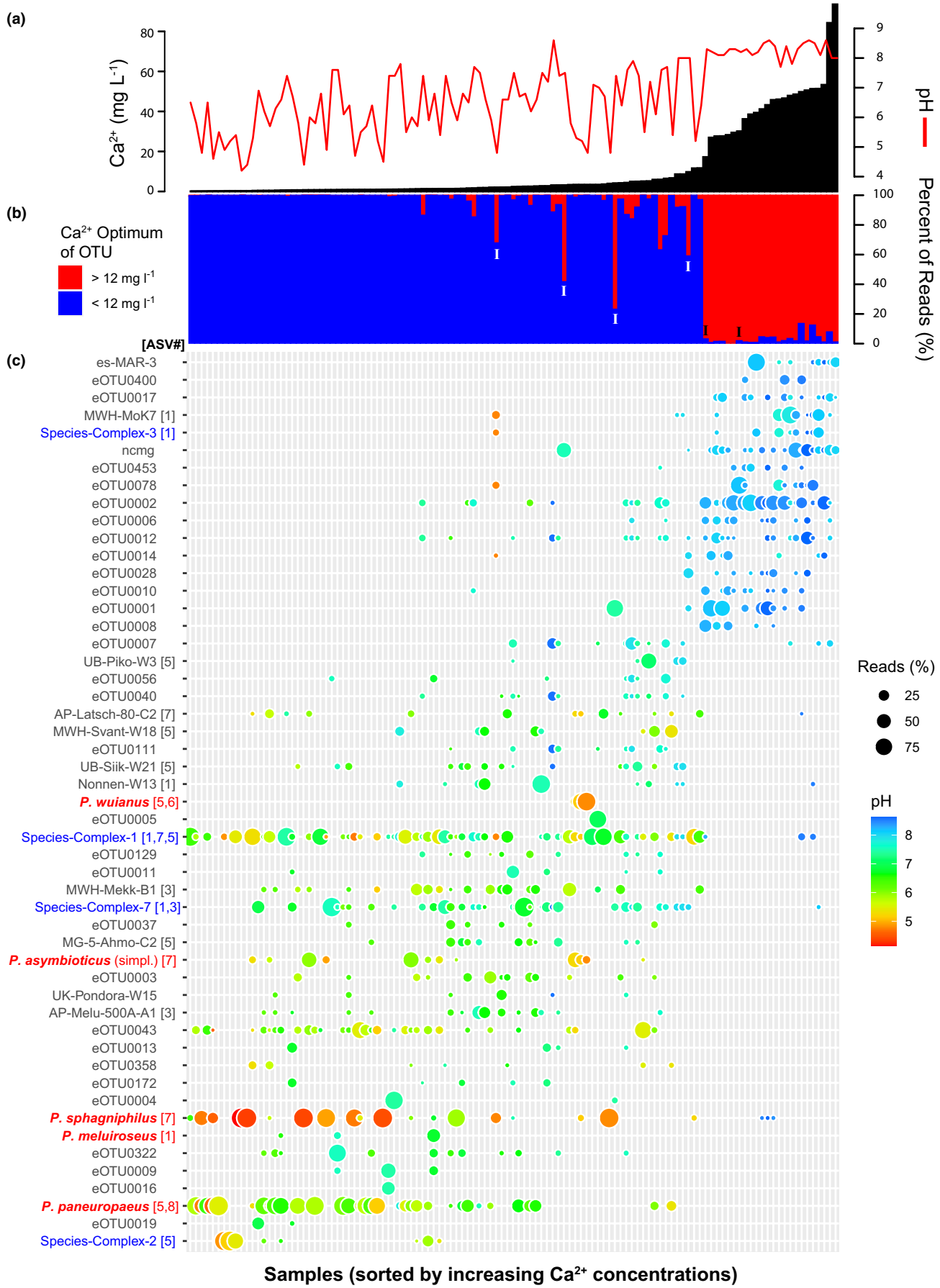
The pH-specific occupancy of OTUs tended to decrease with increasing pH prefered by the respective taxa (Figure S6A). This trend is linked to a trend of increasing community dissimilarities among communities of the same pH class with increasing pH (Figure 5a). This means that differences in composition among communities present in habitats with similar pH are increasing with pH. This is also obvious in the NMDS ordination where the communities from habitats with similar pH are spread over larger ordination space if pH values of their habitats are higher (Figure 3b). Interestingly, in the case of high-Ca^{2+} communities (mainly represented by the pH class 8–9), the link between occupancy of OTUs and community dissimilarities seems to be less strict than in the low Ca^{2+} communities (Figure 5a and Fig. S6A).

3.5 | Biogeography of *Polynucleobacter* communities

Of the set of 123 reference taxa (Table S2) used for taxonomic classification of the detected $OTU_{98\%}$, only 64.2% were detected in the 114 investigated samples; however, the detections differed in relation to the latitudinal origin of the strains representing reference taxa (Figure 7). We compared the latitudinal origin of the strains representing reference taxa (62°S–78°N) with the range of geographic origin of the 114 investigated samples (42°N–71°N). While 80% of the reference taxa represented by strains obtained from sites located at latitudes >40°N were detected, only 19% (0%–30% in particular latitude classes) of the reference taxa represented by strains obtained from latitudes <40°N were detected (Figure 7a). Importantly, all detected reference taxa with strains obtained from latitudes <40°N recruited only very small numbers of reads.

Despite a latitudinal range spanning almost 30° and a maximum distance between habitats of about 3,400 km (Figure S1), a Mantel test did not suggest that differences in community composition increased with geographic distance between the sampled habitats (Mantel $R = -0.01$, $p = .66$; Table S3). Even when controlling for environmental influences including differences in pH or Ca^{2+} concentration (partial Mantel tests), no significant correlations between community composition and geographic distance were observed. Different results were obtained when only *Polynucleobacter* communities from low Ca^{2+} conditions were

FIGURE 4 Operational taxonomic unit (OTU) and community distribution along the Ca^{2+} gradient of the 114 investigated samples. All three plots show the samples sorted by increasing Ca^{2+} concentrations. (a) Ca^{2+} concentrations (black bars) and pH (red line). (b) Community compositions regarding Ca^{2+} preferences of the OTUs (< or >12 mg Ca^{2+} L^{-1}) constituting the particular communities. Communities labelled by an "I" represent communities showing an intermediate position in the NMDS ordination (Figure 3) regarding the Ca^{2+} vector (Figure S3F). (c) Detection of the most abundant high and low Ca^{2+} OTUs. The colour code applied to the taxon names indicates described species (red), species complexes (blue), and other species-like OTUs (black). The numbers of the eight 16S rRNA ASVs (compare Figure 8) are given after the OTU names in squared brackets if known. The colours and the diameters of the bubbles indicate the pH of the samples and the relative read abundances in the respective sample, respectively


(a)
Ca2+ (mg L-1)
80
60
40
20
0
9
8
7
6
5
4
pH
(b)
Ca2+ Optimum of OTU
> 12 mg l-1
< 12 mg l-1
100
80
60
40
20
0
Percent of Reads (%)
[ASV#]
(c)
es-MAR-3
eOTU0400
eOTU0017
MWH-MoK7 [1]
Species-Complex-3 [1]
ncmg
eOTU0453
eOTU0078
eOTU0002
eOTU0006
eOTU0012
eOTU0014
eOTU0028
eOTU0010
eOTU0001
eOTU0008
eOTU0007
UB-Piko-W3 [5]
eOTU0056
eOTU0040
AP-Latsch-80-C2 [7]
MWH-Svant-W18 [5]
eOTU0111
UB-Siik-W21 [5]
Nonnen-W13 [1]
P. wuianus [5,6]
eOTU0005
Species-Complex-1 [1,7,5]
eOTU0129
eOTU0011
MWH-Mekk-B1 [3]
Species-Complex-7 [1,3]
eOTU0037
MG-5-Ahmo-C2 [5]
P. asymbioticus (simpl.) [7]
eOTU0003
UK-Pondora-W15
AP-Melu-500A-A1 [3]
eOTU0043
eOTU0013
eOTU0358
eOTU0172
eOTU0004
P. sphagniphilus [7]
P. meluiroseus [1]
eOTU0322
eOTU0009
eOTU0016
P. paneuropaeus [5,8]
eOTU0019
Species-Complex-2 [5]
Reads (%)
25
50
75
pH
8
7
6
5
Samples (sorted by increasing Ca2+ concentrations)

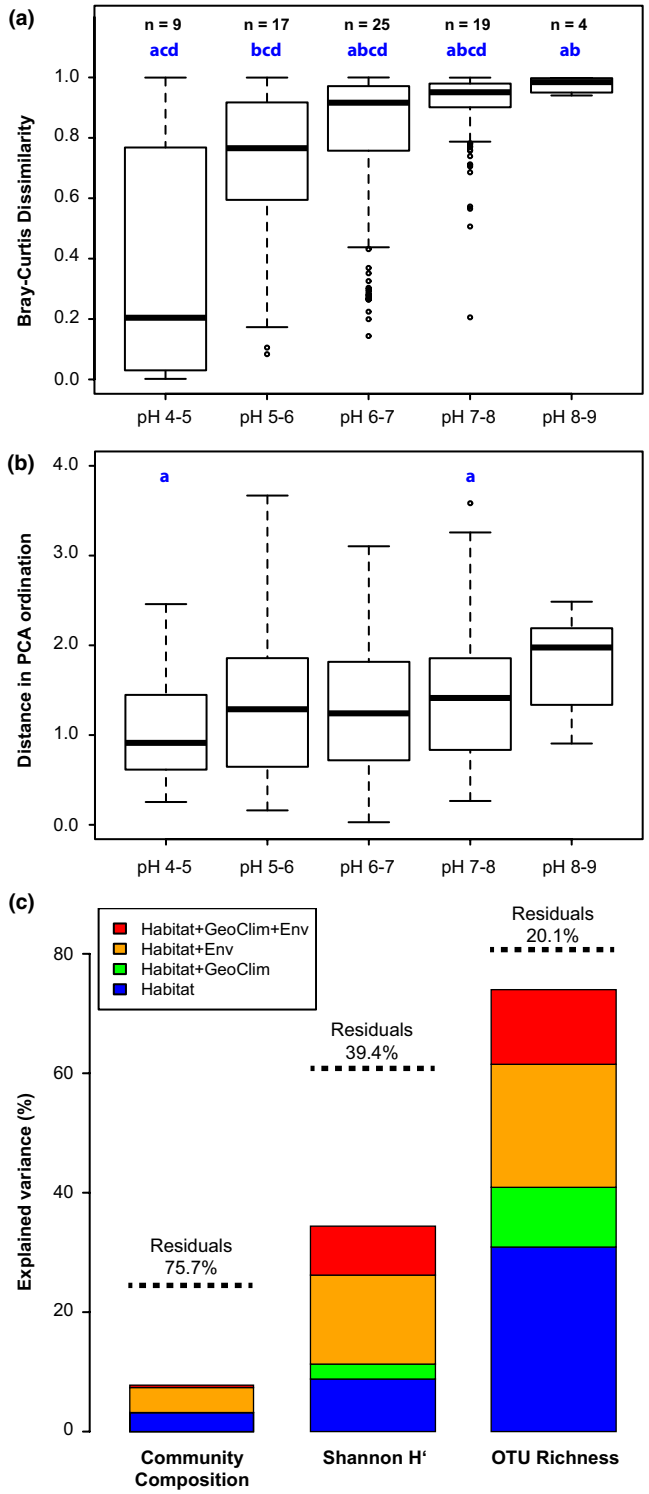


FIGURE 5 (a) Boxplots of pairwise community dissimilarities (Bray-Curtis) within pH classes. Only samples of standing waters with low Ca^{2+} concentrations (<12 mg/L) were considered. The numbers (n) of samples per pH class are given above the bars. Classes with significantly ($p < .025$) different data (Kruskal-Wallis test with Dunn's post hoc test with Holm correction for multiple comparisons) are labelled with the same blue letter. (b) Boxplots of pairwise comparisons of environmental distance among samples within pH classes. The environmental distance was calculated as Euclidian distance between coordinates of a principal component analysis (PCA) ordination of 20 variables. The same set of samples as in the above analysis of Bray-Curtis dissimilarities was used. A test for significant differences between groups was performed as above. (c) Results of variation partitioning analyses on community composition (Bray-Curtis dissimilarity), Shannon H' and operational taxonomic unit (OTU) richness (OTUs >1% of priB reads). Three sets of explanatory variables were used. Env, eight environmental variables; GeoClim, five geographic and climatic variables; Habitat, two variables characterizing habitat properties (surface area and type of habitat, i.e., running or standing waters). The stacked bars show only partitions of explained variance including habitat properties as explanatory variables. The total explained variance is indicated by dotted lines

considered. Weak but significant correlations (Mantel $R < 0.13$, $p < .01$) were observed with geographic distance, even when controlling for distances in pH, Ca^{2+} and other chemical parameters. However, when controlling for a broader set of environmental variables (including habitat type and climatic variables), no significant correlation between community composition and geographic distance was observed (Mantel $R = 0.0514$ $p = .0602$). By contrast, Mantel tests and partial Mantel tests on correlations between community composition and environmental factors yielded in all cases significant correlations. The highest correlation coefficient (Mantel $R = 0.53$, $p = .0001$) was observed for community composition and pH distance between habitats.

Despite lacking indications for an isolation by distance pattern for the investigated *Polynucleobacter* communities, geographic structuring was evident. More than 50% of the detected OTUs were exclusively detected in only one of the three sampled regions (Figure 7b). Such OTUs only detected in single regions tended to be characterized by low average relative abundances and were detected in only a few samples (Figure 7c).

3.6 | Predictive power of 16S rRNA sequence ASVs of *Polynucleobacter* bacteria

We evaluated if 16S rRNA based ASVs of *Polynucleobacter* bacteria possess a predictive power regarding environmental preferences of ASVs. A set of 226 strains (Table S2 plus strains and genomes recently published by Hoetzinger et al., 2021) affiliated with subcluster PnecC was represented by only eight V3–V4 region ASVs, although these strains represented 80 different species according to genome similarities (95% ANI threshold) (Figure 8). Six ASVs represented more than one species, respectively, and three of these ASVs consisted of strains with distinct Ca^{2+} preferences. Thus, bacteria with markedly different environmental preferences (Figure 4) were considered together within single 16S rRNA ASVs.

4 | DISCUSSION

The limited taxonomic and ecological resolution of the 16S rRNA marker is well known (Hahn et al., 2016; Jaspers & Overmann, 2004;

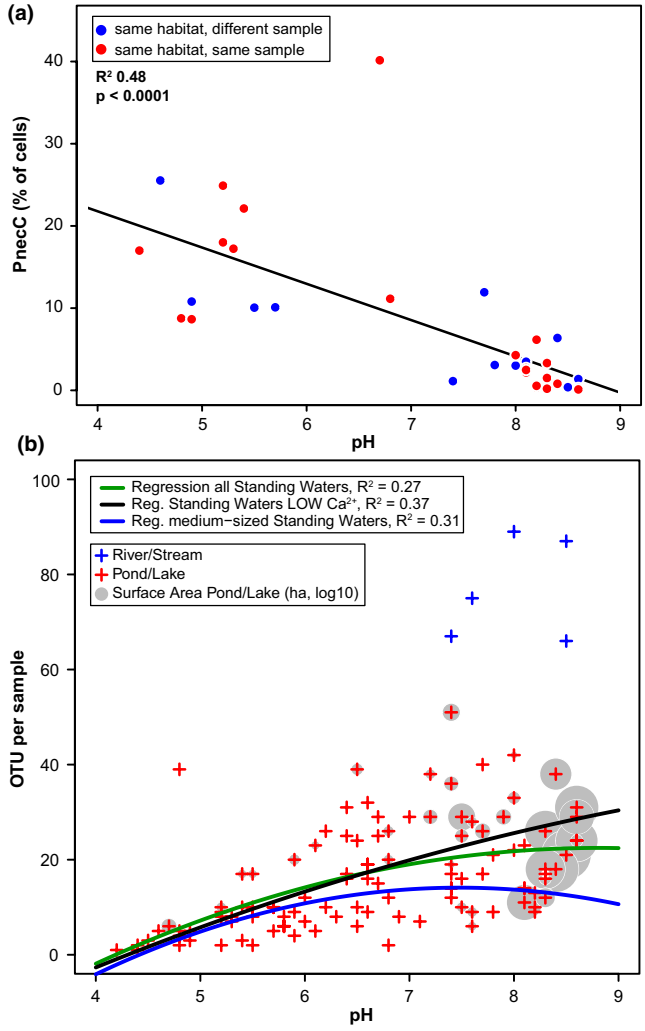


FIGURE 6 (a) Relative abundance of PnecC bacteria determined by FISH (data from Jezbera et al., 2012). Some of the shown data represents samples included in the priB amplicon sequencing (red dots), other samples were obtained from habitats included in the priB sequencing but taken at other dates (blue dots). (b) Polynomial regressions on relationship between pH and operational taxonomic unit (OTU) richness. OTU numbers represent only detections >1% of reads per sample. Samples from running waters were completely excluded from the analyses. Regressions were performed on all remaining samples, all remaining samples with low Ca^{2+} communities, and remaining samples from habitats with medium-sized surface area (0.018–0.64 ha). The surface size of all standing water habitats is indicated by grey bubbles (log transformed data)

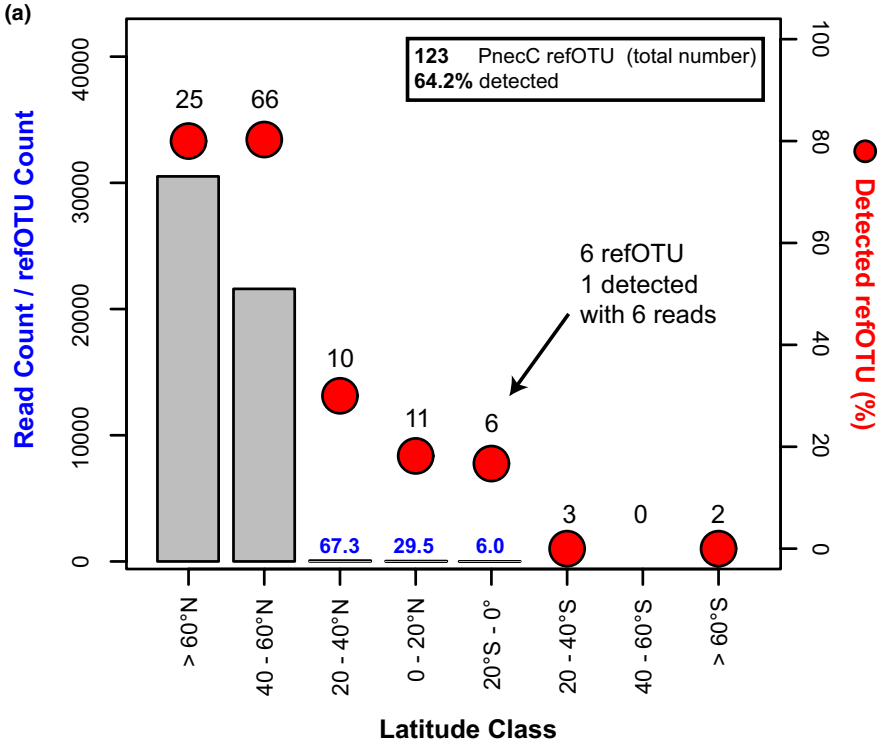
Stackebrandt & Ebers, 2006). An alternative "universal" marker for diversity investigations on bacterial communities is available (Hill et al., 2002) but requires the use of highly degenerated primers strongly substituted with inosine bases. This is potentially biasing comparative compositional analyses of bacterial communities. High degeneration of primers can be avoided if the taxonomic focus of diversity studies is narrowed to genus-like taxa (Pereira et al., 2018; Sánchez et al., 2014).

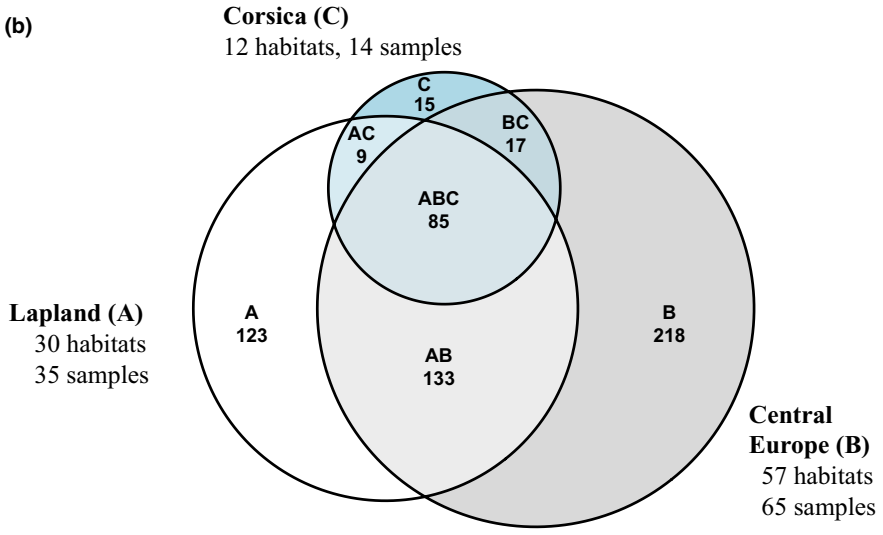
We developed a new marker and investigated the structure of *Polynucleobacter* communities along environmental gradients characterized, amongst other parameters, by a pH range of 4.2–8.6 and Ca^{2+} concentrations of 0.1–94 mg/L. To maximize the studied environmental gradients, a broad variety of freshwater systems were investigated, which ranged from very small, shallow ponds to large lakes, rivers and streams. The sampled habitats were located in different climate zones and at altitudes ranging from about sea level to more than 2,000 m. The applied method for determination of the community composition provided a resolution largely at the species level, but only covered the multispecies subcluster PnecC of the genus *Polynucleobacter*. Remarkably, in typical diversity studies based on 16S rRNA amplicon sequences, the targeted *Polynucleobacter* diversity is represented by only a single $OTU_{99\%}$ harbouring a rather small number of ASVs (Figure 8). In 16S rRNA based studies comparing, for instance, acidic and alkaline habitats or systems with low and high Ca^{2+} concentrations, the $OTU_{99\%}$ representing subcluster PnecC of the genus *Polynucleobacter* is harbouring different organisms across the investigated samples. The same $OTU_{99\%}$ found, for instance, in samples from acidic and alkaline habitats is comprised of species differing pronouncedly in their respective ecological traits. This heterogenous composition of OTUs potentially results in masking of ecological trends and patterns, and may also blur dispersal and community assembly processes. It is known that the studied group of *Polynucleobacter* bacteria is not exceptional among free-living bacteria regarding the limited taxonomic and ecological resolution of their 16S rRNA genes (Chevrette et al., 2019; Jaspers & Overmann, 2004; Rodriguez-R et al., 2018).

The priB gene could also be used for biodiversity survey or experimental studies on other genus-like bacterial taxa. However, this would need sufficient knowledge on sequence diversity at the potential primer binding sites (Supporting Information S1) and the quality of the taxonomic resolution at the species level would largely depend on a suitable collection of reference genomes enabling a sound search for a species discrimination threshold (Figure 1).

4.1 | Enormous but still incompletely covered OTU richness

Our priB-based investigation of 99 freshwater habitats revealed an astonishing total number of 600 species-like *Polynucleobacter* OTUs. We cannot be sure that the performed read processing and filtering removed all erroneous sequences; however, the strict sequence length filtering and the removal of sequences containing additional stop codons should have helped to exclude many sequences representing PCR artefacts. Especially the search for additional stop codons increased the confidence in the established sequence data, because their appearance in the single-copy, essential house-keeping gene priB clearly indicates erroneous sequences. Importantly, rather few such sequences were found and all were present in very low copy numbers (Supporting Information S1). We did not perform chimera filtering due to the lack of a suitable priB reference database. However, we used a phylogenetic tree calculated with all reference and eOTU priB sequences to search for eOTUs displaying unusually





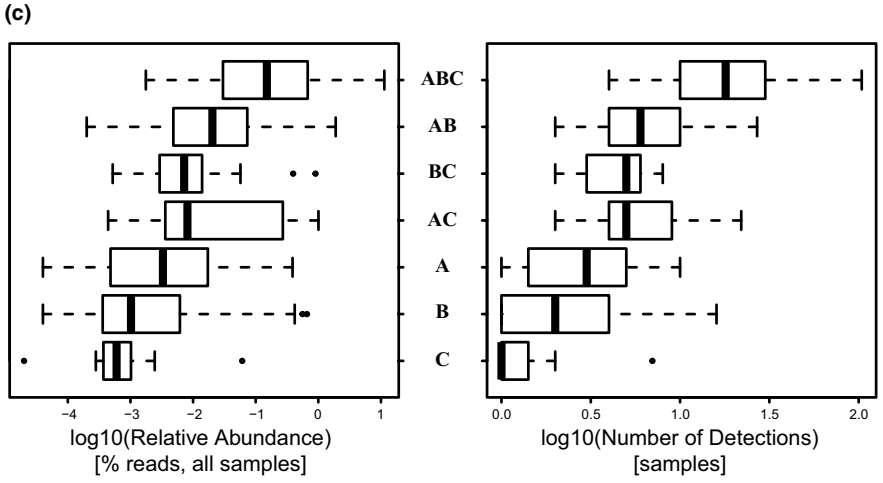


FIGURE 7 Biogeography of refOTUs and eOTUs. (a) Detection of the 123 reference taxa (total number, including undetected) represented by cultured strains obtained from locations of various latitudes. The bars indicate the average number of reads per refOTU assigned to reference taxa grouped according to their respective origin in latitude classes. The numbers above the red dots show the absolute number of reference taxa in a particular latitude class, and the dots indicate the fraction of these reference taxa that was detected in the priB amplicon data set. For instance, the latitude class 0°N–20°N harbours in total 11 reference taxa of which only 18.2% were detected (2 taxa), with average read numbers of 29.5 reads per refOTU. Note that the latitude range of the habitats investigated by priB amplicon sequencing was 42°N–71°N. (b) VENN diagram depicting the number of detected OTUs (eOTUs and refOTUs) shared or not shared between the three investigated geographic regions. (c) (Left graph) Boxplot of total relative abundance (in all samples) of detected OTUs grouped in geographic classes according to the VENN diagram shown in (b), and (right graph) boxplot of the number of samples with detections for the same geographic OTU classes. Note the log transformation of the plotted data in both boxplots

long branch length. Long branch lengths are expected if two sequences with low similarity contribute larger fractions to a chimeric sequence; however, suspicious long branches were not observed for any eOTU. The increasing percentage of eOTUs towards the rare species end of the rank-abundance distribution (Figure 2) could indicate erroneous sequences; however, an alternative explanation for the decreasing percentage of refOTUs could be that rare species tend to be underrepresented in collections of cultured strains.

Even if we assume that 10%, 20%, or even 30% of the detected 600 OTUs were based on erroneous sequences, an impressive number of detected OTUs would remain. In addition, rarefaction analyses suggested that not even all of the abundant taxa (>10% of priB reads of a particular sample) could be detected by the variety of samples we investigated (Figure S2B). This indicated that further sampling would increase the detected number of both abundant and rare OTUs. This is not surprising given the observation of taxa with overall low occupancy but locally abundant populations such as *P. wuianus* and *P. meluiroseus*. In addition, our study could certainly not cover the entire diversity of freshwater systems in the investigated regions. Running water systems, for instance, were only marginally covered and anoxic hypolimnia of lakes known to be rich in *Polynucleobacter* bacteria (Diao et al., 2017; Jezbera et al., 2011) were completely omitted. The indicated incomplete coverage of OTUs present in the investigated area is further supported by the lack of detection of 20% of the reference taxa originating from this area (Figure 7a). Consequently, ecologically broader sampling and inclusion of seasonal aspects are both expected to increase the total number of OTUs detected in the investigated regions.

4.2 | Biogeography of taxa mainly reflected regional differences in ecological conditions

Environmental filtering resulted in a strong biogeographic structuring. For instance, OTUs abundant in limestone areas in Central Europe were not detected in other sampled regions, which lack habitats with high Ca^{2+} concentrations (Figure S1), and low pH preferring OTUs were almost absent from the investigated habitats in Scandinavia with mainly circum-neutral pH. On the other hand, hints on biogeographic structuring caused by an isolation by distance mechanism were scarce. Partial Mantel tests controlling for environmental influences did not suggest that dissimilarity of *Polynucleobacter* communities increased with geographic distance (Table S3). This was in line with a recent study on population structure of *P. paneuropaeus* along the same South-North range studied here, which suggested a lack of dispersal barriers along this 3,400 km latitudinal range

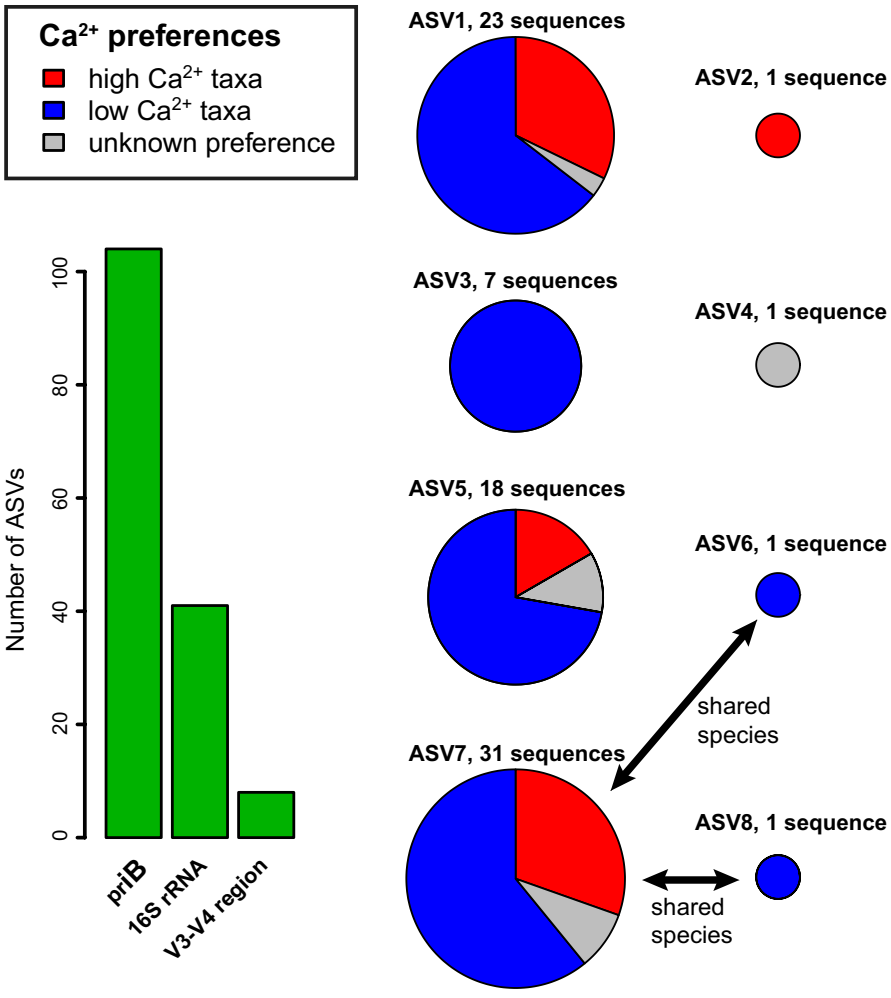


FIGURE 8 Bar chart, theoretical number of ASVs represented by different marker sequences of a set of 226 *Polynucleobacter* strains all affiliated with subcluster PnecC (excluding endosymbionts) representing 81 different *Polynucleobacter* species (>95% ANI). Pie charts, Ca^{2+}concentration preferences of priB OTUs represented by the eight V3–V4 region ASVs (16S rRNA gene). Each species is only represented by a single sequence if no intraspecific sequence polymorphism was present. The Ca^{2+} preferences of the priB OTUs reflect the Ca^{2+} optima determined through analyses of the investigated environmental samples (Figure 4). ASVs with unknown preferences were either not detected or were detected with too low read numbers

(Hoetzinger et al., 2021). The detection of OTUs exclusively found in only one of the three investigated regions (Figure 7b) is well explained by the abundance-occupancy relationship documented in macroecology (Gaston et al., 2000), which predicts a positive relationship between the abundance of a taxon and its range occupancy (Figure 7c). However, in the case of *Polynucleobacter* bacteria, it is not known if this relationship simply resulted from undersampling of rare OTUs, or if it really reflects restricted biogeographic distributions. Nevertheless, the former explanation seems to be more likely. The rather small fraction of reference taxa originating from lower latitudes and the southern hemisphere detected in the investigated habitats, as well as the very low relative abundance of the detected southern taxa (Figure 7a) confirmed a previously revealed biogeographic pattern (Hahn et al., 2015). Currently, it is still unknown if these patterns result from a distance mechanism (dispersal limitation), or from environmental filtering of *Polynucleobacter* taxa differing in thermal adaptation (Hahn et al., 2015). In any case, a pronounced further increase in numbers of species-like *Polynucleobacter* OTUs has to be expected if future cultivation-independent studies investigate habitats located south of 40°N.

4.3 | Complex diversity trends along environmental gradients

An uneven distribution of *Polynucleobacter* subclusters along pH gradients has been previously reported (Jezbera et al., 2012; Nuy et al., 2020), and based on previous investigations, even within subcluster PnecC, differences in distribution of species along pH gradients are expected (Hahn et al., 2016; Jezbera et al., 2011). Due to the pronounced increased taxonomic resolution provided by the priB marker, much deeper insights into the structuring of PnecC communities by environmental factors are possible. This has revealed a couple of diversity trends.

Importantly, the composition of the PnecC communities did not change continuously along all environmental gradients. A previous study suggested that the composition of PnecC communities is mainly controlled by pH (Jezbera et al., 2011); however, in our study we observed that the majority of OTUs preferred either low or high Ca^{2+} concentrations (Figure 4). This trend seemed to be at least partially independent of pH, since alkaline habitats with low and high Ca^{2+} concentrations rarely share their inhabitants. Ca^{2+} concentrations were tightly correlated with conductivity ($R^2 = 0.93$, $p < .0001$), therefore, it is not known if Ca^{2+} concentrations or salinity was specifically controlling the composition of the communities. However, coastal habitats with increased NaCl concentrations shared community compositions with low but not high Ca^{2+} communities, suggesting that salinity is not the major driver of this distribution. The Ca^{2+} concentrations of aquatic systems are mainly controlled by their geological background. Therefore, high Ca^{2+} *Polynucleobacter* communities were restricted to habitats located in limestone areas and characterized by higher Ca^{2+} concentrations (Figure S1). However, even within limestone areas, smaller habitats with low Ca^{2+} concentrations inhabited by low Ca^{2+} communities were found. Such habitats are limited to systems influenced by peat bogs or, at least, influenced by peat moss (*Sphagnum* spp.) vegetation. In addition to Ca^{2+} concentration, pH had, as expected, a strong influence on the PnecC community composition (Figure 3); however, OTU composition changed more continuously along the pH gradient.

In botany, it is well known that silicate and limestone soils basically differ in their plant community compositions, at least regarding the nontree species (Bothe, 2015). These two soil types differ in many variables including pH and $CaCO_3$ content. Due to the manifold factors distinguishing these two soil types, the major drivers of the distinct differences in vegetation composition are unknown (Bothe, 2015). On the other hand, it is well known that in many plant genera pairs of vicarious species evolved, which either dwell on silicate or on limestone soils. Similar vicariances seemed to be given among species of *Polynucleobacter* subcluster PnecC. Another case of vicariance has been previously reported (Schauer et al., 2005) for planktonic freshwater bacteria affiliated with the two related taxa *Candidatus* Aquirestis calciphila (aka subcluster LD2) and *Candidatus* Haliscomenobacter calcifugiens (aka subcluster GKS2-217).

In soils, bacterial OTU richness shows a unimodal distribution along pH gradients with a richness peak at about neutral pH and a 3- to 4-fold change of richness across the pH gradient (Bickel et al., 2019; Fierer & Jackson, 2006). In comparison, the increase in PnecC OTU richness with pH was huge (Figure 6b). The observed maximum increase was about 50-fold for standing waters, and even around 90-fold if running waters were also considered. It is not clear if richness also follows a unimodal trend in PnecC bacteria. Richness is obviously strongly influenced by habitat size and type, as well as several other environmental factors (Figure 5c). However, if only medium-sized standing waters were considered, a unimodal model of richness along the pH gradient was suggested (Figure 6b). Probably, a unimodal shape could be clearly confirmed if the alkaline side of the sample distribution could be extended towards higher pH values.

Unexpectedly, OTU richness and relative abundance of PnecC bacteria showed opposing trends along the investigated pH gradient (Figure 6). It is well known that acidic lakes and ponds tend to possess higher relative abundances of PnecC bacteria (Jezbera et al., 2012; Jezberova et al., 2010; Figure 6a). Thus, lower numbers of PnecC OTUs present in acidic habitats contribute larger fractions to total bacterial numbers than PnecC communities in alkaline habitats with manifold higher OTU richness. Obviously, richness and relative abundance are at least partially decoupled across the investigated habitats. This might suggest differences in niche partitioning in acidic and alkaline waters; however, it is unknown which factors are involved in this unexpected diversity pattern.

Along with OTU richness, Shannon diversity increased with pH (Figure 3b). This increase in diversity was accompanied by an increase of community dissimilarity among communities dwelling in habitats of similar pH (Figure 3b). This trend was obvious even if habitats with high Ca^{2+} concentrations were excluded (Figure 5a). We found no suggestion of a general increase of environmental diversity between different pH classes along the gradient towards higher pH

values (Figure 5b); however, the measured environmental variables seemed to be only poor predictors of variance of composition of PnecC communities (Figure 5c). The increase of dissimilarity among communities of the same pH category is associated with a decreasing trend in occupancy of taxa (Figure S6A). This could be explained by a more stochastic community assembly in habitats with higher pH (Nemergut et al., 2013), combined with a higher number of OTUs able to dwell in systems with higher pH (Figure 6b). Persistence of taxa with rather low occupancy in particular habitats over periods of more than one year (e.g., *P. wuianus* and *P. meluiroseus*) may be suggestive of historical contingencies (Langenheder & Lindström, 2019) combined with potential local adaptation (Kraemer & Boynton, 2017). However, the observed phenomenon could also be linked to unmeasured abiotic environmental variables or to only locally occurring specific biotic interactions (Zhou & Ning, 2017). Time series and broader sets of measured variables are necessary to obtain insights into the mechanisms responsible for this phenomenon.

Obviously, *Polynucleobacter* species strongly differ in ecophysiological adaptations (e.g., pH, Ca^{2+}-related adaptation) but also in other ecological characteristics such as occupancy. The resolution of the priB marker was high enough to reveal these species-specific differences in adaptation and ecological success among *Polynucleobacter* bacteria, which are undetectable with 16S rRNA sequence-based methods (Figure 8).

4.4 | Conclusions

Amplicon sequencing of the priB gene provided an unprecedented insight into the diversity of *Polynucleobacter* bacteria and structuring of their local communities by environmental factors. The used marker gene revealed patterns and trends invisible to 16S rRNA sequence-based methods. An astonishingly high, yet incompletely covered, species richness was found in the studied area. The observed high richness could indicate a general huge underestimation of bacterial species richness by 16S rRNA-based methods, if the observed high degree of diversification is also present in other bacterial $OTUs_{99\%}$. Importantly, *Polynucleobacter* communities showed several patterns well known from macroecological theory, which were previously only observed in phylogenetically much broader microbial taxa and communities (Horner-Devine et al., 2004; Reche et al., 2005; Sogin et al., 2006). This includes species-area and geographic abundance-occupancy relationships, as well as the organization of communities in a few abundant and many rare taxa (rank-abundance curves, (Sogin et al., 2006)). In contrast, the observed opposing trends of abundance and diversity of *Polynucleobacter* communities along the pH gradient, as well as differences in pH-specific occupancy of taxa along this gradient were unexpected. Obviously, priB amplicon sequencing provides a possibility to study the mechanisms of community assembly in great detail. Furthermore, this method may provide an opportunity to measure the response of some important freshwater bacteria to environmental changes caused by anthropogenic impact (Kraemer et al., 2020) with higher sensitivity than synecological methods based on ribosomal markers. However, detailed studies on the influence of various environmental factors and time series are needed to better understand the mechanisms structuring *Polynucleobacter* communities and influencing the occupancy of particular taxa.

ACKNOWLEDGEMENTS

We thank Ulrike Koll and Johanna Schmidt for isolation of strains, DNA isolation, and processing of half of the samples for priB amplicon sequencing, and Johanna Schmidt for determination of major ion concentrations. We thank "Le Syndicat Mixte du Parc Naturel Régional de Corse et le gestionnaire des lacs d'altitude sur son territoire" for permission to take samples in the Corsica Natural Park (France), and we thank the National Park Hohe Tauern (Austria) and the owners of lakes for sampling permissions.

AUTHOR CONTRIBUTIONS

Martin W. Hahn designed research; all authors performed research and analysed data; Martin W. Hahn wrote the manuscript and Matthias Hoetzinger and Alexandra Pitt commented on, and edited, the draft manuscript.

DATA AVAILABILITY STATEMENT

Details on the sampled habitats and the measured environmental variable are provided in Supporting Information (Table S1 and Figure S1) The complete set of environmental variables have been deposited in the DRYAD repository (https://doi.org/10.5061/dryad.hhmgqnkgq). The nucleotide sequences of priB genes (MT988562-MT989336), genome sequences of *Polynucleobacter* strains (CP000655; CP007501; CP015017; CP015922; CP023276; CP023277; CP028940-CP028942; CP030085; CP049628; CP049637; CP049645; CP061288; CP061289; CP061291-CP061293; CP061295-CP061300; CP061302; CP061304-CP061306; CP061308-CP061319; JAANGD000000000; JAANHG000000000; JACVOK000000000; JACVOL000000000; JACVOM000000000; JACVON000000000; JACVOO000000000; JACVOP000000000; JACVOQ000000000; JACVOR000000000; JACVOS000000000; JACVOT000000000; JACVOU000000000; JACVOX000000000; JACVOY000000000; JACVOZ000000000; JACVPA000000000; JACVPD000000000; JACVPE000000000; JACVPF000000000; JACVPG000000000; JACVPH000000000; JACVPI000000000; JACVPJ000000000; JACVPM000000000; JACVPN000000000; JACVPP000000000; JACVPQ000000000; JACVPR000000000; JACVPS000000000; JACVPU000000000; JACVPW000000000; JACVPX000000000; JACVPY000000000; JACVPZ000000000; JACVQA000000000; JACVQB000000000; JACVQC000000000; JACVQD000000000; LOJI00000000; LOJJ00000000; LZFI00000000; LZMQ00000000; MPIY00000000; NAIA00000000; NGUO00000000; NGUP00000000; NJGG00000000; NTGB00000000; OANS00000000; PGTX00000000; QMCG00000000), reads obtained by Illumina amplicon sequencing (SRR11117533-SRR11117652), BioProject data (PRJNA607194), and BioSamples data (SAMN02724733; SAMN03430691; SAMN03430798;

SAMN04080026; SAMN04086652; SAMN04086667-SAMN04086669; SAMN06014615; SAMN07200920; SAMN08383909; SAMN08383917-SAMN08383921; SAMN14212605-SAMN14212701) associated with this study have been deposited in public databases curated by NCBI. Table S2 links reference strains and reference environmental reads to refOTU und eOTU (respectively), accession numbers of priB sequences and genome sequences.

ORCID

Martin W. Hahn https://orcid.org/0000-0003-0501-2556
Alexandra Pitt https://orcid.org/0000-0002-4347-7082
Matthias Hoetzinger https://orcid.org/0000-0002-1932-6479

REFERENCES

Allen, M. A., & Cavicchioli, R. (2017). Microbial communities of aquatic environments on Heard Island characterized by pyrotag sequencing and environmental data. *Scientific Reports*, *7*, 44480. https://doi.org/10.1038/srep44480

Bahr, M., Hobbie, J. E., & Sogin, M. L. (1996). Bacterial diversity in an arctic lake: A freshwater SAR11 cluster. *Aquatic Microbial Ecology*, *11*(3), 271–277. https://doi.org/10.3354/ame011271

Bickel, S., Chen, X., Papritz, A., & Or, D. (2019). A hierarchy of environmental covariates control the global biogeography of soil bacterial richness. *Scientific Reports*, *9*(1), 12129. https://doi.org/10.1038/s41598-019-48571-w

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Boscaro, V., Felletti, M., Vannini, C., Ackerman, M. S., Chain, P. S. G., Malfatti, S., Vergez, L. M., Shin, M., Doak, T. G., Lynch, M., & Petroni, G. (2013). *Polynucleobacter necessarius*, a model for genome reduction in both free-living and symbiotic bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(46), 18590–18595. https://doi.org/10.1073/pnas.1316687110

Bothe, H. (2015). The lime–silicate question. *Soil Biology and Biochemistry*, *89*, 172–183. https://doi.org/10.1016/j.soilbio.2015.07.004

Chevrette, M. G., Carlos-Shanley, C., Louie, K. B., Bowen, B. P., Northen, T. R., & Currie, C. R. (2019). Taxonomic and Metabolic Incongruence in the Ancient Genus Streptomyces. *Frontiers in Microbiology*, *10*(2170), https://doi.org/10.3389/fmicb.2019.02170

Comte, J., Monier, A., Crevecoeur, S., Lovejoy, C., & Vincent, W. F. (2016). Microbial biogeography of permafrost thaw ponds across the changing northern landscape. *Ecography*, *39*(7), 609–618. https://doi.org/10.1111/ecog.01667

Diao, M., Sinnige, R., Kalbitz, K., Huisman, J., & Muyzer, G. (2017). Succession of bacterial communities in a seasonally stratified lake with an anoxic and sulfidic hypolimnion. *Frontiers in Microbiology*, *8*(2511), https://doi.org/10.3389/fmicb.2017.02511

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(3), 626–631. https://doi.org/10.1073/pnas.0507535103

García-García, N., Tamames, J., Linz, A. M., Pedrós-Alió, C., & Puente-Sánchez, F. (2019). Microdiversity ensures the maintenance of functional microbial communities under changing environmental conditions. *The ISME Journal*, *13*(12), 2969–2983. https://doi.org/10.1038/s41396-019-0487-8

Gaston, K. J., Blackburn, T. M., Greenwood, J. J. D., Gregory, R. D., Quinn, R. M., & Lawton, J. H. (2000). Abundance–occupancy relationships. *Journal of Applied Ecology*, *37*(s1), 39–59. https://doi.org/10.1046/j.1365-2664.2000.00485.x

Hahn, M. W. (2003). Isolation of strains belonging to the cosmopolitan *Polynucleobacter necessarius* cluster from freshwater habitats located in three climatic zones. *Applied and Environmental Microbiology*, *69*(9), 5248–5254.

Hahn, M. W., Jezberova, 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. *The ISME Journal*, *10*(7), 1642–1655. https://doi.org/10.1038/ismej.2015.237

Hahn, M. W., Koll, U., Jezberova, J., & Camacho, A. (2015). Global phylogeography of pelagic *Polynucleobacter* bacteria: Restricted geographic distribution of subgroups, isolation by distance, and influence of climate. *Environmental Microbiology*, *17*(3), 829–840.

Hahn, M. W., Pöckl, M., & Wu, Q. L. (2005). Low intraspecific diversity in a *Polynucleobacter* subcluster population numerically dominating bacterioplankton of a freshwater pond. [yes]. *Applied and Environmental Microbiology*, *71*(8), 4539–4547.

Heckmann, K., & Schmidt, H. J. (1987). *Polynucleobacter necessarius* gen. nov., sp. nov., an obligately endosymbiotic bacterium living in the cytoplasm of *Euplotes aediculatus*. *International Journal of Systematic Bacteriology*, *37*(4), 456–457.

Hijmans, R. J. (2019). *Geosphere: Spherical trigonometry. R package version 1.5-10*. https://CRAN.R-project.org/package=geosphere

Hijmans, R. J., Guarino, L., Cruz, M., & Rojas, E. (2001). Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter*, *127*, 15–19.

Hill, J. E., Seipp, R. P., Betts, M., Hawkins, L., Van Kessel, A. G., Crosby, W. L., & Hemmingsen, S. M. (2002). Extensive profiling of a complex microbial community by high-throughput sequencing. *Applied and Environmental Microbiology*, *68*(6), 3055–3066. https://doi.org/10.1128/aem.68.6.3055-3066.2002

Hoetzinger, M., & Hahn, M. W. (2017). Genomic divergence and cohesion in a species of pelagic freshwater bacteria. *BMC Genomics*, *18*(1), 794. https://doi.org/10.1186/s12864-017-4199-z

Hoetzinger, M., Pitt, A., Huemer, A., & Hahn, M. W. (2021). Continental-scale gene flow prevents allopatric divergence of pelagic freshwater bacteria. *Genome Biol Evol*, *13*(3), evab019. https://doi.org/10.1093/gbe/evab019

Hoetzinger, M., Schmidt, J., Jezberová, J., Koll, U., & Hahn, M. W. (2017). Microdiversification of a pelagic *Polynucleobacter* species is mainly driven by acquisition of genomic islands from a partially interspecific gene pool. *Applied and Environmental Microbiology*, *83*(3), e02266–e2216. https://doi.org/10.1128/aem.02266-16

Horner-Devine, M. C., Lage, M., Hughes, J. B., & Bohannan, B. J. (2004). A taxa-area relationship for bacteria. *Nature*, *432*(7018), 750–753. https://doi.org/10.1038/nature03073

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114. https://doi.org/10.1038/s41467-018-07641-9

Jaspers, E., & Overmann, J. (2004). Ecological significance of microdiversity: Identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. *Applied and Environmental Microbiology*, *70*(8), 4831–4839. https://doi.org/10.1128/AEM.70.8.4831-4839.2004

Jezbera, J., Jezberova, J., Brandt, U., & Hahn, M. W. (2011). Ubiquity of *Polynucleobacter necessarius* subspecies *asymbioticus* results from ecological diversification. *Environmental Microbiology*, *13*(4), 922–931. https://doi.org/10.1111/j.1462-2920.2010.02396.x

Jezbera, J., Jezberova, J., Koll, U., Hornak, K., Simek, K., & Hahn, M. W. (2012). Contrasting trends in distribution of four major planktonic betaproteobacterial groups along a pH gradient of epilimnia of 72 freshwater habitats. *FEMS Microbiology Ecology*, *81*(2), 467–479. https://doi.org/10.1111/j.1574-6941.2012.01372.x

Jezberova, J., Jezbera, J., Brandt, U., Lindstrom, E. S., Langenheder, S., & Hahn, M. W. (2010). Ubiquity of *Polynucleobacter necessarius* ssp. *asymbioticus* in lentic freshwater habitats of a heterogenous 2000 km^2 area. *Environmental Microbiology*, *12*(3), 658–669.

Konstantinidis, K. T., Ramette, A., & Tiedje, J. M. (2006). The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1475), 1929–1940. https://doi.org/10.1098/rstb.2006.1920

Kraemer, S. A., Barbosa da Costa, N., Shapiro, B. J., Fradette, M., Huot, Y., & Walsh, D. A. (2020). A large-scale assessment of lakes reveals a pervasive signal of land use on bacterial communities. *The ISME Journal*, *14*(12), 3011–3023. https://doi.org/10.1038/s41396-020-0733-0

Kraemer, S. A., & Boynton, P. J. (2017). Evidence for microbial local adaptation in nature. *Molecular Ecology*, *26*(7), 1860–1876. https://doi.org/10.1111/mec.13958

Langenheder, S., & Lindström, E. S. (2019). Factors influencing aquatic and terrestrial bacterial community assembly. *Environmental Microbiology Reports*, *11*(3), 306–315. https://doi.org/10.1111/1758-2229.12731

Minka, T. P., & Deckmyn, A. (2018). *Maps: Draw geographical maps. R package version 3.3.0*. https://CRAN.R-project.org/package=maps

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., Knelman, J. E., Darcy, J. L., Lynch, R. C., Wickey, P., & Ferrenberg, S. (2013). Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews*, *77*(3), 342–356. https://doi.org/10.1128/MMBR.00051-12

Newton, R. J., & McLellan, S. L. (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. https://doi.org/10.3389/fmicb.2015.01028

Nuy, J. K., Hoetzinger, M., Hahn, M. W., Beisser, D., & Boenigk, J. (2020). Ecological differentiation in two major freshwater bacterial taxa along environmental gradients. *Frontiers in Microbiology*, *11*(154), https://doi.org/10.3389/fmicb.2020.00154

Ogier, J.-C., Pagès, S., Galan, M., Barret, M., & Gaudriault, S. (2019). rpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. *BMC Microbiology*, *19*(1), 171. https://doi.org/10.1186/s12866-019-1546-z

Oksanen, F. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B. & Wagner, H. (2019). *vegan: Community Ecology Package. R package version 2.5-6*. https://CRAN.R-project.org/package=vegan

Peixoto, J. C., Leomil, L., Souza, J. V., Peixoto, F. B., & Astolfi-Filho, S. (2011). Comparison of bacterial communities in the Solimoes and Negro River tributaries of the Amazon River based on small subunit rRNA gene sequences. *Genetics and Molecular Research*, *10*(4), 3783–3793. https://doi.org/10.4238/2011.December.8.8

Percent, S. F., Frischer, M. E., Vescio, P. A., Duffy, E. B., Milano, V., McLellan, M., Stevens, B. M., Boylen, C. W., & Nierzwicki-Bauer, S. A. (2008). Bacterial community structure of acid-impacted lakes: What controls diversity? *Applied and Environmental Microbiology*, *74*(6), 1856–1868. https://doi.org/10.1128/aem.01719-07

Pereira, R. P. A., Peplies, J., Mushi, D., Brettar, I., & Hofle, M. G. (2018). Pseudomonas-specific NGS assay provides insight into abundance and dynamics of pseudomonas species including *P. aeruginosa* in a cooling tower. *Frontiers in Microbiology*, *9*, https://doi.org/10.3389/fmicb.2018.01958

Pitt, A., Schmidt, J., Lang, E., Whitman, W. B., Woyke, T., & Hahn, M. W. (2018). *Polynucleobacter meluiroseus* sp. nov. a bacterium isolated from a lake located in the mountains of the Mediterranean island of Corsica. *International Journal of Systematic and Evolutionary Microbiology*, *68*(6), 1975–1985.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reche, I., Pulido-Villena, E., Morales-Baquero, R., & Casamayor, E. O. (2005). Does ecosystem size determine aquatic bacterial richness? *Ecology*, *86*(7), 1715–1722. https://doi.org/10.1890/04-1587

Rodriguez-R, L. M., Castro, J. C., Kyrpides, N. C., Cole, J. R., Tiedje, J. M., & Konstantinidis, K. T. (2018). How much do rRNA gene surveys underestimate extant bacterial diversity? *Applied and Environmental Microbiology*, *84*(6), e00014–00018. https://doi.org/10.1128/aem.00014-18

Sánchez, D., Matthijs, S., Gomila, M., Tricot, C., Mulet, M., García-Valdés, E., & Lalucat, J. (2014). *rpoD* gene pyrosequencing for the assessment of *Pseudomonas* diversity in a water sample from the Woluwe River. *Applied and Environmental Microbiology*, *80*(15), 4738–4744. https://doi.org/10.1128/aem.00412-14

Schauer, M., Kamenik, C., & Hahn, M. W. (2005). Ecological differentiation within a cosmopolitan group of planktonic freshwater bacteria (SOL cluster, Saprospiraceae, Bacteroidetes). *Applied and Environmental Microbiology*, *71*(10), 5900–5907. https://doi.org/10.1128/aem.71.10.5900-5907.2005

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., Arrieta, J. M., & Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences*, *103*(32), 12115–12120. https://doi.org/10.1073/pnas.0605127103

Stackebrandt, E., & Ebers, J. (2006). Taxonomic parameters revisited: tarnished gold standards. *Microbiology Today*, *33*, 152–155.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

Zhou, J., & Ning, D. (2017). Stochastic community assembly: Does it matter in Microbial Ecology? *Microbiology and Molecular Biology Reviews*, *81*(4), e00002–00017. https://doi.org/10.1128/mmbr.00002-17

Zwart, G., Crump, B. C., Agterveld, M., Hagen, F., & Han, S. K. (2002). Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquatic Microbial Ecology*, *28*(2), 141–155. https://doi.org/10.3354/ame028141

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Hahn, M. W., Huemer, A., Pitt, A., & Hoetzinger, M. (2021). Opening a next-generation black box: Ecological trends for hundreds of species-like taxa uncovered within a single bacterial >99% 16S rRNA operational taxonomic unit. *Molecular Ecology Resources*, 21, 2471–2485. https://doi.org/10.1111/1755-0998.13444