

Assessing biogeographic survey gaps in bacterial diversity knowledge: A global synthesis of freshwaters

Allison M. Veach¹  | Matthew J. Troia¹ | Melissa A. Cregger^{2,3}

¹Department of Environmental Science and Ecology, University of Texas at San Antonio, San Antonio, TX, U.S.A.

²Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, U.S.A.

³Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, U.S.A.

Correspondence

Allison M. Veach, Department of Environmental Science and Ecology, University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, U.S.A.
Email: allison.veach@utsa.edu

Funding information

Office of Biological and Environmental Research, Office of Science, U.S. Department of Energy (DOE), Grant/Award Number: DE-AC05-00OR22725

Abstract

1. Freshwaters account for 0.8% of Earth's surface area, yet support >10% of known plant and animal species making them disproportionately biodiverse. Modern molecular techniques have begun to reveal microbial diversity, but application of these approaches to address global microbial biogeography is relatively unknown in freshwaters.
2. Our aim was to identify gaps in microbial data coverage along climatic and landscape disturbance gradients and among terrestrial biomes and hydrographic regions for all freshwater ecosystems and three freshwater habitat types: lakes and reservoirs (lentic); streams and rivers (lotic); and wetlands.
3. We reviewed literature on microbial diversity in freshwaters surveyed using 16S ribosomal RNA sequencing which identify microbial taxa. We georeferenced survey locations and used a geographic information system to identify and map gaps in survey coverage using open-source data for climate, landscape disturbance, terrestrial biomes, and freshwater ecoregions.
4. We compiled 3,425 georeferenced survey locations reported from 963 studies. Streams were surveyed most frequently (60.8% of survey locations), followed by lakes (33.5%) and wetlands (5.6%). Surveys were concentrated in North America, central and western Europe, and Southeast Asia; 35% of freshwater ecoregions were surveyed at least once across freshwater habitat types, whereas 23%, 23%, and 12% were surveyed at least once for lentic, lotic, and wetland habitat types, respectively. The climatic gap analysis indicated coverage is high for temperate regions but lacking in the tropics and Arctic, particularly for wetland ecosystems.
5. Our assessment revealed high climatic coverage of freshwater microbial diversity knowledge, but expansive ecoregional gaps attributable to biased sampling near research institutions in North America, western Europe, and China. Future surveys should target ecoregions in Africa, South America, Central Asia, Australia, and Antarctica. An essential next step will be to curate and disseminate sequencing efforts to facilitate the study of processes driving global diversity patterns.

KEYWORDS

16S rRNA, climatic gaps, lotic, lentic, wetlands

1 | INTRODUCTION

Documenting global biodiversity is a fundamental goal of biogeography and essential for predicting and mitigating impacts of global environmental change (Reyers et al., 2012; Wheeler et al., 2012). Biodiversity not only enhances cultural services, such as recreation and landscape aesthetics (Harrison et al., 2014), but also mediates ecosystem processes, such as nutrient cycling and climate regulation, and supports ecosystem provisioning (e.g. timber production and commercial fisheries; Díaz et al., 2006). For centuries, naturalists and scientists have compiled species occurrence records for macrobes (Wallace, 1876) and, within the last decade, a wealth of biogeographic information has been compiled and made available digitally. For example, the Global Biodiversity Information Facility provides online access to 1.6 billion records from over 54,000 datasets globally (GBIF.org). Open-access biodiversity databases have contributed to significant advancements in basic ecology and biodiversity research (Beck et al., 2013) and have informed policymakers on the impacts of anthropogenic change (Harrison et al., 2014; Yang, 2011).

A challenge of using open-source biodiversity data is that occurrence records are unevenly distributed globally (Wilson et al., 2016). Such geographic biases arise from numerous investigators conducting surveys with different research objectives resulting in biodiversity that remains undescribed—the Linnean shortfall—and geographic distributions of species that are poorly delineated—the Wallacean shortfall (Bini et al., 2006). Assessments characterising the geographic and environmental coverage of open-source biodiversity databases have highlighted where gaps in knowledge occur, why they occur, and how these gaps can be filled. For example, national and global assessments have been conducted for terrestrial vertebrates (Meyer et al., 2015), plants (Sousa-Baena et al., 2014), and freshwater fishes (Pelayo-Villamil et al., 2018). A critical element of the Linnean shortfall is the lack of microbial biodiversity knowledge because: (1) the molecular technology necessary to describe their diversity have only become available recently; and (2) microbial distributions are viewed as ubiquitous and of low conservation need. Indeed, microbial taxa among free-living habitats, such as soils, marine waters, and freshwaters (Angel et al., 2010; Fierer & Jackson, 2006; Fierer et al., 2012; Ibarbalz et al., 2019; Martiny et al., 2006; Nemergut et al., 2011; Newton et al., 2011) and host-associated habitats probably exhibit biogeographic patterns (Thompson et al., 2017). The accessibility of advanced molecular methodologies has allowed microbial diversity surveys to accumulate in the last 2 decades, notably archived sequence repositories (e.g. NCBI Sequence Read Archives) but it is largely unknown how well surveying efforts cover the geographic and environmental diversity of the Earth's surface.

Freshwater biodiversity is disproportionately diverse globally: freshwaters cover approximately 0.8% of Earth's surface yet support approximately 10% of all known species (Dudgeon et al., 2006); therefore, understanding biodiversity knowledge gaps in freshwaters are imperative in the face of environmental change. Freshwater

ecosystems are fragmented throughout landscapes, are heavily impacted due to exploitation for human goods and services (Woodward et al., 2010) and are more impacted by global environmental change compared to terrestrial ecosystems (Dudgeon et al., 2006). Due to the heightened threat to freshwater biodiversity in the face of climate change, it is critical that ecologists begin to recognise the global extent of microbial diversity knowledge within these systems. This study aimed to characterise geographic coverage of freshwater microbial diversity surveys among biogeographic realms and along gradients of landscape disturbance and climate. This work will help prioritise future surveying efforts to maximise discovery of novel microbiomes.

2 | METHODS

2.1 | Literature search

During September 2017–February 2021, we performed multiple literature searches through the Web of Science database to find peer-reviewed scientific studies investigating microbial diversity via community-fingerprinting, next-generation sequencing, and metagenomic sequencing approaches in freshwaters. Because different types of freshwater habitat harbour different microbial communities (Monard et al., 2016; Zeglin, 2015), we grouped surveys into three freshwater habitat types: (1) flowing streams and rivers (hereafter *lotic*); (2) non-flowing natural lakes, ponds, and human-created reservoirs (hereafter *lentic*); and (3) wetlands (hereafter *wetland*). For all freshwater ecosystems we included the terms “16S” and “bacteria*”, but for specific habitat types, we used the search terms “stream” or “river” or “lotic” for streams and rivers, “lake” or “reservoir” or “pond” or “lentic” or “lacustrine” for lakes, and “wetland” or “bog” or “mire” for wetlands. All papers found through literature searches were retained for geospatial analyses if a study: (1) targeted 16S ribosomal RNA (rRNA) or rRNA genes; (2) sampled an aquatic ecosystem that had a salinity <3 ppm thus excluding brackish and estuarine waters; (3) sampled a habitat that was not *extreme* such as extremely acidic or alkaline waters; (4) included representative samples of a natural ecosystem (inclusive of field and laboratory studies, exclusive of bioreactor studies); and (5) authors either provided latitude and longitude data in the publication, in correspondence for this study's purposes, or was feasible to hand-annotate by site descriptions.

Latitude and longitude data were extracted from publications when provided. If a study performed sampling within a narrow spatial extent (<10 km range) and only provided one latitude and longitude, these points were extracted. If multiple latitudes and longitudes were given, these were extracted. In some instances, a study did not report latitude and longitude data or representative points when sampling had large spatial extents, so authors were contacted for latitude and longitude data. If site descriptions were adequate for recording sampling locations, latitude and longitude were hand-annotated and extracted via Google Maps. We also recorded

relevant metadata: year of publication, country, 16S rRNA methodology, field or laboratory-based, and specific habitat surveyed (e.g. biofilm, sediments, water column). The specific habitat surveyed included differentiating biofilms and microbial mats (referred to as *biofilms*), sampling of the water column (*water*), soils and sediments (*sediments*), microbiomes associated with plants or moss (*plant-associated*), leaf litter, ice, and garbage and microplastics (*human litter*). Note that we include studies which use shotgun metagenomics and extracted 16S rRNA data.

2.2 | Geospatial analyses

We implemented a geospatial analysis to characterise the geographic coverage of freshwater microbial surveys and identify gaps in the context of biogeography, anthropogenic landscape alteration, and climate. All subsequent geospatial analyses were performed on all three habitat types (hereafter *all freshwaters*) and each habitat type separately. First, we characterised the distribution of surveys among the 426 freshwater ecoregions of the world (Abell et al., 2008) and 16 terrestrial biomes (Olson et al., 2001) using the spatial join tool in ArcMap (ESRI, Inc.; Version 10.5). The freshwater ecoregions dataset is delineated along drainage divides that distinguish freshwater faunas (primarily freshwater fishes) with distinct phylogenetic history, paleogeography, and ecology (Abell et al., 2008). Although many freshwater bacterial groups are not obligate aquatic and probably not dispersal limited across drainage divides (Monard et al., 2016; Nemergut et al., 2013; Padial et al., 2014), we assumed that freshwater ecoregions are as informative as terrestrial biomes for microbes because freshwater processes are likely to drive their occurrence regionally as much as coarser-scale vegetation-defined terrestrial biomes.

Second, we characterised anthropogenic landscape alteration at survey locations using the 2009 Human Footprint dataset, which we acquired from the Socioeconomic Data and Applications Center (Venter et al., 2018). This dataset provides a landscape disturbance index ranging from 0 (low disturbance) to 50 (high disturbance) at a 1-km² resolution. The index is computed from eight variables: built-up environments, population density, electric power infrastructure, crop land use, pasture land use, road corridors, railway corridors, and navigable waterways. We extracted landscape disturbance values to all survey locations representing bodies of water on larger land masses outside of Antarctica; however, this method yielded missing data due to three geospatial disparities. First, because the Human Footprint dataset does not provide landscape disturbance values for Antarctica, we assumed landscape disturbance approximates mean landscape disturbance in the Arctic, which we calculated using gridded points within the Arctic Circle (i.e. north of 66.5°N latitude). Second, landscape disturbance values are unavailable for small landmasses (e.g. American Samoa), so we assumed landscape disturbance at these survey locations approximates those of the nearest small landmass for which the landscape disturbance was available (e.g. French Polynesia). Third, for survey locations from

large waterbodies (e.g. Lake Michigan, Lake Baikal, Lake Tanganyika) for which landscape disturbance was not available, we computed the mean landscape disturbance of a 50-km buffer surrounding the perimeter of each waterbody. Perimeters of these waterbodies were acquired from the global reservoirs and dams dataset (Lehner et al., 2011) if available, or were traced manually using satellite imagery in ArcMap.

Third, we characterised climatic coverage across all survey locations and use this to estimate a climatic gap index (CGI) to identify underrepresented climates. To calculate CGI, we extracted the 19 bioclim variables provided by the WorldClim 2 dataset (Fick & Hijmans, 2017) via the extract values to points tool in ArcMap. Because this dataset does not include Antarctica, we extracted the same 19 bioclim variables for Antarctic survey locations using the MerraClim dataset (Vega et al., 2017). We generated a grid of equal area cells (44 × 44 km separation between adjacent points) across the global terrestrial surface and extracted values of the 19 bioclim variables for each of these 61,607 grid cells. For each grid cell, we computed CGI using the following equation:

$$CGI_i = \sum_{j=1}^{19} x_{ij}$$

where x_{ij} is an index of climate dissimilarity between grid cell i and all surveyed grid cells for bioclim variable j . This index, x , ranged from 0 (indicating identical climate) to 10 (no climatic similarity). The value of index, x , was based on percent deviation from climatic conditions at survey points and was scaled according to the gradient length of the bioclim variable (Table S1). This followed the assumption that community composition turns over along each bioclim gradient and increasing climatic dissimilarity drives increasing dissimilarity in community composition for microbes (Bryant et al., 2008; Currie et al., 2004; Davidar et al., 2007; Fierer et al., 2012). Under this assumption, grid cells (or geospatially delineated locations on Earth's surface) with increasing climatic dissimilarity represent increasingly important gaps in knowledge of microbial diversity (Jetz et al., 2012; Troia & McManamay, 2016) and will have greater CGI values. Lastly, we mapped this multivariate climatic dissimilarity index (ranging from 0 to 190) to the global terrestrial surface.

3 | RESULTS

3.1 | Metadata summary

We reviewed 963 peer-reviewed articles published, from which 3,425 unique georeferenced survey locations were described in the article or provided by authors upon request. The majority (c. 95%) of surveys were field-based ($n = 3,271$) and a small proportion (c. 5%) were exclusively mesocosm or microcosm-based ($n = 154$). These surveys represent lotic habitats first ($n = 2,068$), followed by lentic habitats ($n = 1,156$) and wetlands ($n = 201$). The first surveys occurred in 1997 and the number of surveys has accelerated

in subsequent years. Lentic and lotic survey number have increased at a greater rate than wetland surveys (Figure 1). Lentic and lotic surveys primarily surveyed the water column (c. 79% and 69% of all surveys, respectively), and secondarily, sediment (both c. 16%). Lotic studies also surveyed biofilms (c. 11%) more commonly than other freshwater habitats (c. 2% in both lentic and wetland surveys; Figure S1). Wetland surveys primarily surveyed sediments (c. 75% of all wetland surveys) and water secondarily (c. 23%; Figure S1). Lakes had a greater proportion and number of algal-associated surveys ($n = 12$, 1.0%) compared to streams ($n = 1$) and wetlands ($n = 0$) whereas streams had a greater proportion and number of both plant litter ($n = 23$, c. 1.0%) and human litter ($n = 36$, c. 1.7%) surveys compared to lakes ($n = 1$, 0 respectively) and wetlands ($n = 1$, 0 respectively; Figure S1).

3.2 | Geospatial analyses

The 3,425 georeferenced surveys were distributed across all seven continents—including 50 surveys in Antarctica—but were most densely distributed in eastern North America, western Europe, and eastern Asia (Figure 2a). This geographic pattern was similar for lentic, lotic, and wetland habitats, except Africa and Antarctica are devoid of wetland surveys (Figure 2b–d). We identified surveys in 74 of 195 countries, with most surveys in the U.S.A. ($n = 710$), China ($n = 602$), Canada ($n = 490$), and France ($n = 178$). Specifically, lentic habitats were most densely surveyed in China ($n = 231$), U.S.A. ($n = 207$), and France ($n = 118$) whereas lotic habitats were most dense in U.S.A. ($n = 433$), Canada ($n = 415$), and China ($n = 306$).

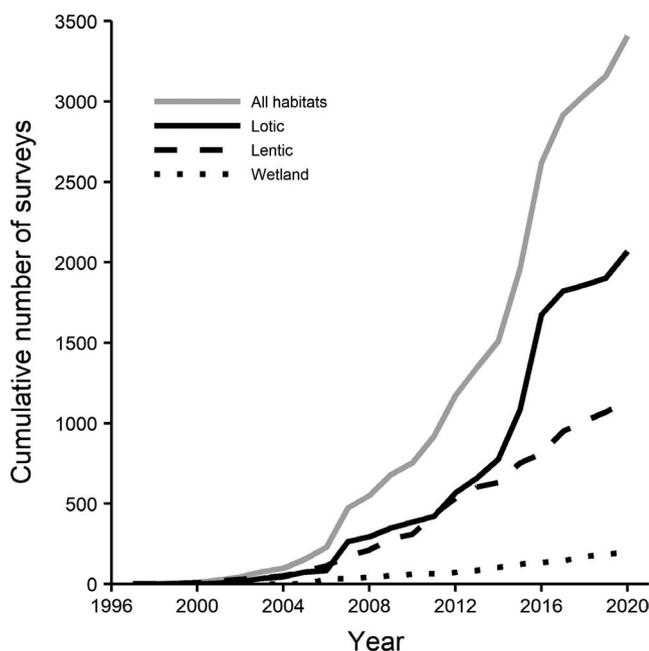


FIGURE 1 Accumulation of the number of georeferenced surveys of freshwater microbial communities from 1997 through 2020. Lines represent all habitats (solid grey), lotic (solid black), lentic (dashed black), and wetland (dotted black) habitats

Wetlands were most densely surveyed in U.S.A. ($n = 68$), China ($n = 561$), and Mexico ($n = 13$; Figure 3a,d,g,i).

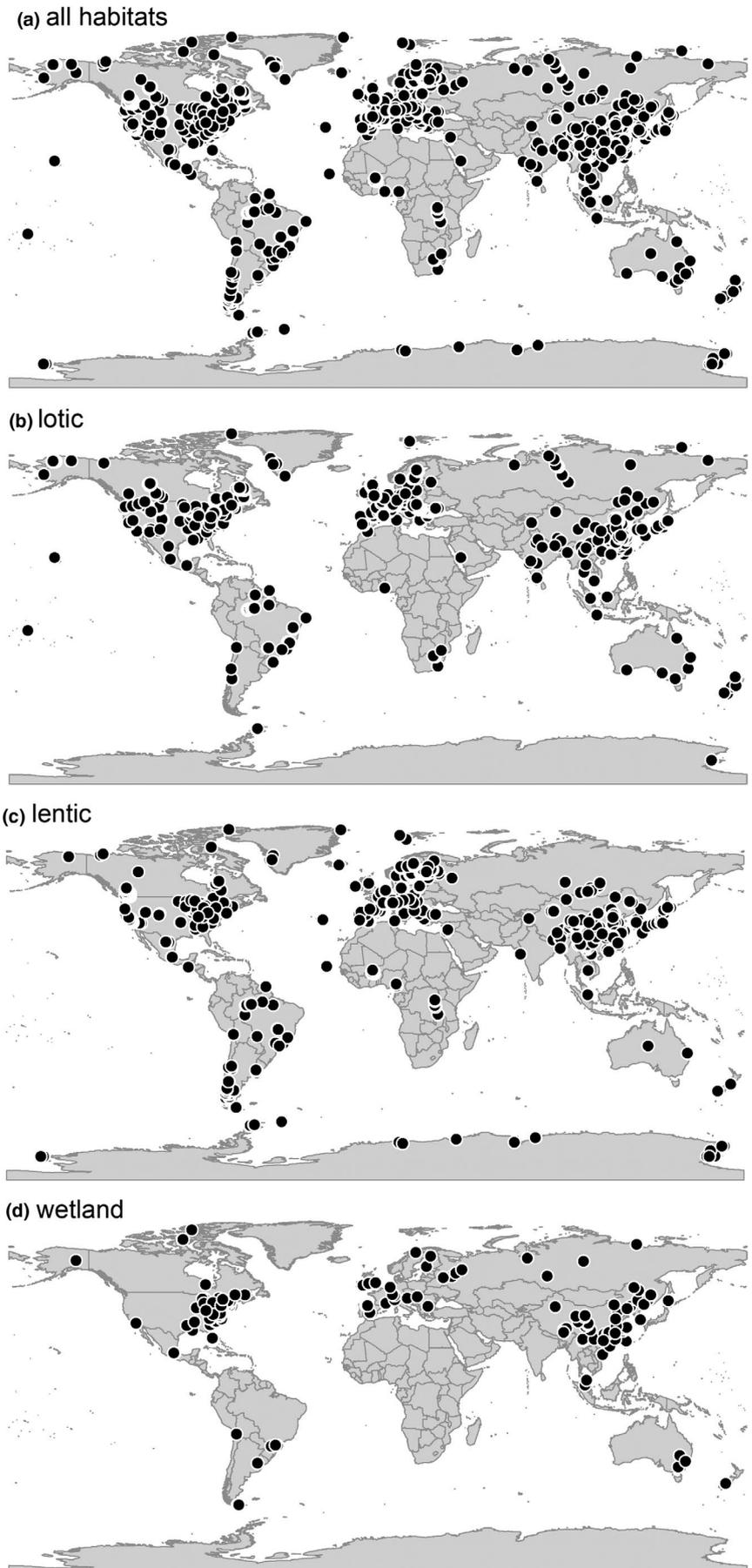
Terrestrial biomes were moderately represented across freshwater ecosystems with temperate broadleaf and mixed forests ($n = 1,649$, 48%), boreal forests ($n = 477$, 14%), tropical and subtropical moist broadleaf forests ($n = 222$, 6.5%), temperate conifer forests ($n = 203$, c. 6%), Mediterranean forests, woodlands, shrubs ($n = 202$, c. 6%), and tundra ($n = 173$, 5%) representing the greatest number of surveys across all freshwaters (Figure 3b). All grassland types ($n = 272$, 8%), deserts ($n = 82$, c. 2.5%), and Arctic and Antarctic ice ($n = 8$, 0.2%) were not well surveyed across freshwaters. Lentic and lotic ecosystems had a similar trend of survey proportion across these terrestrial biomes (Figure 3e,h) except lentic habitats had a greater proportion of tundra ($n = 87$, c. 8%) and desert ($n = 45$, c. 4%) surveys and lotic habitats had a greater proportion of temperate grasslands ($n = 104$, c. 5%). Wetlands had the majority of surveys within temperate broadleaf and mixed forests ($n = 100$, 50%), tropical and subtropical moist broadleaf forests ($n = 21$, 10%), flooded grasslands/savannas ($n = 16$, 8%), and deserts ($n = 15$, 7%; Figure 3k).

Freshwater ecoregions were modestly represented with 172 (40%), 124 (29%), 118 (28%), and 66 (15%) different freshwater ecoregions containing one or more surveys of any habitat, lotic habitat, lentic habitat, and wetland habitat, respectively (Figure 3c,f,i,l). The most-surveyed freshwater ecoregions include central and western Europe (Freshwater Ecoregion (FE) 404, $n = 298$), lower Yangtze (FE 766, $n = 238$), upper Danube (FE 417, $n = 219$), Laurentian Great Lakes (FE 116, $n = 176$), and northern Baltic Drainages (FE 406, $n = 176$). Lentic and lotic ecosystems followed this trend (Figure 3f,i), except lentic systems had a high proportion of surveys in the Cantabric Coast–Languedoc ecoregion (FE 403, 4%) and lotic ecosystems had a high proportion of surveys in Dniester–lower Danube (FE 418, 6%) and Eastern Hudson Bay–Ungava (FE 113, 5%). Alternatively, wetlands greatest number of surveys were in the north-east U.S. and south-east Canada Atlantic Drainages (FE 118, 9%), Laurentian Great Lakes (FE 116, 7%), lower Yangtze (FE 766, 7%), and Lerma–Chapala (FE 165, 5%; Figure 3l).

Relative to the background environment, surveys encompassed a range of anthropogenic landscape disturbance gradients for freshwaters (Figure S2). Both lotic and lentic surveys had greater frequencies of low landscape disturbance (15% and 13% surveys of no disturbance, respectively) relative to high disturbance (Figure S2). High disturbance surveys were a greater proportion of surveys for lakes compared to streams (Figure S2).

Climatic gap index values close to 0 indicate high coverage of climate whereas CGI values close to 190 indicate low coverage of climate. Based on the CGI metric, few climatic gaps existed for all freshwaters regardless of disturbance intensity (Figures 4 and 5). All freshwater surveys had CGI values below or at 100 (with 190 being the highest rating of climatic gaps possible). Climatic gaps did exist for portions of high latitude Palearctic and Nearctic biogeographic realms, Indo-Malayan realms, and the

FIGURE 2 Global distribution of freshwater microbial taxonomy surveys for all (a), (b) lotic, (c) lentic, and (d) wetlands habitats. Each point represents an individual 16S rRNA survey derived from the Web of Science meta-analysis



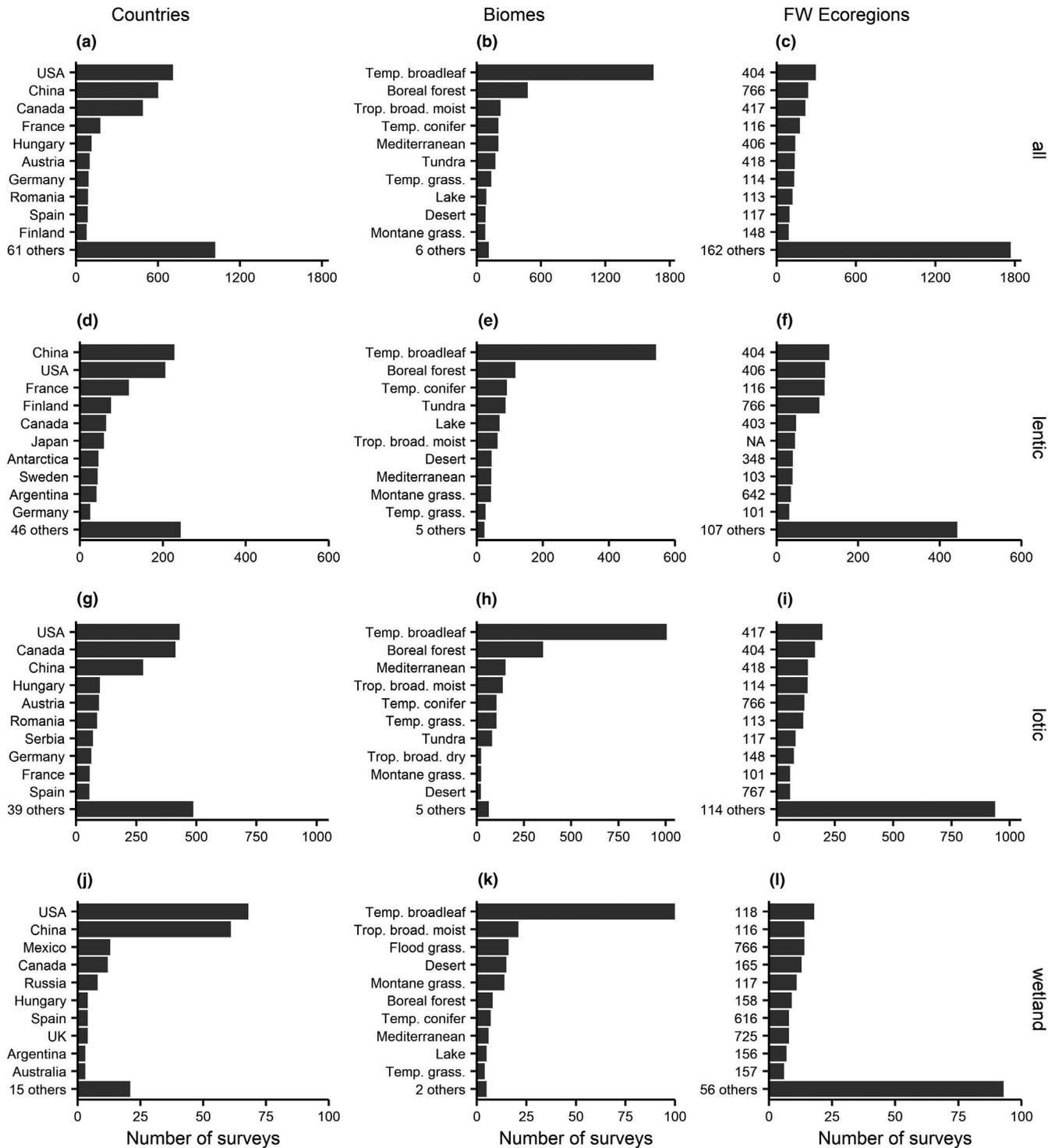


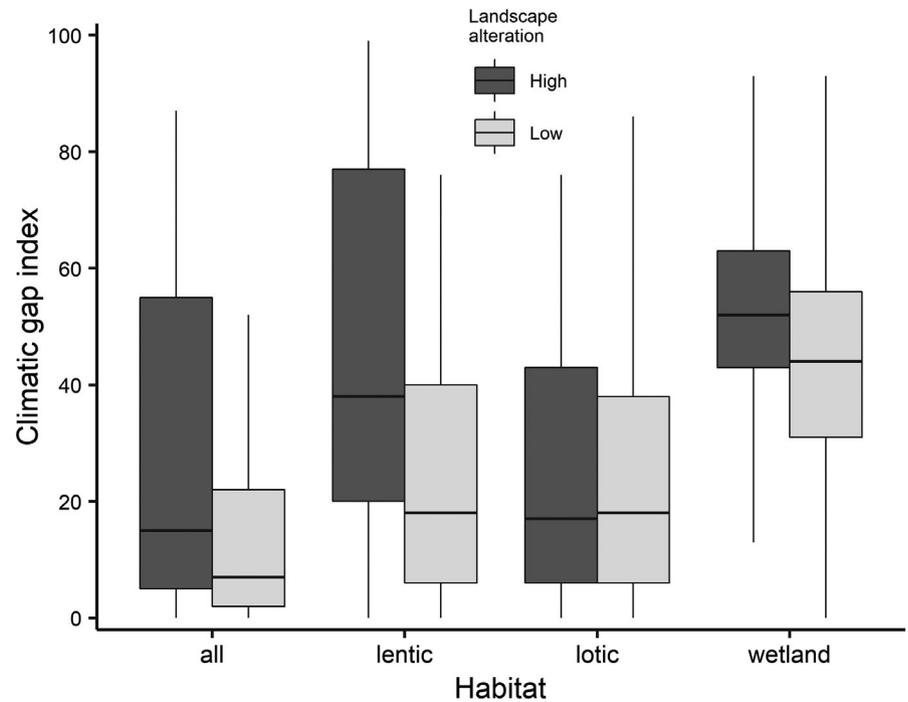
FIGURE 3 The frequency, or number of surveys, of freshwater microbial taxonomy surveys for all (a,b,c), lentic (d,e,f), lotic (g,h,i), and wetland (j,k,l) habitats grouped by countries (a,d,g,j), terrestrial biomes (b,e,h,k), and freshwater ecoregions (c,f,i,l). Freshwater ecoregion codes follow Abell et al. (2008); NA represents Antarctica

Antarctic, particularly surveys of high-disturbance (Figure 5a-h and Figure S3). Climatic gaps were more pronounced for wetlands that had the highest measures of CGI regardless of landscape disturbance (Figures 4 and 5). Specifically, Neotropical and Indo-Malayan biogeographic realms had the greatest CGI for wetlands (Figure 5g,h and Figure S3).

4 | DISCUSSION

Our results demonstrated that global coverage of freshwater microbial diversity surveys exhibits bias dependent on habitat type. These data confirm that peer-reviewed research is biased due to: (1) research institution proximity and availability of funding and resources

FIGURE 4 Boxplots showing distribution of the climatic gap index of freshwater microbial diversity for the four habitat types (all freshwater, lentic, lotic, and wetlands) and in areas of low (light grey) versus high (dark grey) anthropogenic landscape alteration. Landscape alteration was determined by the 2009 Human Footprint database (Venter et al., 2018)



or logistical ease of sampling of geographic regions (e.g. U.S. Great Lakes vs. Amazonian and Andean lakes; Tydecks *et al.*, 2018); and (2) traditional methods that are discipline specific (i.e. stream ecology and limnology vs. wetlands), which suggests that paradigm biases among ecological disciplines dictate how commonly research questions address microbial diversity in different types of freshwater ecosystems (Graham & Dayton, 2002).

The rate of increase for microbial diversity surveys over time are high for lotic and lentic habitats relative to wetlands indicative that wetland research has lagged in targeted 16S rRNA or metagenomic methodologies to identify bacterial/archaeal communities. Notably, a moderate percentage of wetland surveys (c. 27%) also measured greenhouse gas emissions (e.g. methane production), due to the prevalence of anoxic conditions in wetland sediments. Numerous publications in wetlands were not included in this study due to several exclusively targeting functional genes of methanogenic archaea (methyl-coenzyme reductase subunit A gene; *mcrA*) instead of all bacterial and archaeal taxa (16S rRNA surveys). This suggests that microbial ecologists often frame research questions and design field studies to understand biodiversity–ecosystem function, unlike microbial biodiversity researchers (Liu *et al.*, 2011). Liu *et al.*, 2011 performed a bibliometric meta-analysis and demonstrated that the most frequent research topics in biodiversity research are related to structure (terms: “populations”, “diversity”, etc.) more frequently than function (terms: “productivity”, “ecosystems”, etc.). This suggests that biodiversity–function relationships constitute a lower proportion of biodiversity research. Microbes have been mostly excluded in conservation research and, likewise, environmental microbiology historically lacked a biodiversity cataloguing perspective. Microbes are likely to be excluded due to their perceived low chance of extinction among habitats based on relatively large population sizes and assumed cosmopolitan distribution (Casamayor, 2017). Despite this,

microorganisms are the engines of biogeochemistry and have a large effect on ecosystem function and preservation so it is unsurprising that biodiversity–function is a dominant research theme in current microbial biodiversity studies (Cavicchioli *et al.*, 2019).

Surveys encompassed a large geographic extent, but density was greatest for North America, Europe, and East Asia and freshwater ecoregions and terrestrial biomes spanning these geographic locations were most densely surveyed. Likewise, temperate and boreal forests were most heavily surveyed whereas grasslands, shrublands, taiga, and particularly, Arctic rock and ice were infrequently surveyed. Survey proportions were biased towards developed countries with high numbers of research institutions. Biases towards research institution density has been observed for animal and plant biodiversity databases and sampling campaigns (Tittley *et al.*, 2017; Trimble & van Aarde, 2012; Wilson *et al.*, 2016). Conservation and biodiversity research goals partially are to generate complete coverage of threatened populations, many of which are non-uniformly distributed and are under-surveyed due to species prevalence in developing countries. Although we do not equate conservation issues and microbes to threatened and endangered macrobial species, we do assert that freshwater bacterial community sampling coverage needs to expand to understand biogeographical patterns and processes in freshwater ecosystems (Dudgeon *et al.*, 2006), particularly among human-altered landscapes, which will be highly impacted by climate change (Wenger *et al.*, 2009).

The CGI metric for all freshwaters indicated there were few gaps in climatic diversity globally. Ecosystem types that had fewer surveys (primarily wetlands and secondarily lakes) had greater climatic gaps. Surprisingly, surveys have somewhat lower CGI (greater climatic coverage) under low disturbance versus high disturbance. A strong pattern exists with climate and taxonomic richness across geographic regions for most taxonomic groups

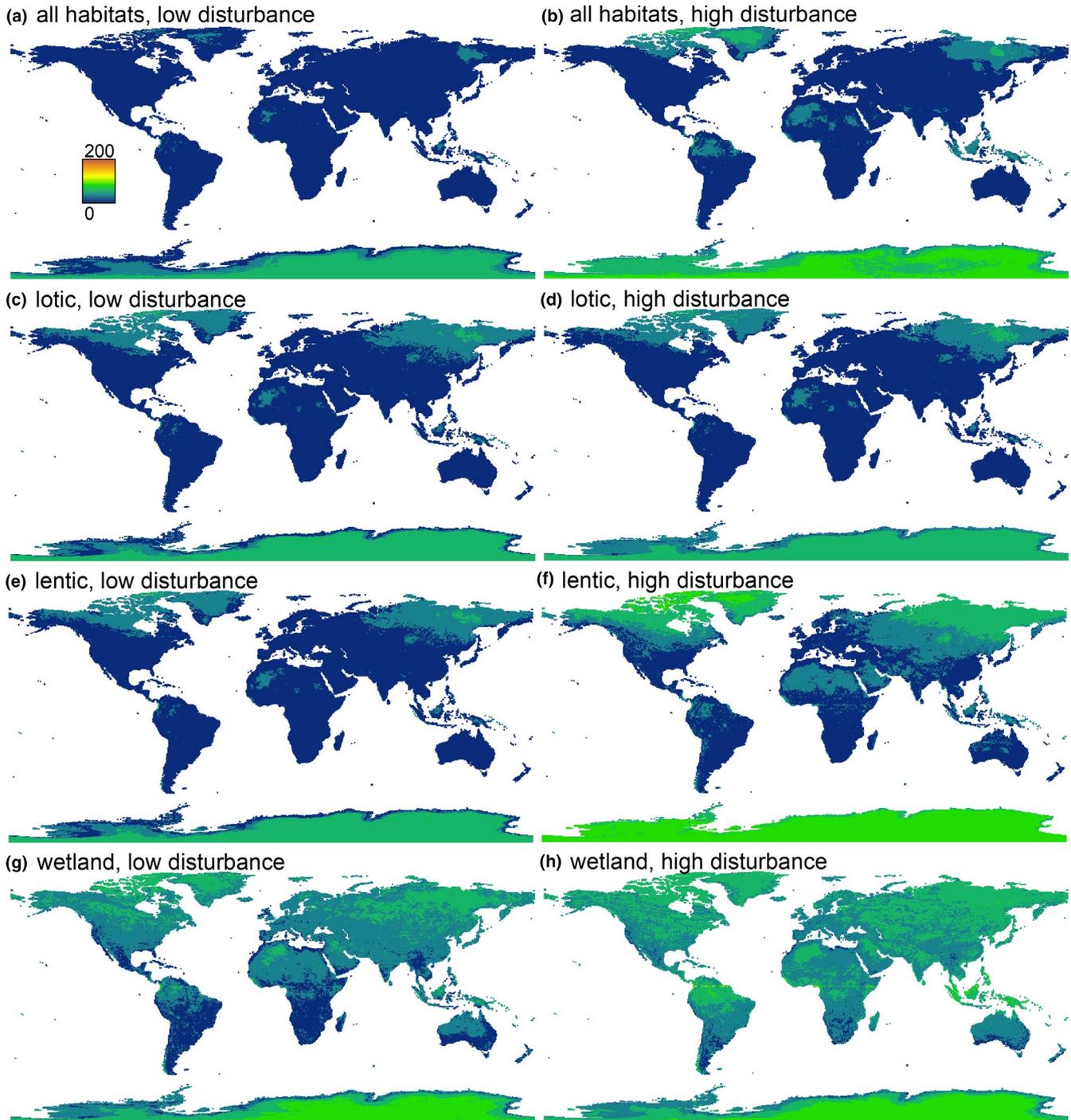


FIGURE 5 Geographic distribution of the climatic gap index (CGI) of freshwater microbial diversity for all (a and b), lotic (c and d), lentic (e and f), and wetland (g and h) surveys in areas of low (a,c,e,g) versus high (b,d,f,h) anthropogenic landscape alteration. Scales range between 0 and 200 for ease of viewing but CGI metric is between 9 and 190. Greater CGI values represent greater climatic gaps (i.e. greater climatic dissimilarity) and low CGI values represent low climatic gaps

partially due to variation in edaphic conditions, such as water availability and temperature, but also due to physiographic barriers (Currie et al., 2004; Hawkins et al., 2003; Liu et al., 2018). Microbial communities in soils (Angel et al., 2010; Fierer & Jackson, 2006; Martiny et al., 2006) and marine waters (Ghiglione et al., 2012; Pommier et al., 2006) have demonstrated distance–decay relationships and associations with edaphic variables (e.g.

pH, soil type) and suggest environmental selection or dispersal as mechanisms for biogeographic patterns. Discerning microbial survey completeness in association with climatic variables, including not just mean annual precipitation and temperature, but also seasonality is likely to be essential for understanding spatial overlap and turnover in microbial diversity. Our results imply that although surveys are concentrated in developed countries and near

research institutions, much of the climatic diversity globally for all freshwaters has been well surveyed, although wetlands do have gaps within mid-latitude, tropical regions indicative that future sampling should concentrate more so on these locales.

Other studies have provided diversity coverage censuses based on sequencing efforts (Schloss et al., 2016) or projected estimates of global microbial diversity (Locey & Lennon, 2016) and found the vast amount of diversity has probably been undocumented, particularly for aquatic habitats (Schloss et al., 2016). Our data demonstrate that, based on sampling effort spanning c. 20 years, freshwater studies have under-surveyed several biogeographic realms and biomes that harbour highly diverse animal and plant taxa, but that climatic diversity is fairly well represented among surveyed regions. The CGI metric is dependent on whether a specific climate has been surveyed at all, and not the *number* of surveys within that unique climate. We recommend additional future studies strive to address microbial ecological research questions in under-surveyed biomes and freshwater ecoregions that were poorly represented, such as tropical and arctic habitats.

Our assessment highlights key geographic and ecoregional gaps that should be targeted for microbial diversity surveying. These priority regions include Africa, South America, central Europe, Australia, and Antarctica. Although we present a comprehensive and global evaluation of *survey distribution*, we stop short of describing microbial *community composition* at a global scale. This latter objective has been limited by the scientific community's inability to standardise and broadly disseminate sequence libraries developed from many investigators using different community fingerprinting techniques and sequencing platforms over the last 2 decades. Access to such a global database of microbial community composition would facilitate the study of broad scale processes driving global patterns in richness and turnover (although certain large-scale initiatives have been implemented—the Earth Microbiome Project (Gilbert & Knight, 2014)). Moreover, such macro-ecological studies would help to refine priority regions by clarifying the strength of within-region environmental filtering versus among-region dispersal limitation. Specifically, if contemporary climate is the overriding filter of microbial taxa, then it may not be essential to survey all remaining unsurveyed freshwater ecoregions (279 ecoregions) as these are likely to have similar microbial taxa to already-surveyed ecoregions with similar climates. Alternatively, if freshwater microbial communities' exhibit turnover associated with drainage divides, then it will be important to survey geographically disparate ecoregions because these are likely to harbour undescribed microbial diversity associated with these regions' unique faunas or evolutionary histories.

ACKNOWLEDGMENTS

This research was sponsored in part by the Office of Biological and Environmental Research, Office of Science, U.S. Department of Energy (DOE) as part of the Mercury Science Focus Area at Oak Ridge National Laboratory, which is managed by UT-Battelle LLC for the DOE under contract DE-AC05-00OR22725.

DATA AVAILABILITY STATEMENT

The meta-analysis database and R code for the CGI metric and other results are available on Zenodo (<https://doi.org/10.5281/zenodo.4905176>). Metadata and information regarding the 963 individual studies is available to readers upon request.

ORCID

Allison M. Veach  <https://orcid.org/0000-0001-9845-3558>

REFERENCES

- Abell, R., Thieme, M. L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., ... Petry, P. (2008). Freshwater ecoregions of the world: A new map of biogeographic units for freshwater biodiversity conservation. *BioScience*, 58, 403–414. <https://doi.org/10.1641/B580507>
- Angel, R., Soares, M. I. M., Ungar, E. D., & Gillor, O. (2010). Biogeography of soil archaea and bacteria along a steep precipitation gradient. *The ISME Journal*, 4, 553–563. <https://doi.org/10.1038/ismej.2009.136>
- Beck, J., Ballesteros-Mejia, L., Nagel, P., & Kitching, I. J. (2013). Online solutions and the 'Wallacean shortfall': What does GBIF contribute to our knowledge of species' ranges? *Diversity and Distributions*, 19, 1043–1050. <https://doi.org/10.1111/ddi.12083>
- Bini, L. M., Diniz-Filho, J. A. F., Rangel, T. F. L. V. B., Bastos, R. P., & Pinto, M. P. (2006). Challenging Wallacean and Linnean shortfalls: Knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distributions*, 12, 475–482. <https://doi.org/10.1111/j.1366-9516.2006.00286.x>
- Bryant, J. A., Lamanna, C., Morlon, H., Kerkhoff, A. J., Enquist, B. J., & Green, J. L. (2008). Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 11505–11511. <https://doi.org/10.1073/pnas.0801920105>
- Casamayor, E. (2017). *Towards a microbial conservation perspective in high Mountain Lakes. High Mountain conservation in a changing world* (pp. 157–180). Springer.
- Cavicchioli, R., Ripple, W. J., Timmis, K. N., Azam, F., Bakken, L. R., Baylis, M., ... Webster, N. S. (2019). Scientists' warning to humanity: Microorganisms and climate change. *Nature Reviews Microbiology*, 17, 569–586. <https://doi.org/10.1038/s41579-019-0222-5>
- Currie, D. J., Mittelbach, G. G., Cornell, H. V., Field, R., Guegan, J.-F., Hawkins, B. A., ... Turner, J. R. G. (2004). Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, 7, 1121–1134. <https://doi.org/10.1111/j.1461-0248.2004.00671.x>
- Davidar, P., Rajagopal, B., Mohandass, D., Puyravaud, J.-P., Condit, R., Wright, S. J., & Leigh, E. G. (2007). The effect of climatic gradients, topographic variation and species traits on the beta diversity of rain forest trees. *Global Ecology and Biogeography*, 16, 510–518. <https://doi.org/10.1111/j.1466-8238.2007.00307.x>
- Díaz, S., Fargione, J., Chapin, F. S., & Tilman, D. (2006). Biodiversity loss threatens human well-being. *PLoS Biology*, 4, e277. <https://doi.org/10.1371/journal.pbio.0040277>
- Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., ... Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, 81, 163. <https://doi.org/10.1017/S1464793105006950>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas: New climate surfaces for global land areas. *International Journal of Climatology*, 37, 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, 626–631. <https://doi.org/10.1073/pnas.0507535103>

- Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., ... Caporaso, J. G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 21390–21395. <https://doi.org/10.1073/pnas.1215210110>
- GBIF.org GBIF Home Page.
- Ghiglione, J.-F., Galand, P. E., Pommier, T., Pedros-Alio, C., Maas, E. W., Bakker, K., ... Murray, A. E. (2012). Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 17633–17638. <https://doi.org/10.1073/pnas.1208160109>
- Gilbert, J. A., & Knight, R. (2014). The Earth Microbiome project: Successes and aspirations. *BMC Biology*, *12*, 69. <https://doi.org/10.1186/s12915-014-0069-1>
- Graham, M. H., & Dayton, P. K. (2002). On the evolution of ecological ideas: Paradigms and scientific progress. *Ecology*, *83*(6), 1481–1489.
- Harrison, P. A., Berry, P. M., Simpson, G., Haslett, J. R., Blicharska, M., Bucur, M., ... Turkelboom, F. (2014). Linkages between biodiversity attributes and ecosystem services: A systematic review. *Ecosystem Services*, *9*, 191–203. <https://doi.org/10.1016/j.ecoser.2014.05.006>
- Hawkins, B. A., Porter, E. E., & Felizola Diniz-Filho, J. A. (2003). Productivity and history of predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology*, *84*, 1608–1623.
- Ibarbalz, F. M., Henry, N., Brandão, M. C., Martini, S., Busseni, G., Byrne, H., Coelho, L. P., ... Wincker, P. (2019). Global trends in marine plankton diversity across kingdoms of life. *Cell*, *179*, 1084–1097.e21. <https://doi.org/10.1016/j.cell.2019.10.008>
- Jetz, W., McPherson, J. M., & Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: Toward a global map of life. *Trends in Ecology & Evolution*, *27*, 151–159. <https://doi.org/10.1016/j.tree.2011.09.007>
- Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., ... Wisser, D. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*, 494–502. <https://doi.org/10.1890/100125>
- Liu, C., Dudley, K. L., Xu, Z.-H., & Economo, E. P. (2018). Mountain metacommunities: Climate and spatial connectivity shape ant diversity in a complex landscape. *Ecography*, *41*, 101–112. <https://doi.org/10.1111/ecog.03067>
- Liu, X., Zhang, L., & Hong, S. (2011). Global biodiversity research during 1900–2009: A bibliometric analysis. *Biodiversity and Conservation*, *20*, 807–826. <https://doi.org/10.1007/s10531-010-9981-z>
- Locey, K. J., & Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 5970–5975. <https://doi.org/10.1073/pnas.1521291113>
- Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., ... Staley, J. T. (2006). Microbial biogeography: Putting microorganisms on the map. *Nature Reviews Microbiology*, *4*, 102–112. <https://doi.org/10.1038/nrmicro1341>
- Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, *6*. <https://doi.org/10.1038/ncomms9221>
- Monard, C., Gantner, S., Bertilsson, S., Hallin, S., & Stenlid, J. (2016). Habitat generalists and specialists in microbial communities across a terrestrial-freshwater gradient. *Scientific Reports*, *6*. <https://doi.org/10.1038/srep37719>
- Nemergut, D. R., Costello, E. K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S. K., ... Knight, R. (2011). Global patterns in the biogeography of bacterial taxa: Global bacterial biogeography. *Environmental Microbiology*, *13*, 135–144. <https://doi.org/10.1111/j.1462-2920.2010.02315.x>
- Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., ... Ferrenberg, S. (2013). Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews*, *77*, 342–356. <https://doi.org/10.1128/MMBR.00051-12>
- Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., & Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. *Microbiology and Molecular Biology Reviews*, *75*, 14–49. <https://doi.org/10.1128/MMBR.00028-10>
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., ... Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on earth. *BioScience*, *51*, 933.
- Padial, A. A., Ceschin, F., Declerck, S. A. J., De Meester, L., Bonecker, C. C., Lansac-Tôha, F. A., ... Bini, L. M. (2014). Dispersal ability determines the role of environmental, spatial and temporal drivers of metacommunity structure. *PLoS One*, *9*, e111227. <https://doi.org/10.1371/journal.pone.0111227>
- Pelayo-Villamil, P., Guisande, C., Manjarrés-Hernández, A., Jiménez, L. F., Granado-Lorencio, C., García-Roselló, E., ... Lobo, J. M. (2018). Completeness of national freshwater fish species inventories around the world. *Biodiversity and Conservation*, *27*, 3807–3817. <https://doi.org/10.1007/s10531-018-1630-y>
- Pommier, T., Canbäck, B., Riemann, L., Boström, K. H., Simu, K., Lundberg, P., ... Hagström, Å. (2006). Global patterns of diversity and community structure in marine bacterioplankton: Marine bacteria diversity and distribution. *Molecular Ecology*, *16*, 867–880. <https://doi.org/10.1111/j.1365-294X.2006.03189.x>
- Reyers, B., Polasky, S., Tallis, H., Mooney, H. A., & Larigauderie, A. (2012). Finding common ground for biodiversity and ecosystem services. *BioScience*, *62*, 503–507.
- Schloss, P. D., Girard, R. A., Martin, T., Edwards, J., & Thrash, J. C. (2016). Status of the archaeal and bacterial census: An update. *mBio*, *7*(3), e00201-16. <https://doi.org/10.1128/mBio.00201-16>
- Sousa-Baena, M. S., Garcia, L. C., & Peterson, A. T. (2014). Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions*, *20*, 369–381. <https://doi.org/10.1111/ddi.12136>
- Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., ... Zhao, H. (2017). A communal catalogue reveals Earth's multi-scale microbial diversity. *Nature*, *551*, 457–463.
- Titley, M. A., Snaddon, J. L., & Turner, E. C. (2017). Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. *PLoS One*, *12*, e0189577. <https://doi.org/10.1371/journal.pone.0189577>
- Trimble, M. J., & van Aarde, R. J. (2012). Geographical and taxonomic biases in research on biodiversity in human-modified landscapes. *Ecosphere*, *3*, art119. <https://doi.org/10.1890/ES12-00299.1>
- Troia, M. J., & McManamay, R. A. (2016). Filling in the GAPS: Evaluating completeness and coverage of open-access biodiversity databases in the United States. *Ecology and Evolution*, *6*, 4654–4669.
- Tydecks, L., Jeschke, J. M., Wolf, M., Singer, G., & Tockner, K. (2018). Spatial and topical imbalances in biodiversity research. *PLoS one*, *13*(7), e0199327.
- Vega, G., Pertierra, L. R., & Olalla-Tárraga, M. Á. (2017). MERRAclim, a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling. *Scientific Data*, *4*, 170078. <https://doi.org/10.1038/sdata.2017.78>
- Venter, O., Sanderson, E. W., Magrath, A., Allan, J. R., Beher, J., Jones, K. R., ... Watson, J. E. (2018). Last of the Wild Project, Version 3 (LWP-3): 2009 Human Footprint, 2018 Release.
- Wallace, A. (1876). *The geographical distribution of animals*. Haper & Brothers.
- Wenger, S. J., Roy, A. H., Jackson, C. R., Bernhardt, E. S., Carter, T. L., Filoso, S., ... Walsh, C. J. (2009). Twenty-six key research questions in urban stream ecology: An assessment of the state of the science. *Journal of the North American Benthological Society*, *28*, 1080–1098. <https://doi.org/10.1899/08-186.1>

- Wheeler, Q. D., Knapp, S., Stevenson, D. W., Stevenson, J., Blum, S. D., Boom, B. M., ... Woolley, J. B. (2012). Mapping the biosphere: Exploring species to understand the origin, organization and sustainability of biodiversity. *Systematics and Biodiversity*, 10, 1–20. <https://doi.org/10.1080/14772000.2012.665095>
- Wilson, K. A., Auerbach, N. A., Sam, K., Magini, A. G., Moss, A., St, L., Langhans, S. D., Budiharta, S., Terzano, D., & Meijaard, E. (2016). Conservation research is not happening where it is most needed. *PLOS Biology*, 14, e1002413. <https://doi.org/10.1371/journal.pbio.1002413>
- Woodward, G., Perkins, D. M., & Brown, L. E. (2010). Climate change and freshwater ecosystems: Impacts across multiple levels of organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 2093–2106.
- Yang, W. (2011). Developing open access in conservation research. *Journal of Soil and Water Conservation*, 66, 6A–8A. <https://doi.org/10.2489/jswc.66.1.6A>
- Zeglin, L. H. (2015). Stream microbial diversity in response to environmental changes: Review and synthesis of existing research. *Frontiers in Microbiology*, 6. <https://doi.org/10.3389/fmicb.2015.00454>

SUPPORTING INFORMATION

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

How to cite this article: Veach AM, Troia MJ, Cregger MA. Assessing biogeographic survey gaps in bacterial diversity knowledge: A global synthesis of freshwaters. *Freshw Biol.* 2021;00:1–11. <https://doi.org/10.1111/fwb.13777>