



Microbial Community Composition in Municipal Wastewater Treatment Bioreactors Follows a Distance Decay Pattern Primarily Controlled by Environmental Heterogeneity

Taeyu Kim,^a  Sebastian Behrens,^{a,b}  Timothy M. LaPara^{a,b}

^aDepartment of Civil, Environmental, and Geo-Engineering, University of Minnesota Twin Cities, Minneapolis, Minnesota, USA

^bBiotechnology Institute, University of Minnesota Twin Cities, St. Paul, Minnesota, USA

ABSTRACT Understanding spatiotemporal patterns in microbial community composition is a central goal of microbial ecology. The objective of this study was to better understand the biogeography of activated sludge microbial communities, which are important for the protection of surface water quality. Monthly samples were collected from 20 facilities (25 bioreactors) within 442 km of each other for 1 year. Microbial community composition was characterized by sequencing of PCR-amplified 16S rRNA gene fragments. Statistically significant distance decay of community similarity was observed in these bioreactors independent of clustering method (operational taxonomic units [OTUs] at 97% similarity, genus-level phylotypes) and community dissimilarity metric (Sørensen, Bray-Curtis, and weighted Unifrac). Universal colonizers (i.e., detected in all samples) and ubiquitous genus-level phylotypes (i.e., detected in every facility at least once) also exhibited a significant distance decay relationship. Variation partitioning analysis of community composition showed that environmental characteristics (temperature, influent characteristics, etc.) explained more of the variance in community composition than geographic distance did, suggesting that environmental heterogeneity is more important than dispersal limitation as a mechanism for determining microbial community composition. Distance decay relationships also became stronger with increasing distance between facilities. Seasonal variation in community composition was also observed from selected bioreactors, but there was no clear seasonal pattern in the distance decay relationships.

IMPORTANCE Understanding the spatiotemporal patterns of biodiversity is a central goal of ecology. The distance decay of community similarity is one of the spatial scaling patterns observed in many forms of life, including plants, animals, and microbial communities. Municipal wastewater treatment relies on microorganisms to prevent the release of excessive quantities of nutrients and other pollutants, but relatively few studies have explored distance decay relationships in wastewater treatment bioreactors. Our results demonstrate a strong distance decay pattern in wastewater treatment bioreactors, regardless of the sequence clustering method or the community dissimilarity metric. Our results suggest that microbial communities in wastewater treatment bioreactors are not randomly assembled but rather exhibit a statistically significant spatial pattern.

KEYWORDS wastewater treatment, activated sludge, distance decay, microbial biogeography pattern, microbiome

Distance decay relationships are well-known, fundamental spatial patterns of biodiversity and have been recognized by ecologists for several decades (1–3). The bulk of our knowledge on distance decay relationships is based on plants (2, 4) and animals (1, 4), but the spatial distribution of microbes has become of interest. Recent empirical analyses of some patterns for microbial communities suggest that there are biodiversity

Citation Kim T, Behrens S, LaPara TM. 2021. Microbial community composition in municipal wastewater treatment bioreactors follows a distance decay pattern primarily controlled by environmental heterogeneity. *mSphere* 6:e00648-21. <https://doi.org/10.1128/mSphere.00648-21>.

Editor Michael J. Imperiale, University of Michigan—Ann Arbor

Copyright © 2021 Kim et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Timothy M. LaPara, lapar001@umn.edu.

Received 23 July 2021

Accepted 11 October 2021

Published 20 October 2021

scaling rules common to all forms of life (5); in fact, recent studies indicate that microbes also have a substantial distance decay pattern (5–12).

Two different mechanisms are used to explain the distance decay of community similarity: environmental heterogeneity and dispersal limitation (2, 13). As geographic distance increases, environmental conditions tend to be more different but also the dispersal limitation becomes stronger. Previous studies have attempted to determine which of these two interconnected variables is the main driver of microbial community distance decay patterns (environmental heterogeneity versus geographic distance) using experimentally manipulated enrichment cultures (14) and variation partitioning statistical analyses of natural ecosystems (4, 15). These studies showed that the distance decay of microbial community composition was caused more by environmental heterogeneity than by limited dispersal.

Wastewater treatment bioreactors support one of the most complex microbial ecosystems ever applied for a specific purpose (16). Many efforts have been made to understand the microbial communities in these bioreactors, and understanding their spatial patterning and underlying mechanisms controlling microbial community composition is emerging. Recently, a weak distance decay relationship was observed with microbial communities from 26 wastewater treatment bioreactors (as far as ~3,000 km between two bioreactors) (17); another study observed a statistically significant distance decay relationship in wastewater bioreactor communities at a global scale (18).

The objective of this study was to characterize the distance decay relationships among the complex bacterial communities growing in municipal wastewater treatment bioreactors at relatively short distances (<500 km). Because prior research has demonstrated that bioreactor community assembly is largely deterministic (19, 20), we hypothesize that wastewater treatment bioreactors would exhibit a strong distance decay relationship, particularly at shorter distances than has been previously studied. To test this hypothesis, samples from 20 full-scale wastewater treatment facilities from within the state of Minnesota were collected each month for a year. Microbial community composition was determined by sequencing of PCR-amplified 16S rRNA gene fragments. Community dissimilarity, richness, and specific genus-level phylotype abundances were then studied along with bioreactor characteristics and geographic locations. The spatial community shift rate (β) was calculated for the whole community and the *Archaea*. The β was also calculated only considering universal and ubiquitous genus-level phylotypes to ascertain distance decay relationships excluding the effects of limited dispersal (i.e., if these microbes were detected at all locations, then dispersion was not a pertinent factor).

RESULTS

Community richness and dissimilarity. Samples were collected monthly from 20 different wastewater treatment facilities (25 bioreactors) for 1 year. Annual average operating parameters and facility characteristics are summarized in Tables S1 and S2 in the supplemental material, respectively. Distances between the facilities ranged from 11 km to 442 km; the relative distances among the facilities are depicted in Fig. S1 in the supplemental material. A total of 292 activated sludge samples were collected, and their extracted DNA was used as the template for sequence analysis of PCR-amplified 16S rRNA gene fragments. An average of 24,419 (standard deviation [SD] = 9,626) quality sequences were obtained per sample. Three samples that had less than 4,336 quality sequences were discarded prior to operational taxonomic unit (OTU)-based analysis; 8 samples that had less than 7,805 quality sequences were discarded prior to analysis of genus-level phylotypes. This analysis generated an average of 2,030 OTUs per sample (SD = 302) and 262 genus-level phylotypes per sample (SD = 25).

The diversity of bacterial communities growing in these wastewater treatment bioreactors varied as a function of several key parameters (Table 1). The number of observed OTUs had a statistically significant ($P = 1.2 \times 10^{-11}$) positive correlation with the average flow rate (~facility size), although the number of observed genus-level

TABLE 1 Spearman's correlation coefficients between environmental variables and bioreactor community alpha-diversity indices^a

Parameter or type of water	Environmental variable ^b	No. of observed OTUs	OTU Shannon diversity index	No. of observed phylotypes	Phylotype Shannon diversity index
Community richness	No. of observed OTUs		0.97*		0.25*
	No. of observed phylotypes	0.47*	0.47*	0.47*	0.61*
Operational parameter	Avg flow rate	0.39*	0.34*	0.04	-0.14
	pH	-0.22	-0.19	-0.16	-0.03
	Mean cell residence time	0.37*	0.35*	0.18	0.03
Untreated wastewater	CBOD	-0.32*	-0.32*	-0.19	-0.11
	TSS	0.11	0.12	-0.11	-0.15
	Phosphorus	-0.09	-0.08	-0.02	-0.14
	TKN	-0.02	-0.06	0.00	-0.04
Treated wastewater	pH	0.09	0.15	0.19	0.08
	CBOD	0.05	0.01	0.10	0.22
	TSS	-0.07	-0.07	-0.07	-0.06
	Phosphorus	-0.07	-0.04	-0.01	0.08
	NH ₃	-0.19	-0.20	0.01	0.11
	NO ₃ + NO ₂	0.17	0.16	-0.11	-0.15
	TKN	-0.33	-0.35*	0.01	0.12

^aAn asterisk indicates a statistically significant correlation ($P < 0.001$).

^bOTU, operational taxonomic unit; CBOD, carbonaceous biochemical oxygen demand; TSS, total suspended solid; TKN, total Kjeldahl nitrogen.

phylotypes did not correlate with the average flow rate ($P = 0.54$) (Fig. 1). A statistically significant taxon-volume relationship was shown when the number of observed OTUs and the facility size were correlated in log scale ($z = 0.03$); however, this relationship was not statistically significant when genus-level phylotype-based community profiles were considered (Fig. S2).

The number of observed OTUs also had a significant positive correlation with mean cell residence time and exhibited a significant negative correlation with carbonaceous biochemical oxygen demand (CBOD) in the untreated wastewater (Table 1). However, the correlation between the number of observed OTUs and these two variables became insignificant ($P = 0.004$ and 0.002 , respectively) when partial correlations were performed with average flow rate as a control variable (Table S3). The average flow rate did not lose statistical significance in the partial correlation performed by controlling other variables. When correlations between the number of observed OTUs and these two environmental variables were made within each facility, no significant corre-

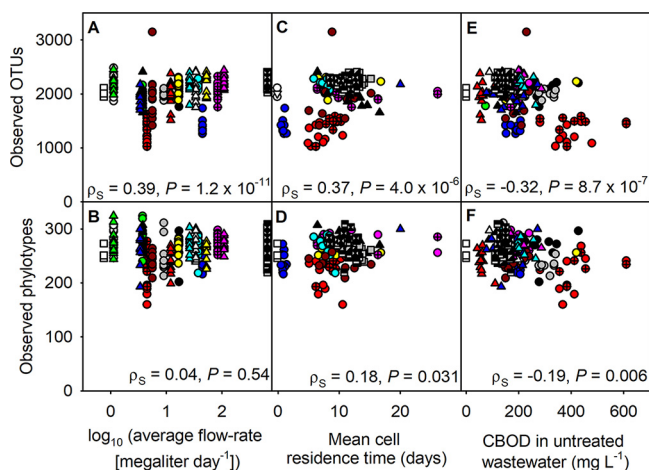


FIG 1 The number of observed operational taxonomic units (OTUs) and genus-level phylotypes as a function of wastewater flow rate (A and B), mean cell residence time (C and D), and carbonaceous biochemical oxygen demand (CBOD) in the untreated wastewater (E and F). Spearman's correlation coefficients (ρ_s) and P values are shown in each plot.

lation was observed, possibly because of the low sample number (typically 11 or 12 samples were collected at each facility) and/or the variance of environmental characteristics was not substantial within each facility. Variance of the mean cell residence time and of the CBOD were generally larger between the facilities than within each facility. A visual inspection of the correlation between community richness and these two variables (Fig. 1) suggests that the correlations were driven by differences among the facilities, not by temporal variation within each facility.

The bioreactors studied here generally did not show seasonal cyclic patterns in alpha-diversity; monthly community alpha-diversity did not correlate with monthly average temperatures except for possibly one facility. Facility L showed seasonal cyclic changes in all four alpha-diversity measurements (i.e., observed number of OTUs, observed number of genus-level phylotypes, OTU-based Shannon diversity index, and Shannon diversity index using genus-level phylotypes) with strong correlations with monthly average temperature (Spearman's rho [ρ_s] = 0.66, 0.76, 0.78, and 0.81, respectively).

Bray-Curtis dissimilarity indices were computed using the community profiles of genus-level phylotypes (Fig. S3). Analysis of molecular variance (AMOVA) showed that bacterial community composition from different facilities had statistically different community compositions (all $P < 0.001$), which suggests that facility-to-facility variation in bacterial community composition is bigger than the temporal variance within an individual facility. In contrast, bacterial communities in parallel bioreactors within the same facility were not statistically distinguishable (all $P > 0.001$).

Distance decay of the whole community similarity. There is a strong linear correlation between the \log_{10} -transformed annual average community dissimilarities (Sørensen, Bray-Curtis, and weighted Unifrac) and the \log_{10} -transformed geographic distances between paired facilities (Fig. 2). This suggests that community dissimilarity metrics based on presence/absence (Sørensen), relative abundance (Bray-Curtis), and phylogenetic distance (weighted Unifrac) exhibit distance decay relationships. The distance decay relationship calculated based on the OTU-based community profile had a lower β than the genus-level phylotype-based community profile by ~ 10 -fold, with both the Sørensen (0.011 versus 0.12) and Bray-Curtis (0.016 versus 0.12) metrics. The β calculated using phylogenetic distance (0.074) fell between the calculations based on OTUs and on genus-level phylotypes. Nonetheless, all sequence clustering methods exhibited statistically significant distance decay relationships ($P < 10^{-16}$). Therefore, the qualitative conclusion (i.e., significant distance decay of community similarity) did not depend on either the dissimilarity metric or how the sequences were clustered.

Distance decay of specific microbial groups. Highly abundant members of the microbial communities were generally ubiquitous and persistent, whereas rare members were typically sparse and transient (Fig. S4). Universal colonizers were defined as genus-level phylotypes that were detected in all bioreactor samples analyzed in this study. There were 22 universal colonizers; the average relative abundance of universal colonizers was 37.9% (SD = 5.3%) of the total community in each sample. Ubiquitous phylotypes were defined as genus-level phylotypes that were detected in every facility on at least one occasion. There were 181 ubiquitous phylotypes, comprising an average of 72.3% (SD = 6.2%) of the microbial community in each sample. Finally, *Archaea* were examined in more detail; 34 different archaeal phylotypes were detected (arithmetic mean = 0.31%, SD = 0.23% in each sample as the relative abundance). Community dissimilarity matrices were constructed separately using each of these groups of organisms.

All three of these groups of organisms (i.e., universal colonizers, ubiquitous phylotypes, and *Archaea*) exhibited statistically significant distance decay relationships (Fig. 3). The β of the universal colonizers and of the ubiquitous phylotypes were similar to the β of the whole community. When β was calculated for each month, the β of the universal colonizers (median = 0.13) and of the ubiquitous phylotypes (median = 0.12) were not significantly different than the β of the whole community (median = 0.12) (Wilcoxon $Z = 0.92$ and 0.32 and $P = 0.36$ and 0.75 , respectively). The *Archaea* had a smaller β (0.05) compared to the whole community; the statistical significance of the

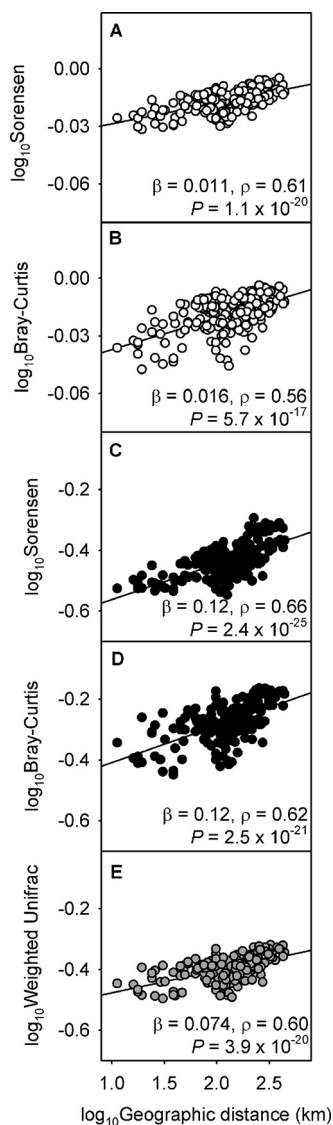


FIG 2 \log_{10} -transformed distance and annual average operational taxonomic unit (OTU)-based community dissimilarities when calculated by using Sørensen index (A) and Bray-Curtis index (B). \log_{10} -transformed distance and annual average community dissimilarities based on genus-level phylotypes when calculated by using Sørensen index (C) and Bray-Curtis index (D). (E) \log_{10} -transformed distance and annual average weighted Unifrac distance. The slope of each linear regression determines the spatial community shift rate (β). The linearity and the statistical significance of each model were shown using the Pearson's correlation coefficient (ρ) and P value, respectively.

distance decay relationship for the *Archaea* was weaker ($P = 1.9 \times 10^{-5}$) than for the whole community.

Environmental heterogeneity and geographic distance. Distance decay relationships can be driven by both changes in environmental conditions as well as dispersal. To better understand the relative importance of these two mechanisms, the annual average Bray-Curtis community dissimilarities between the facilities (community profiles based on genus-level phylotypes) were correlated with analogous distance matrices of facility characteristics such as geographic location, operational parameters, untreated wastewater characteristics, and treated wastewater characteristics (Table 2; raw data available in Table S1). In addition to the strong correlation between geographic location and Bray-Curtis community dissimilarities that has been previously described, the whole community dissimilarities also exhibited statistically significant correlations with total Kjeldahl nitrogen (TKN) concentrations in the treated and untreated wastewater

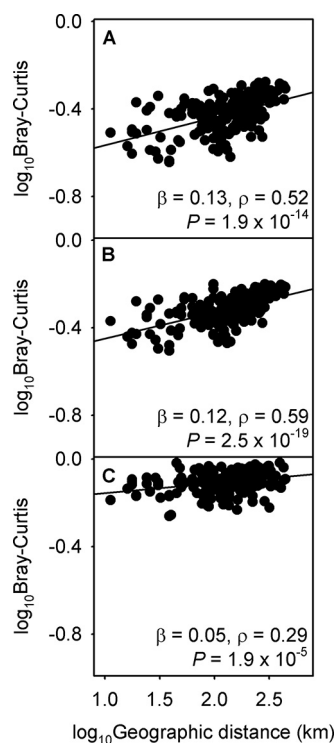


FIG 3 Log₁₀-transformed distance and annual average dissimilarities calculated using the Bray-Curtis index with various subcommunities: universal colonizers (A), ubiquitous phylotypes (B), and *Archaea* (C). The slope of each linear regression determines the spatial community shift rate (β). The linearity and the statistical significance of each model were shown using the Pearson's correlation coefficient (ρ) and P value, respectively.

as well as the ammonia concentrations and CBOD concentrations in the treated wastewater. The subgroups of universal colonizers and ubiquitous phylotypes generally exhibited the same correlation with the whole community, while the *Archaea* had a stronger correlation with mean cell residence time than with geographic distance.

TABLE 2 Spearman's correlation coefficients (from Mantel test) between annual average environmental characteristic differences and annual average Bray-Curtis community dissimilarities or geographic distances^a

Characteristic	Environmental variable ^b	Whole community dissimilarity	Universal colonizer dissimilarity	Ubiquitous phylotype dissimilarity	<i>Archaea</i> dissimilarity	Geographic distance
Geographic distance		0.62*	0.51*	0.61*	0.29*	
Operational parameter	Avg treatment	-0.04	0.01	-0.08	0.16	-0.11
	pH	-0.04	-0.09	-0.08	0.10	0.03
	Mean cell residence time	0.33	0.38	0.29	0.70*	0.36
Untreated wastewater characteristics	CBOD	0.18	0.19	0.19	0.10	0.15
	TSS	0.13	0.18	0.09	0.40*	0.17
	Phosphorus	0.27	0.24	0.26	0.12	0.19
	TKN	0.51*	0.47*	0.53*	0.11	0.58*
Treated wastewater characteristics	pH	0.04	-0.01	0.04	0.03	0.04
	CBOD	0.37*	0.26	0.34*	0.22	0.22
	TSS	0.19	0.24	0.21	0.17	0.16
	Phosphorus	0.24	0.22	0.31*	-0.16	0.06
	NH ₃	0.55*	0.47*	0.56*	0.13	0.43*
	NO ₃ + NO ₂	0.25	0.10	0.17	0.16	0.27
	TKN	0.64*	0.58*	0.66*	0.15	0.62*

^aThe arithmetic mean value was used for facilities with multiple bioreactors. The amount of data varied by facility (details in Table S1 in the supplemental material), such that P can vary even if the correlation coefficient is the same. An asterisk indicates statistically significant correlation ($P < 0.001$).

^bCBOD, carbonaceous biochemical oxygen demand; TSS, total suspended solids; TKN, total Kjeldahl nitrogen.

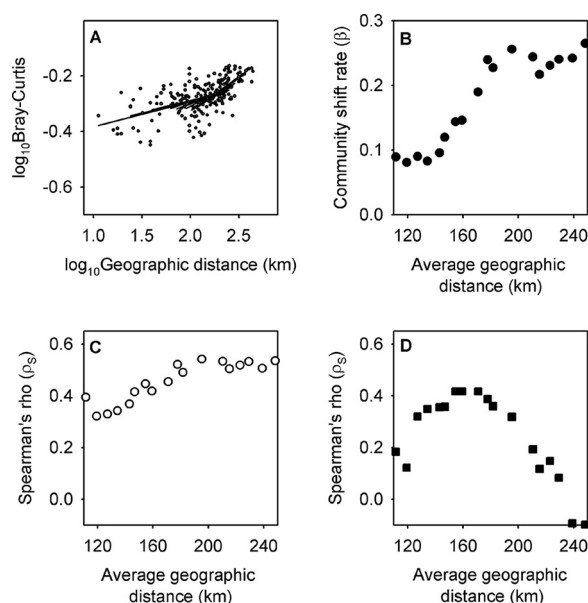


FIG 4 (A) Changes in spatial community shift rate (β) investigated by a sliding spatial scale ($P < 0.001$). (B) Changes in β versus average geographic distance between the facilities (i.e., the slope of the lines shown in panel A), (C) Spearman's correlation coefficient (ρ_s) between the community dissimilarity and geographic distance, and (D) ρ_s between the community dissimilarity and differences in total Kjeldahl nitrogen concentrations (TKN) in the untreated wastewater. All four analyses used the geographic distance between facility pairs on a sliding scale from 10 to 210 km to 180 to 380 km, in 10-km increments, and community dissimilarity was calculated using community profiles based on genus-level phylotypes.

Three of the environmental parameters (i.e., TKN concentrations in the treated and untreated wastewater and ammonia concentrations in the treated wastewater), which were strongly correlated with the whole community composition, also significantly correlated with geographic distance (Table 2). Partial canonical correspondence analysis (CCA) was therefore performed to separate the importance of geographic distance from facility characteristics to the whole community composition. For this analysis, only 11 of the 20 facilities participating in this study were analyzed because only these facilities provided a full set of metadata. Geographic distance explained 22.4% of the variance in microbial community composition, when environmental factors were controlled. In contrast, three environmental parameters explained 35.6% of the variance when geographic distance was controlled. Spatially structured environmental variation (i.e., the combination of geographic distance and environmental factors) explained 76.0% of the variance, which leaves 24.0% of the variance unexplained.

Impacts of geographic scale and site selection. Because the previous section focused on only 11 of the 20 facilities participating in this investigation (i.e., those facilities that had a full set of metadata), the effect of spatial scale on the distance decay relationships was evaluated further using community profiles based on genus-level phylotypes. That is, the partial CCA could have been biased by the specific facilities incorporated into that analysis. Specifically, seven of the facilities were located in a single metropolitan region and were of close proximity (11 km to ~62 km; average, 30 km).

This study included 20 different wastewater treatment facilities, allowing 190 pairwise calculations of microbial community dissimilarities. Using a sliding window of facility pairs ranked from shortest to longest geographic distance (10-km increment each time from 10- to ~210-km scale to 180- to ~380-km scale), it was observed that β generally increased as the geographic scale of the study became larger (Fig. 4A and B), while all the distance decay relationships constructed within each geographic scale were statistically significant (all $P < 0.001$).

Previously, it was observed that bioreactor community dissimilarity was well correlated with both geographic distance and environmental characteristics such as TKN in the untreated wastewater. The effect of geographic-scale changes on these correlations were studied as well. The Spearman's correlation coefficient between the bioreactor community dissimilarities and geographic distance also increased as geographic scale became bigger (Fig. 4C). In contrast, the correlation between bioreactor community dissimilarity and TKN in the untreated wastewater did not become stronger as geographic scale became bigger (Fig. 4D). This suggests that the partial CCA results reported in the previous section (i.e., bioreactor community dissimilarity was better explained by environmental heterogeneity than geographic distance) could have been biased by the specific wastewater facilities incorporated into that analysis and might not reflect the results if all 20 facilities had been analyzed.

Seasonal changes in microbial community composition. We hypothesized that seasonal changes might also impact distance decay relationships. To test seasonal impacts on the microbial community composition in bioreactors (e.g., cyclic changes in beta-diversity), community composition from 24 different bioreactors (19 facilities; facility T was excluded from this analysis due to low sample numbers) was investigated as a function of the time interval between sample collection events using genus-level phylotype community profiles (Fig. 5 and Table S4). About half of the bioreactors (13 out of 24) had a statistically significant fit to a quadratic model with an axis of symmetry between 5 and 8 months, suggesting that these bioreactors had cyclic seasonal changes in microbial community composition. In contrast, about half of the bioreactors (15 out of 24) exhibited a statistically significant fit to a linear model, and two bioreactors did not exhibit a statistically significant fit to either the linear model or the quadratic model. Six of the bioreactors simultaneously had a statistically significant fit to both the quadratic and linear models.

Similarly, we also tested the hypothesis that distance decay relationships would be less pronounced in the winter than during the summer. When the distance decay relationships were constructed using the Bray-Curtis dissimilarities from the entire microbial community, distance decay relationships were significant ($P < 0.001$) every month (similar results were obtained for both Sørensen and weighted Unifrac indices) (Table S5), and no clear seasonal trend in β was observed (Fig. S5). Similarly, the universal colonizers and ubiquitous phylotypes also showed significant distance decay relationships for 9 and 10 months, respectively. In contrast, the *Archaea* exhibited a significant distance decay relationship during only 2 months (August 2017 and January 2018).

DISCUSSION

Understanding the spatiotemporal patterns of biodiversity is a central goal of ecology. The distance decay of community similarity is one of the spatial scaling patterns observed in many forms of life, including plants (2, 4), animals (1, 4), and microbial communities (4, 9–11). However, relatively few studies have explored distance decay relationships in municipal wastewater treatment bioreactors (17, 18). Municipal wastewater treatment is critically important for protecting surface water quality; it relies on microorganisms to reduce the release of excessive nutrients (biodegradable organic carbon, nitrogen, and phosphorus) as well as other priority pollutants to the environment. Within the studied geographic scale (<442 km), a significant distance decay pattern was observed among microbial communities in wastewater treatment bioreactors. These distance decay relationships were also robust, exhibiting statistical significance regardless of the sequence clustering method or the community dissimilarity metric. Our results strongly suggest that microbial communities in wastewater treatment bioreactors are not randomly assembled but rather exhibit a spatial pattern. This novel understanding of the mechanisms of microbial community assembly in wastewater treatment bioreactors builds on prior work that demonstrated synchrony in wastewater treatment bioreactor communities in close proximity to each other (19) and weak distance decay relationships over much larger spatial scales (~3,000 km) (17).

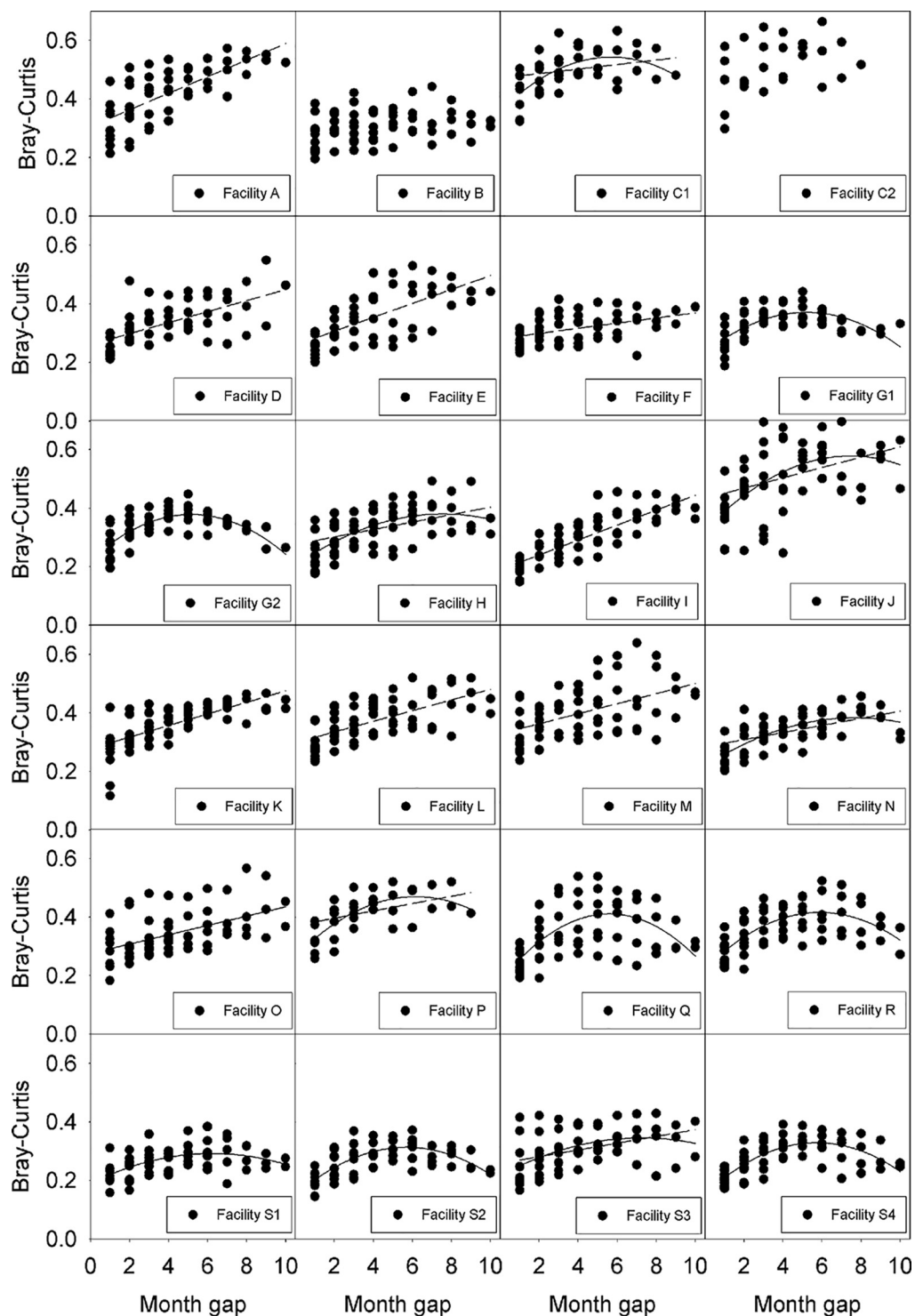


FIG 5 Changes in community dissimilarity as a function of the time difference between two sampling events observed in 24 activated sludge bioreactors from 19 facilities. Dashed lines indicate a statistically significant linear change in dissimilarity over time. Solid lines indicate a statistically significant fit to a quadratic regression model with an axis of symmetry between 5 and 8 months. Facility T is not shown due to a low number of samples. Community dissimilarity was calculated using community profiles using genus-level phylotypes.

The distance decay of community composition can be explained by two different factors: gradients in environmental conditions and dispersal limitation. Our research makes a novel contribution by demonstrating that environmental heterogeneity is a more important driver of the observed distance decay relationships than dispersal limitation in

wastewater treatment bioreactors. Distinguishing these two mechanisms is challenging because differences in environmental conditions are generally confounded by geographic distance (21). Universal colonizers (detected in all bioreactors during all sampling times) and ubiquitous phylotypes (detected in all facilities on at least one occasion) were studied in more detail because we hypothesized that these organisms were unaffected by dispersal limitation. Both subgroups exhibited strong distance decay relationships when calculated using Bray-Curtis dissimilarity, demonstrating that environmental gradients were sufficient to produce a statistically significant distance decay relationship. Furthermore, the β values for both subgroups were approximately the same as the β value for the entire microbial community, suggesting that environmental gradients were the dominant factor for the observed distance decay relationship of microbial communities in wastewater treatment bioreactors. Previous researchers concluded that environmental filtering is a more important driver of biogeographic patterns at smaller spatial scales and dispersal limitation is important at large-scale spatial patterns (11, 22); it is possible, therefore, that the distances studied herein (<442 km) are insufficient to observe a stronger dispersal limitation. That is, at least for this distance scale, “everything is more or less everywhere, but the environment selects” (23–25).

We hypothesized that the distance decay relationship of the archaeal community would largely be driven by dispersal rather than an active response to environmental conditions. Previous research has demonstrated that only a small amount of methanogenic activity occurs in full-scale activated sludge bioreactors, resulting in only a minor fraction of the overall carbon cycling (26). The majority of the *Archaea* detected in our study were methanogens that are well-known to be obligate anaerobes; ammonia-oxidizing archaea were not detected. Although the archaeal community exhibited a statistically significant distance decay relationship, the β for the archaeal community was much smaller than the β for universal colonizers and ubiquitous phylotypes. This suggests that the effect of the dispersal ($\beta = 0.05$) is substantially smaller than environmental heterogeneity ($\beta = 0.12$ to 0.13) in the distance decay relationships at this scale.

Partial CCA also suggested that environmental heterogeneity was a more important factor in the observed distance decay relationships than dispersal limitation. This is consistent with previous researchers who concluded that bacterial community assembly in wastewater treatment bioreactors is controlled by deterministic (i.e., niche-based community assembly) rather than stochastic mechanisms (19). However, quantitative correlations can change based on geographic scale and/or site selection. In this study, bioreactor community dissimilarity was more strongly correlated with geographic distance as geographic distance increased. In contrast, the correlation between community dissimilarity and the TKN in the untreated wastewater exhibited a hump-shaped profile as geographic distance increased. That is, there is an incongruity between the mathematical importance of TKN in the untreated wastewater, which had a strong Spearman’s correlation with community dissimilarity, and bioreactor community dissimilarity as distance increases. Admittedly, our analysis is implicitly limited by the number of environmental parameters considered; numerous other deterministic factors (e.g., concentrations of specific chemicals) were not considered in our CCA.

Our experimental design (20 facilities; 12 months) afforded us the opportunity to observe temporal changes in community richness and temporal changes in distance decay relationships. Previous studies have found that bioreactors showed seasonal cyclic patterns in alpha-diversity, with higher diversity (or richness) during warmer seasons and lower diversity during colder seasons (19, 27). However, cyclic fluctuations in community alpha-diversity driven by seasonal changes were not observed and alpha-diversity did not correlate with the monthly average temperature in our study. Similarly, the distance decay relationships for the entire microbial community, the universal colonizers, and the ubiquitous phylotypes were statistically significant as a function of time. This suggests that the distance decay relationships observed herein are robust and consistent with deterministic mechanisms of community assembly.

Our experimental design also afforded us the opportunity to observe cyclical patterns in microbial community composition at numerous wastewater treatment facilities over the same time period. Previous studies have shown cyclic seasonal patterns in beta-diversity (19, 28, 29); that is, sample pairs collected from the same bioreactor showed increasing community dissimilarity with increasing time gaps for ~7 months but then subsequently showed decreasing dissimilarity (19, 28). However, these claims may have been spurious because of the relatively small number of facilities (<5) and the short durations (<14 months) of these studies. In our study, seasonal variations in community composition (i.e., a cyclic pattern in beta-diversity) was observed in about half of the bioreactors, similar to the number of bioreactors that exhibited a linear (i.e., nonreturning) pattern in community dissimilarity. This suggests that the cyclic community changes may be a site-specific or spurious observation. Additional research is needed for longer time periods to properly address this question.

A statistically significant taxon-volume relationship was found between the number of observed OTUs and the relative size of the treatment facilities based on annual average flow rate. This suggests that island biogeography theory, which predicts more species in larger ecosystems (30), may also apply to municipal wastewater treatment bioreactors. A previous study made a similar observation during the investigation of membrane bioreactors used for municipal wastewater (31). Both taxon-volume relationships and distance decay relationships are pertinent ecological theories that explain spatial biodiversity patterns; several studies have predicted that the taxon-area relationship exponent z would be equal to one-half the spatial community shift rate β , as calculated via the Sørensen index (6, 15, 17, 32). In our study, however, z and β did not correlate as predicted. Furthermore, genus-level phylotype-based community profiles did not exhibit a statistically significant taxon-volume relationship, whereas a strong distance decay relationship was observed. Further research is needed to reconcile the relationships between island biogeography and distance decay in controlling microbial community composition in municipal wastewater treatment bioreactors.

Although we report values for β , we caution against comparing these values with those from past and future studies. Specifically, biodiversity patterns vary systematically as a function of the taxonomic resolution used to describe microbial community composition (33). Even this conclusion is controversial because a previous study concluded that the OTU similarity cutoff did not impact β in soil samples (11), whereas another study found that narrowing the similarity cutoff increased z (estimated from β) in salt marsh sediments (15). In our study, narrowing the similarity cutoff (97% sequences base similarity [OTU] versus genus-level phylotypes) effectively reduced β in wastewater treatment bioreactors. In addition, removing rare sequences (or low-abundance OTUs) affected β (34) as did changing sequencing depth (11). Previous studies suggested that different spatial scales also can affect β (1, 18, 35); our results support these conclusions because β increased as the spatial scale became larger in the range of 10 to 380 km.

In conclusion, this study showed that microbial community composition exhibited a strong distance decay relationship in community similarity in municipal wastewater treatment facilities within 442 km of each other. Our results suggest that environmental heterogeneity is a more important factor than dispersal limitation, which strongly implies that deterministic community assembly mechanisms dominate municipal wastewater treatment bioreactors. Our results demonstrate a significant effect of geographic scale on distance decay relationships, suggesting that caution should be used before generalizing our quantitative results. We observed that the ecological theory that had been developed for plants and animals could be applied to microbial communities in engineered systems. Additional research is needed to identify and understand how specific factors affect specific groups of microorganisms during municipal wastewater treatment.

MATERIALS AND METHODS

Wastewater treatment facility description. Twenty wastewater treatment facilities (facility A to facility T) in the state of Minnesota participated in this study. All the facilities utilize continuous aeration

and recirculation of settled/concentrated biomass with some variation in process design and operation (see Table S2 in the supplemental material). Facilities with unique system designs (e.g., adsorption/bio-oxidation systems, membrane bioreactors, and sequencing batch reactors) were purposely excluded from this study because such system designs could confound the interpretation of spatial biogeographic patterns. Facilities C and G were equipped with two independent parallel bioreactors (C1 and C2 and G1 and G2), and facility S had four parallel bioreactors (S1 to S4). Untreated and treated wastewater characteristics were measured at each facility (Table S2). Each of the laboratories at the wastewater treatment facilities are certified by the U.S. EPA, and all of their methods are described in *Standard Methods for the Examination of Water and Wastewater* (36).

Sample collection and DNA extraction. Grab samples (volume, 10 to 15 ml) were collected by facility operators on the same week and kept frozen until processed. Samples were collected 12 times (approximately monthly) from July 2017 until July 2018. Aliquots of bioreactor samples (0.1 ml) were mixed with 0.9 ml of lysis buffer (5% sodium dodecyl sulfate, 120 mM sodium phosphate buffer [pH 8]). Mixed samples underwent three freeze-thaw cycles followed by 90 min of incubation at 70°C. DNA was then purified using the FastDNA kit (MP Biomedicals, Santa Ana, CA, USA) per the manufacturer's instructions.

Amplicon sequencing. From each sample, amplicons of the V4 region of the 16S rRNA gene were generated using primer set Meta_V4_515F and Meta_V4_806R and subsequently sequenced using Illumina MiSeq at the University of Minnesota Genomics Center as described previously (37). Each MiSeq run was loaded with a pooled library of amplicons (8 pM) with 15% PhiX. Cluster densities ranged from 824 to 1,104 K mm⁻²; 83.6% ± 5.2% of the bases had quality scores higher than 30.

DNA sequence analysis. For sequence analysis, mothur (version 1.40.0) was used (38), with a protocol similar to the MiSeq standard operating procedure (SOP) (39). Paired-end reads were merged, and sequences with one or more ambiguous bases or with a homopolymer longer than 8 nucleotides were removed. Sequences were aligned to SILVA v.128 reference file. Aligned sequences were filtered to remove columns that contained gaps and sequences with more than 300 bases or fewer than 280 bases were discarded. Chimeric sequences were identified and discarded with Vsearch (40). Sequences were classified with the Bayesian classifier using RDP v.16 reference files (41); sequences that did not classify as *Bacteria* or *Archaea* were discarded (i.e., unknown, chloroplast, mitochondria, or Eukaryota). OTU-based analysis was done using 4,336 sequences per sample to simultaneously maximize the number of microbiome profiles analyzed and the depth of sequences per microbiome profile. For OTU-based analysis, the sequences were clustered into OTUs at 97% similarity using the OptiClust method (42). For phylotype-based analysis, sequences were binned into phylotypes according to their genus-level taxonomic classification. Genus-level phylotypes show communities at a lower resolution (i.e., less richness and diversity) than OTUs; therefore, more sequences per sample were used. Phylotype-based analysis was performed using 7,805 sequences per sample in an attempt to optimize the number of qualifying microbiome profiles and sequencing depth. A phylogenetic tree was constructed with clear-cut using 1,000 random sequences per sample (43). AMOVA was performed to determine whether there was a significant difference in the microbial community composition in different bioreactors (44–46).

Data analysis. The taxon-volume relationship exponent z was calculated using the following equation (47):

$$\log_{10}(R) = z \log_{10}(V) + c$$

where R is community richness (e.g., the number of observed OTUs) in each bioreactor, V is the flow rate of the bioreactor, and c is an empirically derived constant. The spatial community shift rate (β) was calculated using the following equation (2):

$$\log_{10}(DS) = \beta \log_{10}(D) + c$$

where DS is the community dissimilarity between two bioreactors, D is the geographic distance between two bioreactors, and c is an empirically derived constant. Linear least-squares regressions were performed to estimate z and β . The arithmetic means from all the possible bioreactor pairs were used when dissimilarity calculation includes facilities with multiple bioreactors (e.g., dissimilarity between facilities B and C was calculated as the arithmetic mean of B-C1 and B-C2 dissimilarities). The linearity of both models was tested using Pearson's linear correlation.

Spearman's rank correlation was used to determine whether two parameters had a monotonic relationship. Two-tailed Wilcoxon rank sum tests were used to determine whether a metric for one group was significantly different than another group. Linear and quadratic regressions were performed using the time interval between two sampling events versus community dissimilarity (phylotype-based); analysis of variance (ANOVA) and axis of symmetry were then tested to determine which bioreactors exhibited linear and/or cyclic changes in community composition. Community dissimilarities were correlated with geographic distance and environmental differences (operational parameters and wastewater characteristics) using Mantel tests with Spearman's rank correlation. Using variables identified as having statistically significant correlations with community dissimilarities, CCA was performed to estimate the fraction that each environmental variable explained community dissimilarities. Mantel tests and CCA were performed using the package "vegan" in R (48); the significance of the CCA model was determined by analysis of variance (ANOVA) with 999 permutations. All the other statistical analyses were performed using MATLAB R2019a (MathWorks, Natick, MA, USA).

A significance level of $\alpha = 0.001$ was assumed throughout this study. This value was selected because the relatively high number of microbiome profiles ($n = 190$) could lead to "statistically

significant” results even when the correlations were not strong; in addition, computed correlation coefficients and *P* values were reported throughout.

Accession number(s). The sequences of this study have been deposited in NCBI Sequence Read Archive under accession number [PRJNA560576](https://www.ncbi.nlm.nih.gov/sra/PRJNA560576).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, DOCX file, 0.04 MB.

FIG S2, DOCX file, 0.1 MB.

FIG S3, DOCX file, 0.1 MB.

FIG S4, DOCX file, 0.1 MB.

FIG S5, DOCX file, 0.1 MB.

TABLE S1, DOCX file, 0.1 MB.

TABLE S2, DOCX file, 0.1 MB.

TABLE S3, DOCX file, 0.04 MB.

TABLE S4, DOCX file, 0.1 MB.

TABLE S5, DOCX file, 0.1 MB.

ACKNOWLEDGMENTS

Financial support was provided by the Minnesota Environment and Natural Resources Trust Fund.

We thank the operators at the treatment facilities for collecting samples and Elizabeth Hill for technical assistance.

REFERENCES

- Soininen J, McDonald R, Hillebrand H. 2007. The distance decay of similarity in ecological communities. *Ecography* 30:3–12. <https://doi.org/10.1111/j.0906-7590.2007.04817.x>.
- Nekola JC, White PS. 1999. The distance decay of similarity in biogeography and ecology. *J Biogeogr* 26:867–878. <https://doi.org/10.1046/j.1365-2699.1999.00305.x>.
- Whittaker R. 1975. *Communities and ecosystems*. Macmillan Publishing Co Inc, New York, NY.
- Astorga A, Oksanen J, Luoto M, Soininen J, Virtanen R, Muotka T. 2012. Distance decay of similarity in freshwater communities: do macro- and microorganisms follow the same rules? *Global Ecol Biogeogr* 21:365–375. <https://doi.org/10.1111/j.1466-8238.2011.00681.x>.
- Green J, Bohannan BJ. 2006. Spatial scaling of microbial biodiversity. *Trends Ecol Evol* 21:501–507. <https://doi.org/10.1016/j.tree.2006.06.012>.
- Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M, Gillings M, Beattie AJ. 2004. Spatial scaling of microbial eukaryote diversity. *Nature* 432:747–750. <https://doi.org/10.1038/nature03034>.
- Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A* 103:626–631. <https://doi.org/10.1073/pnas.0507535103>.
- Bahram M, Koljalg U, Courty PE, Diedhiou AG, Kjølner R, Polme S, Ryberg M, Veldre V, Tedersoo L. 2013. The distance decay of similarity in communities of ectomycorrhizal fungi in different ecosystems and scales. *J Ecol* 101:1335–1344. <https://doi.org/10.1111/1365-2745.12120>.
- Miura T, Sánchez R, Castañeda LE, Godoy K, Barbosa O. 2017. Is microbial terroir related to geographic distance between vineyards? *Environ Microbiol Rep* 9:742–749. <https://doi.org/10.1111/1758-2229.12589>.
- Feng M, Tripathi BM, Shi Y, Adams JM, Zhu YG, Chu H. 2019. Interpreting distance-decay pattern of soil bacteria via quantifying the assembly processes at multiple spatial scales. *MicrobiologyOpen* 8:e851. <https://doi.org/10.1002/mbo3.851>.
- Meyer KM, Memiaghe H, Korte L, Kenfack D, Alonso A, Bohannan BJ. 2018. Why do microbes exhibit weak biogeographic patterns? *ISME J* 12:1404–1413. <https://doi.org/10.1038/s41396-018-0103-3>.
- Ma B, LaPara TM, Evans AN, Hozalski RM. 2020. Effects of geographic location and water quality on bacterial communities in full-scale biofilters across North America. *FEMS Microbiol Ecol* 96:fiz210. <https://doi.org/10.1093/femsec/fiz210>.
- Zhou J, Ning D. 2017. Stochastic community assembly: does it matter in microbial ecology? *Microbiol Mol Biol Rev* 81:e00002-17. <https://doi.org/10.1128/MMBR.00002-17>.
- Bell T. 2010. Experimental tests of the bacterial distance–decay relationship. *ISME J* 4:1357–1365. <https://doi.org/10.1038/ismej.2010.77>.
- Horner-Devine MC, Lage M, Hughes JB, Bohannan BJ. 2004. A taxa–area relationship for bacteria. *Nature* 432:750–753. <https://doi.org/10.1038/nature03073>.
- Orhon D. 2015. Evolution of the activated sludge process: the first 50 years. *J Chem Technol Biotechnol* 90:608–640. <https://doi.org/10.1002/jctb.4565>.
- Wang X, Wen X, Deng Y, Xia Y, Yang Y, Zhou J. 2016. Distance–decay relationship for biological wastewater treatment plants. *Appl Environ Microbiol* 82:4860–4866. <https://doi.org/10.1128/AEM.01071-16>.
- Wu L, Ning D, Zhang B, Li Y, Zhang P, Shan X, Zhang Q, Brown MR, Li Z, Van Nostrand JD, Ling F, Xiao N, Zhang Y, Vierheilig J, Wells GF, Yang Y, Deng Y, Tu Q, Wang A, Global Water Microbiome Consortium, Zhang T, He Z, Keller J, Nielsen PH, Alvarez PJJ, Criddle CS, Wagner M, Tiedje JM, He Q, Curtis TP, Stahl DA, Alvarez-Cohen L, Rittmann BE, Wen X, Zhou J. 2019. Global diversity and biogeography of bacterial communities in wastewater treatment plants. *Nat Microbiol* 4:1183–1195. <https://doi.org/10.1038/s41564-019-0426-5>.
- Griffin JS, Wells GF. 2017. Regional synchrony in full-scale activated sludge bioreactors due to deterministic microbial community assembly. *ISME J* 11:500–511. <https://doi.org/10.1038/ismej.2016.121>.
- Kim T, Behrens S, LaPara TM. 2021. Direct evidence for deterministic assembly of bacterial communities in full-scale municipal wastewater treatment facilities. *Appl Environ Microbiol* 87:e01086-21. <https://doi.org/10.1128/AEM.01086-21>.
- Gilbert B, Lechowicz MJ. 2004. Neutrality, niches, and dispersal in a temperate forest understory. *Proc Natl Acad Sci U S A* 101:7651–7656. <https://doi.org/10.1073/pnas.0400814101>.
- Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumsin JA, Kuske CR, Morin PJ, Naeem S, Ovreås L, Reysenbach A-L, Smith VH, Staley JT. 2006. Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* 4:102–112. <https://doi.org/10.1038/nrmicro1341>.
- Baas Becking LGM. 1934. *Geobiologie of inleiding tot de milieukunde*. WP Van Stockum & Zoon, The Hague, The Netherlands.

24. O'Malley MA. 2008. 'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography. *Stud Hist Philos Biol Biomed Sci* 39:314–325. <https://doi.org/10.1016/j.shpsc.2008.06.005>.
25. Beijerinck M. 1913. De infusies en de ontdekking der bacteriën. Jaarboek van de Koninklijke Akademie voor Wetenschappen. Müller, Amsterdam, The Netherlands.
26. Gray ND, Miskin IP, Kornilova O, Curtis TP, Head IM. 2002. Occurrence and activity of Archaea in aerated activated sludge wastewater treatment plants. *Environ Microbiol* 4:158–168. <https://doi.org/10.1046/j.1462-2920.2002.00280.x>.
27. Kim T-S, Jeong J-Y, Wells GF, Park H-D. 2013. General and rare bacterial taxa demonstrating different temporal dynamic patterns in an activated sludge bioreactor. *Appl Microbiol Biotechnol* 97:1755–1765. <https://doi.org/10.1007/s00253-012-4002-7>.
28. Jiang X-T, Ye L, Ju F, Wang Y-L, Zhang T. 2018. Toward an intensive longitudinal understanding of activated sludge bacterial assembly and dynamics. *Environ Sci Technol* 52:8224–8232. <https://doi.org/10.1021/acs.est.7b05579>.
29. Johnston J, LaPara T, Behrens S. 2019. Composition and dynamics of the activated sludge microbiome during seasonal nitrification failure. *Sci Rep* 9:4565. <https://doi.org/10.1038/s41598-019-40872-4>.
30. MacArthur RH, Wilson EO. 2001. *The theory of island biogeography*, vol 1. Princeton University Press, Princeton, NJ.
31. Van Der Gast CJ, Jefferson B, Reid E, Robinson T, Bailey MJ, Judd SJ, Thompson IP. 2006. Bacterial diversity is determined by volume in membrane bioreactors. *Environ Microbiol* 8:1048–1055. <https://doi.org/10.1111/j.1462-2920.2006.00996.x>.
32. Harte J, McCarthy S, Taylor K, Kinzig A, Fischer ML. 1999. Estimating species-area relationships from plot to landscape scale using species spatial-turnover data. *Oikos* 86:45–54. <https://doi.org/10.2307/3546568>.
33. Storch D, Šizling AL. 2008. The concept of taxon invariance in ecology: do diversity patterns vary with changes in taxonomic resolution? *Folia Geobot* 43:329–344. <https://doi.org/10.1007/s12224-008-9015-8>.
34. Zinger L, Boetius A, Ramette A. 2014. Bacterial taxa–area and distance–decay relationships in marine environments. *Mol Ecol* 23:954–964. <https://doi.org/10.1111/mec.12640>.
35. Turner WR, Tjørve E. 2005. Scale-dependence in species-area relationships. *Ecography* 28:721–730. <https://doi.org/10.1111/j.2005.0906-7590.04273.x>.
36. Clesceri L, Greenberg A, Easton A. 1998. *Standard methods for the examination of water and wastewater*. American Public Health Association, Washington, DC.
37. Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, Gould TJ, Clayton JB, Johnson TJ, Hunter R, Knights D, Beckman KB. 2016. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat Biotechnol* 34:942–949. <https://doi.org/10.1038/nbt.3601>.
38. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
39. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79:5112–5120. <https://doi.org/10.1128/AEM.01043-13>.
40. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. <https://doi.org/10.7717/peerj.2584>.
41. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267. <https://doi.org/10.1128/AEM.00062-07>.
42. Westcott SL, Schloss PD. 2017. OptiClust, an improved method for assigning amplicon-based sequence data to operational taxonomic units. *mSphere* 2:e00073-17. <https://doi.org/10.1128/mSphereDirect.00073-17>.
43. Sheneman L, Evans J, Foster JA. 2006. Clearcut: a fast implementation of relaxed neighbor joining. *Bioinformatics* 22:2823–2824. <https://doi.org/10.1093/bioinformatics/btl478>.
44. Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491. <https://doi.org/10.1093/genetics/131.2.479>.
45. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26:32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>.
46. Martin AP. 2002. Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Appl Environ Microbiol* 68:3673–3682. <https://doi.org/10.1128/AEM.68.8.3673-3682.2002>.
47. Arrhenius O. 1921. Species and area. *J Ecol* 9:95–99. <https://doi.org/10.2307/2255763>.
48. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2015. Package 'vegan'. Community ecology package, version 2.