Freshwater Faculty Articles   Freshwater Sciences (School of)

9-8-2021

Human fecal contamination corresponds to changes in the freshwater bacterial communities of a large river basin

Jill McClary-Gutierrez
University of Wisconsin-Milwaukee, newtonr@uwm.edu

Zac Driscoll
Milwaukee Riverkeeper

Cheryl Nenn
Milwaukee Riverkeeper

Recommended Citation

Human Fecal Contamination Corresponds to Changes in the Freshwater Bacterial Communities of a Large River Basin

Jill S. McClary-Gutierrez,[a]* Zac Driscoll,[b] Cheryl Nenn,[b] Ryan J. Newton[a]

[a]School of Freshwater Sciences, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin, USA
[b]Milwaukee Riverkeeper, Milwaukee, Wisconsin, USA

ABSTRACT Microbial water quality is generally monitored by culturable fecal indicator bacteria (FIB), which are intended to signal human health risk due to fecal pollution. However, FIB have limited utility in most urbanized watersheds as they do not discriminate among fecal pollution sources, tend to make up a small fraction of the total microbial community, and do not inform on pollution impacts on the native ecosystem. To move beyond these limitations, we assessed entire bacterial communities and investigated how bacterial diversity relates to traditional ecological and human health-relevant water quality indicators throughout the Milwaukee River Basin. Samples were collected from 16 sites on 5 days during the summer, including both wet and dry weather events, and were processed by 16S rRNA gene amplicon sequencing. Historical water quality at each sampling location, as opposed to upstream land use, was associated significantly with bacterial community alpha diversity. Source partitioning the sequence data was important for determining water quality relationships. Sewage-associated bacterial sequences were detected in all samples, and the relative abundance of sewage sequences was strongly associated with the human *Bacteroides* fecal marker. From this relationship, we developed a preliminary threshold for human sewage pollution when using bacterial community sequence data. Certain abundant freshwater bacterial sequences were also associated with human fecal pollution, suggesting their possible utility in water quality monitoring. This study sheds light on how bacterial community analysis can be used to supplement current water quality monitoring techniques to better understand interactions between ecological water quality and human health indicators.

IMPORTANCE Surface waters in highly developed mixed-use watersheds are frequently impacted by a wide variety of pollutants, leading to a range of impairments that must be monitored and remediated. With advancing technologies, microbial community sequencing may soon become a feasible method for routine evaluation of the ecological quality and human health risk of a water body. In this study, we partnered with a local citizen science organization to evaluate the utility of microbial community sequencing for identifying pollution sources and ecological impairments in a large mixed-use watershed. We show that changes in microbial community diversity and composition are indicative of both long-term ecological impairments and short-term fecal pollution impacts. By source partitioning the sequence data, we also estimate a threshold target for human sewage pollution, which may be useful as a starting point for future development of sequencing-based water quality monitoring techniques.

KEYWORDS bacterial community, fecal pollution, rivers, sewage

Mixed-use watersheds can be impaired by pollutants from many different sources, such as urban runoff, sewage overflows, agricultural runoff, livestock, or wildlife. These sources can contribute a range of pollutants, including pathogens, sediment, or nutrients, which can lead to deteriorating ecological water quality and/or increased

human health risk. Current microbial water quality monitoring methods based on detection of culturable fecal indicator bacteria (FIB) can produce highly variable results, suggesting that evaluating a single indicator may not be appropriate for determining health risk (1, 2). Identifying the causes of specific pollution events or persistent water quality impairments is an important step in initiating new watershed-scale protection efforts (3–6).

Microbial source tracking (MST) is one technique used to identify the sources of fecal pollution in a water body. MST typically uses quantitative PCR methods to detect 16S rRNA gene fragments from bacteria that are sufficiently specific to an environmental source of interest (7–9). The majority of MST assays for fecal pollution tracking are designed to identify bacteria specific to the gastrointestinal tracts of certain animals (such as dogs or ruminants) (10, 11) or humans (12), though additional MST assays for identifying sewage as opposed to human feces are also being developed (13). MST assays have been successfully applied to identify pollution sources in various locations (14–19). Despite their demonstrated utility, MST requires sufficient prior knowledge of the pollution source of interest in order to design an appropriately specific assay (20–22). Additionally, it is highly likely that no MST assay will ever demonstrate perfect specificity with its intended source, leading many to recommend the use of multiple MST assays (21).

As DNA sequencing can identify many potential microbes of concern in a single sample, it may prove advantageous over single MST assays for profiling risks and impairments in a broad range of water environments. Most previous applications of DNA sequencing in water environments have focused on broadly describing the impact of pollutants or processes on bulk microbial communities in particular sample types (23–27). However, as water environments are often highly complex, with a mix of active or inactive community members and contributions from various natural or anthropogenic sources, observations of bulk microbial community composition may not be appropriate for identifying pollutant monitoring targets, especially when pollutants may be present at low levels. Methods to utilize sequencing data for pollution tracking, such as SourceTracker (28) and FORENSIC (29, 30), are being developed. While these approaches are useful, they are limited by a need for input of both pollutant and environmental community data (SourceTracker) or limited fecal pollution targets (FORENSIC), and neither method has been connected to health risk. With the plethora of sequencing data from a wide range of sample types and environments now available in public repositories, additional improvements can be made to identify pollutant thresholds related to human health or environmental health risks. Furthermore, coupling assessment of ecological and environmental conditions with specific pollution indicators could lead to the identification of new indicators or conditions that promote pollutant persistence (12). While DNA sequencing is not yet feasible for routine monitoring, new technologies that reduce both cost and time to generate results are actively being developed and validated (31, 32). In parallel with developments in technology, we need further developments in data analysis if we are to capitalize on the advantages that sequencing microbial communities provide for source tracking.

The goal of this research is to evaluate the use of microbial community diversity and composition as parameters for evaluating river and stream water quality. By assessing changes in microbial community diversity and partitioning communities against curated databases, we provide insights into how DNA sequencing data can be incorporated into monitoring frameworks for both ecological and human health.

RESULTS

Location description. The Milwaukee River watershed is a large mixed-land-use watershed containing a total of 875 miles of stream reaches that divide into three main rivers, the Milwaukee, Menomonee, and Kinnickinnic Rivers, which flow into the Milwaukee harbor estuary and Lake Michigan. Sampling sites included different stretches of the rivers themselves but mostly consisted of smaller tributary streams. Land uses range
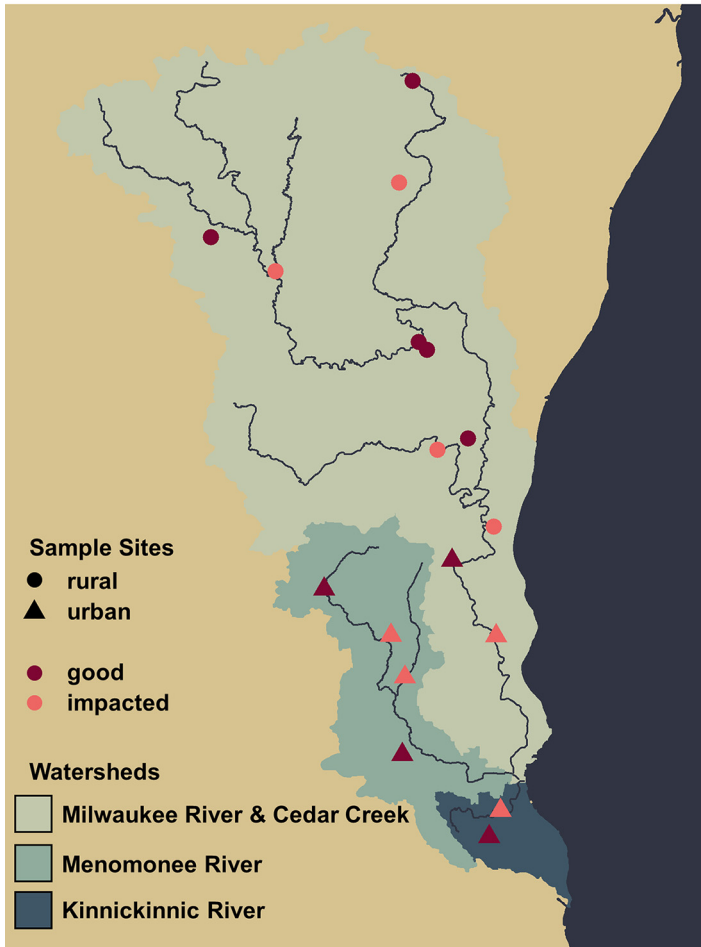
FIG 1 Map of sampling sites and watersheds. Sampling sites are designated urban or rural and good or impacted water quality.

from largely agricultural in the north to densely urban in the south, and 46% of stream reaches within the basin are considered impaired by at least one pollutant (33). Overall, sampling sites represented first- through fourth-order streams, with upstream watershed areas ranging from 1.8 to 117 square miles. Upstream watersheds of each sampling site also ranged in land use from 3% to 100% developed land based on U.S. Geological Survey (USGS) data. In general, sites were considered "urban" in the south and "rural" in the north of the Milwaukee River Basin. In the individual sites' upstream watersheds, rural sites contained 3 to 23% developed land and urban sites contained 24 to 100% developed land (see Table S1 in the supplemental material). Samples were collected from 16 different sites in the Milwaukee River watershed on five separate days between June and August 2017 (Fig. 1).

Bacterial community diversity is higher in historically high-quality streams and correlates with physicochemical measurements. In total, we detected more than 10,797 unique amplicon sequence variants (ASVs) in our DNA sequencing data. Each individual sample contained no more than a maximum of 2,402 ASVs, and 268 ASVs were only detected at one sampling site. Across all samples, only 1,488 ASVs were detected in more than half of sampling sites, demonstrating the diversity of bacterial communities from site to site and the complexity in determining indicators of water quality and human health risks throughout the watershed. We found a statistically significant decrease in bacterial community diversity in impacted streams compared to unimpacted streams using two different metrics for diversity (Wilcoxon rank sum test, $P < 0.01$) (Fig. 2A). In contrast, there was no significant difference in alpha diversity
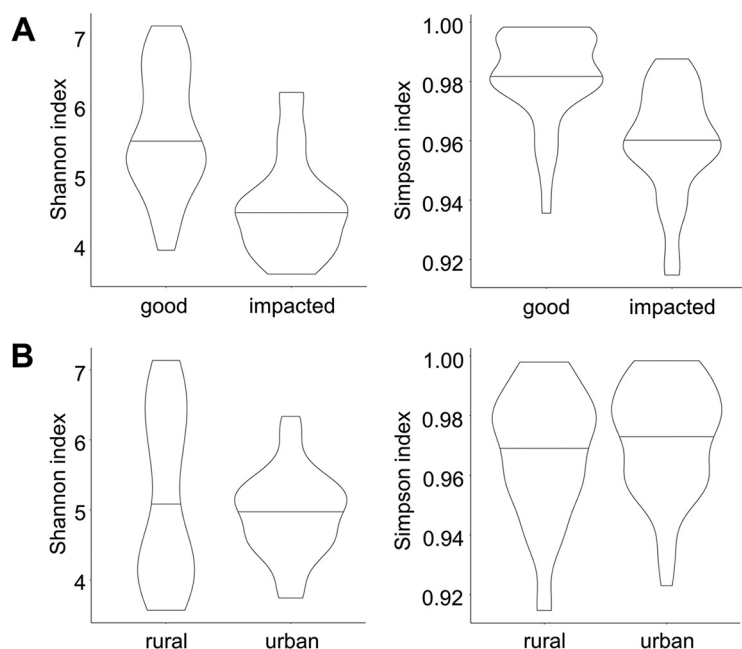
FIG 2 Alpha diversities (calculated using Shannon and Simpson indices) of microbial communities at good versus impacted sites (A) and urban versus rural sites (B).

metrics for urban versus rural sampling sites (Fig. 2B), but we did observe a statistically significant correlation between developed land area and Shannon diversity across the sites sampled (Spearman rank sum test, $\rho = 0.50$, $P < 0.05$) (Fig. S1 in the supplemental material), with higher Shannon diversity associated with a lower proportion of developed land area in the upstream watershed.

Bacterial community alpha diversity was significantly correlated with temperature, dissolved oxygen (DO), and pH (Fig. 3), suggesting that streams in this region with lower temperature, higher DO, and higher pH tended to be characterized by more diverse microbial communities. However, correlations between these instantaneous water quality measurements and bacterial community diversity were generally weak, suggesting that these variables do not explain much of the variability in community diversity between sites. DO and pH were also correlated with each other, suggesting that possibly only one of these variables influences bacterial community diversity. In addition, we found that bacterial community alpha diversity was significantly correlated with stream order, with higher diversity in lower-order streams (Fig. 4).

Sewage sequences are detected in all samples. Overall, the percentages of sequences attributed to freshwater or sewage sources varied widely across samples (Fig. 5). Freshwater sequences made up 15% to 90% of total sequences in each sample collected. By visualizing Bray-Curtis dissimilarity using a nonmetric multidimensional scaling (NMDS) ordination, samples with a high percentage of freshwater sequences clustered together and were made up of mostly fourth-order streams (Fig. 6). Sewage sequences were detected in all samples collected, making up between 0.3% and 30% of total sequences in each sample. Within the sewage sequences, we identified those most likely originating from human feces or of sewer origin (sewer associated). Across all water samples, a total of 361 unique ASVs were identified as probable sewer sequences, and an additional 16 ASVs were identified as probable human fecal sequences. Human fecal sequences were detected in 37 of 79 samples, making up between 0.003% and 0.5% of total sequences in samples where they were detected. The proportion of sewage sequences in each sample generally increased with increasing developed land proportions in the watershed, though this relationship was not significant (Spearman rank sum
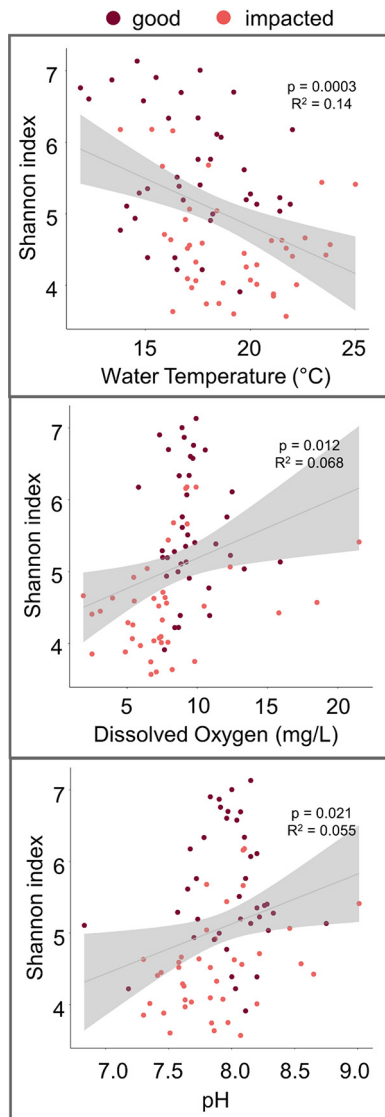
FIG 3 Correlation between alpha diversity (Shannon index) and water temperature, dissolved oxygen, and pH. Shaded regions are 95% confidence intervals around linear model fits.

test, $P = 0.10$) (Fig. 7A). The proportion of sewage sequences appeared to also be influenced by sampling date; nearly half (7/16) of the sites exhibited their highest proportions of sewage sequences on August 17, which was the only sampling date associated with wet weather (Fig. 7A). By considering the 377 total human fecal and sewer-derived ASVs, we found that the beta diversity of this community was significantly associated with both sample site latitude and urban versus rural land use (PERMANOVA, adonis function [34], $P < 0.01$) (Fig. 7B), suggesting significant spatial variation in human sewage sequence distribution throughout the basin. Human sewage community composition was also significantly associated with both water temperature and DO (PERMANOVA, $P < 0.01$), but not pH ($P = 0.12$). However, correlation coefficients were again quite low, indicating weak explanatory power of these variables.

Connection between sequencing-based human sewage and freshwater microbial communities and HB qPCR abundance. Using quantitative PCR (qPCR), human *Bacteroides* (HB) was detected in 66% of all samples collected, and the concentrations in these samples ranged from below the limit of quantification (<180 copy number [CN]/100 ml) to 26,442 CN/100 ml. HB concentrations were significantly associated with the compositions of both human sewage and freshwater microbial communities
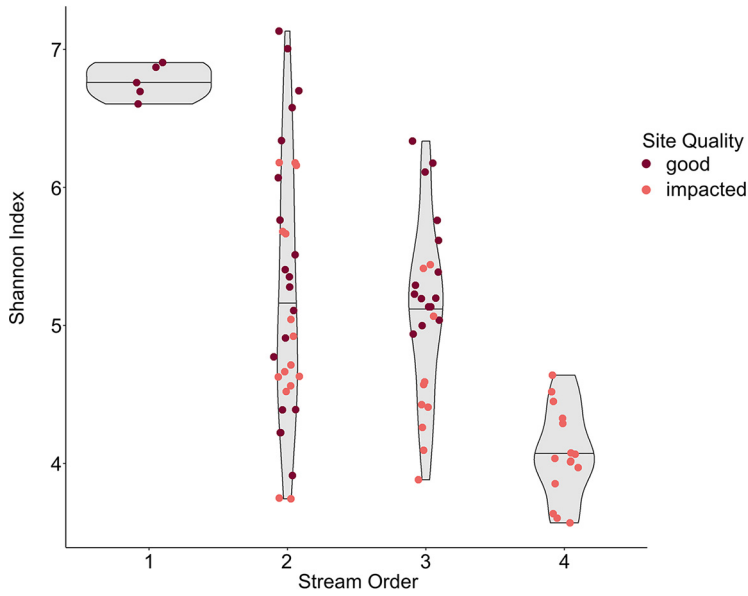
FIG 4 Relationship between stream order and alpha diversity (Shannon index). Points represent individual samples, and the widths of gray shaded violins represent the distributions of the individual samples. Black horizontal lines are medians of point distributions.

(permutational multivariate analysis of variance [PERMANOVA], $P < 0.001$). Additionally, while no association was observed between quantifiable HB concentrations and the proportions of ASVs assigned to freshwater bacteria (Spearman rank sum test, $P = 0.2$), a significant rank correlation was observed between HB concentrations and the proportions of sewer and fecal sequences ($P = 0.02$). This relationship between HB concentrations, which are already used in health risk assessments, and sewer-associated sequences suggests that the relative abundance of sewer-associated organisms detected by sequencing may be a
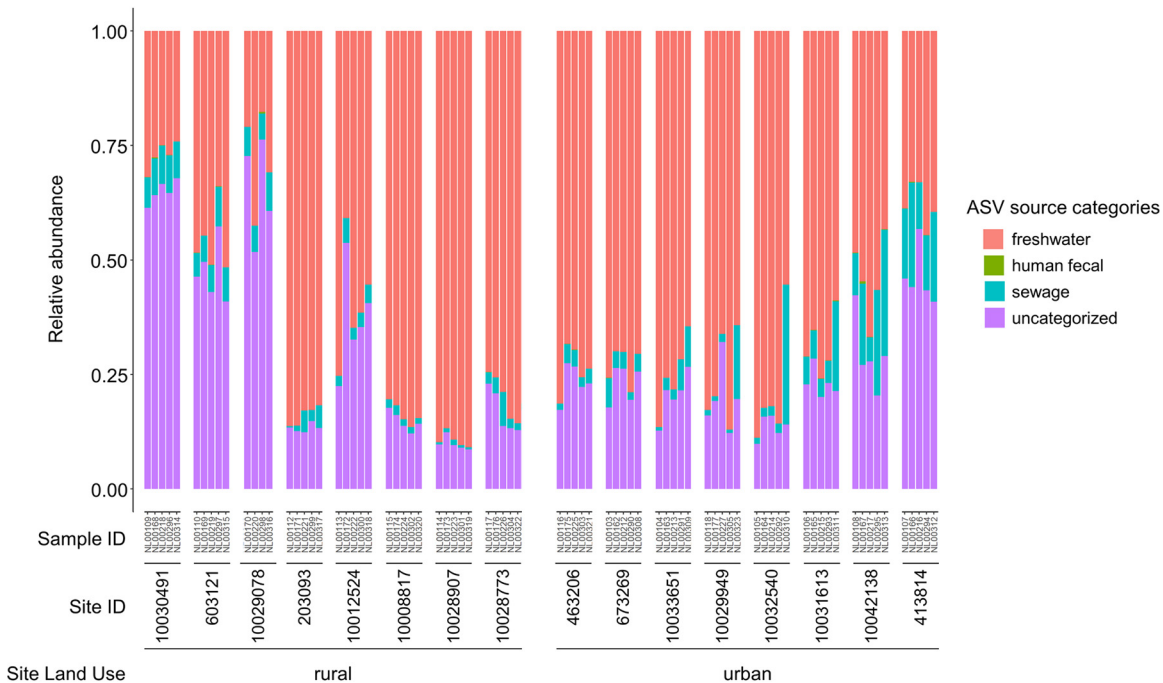

FIG 5 Relative abundances of ASVs across freshwater, human fecal, and sewage source categories. Uncategorized sequences are those that could not be assigned to any of the other categories. Each bar represents a single sample. Sites within rural and urban categories are organized by site location latitude (i.e., north to south).
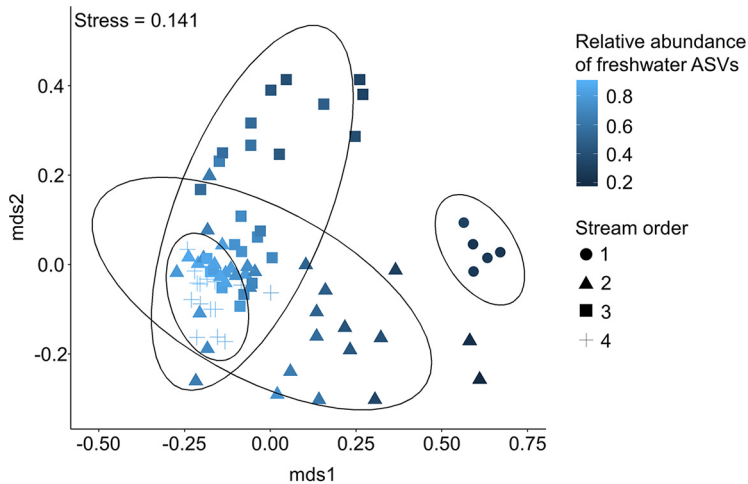
FIG 6 Nonmetric multidimensional scaling of microbial communities based on Bray-Curtis dissimilarity. Data ellipses designate groupings by stream order.

useful indicator of human health risk as sequencing technologies develop and become more appropriate for routine use.

To explore this possibility, we used a Monte Carlo simulation to convert a previously defined risk-based threshold of HB (7,800 CN/100 ml [16]) into an expected corresponding threshold of sewage sequence relative abundance (see supplemental materials for further details). From this simulation, we estimate that a threshold of 18% sewage sequences corresponds to significant human sewage pollution. A comparison of these two thresholds for the samples collected in this study is shown in Fig. 8. Seventy-two of 79 samples demonstrate agreement between the two thresholds, with 71 samples appearing below both thresholds and one sample above both thresholds. Six of 79 samples exceed the sewer sequence threshold only, and one sample exceeds the HB threshold only. Error on our sequence proportion estimate remains relatively large (±1 geometric standard deviation = 2% to 200%) and is primarily due to high variability in the distribution of total cell concentrations in rivers (see the supplemental material).

To determine specifically which, if any, ASVs in these communities were indicative of samples with elevated HB concentrations, we categorized samples based on their HB concentrations in categories of undetectable (below limit of detection), unquantifiable (detected but below limit of quantification), quantifiable (180 to 7,800 CN/100 ml), and risk (≥7,800 CN/100 ml). Since land use was determined previously to impact the microbial community composition, we performed this analysis only on samples from urban catchment areas. Among sewer and fecal ASVs, four ASVs (assigned to *Bacteroides dorei*, *Bacteroides uniformis*, *Trichococcus* sp., and *Leuconostoc* sp.) were identified as indicative of HB contamination above the risk threshold, one ASV was indicative of detectable HB contamination, one ASV of HB contamination below the risk threshold, and one ASV of unquantifiable (Table 1). Of the seven total sewer and stool ASVs identified, only the two *Bacteroides* ASVs associated with HB contamination above the risk threshold were from human stool communities (as opposed to sewer pipe communities). Among freshwater ASVs, 4 ASVs were associated with HB contamination above the risk threshold, 2 ASVs were associated with quantifiable HB contamination, 11 with detectable HB contamination, 10 with HB contamination below the risk threshold, and 3 with unquantifiable HB contamination (Table 1). Of the 30 total freshwater ASVs identified by indicspecies, 8 were assigned to the genus *Flavobacterium*.

DISCUSSION

Large mixed-use watersheds like the Milwaukee River Basin can be affected by many different pollutants, leading to a range of impairments and problems that must
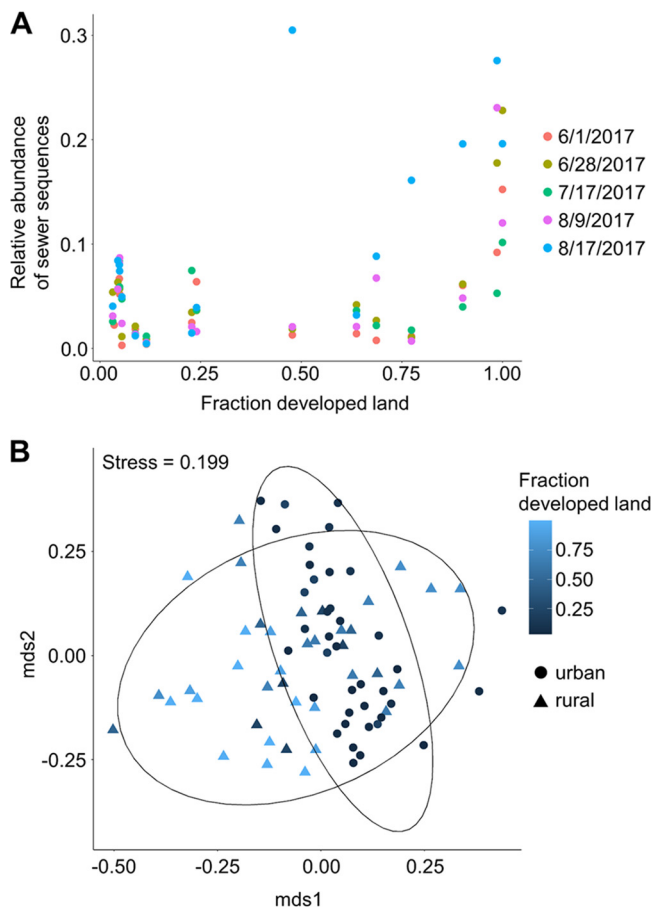
FIG 7 (A) Relationship between proportion of sewage sequences and developed land. (B) Relationship between composition of sewage sequences and land use. Nonmetric multidimensional scaling is based on Bray-Curtis dissimilarity of sewage sequences only. Data ellipses designate groups based on dominant land use.

be addressed by water quality monitoring efforts. Methods that allow combined assessment of ecological quality and human health risk would be valuable to stream-line both monitoring and mitigation. Bacterial community DNA sequencing is a rapidly advancing technique that can provide information on the overall quality of a water-shed and simultaneously detect specific microbes that may be associated with tran-sient or continuous human health-relevant pollution sources. While using DNA sequenc-ing for routine water quality monitoring remains generally prohibitive due to the cost and expertise required, sequencing methods are quickly advancing to address these difficulties. Our goal in this project was to apply DNA sequencing in conjunction with typical water quality monitoring performed by a local citizen science organization to determine its potential future utility in describing the overall quality of the watershed and identifying possible pollution sources to address.

Dissolved oxygen (DO) has been shown to significantly influence microbial commu-nity composition in various water bodies (35, 36). In particular, a previous study identi-fied positive correlations between DO and bacterial community alpha diversity in a Great Lakes watershed (37), and our study corroborates these relationships in the Milwaukee River Basin. As the site-specific upstream watersheds varied widely in size in our study, we hypothesized that bacterial community diversity might be impacted more strongly by the overall size of the developed area as opposed to the percentage of developed area in the upstream watershed. In our system, as streams travel from first to fourth order, changes in water chemistry, including increased temperature and
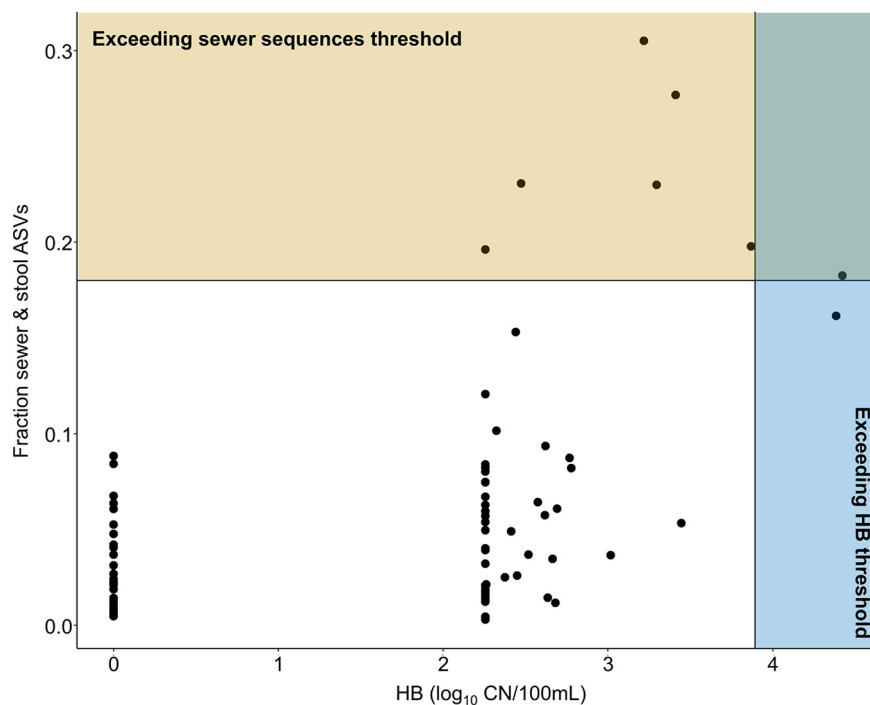
FIG 8 Comparison of sewer sequence relative abundances and human *Bacteroides* (HB) concentrations across all samples. Samples exceeding the HB concentration threshold appear in the blue shaded region; samples exceeding the sewer sequence threshold appear in the yellow shaded region.

decreased pH, appear to also decrease bacterial community diversity. While others have observed positive associations between alpha diversity and temperature or flow rate (38, 39), in this watershed, as communities travel from headwaters to Lake Michigan, they are likely to develop into communities more closely resembling typical freshwater lake communities, which tend to have lower diversity than terrestrial sources (40). Previous studies have also shown declines in bacterial diversity in river systems from upstream to downstream (41, 42), though Jordaan and Bezuidenhout observed the opposite in a highly urbanized river (43). Thus, the relationship between bacterial community diversity and land use may be less a factor of land use impacts on the water body and more related to maturation and development of the freshwater microbial community. This highlights the difficulty in establishing a threshold for alpha diversity to assess water quality, as the "ideal" alpha diversity will vary with other factors, such as stream order, trophic status, or possibly land use. However, the strong correlation observed in our data between alpha diversity and historical water quality suggests promise for this method, and future research should focus on understanding how alpha diversity varies within stream order classifications and across land use and water quality environmental variables.

While the HB marker is considered highly specific for the discrimination of human fecal inputs from other fecal pollutants for microbial source tracking, its widespread utility for risk assessment is hampered by inconsistencies in its presence and stability across global populations (21, 44–46), along with variable decay or transport between the HB marker and human pathogens (47–49). In contrast to this single-marker approach, which is prone to more random source input variation, our sewage community proportion threshold represents a full community approach that essentially incorporates hundreds of individual indicators to create a more robust and sensitive signal of sewage pollution. While detection of DNA via qPCR or sequencing does not discriminate between live and dead organisms, sewage exposure is strongly linked to health risk in epidemiological studies (50–52). Therefore, identification of high proportions of the full sewage community may serve as a more conservative indicator of health risk

TABLE 1 ASVs strongly associated with different human *Bacteroides* contamination categories in urban samples[a]

HB category	Sewer and stool community sequences				Freshwater community sequences			
	ASV ID[b]	Taxonomy	Specificity	Sensitivity	ASV ID	Taxonomy	Specificity	Sensitivity
Risk (≥7,800 CN/100 ml)	5850	*Bacteroides dorei*	1	1	1802	*Flavobacterium* sp.	0.9611	1
	5624	*Bacteroides uniformis*	0.9276	1	3070	*Novosphingobium* sp.	0.8907	1
	955	*Trichococcus* sp.	0.8995	1	4335	Unclassified *Enterobacteriaceae*	0.8515	1
	681	*Leuconostoc* sp.	0.8229	1	508	*Enterobacter* sp.	0.8309	1
Quantifiable (>LOQ)					342	*Flavobacterium* sp.	0.9333	0.8824
					585	*Flavobacterium chungangense*	0.8852	0.8235
Detectable (>LOD)	116	*Hydrogenophaga taeniospiralis*	0.9483	1	35	*Flavobacterium* sp.	0.9901	1
					258	*Flavobacterium* sp.	0.9572	1
					298	*Flavobacterium segetis*	0.9448	0.9643
					341	*Gemmobacter aquatilis*	0.9494	0.9286
					103	*Flavobacterium* sp.	0.9228	0.9286
					96	*Flavobacterium* sp.	0.9892	0.8571
					248	*Pseudorhodobacter* sp.	0.9703	0.8571
					203	*Sediminibacterium* sp.	0.9667	0.8571
					155	*Rhodoferax* sp.	0.9515	0.8571
					142	*Acinetobacter* sp.	0.9845	0.8214
					462	*Lelliottia* sp.	0.9717	0.8214
No risk (<7,800 CN/100 ml)	333	*Reyranella* sp.	1	0.9211	21	*Fluviicola* sp.	1	0.9211
					28	*Rhodoluna* sp.	0.9793	0.9211
					52	"*Candidatus* Planktophila sp."	1	0.8947
					58	"*Candidatus* Planktoluna sp."	1	0.8947
					139	MWH-Ta3 sp.	1	0.8684
					160	*Tabrizicola* sp.	1	0.8684
					204	*Sediminibacterium* sp.	1	0.8684
					22	Unclassified *Burkholderiaceae*	1	0.8684
					233	Unclassified *Sporichthyaceae*	1	0.8158
					344	hgcI_clade sp.	1	0.8158
Unquantifiable (<LOQ)	351	Unclassified *Steroidobacteraceae*	0.8659	0.8696	55	"*Candidatus* Limnoluna sp."	0.8599	0.8696
					144	*Sediminibacterium* sp.	0.8776	0.8261
					187	"*Candidatus* Planktoluna sp."	0.8214	0.8696

[a]Strongly associated ASVs were those with specificity and sensitivity both ≥80%.
[b]ID, identification number.

than qPCR markers and could be applied as a useful benchmark for the development of sequencing-based water quality assessments. Here, we have used a mass balance and Monte Carlo simulation approach to estimate a threshold based on sewage sequence abundance. However, we note that our estimation is relatively simple and should be further refined by considering other factors. For example, as our proportion estimate is based on total cell concentrations in river water being relatively stable, a high influx of another biological contaminant, such as animal feces, at the time of sampling could confound our sewage sequence proportion by contributing related fecal sequences and increasing the background total cell concentration. Additional data collection directed at understanding the effects of other variables (e.g., total bacterial cell concentrations, sequencing biases, flow rates, community composition shifts, etc.) on our sewage community proportion metric are needed to reduce error in our estimations and set more definitive thresholds.

In examining the human sewage community specifically, we observed that two human stool ASVs assigned to *B. dorei* and *B. uniformis* were indicative of high HB contamination. These two ASVs were only detected in three and five samples, respectively, and the combination of both ASVs was only detected in the two samples with high HB contamination above the health risk threshold. As the HB qPCR marker is designed for detection of a particular *B. dorei* strain (9, 11, 12), its association with two *Bacteroides* sequences suggests agreement between sequencing and qPCR methods. Furthermore, the lack of identification of any other human stool bacteria in high-HB samples suggests the possibility of differential decay between these markers and other sewer or human fecal bacteria. Additional sewage bacteria, including *Trichococcus* and *Leuconostoc* spp., were indicative of HB contamination above the risk threshold. *Trichococcus* spp. are typically present at high abundance in sewage (6, 53, 54).

Overall, the freshwater ASVs identified as associated with quantifiable HB levels were more abundant than human sewage sequences in the full community. In particular, 8 of 17 freshwater ASVs associated with HB detection were identified as *Flavobacterium* spp. Flavobacteria are commonly associated with urban environments and sewage-polluted waters (40), presumably because many members of this genus show strong growth responses in aquatic environments containing numerous complex carbon compounds (55). In addition to *Flavobacterium*, other freshwater bacteria typically associated with stormwater, wastewater infrastructure, and urban runoff, including members of *Enterobacter*, *Sediminibacterium*, *Acinetobacter*, and *Enterobacteriaceae* (6, 40, 56, 57), were also associated with HB detection. These freshwater ASVs are not expected to associate with health risk in all cases, as their detection may also be due to other organic matter inputs that do not comprise a human health risk. However, as real-time flow-based sequencing technologies detect higher-abundance sequences faster than rare sequences, these community members may be useful targets for the development of sequencing-based monitoring protocols for general contamination issues that warrant further evaluation. Finally, sequences identified as common freshwater organisms, including "*Candidatus* Planktophila," "*Candidatus* Planktoluna," and *Fluviicola* (40, 58, 59), were associated with no risk or unquantifiable HB. As these sequences are much more abundant in the water column than the *Bacteroides* sequences associated with high-risk HB concentrations, they may be useful for the development of markers of freshwater bodies with no or low impact.

Conclusions. Large mixed-use watersheds can be affected by many different pollution sources, and identifying the causes and effects of these pollutants remains a challenge. Despite their diversity of sizes, land uses, and flow paths, streams in large river basins are often regulated by basin-wide standards for pollutant loads and downstream impacts. In this study, we evaluated bacterial communities across a wide array of streams in a watershed with many water quality impairments, and we showed that higher bacterial community diversity is associated with historically healthy stream reaches. Furthermore, by evaluating subpopulations of bacterial communities associated with curated freshwater and sewage databases, we were able to more easily identify anthropogenic inputs at specific sites and potential impacts of these inputs on the

typical freshwater community. While further technological advancements are needed before bacterial community analysis can be incorporated into routine monitoring, we demonstrate that this approach can provide useful information on both ecological and human health-relevant water quality metrics and that evaluation of disturbances in expected naturally occurring bacterial communities can provide insight into the prevalence of low-level sewage contamination.

MATERIALS AND METHODS

Sample collection. Sample sites were selected to represent both southern (urban) and northern (rural) areas of the watershed, as well as streams with historically impacted or unimpacted quality. Sites were designated as impacted based on atypical dissolved oxygen (DO) concentrations in a long-term-monitoring data set (see the supplemental material) (60). Due to accessibility issues, only 15 sites were visited on the first sampling date (1 June 2017), and one site (number 10012525) was removed and replaced with site number 10012524 after the first sampling date. A total of 79 samples were collected. One sample date (17 August 2017) was during a wet weather event; all other samples were collected during dry weather periods. At the time of sample collection, physicochemical parameters (air temperature, water temperature, DO, pH, transparency, specific conductivity, and total phosphorus) were also measured using Wisconsin Department of Natural Resources recommended protocols for water quality monitoring (61). For each sample, surface water was collected into a 2-liter bottle and stored on ice for transport to the laboratory. Samples were processed immediately upon arrival at the laboratory (within 6 h of collection).

Sample processing. Two hundred fifty milliliters of each sample was vacuum filtered onto a 47-mm-diameter, 0.22-μm-pore-size mixed-cellulose ester filter (Millipore, Billerica, MA). Each filter was aseptically folded, inserted into a 2-ml screw-cap tube, and stored at $-80°C$ until DNA extraction. DNA was extracted using a previously published protocol (62). In brief, samples were removed from the freezer and filters were crushed immediately using a sterile spatula. Crushed filter pieces were then added to FastDNA spin kit for soil bead-beating matrix tubes (MP Biomedicals, Solon, OH, USA). DNA extraction then proceeded following the manufacturer's recommendations. After extraction, samples were stored at $-20°C$ until sequencing.

Library preparation and DNA sequencing. Each DNA extract was amplified by targeting the V4-V5 hypervariable region of the 16S rRNA gene in triplicate using previously published PCR primers (518F and 926R) and conditions (63). A positive-control mock community (HM-783D, final error rate = 0.02% of sequenced nucleotides; BEI Resources, Manassas, VA) and a no-template control were also included in PCRs and carried through to sequencing. After PCR, triplicate amplicons from each sample were pooled and cleaned using the Agencourt AMPure XP kit (Beckman Coulter, Indianapolis, IN). Samples were then submitted to the Great Lakes Genomics Center, where they were barcoded, pooled, and sequenced on an Illumina MiSeq machine, generating 250-bp paired-end reads.

Sequencing data processing. Raw sequencing data were demultiplexed and quality scored by Illumina MiSeq software to generate fastq files for forward and reverse reads for each sample. PCR primers were first trimmed using cutadapt 1.18 from both forward and reverse reads in paired-end mode (64). Next, dada2 version 1.8.0 was used to filter input sequences by quality, generate amplicon sequence variants (ASVs), and remove chimeras (65). Additional quality filtering was performed in mothur version 1.41.1 (66) to screen out ASVs that were more than 5% shorter or longer than the median amplicon sequence length or were identical. Taxonomic assignments were then determined for ASVs using the SILVA version 132 taxonomy database and the assignTaxonomy and addSpecies functions in dada2. Following taxonomy assignment, ASVs that assigned to Eukarya, Chloroplast, or Mitochondria were removed, and any ASV present in fewer than three samples was also removed. Finally, the remaining ASVs were compared to the positive-control mock community ASVs. Of these, four mock community ASVs were identified in the samples, suggesting possible well-to-well contamination. These four ASVs were also removed from the sample data set.

To generate additional taxonomic assignments for typical freshwater organisms, TaxAss was used to assign ASVs to the FreshTrain database (58, 67). For human fecal and sewer organisms, a custom database was used based on a combination of DNA sequencing data from the Human Microbiome Project (68) and a large data set of raw sewage samples from Milwaukee (see the supplemental material) (69). Assignment to the human fecal and sewer ASV database was performed using a custom script to identify exact alignments (as in reference 69) (see the supplemental material), and any ASVs that were previously assigned to the freshwater database were excluded. River sequences that aligned exactly to human fecal or sewer sequences in the custom database were considered probable human fecal or sewer sequences if the 5th-percentile relative abundance in river samples was lower than the 95th-percentile relative abundance in raw sewage. ASVs that did not align with the FreshTrain database or custom human fecal and sewer databases were considered to come from different sources (e.g., terrestrial) and were classified as "uncategorized."

Quantitative PCR for human fecal marker. To further examine relationships between microbial community parameters and potential human fecal contamination, a quantitative PCR (qPCR) assay for the human *Bacteroides* (HB) fecal marker was used (11, 12). Raw DNA extracts from each sample were used as the template for the qPCR, which followed previously published conditions (12). The concentration of HB marker in each sample was determined as the copy number (CN) per 100 ml of original water sample based on the volume of sample per filter and triplicate qPCR standard curves. Standard curves

indicated an assay efficiency and R^2 of 98.1% and 0.997, respectively. The limit of quantification was determined as the lowest quantifiable standard dilution (15 CN per reaction mixture volume), which is equivalent to 180 CN/100-ml sample. No-template controls were included for each qPCR run to ensure lack of contamination. As part of the laboratory's quality assurance procedures, a subset of all samples processed in the laboratory during this time period were also tested for inhibition using a salmon sperm spike-in (16). No inhibition was detected.

Statistical analysis. Additional metadata for each sampling site, including watershed area, land use percentages, and stream order, were obtained from the USGS StreamStats and WiDNR 24K Hydro databases. Alpha diversity for each sample was calculated using both Shannon and Simpson indices in vegan 2.5-6 (34). Correlations between alpha diversity and categorical variables (historical water quality and dominant land use) were tested using Wilcoxon rank sum tests. Correlations between alpha diversity and continuous variables (percent developed land, water temperature, and dissolved oxygen) were tested using Pearson's product-moment test for normally distributed variables or Spearman's rank sum test for nonnormally distributed variables. To evaluate beta diversity, Bray-Curtis dissimilarity was calculated for the full data set and plotted using a nonmetric multidimensional scaling (NMDS) ordination. Depending on the hypotheses tested, Bray-Curtis dissimilarity was also calculated using subsets of data based on their taxonomic assignments in either freshwater or human sewage databases. PERMANOVA was used to test associations between Bray-Curtis dissimilarity and sample metadata (i.e., impacted versus good water quality and rural versus urban land uses).

Previous work using quantitative microbial risk assessment has determined that an HB concentration of 7,800 CN/100 ml indicates an excess illness risk of 0.03 (16). To determine the approximate relative abundance of sewage sequences that would be associated with this risk threshold, we evaluated the following equations using a Monte Carlo simulation:

$$f = \frac{C_{river,HB}}{C_{sewage,HB}}$$

$$r = \frac{C_{sewage,cells} \times f}{C_{river,cells} \times (1-f)}$$

where f is the volumetric fraction of sewage in the river, $C_{river,HB}$ is the concentration of HB in the river, $C_{sewage,HB}$ is the concentration of HB in sewage, $C_{sewage,cells}$ is the total cell concentration in sewage, $C_{river,cells}$ is the total cell concentration in the river, and r is the relative abundance of sewage sequences. These equations are based on an assumption that the relative abundance of an individual ASV generated using amplicon sequencing generally reflects the relative abundance of that sequence in the volume of sample analyzed. $C_{river,HB}$ was set at the health risk threshold of 7,800 CN/100 ml. All other input variables ($C_{sewage,HB}$, $C_{sewage,cells}$, and $C_{river,cells}$) were represented as log-normal distributions based on data collected previously by our laboratory and/or in published literature. Specific input data sources and distributions are described further in the supplemental material.

Samples were also divided categorically into risk ($\geq$7,800 CN/100 ml), quantifiable (180 to 7,800 CN/100 ml), detectable (detected and <180 CN/100 ml), or undetectable (below limit of detection) HB contamination. Between these categories, indicspecies (70) was used with either sewer and stool or freshwater communities to determine the strength of these communities' prediction for high HB abundance. Strongly associated ASVs were considered to be those with both sensitivity and specificity greater than 80%.

All data processing was performed in Linux command line, R version 4.0.0 with Rstudio version 1.2.5042, and Microsoft Excel.

Data availability. Raw sequence data are deposited in the NCBI Sequence Read Archive (BioProject PRJNA665728).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.2 MB.

REFERENCES

1. Aranda D, Lopez JV, Solo-Gabriele HM, Fleisher JM. 2016. Using probabilities of enterococci exceedance and logistic regression to evaluate long term weekly beach monitoring data. J Water Health 14:81–89. https://doi.org/10.2166/wh.2015.030.

2. US Environmental Protection Agency. 2012. 2012 Recreational water quality criteria. US Environmental Protection Agency, Washington, DC.

3. Jones TH, Brassard J, Topp E, Wilkes G, Lapen DR. 2017. Waterborne viruses and F-specific coliphages in mixed-use watersheds: microbial

associations, host specificities, and affinities with environmental/land use factors. Appl Environ Microbiol 83:e02763-16. https://doi.org/10.1128/AEM.02763-16.

4. Dila DK, Corsi SR, Lenaker PL, Baldwin AK, Bootsma MJ, McLellan SL. 2018. Patterns of host-associated fecal indicators driven by hydrology, precipitation and watershed attributes in Great Lakes watersheds. Environ Sci Technol 52:11500–11509. https://doi.org/10.1021/acs.est.8b01945.

5. Weidhaas J, Anderson A, Jamal R. 2018. Elucidating waterborne pathogen presence and aiding source apportionment in an impaired stream. Appl Environ Microbiol 84:e02510-17. https://doi.org/10.1128/AEM.02510-17.

6. Newton RJ, Bootsma MJ, Morrison HG, Sogin ML, McLellan SL. 2013. A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. Microb Ecol 65:1011–1023. https://doi.org/10.1007/s00248-013-0200-9.

7. Ahmed W, Gyawali P, Feng S, McLellan SL. 2019. Host specificity and sensitivity of established and novel sewage-associated marker genes in human and nonhuman fecal samples. Appl Environ Microbiol 85:e00641-19. https://doi.org/10.1128/AEM.00641-19.

8. Harwood VJ, Staley C, Badgley BD, Borges K, Korajkic A. 2014. Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. FEMS Microbiol Rev 38:1–40. https://doi.org/10.1111/1574-6976.12031.

9. Ahmed W, Hughes B, Harwood VJ. 2016. Current status of marker genes of Bacteroides and related taxa for identifying sewage pollution in environmental waters. Water 8:231. https://doi.org/10.3390/w8060231.

10. Reischer GH, Kasper DC, Steinborn R, Mach RL, Farnleitner AH. 2006. Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. Appl Environ Microbiol 72:5610–5614. https://doi.org/10.1128/AEM.00364-06.

11. Kildare BJ, Leutenegger CM, McSwain BS, Bambic DG, Rajal VB, Wuertz S. 2007. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: a Bayesian approach. Water Res 41:3701–3715. https://doi.org/10.1016/j.watres.2007.06.037.

12. Newton RJ, VandeWalle JL, Borchardt MA, Gorelick MH, McLellan SL. 2011. Lachnospiraceae and Bacteroidales alternative fecal indicators reveal chronic human sewage contamination in an urban harbor. Appl Environ Microbiol 77:6972–6981. https://doi.org/10.1128/AEM.05480-11.

13. Feng S, McLellan SL. 2019. Highly specific sewage-derived Bacteroides quantitative PCR assays target sewage-polluted waters. Appl Environ Microbiol 85:e02696-18. https://doi.org/10.1128/AEM.02696-18.

14. Cloutier DD, McLellan SL. 2016. Distribution and differential survival of traditional and alternative indicators of fecal pollution at freshwater beaches. Appl Environ Microbiol 83:e02881-16. https://doi.org/10.1128/AEM.02881-16.

15. Goodwin KD, Gruber S, Vondrak M, Crumpacker A. 2016. Watershed assessment with beach microbial source tracking and outcomes of resulting gull management. Environ Sci Technol 50:9900–9906. https://doi.org/10.1021/acs.est.6b02564.

16. McLellan SL, Sauer EP, Corsi SR, Bootsma MJ, Boehm AB, Spencer SK, Borchardt MA. 2018. Sewage loading and microbial risk in urban waters of the Great Lakes. Elementa (Wash D C) 6:46. https://doi.org/10.1525/elementa.301.

17. Malla B, Ghaju Shrestha R, Tandukar S, Bhandari D, Inoue D, Sei K, Tanaka Y, Sherchand J, Haramoto E. 2018. Identification of human and animal fecal contamination in drinking water sources in the Kathmandu Valley, Nepal, using host-associated Bacteroidales quantitative PCR assays. Water 10:1796. https://doi.org/10.3390/w10121796.

18. Staley ZR, Boyd RJ, Shum P, Edge TA. 2018. Microbial source tracking using quantitative and digital PCR to identify sources of fecal contamination in stormwater, river water, and beach water in a Great Lakes Area of Concern. Appl Environ Microbiol 84:e01634-18. https://doi.org/10.1128/AEM.01634-18.

19. Kirs M, Caffaro-Filho RA, Wong M, Harwood VJ, Moravcik P, Fujioka S. 2016. Human-associated Bacteroides spp. and human polyomaviruses as microbial source tracking markers in Hawaii. Appl Environ Microbiol 82:6757–6767. https://doi.org/10.1128/AEM.01959-16.

20. Soller JA, Schoen ME, Varghese A, Ichida AM, Boehm AB, Eftim S, Ashbolt NJ, Ravenscroft JE. 2014. Human health risk implications of multiple sources of faecal indicator bacteria in a recreational waterbody. Water Res 66:254–264. https://doi.org/10.1016/j.watres.2014.08.026.

21. McLellan SL, Eren AM. 2014. Discovering new indicators of fecal pollution. Trends Microbiol 22:697–706. https://doi.org/10.1016/j.tim.2014.08.002.

22. Ahmed W, Staley C, Sadowsky MJ, Gyawali P, Sidhu JPS, Palmer A, Beale D, Toze S. 2015. Toolbox approaches using molecular markers and 16S rDNA amplicon datasets for the identification of fecal pollution in surface water. Appl Environ Microbiol 81:7067–7077. https://doi.org/10.1128/AEM.02032-15.

23. Nshimyimana JP, Freedman AJE, Shanahan P, Chua LCH, Thompson J. 2017. Variation of bacterial communities with water quality in an urban tropical catchment. Environ Sci Technol 51:5591–5601. https://doi.org/10.1021/acs.est.6b04737.

24. Stamps BW, Leddy MB, Plumlee MH, Hasan NA, Colwell RR, Spear JR. 2018. Characterization of the microbiome at the world's largest potable water reuse facility. Front Microbiol 9:2435. https://doi.org/10.3389/fmicb.2018.02435.

25. Tse TJ, Doig LE, Tang S, Zhang X, Sun W, Wiseman SB, Feng CX, Liu H, Giesy JP, Hecker M, Jones PD. 2018. Combining high-throughput sequencing of sedaDNA and traditional palaeolimnological techniques to infer historical trends in cyanobacterial communities. Environ Sci Technol 52:6842–6853. https://doi.org/10.1021/acs.est.7b06386.

26. Xu Z, Jiang Y, Te S, He Y, Gin K. 2018. The effects of antibiotics on microbial community composition in an estuary reservoir during spring and summer seasons. Water 10:154. https://doi.org/10.3390/w10020154.

27. Peters L, Spatharis S, Dario MA, Dwyer T, Roca IJT, Kintner A, Kanstad-Hanssen Ø, Llewellyn MS, Praebel K. 2018. Environmental DNA: a new low-cost monitoring tool for pathogens in Salmonid aquaculture. Front Microbiol 9:3009. https://doi.org/10.3389/fmicb.2018.03009.

28. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, Bushman FD, Knight R, Kelley ST. 2011. Bayesian community-wide culture-independent microbial source tracking. Nat Methods 8:761–765. https://doi.org/10.1038/nmeth.1650.

29. Roguet A, Eren AM, Newton RJ, McLellan SL. 2018. Fecal source identification using random forest. Microbiome 6:185. https://doi.org/10.1186/s40168-018-0568-3.

30. Roguet A, Esen ÖC, Eren AM, Newton RJ, McLellan SL. 2020. FORENSIC: an online platform for fecal source identification. mSystems 5:e00869-19. https://doi.org/10.1128/mSystems.00869-19.

31. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif 3:1–8. https://doi.org/10.1016/j.bdq.2015.02.001.

32. Hatfield RG, Batista FM, Bean TP, Fonseca VG, Santos A, Turner AD, Lewis A, Dean KJ, Martinez-Urtaza J. 2020. The application of nanopore sequencing technology to the study of Dinoflagellates: a proof of concept study for rapid sequence-based discrimination of potentially harmful algae. Front Microbiol 11:844. https://doi.org/10.3389/fmicb.2020.00844.

33. Milwaukee Riverkeeper. 2018. 2018 Milwaukee River Basin report card. Milwaukee Riverkeeper, Milwaukee, WI.

34. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2019. Community Ecology Package. 2.5-6. https://CRAN.R-project.org/package=vegan.

35. Leight AK, Crump BC, Hood RR. 2018. Assessment of fecal indicator bacteria and potential pathogen co-occurrence at a shellfish growing area. Front Microbiol 9:384. https://doi.org/10.3389/fmicb.2018.00384.

36. Nakatsu CH, Byappanahalli MN, Nevers MB. 2019. Bacterial community 16S rRNA gene sequencing characterizes riverine microbial impact on Lake Michigan. Front Microbiol 10:996. https://doi.org/10.3389/fmicb.2019.00996.

37. Mohiuddin MM, Botts SR, Paschos A, Schellhorn HE. 2019. Temporal and spatial changes in bacterial diversity in mixed use watersheds of the Great Lakes region. J Great Lakes Res 45:109–118. https://doi.org/10.1016/j.jglr.2018.10.007.

38. Wang P, Chen B, Yuan R, Li C, Li Y. 2016. Characteristics of aquatic bacterial community and the influencing factors in an urban river. Sci Total Environ 569-570:382–389. https://doi.org/10.1016/j.scitotenv.2016.06.130.

39. Luo X, Xiang X, Yang Y, Huang G, Fu K, Che R, Chen L. 2020. Seasonal effects of river flow on microbial community coalescence and diversity in a riverine network. FEMS Microbiol Ecol 96:fiaa132. https://doi.org/10.1093/femsec/fiaa132.

40. Newton RJ, McLellan SL. 2015. A unique assemblage of cosmopolitan freshwater bacteria and higher community diversity differentiate an urbanized estuary from oligotrophic Lake Michigan. Front Microbiol 6:1028. https://doi.org/10.3389/fmicb.2015.01028.

41. Savio D, Sinclair L, Ijaz UZ, Parajka J, Reischer GH, Stadler P, Blaschke AP, Blöschl G, Mach RL, Kirschner AKT, Farnleitner AH, Eiler A. 2015. Bacterial

diversity along a 2600km river continuum. Environ Microbiol 17:4994–5007. https://doi.org/10.1111/1462-2920.12886.

42. Staley C, Unno T, Gould TJ, Jarvis B, Phillips J, Cotner JB, Sadowsky MJ. 2013. Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. J Appl Microbiol 115:1147–1158. https://doi.org/10.1111/jam.12323.

43. Jordaan K, Bezuidenhout CC. 2016. Bacterial community composition of an urban river in the North West Province, South Africa, in relation to physico-chemical water quality. Environ Sci Pollut Res Int 23:5868–5880. https://doi.org/10.1007/s11356-015-5786-7.

44. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A 107:14691–14696. https://doi.org/10.1073/pnas.1005963107.

45. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. 2012. Human gut microbiome viewed across age and geography. Nature 486:222–227. https://doi.org/10.1038/nature11053.

46. Reischer GH, Ebdon JE, Bauer JM, Schuster N, Ahmed W, Åström J, Blanch AR, Blöschl G, Byamukama D, Coakley T, Ferguson C, Goshu G, Ko G, De Roda Husman AM, Mushi D, Poma R, Pradhan B, Rajal V, Schade MA, Sommer R, Taylor H, Toth EM, Vrajmasu V, Wuertz S, Mach RL, Farnleitner AH. 2013. Performance characteristics of qPCR assays targeting human- and ruminant-associated Bacteroidetes for microbial source tracking across sixteen countries on six continents. Environ Sci Technol 47:8548–8556. https://doi.org/10.1021/es304367t.

47. Boehm AB, Silverman AI, Schriewer A, Goodwin K. 2019. Systematic review and meta-analysis of decay rates of waterborne mammalian viruses and coliphages in surface waters. Water Res 164:114898. https://doi.org/10.1016/j.watres.2019.114898.

48. Boehm AB, Graham KE, Jennings W. 2018. Can we swim yet? Systematic review, meta-analysis, and risk assessment of aging sewage in surface waters. Environ Sci Technol 52:9634–9645. https://doi.org/10.1021/acs.est.8b01948.

49. Ahmed W, Zhang Q, Kozak S, Beale D, Gyawali P, Sadowsky MJ, Simpson S. 2019. Comparative decay of sewage-associated marker genes in beach water and sediment in a subtropical region. Water Res 149:511–521. https://doi.org/10.1016/j.watres.2018.10.088.

50. Cabelli VJ, Dufour AP, McCabe LJ, Levin MA. 1982. Swimming-associated gastroenteritis and water quality. Am J Epidemiol 115:606–616. https://doi.org/10.1093/oxfordjournals.aje.a113342.

51. Dewailly E, Poirier C, Meyer FM. 1986. Health hazards associated with windsurfing on polluted water. Am J Public Health 76:690–691. https://doi.org/10.2105/ajph.76.6.690.

52. Cheung WHS, Chang KCK, Hung RPS, Kleevens JWL. 1990. Health effects of beach water pollution in Hong Kong. Epidemiol Infect 105:139–162. https://doi.org/10.1017/s0950268800004737.

53. Vandewalle JL, Goetz GW, Huse SM, Morrison HG, Sogin ML, Hoffmann RG, Yan K, McLellan SL. 2012. Acinetobacter, Aeromonas and Trichococcus populations dominate the microbial community within urban sewer infrastructure. Environ Microbiol 14:2538–2552. https://doi.org/10.1111/j.1462-2920.2012.02757.x.

54. McLellan SL, Roguet A. 2019. The unexpected habitat in sewer pipes for the propagation of microbial communities and their imprint on urban

55. McBride M. 2014. The family Flavobacteriaceae, p 643–676. In Rosenberg E, DeLong E, Lory S, Stackebrandt E, Thompson F (ed), The prokaryotes, vol 11, 4th ed. Springer, Berlin, Germany.

56. Fisher JC, Newton RJ, Dila DK, McLellan SL. 2015. Urban microbial ecology of a freshwater estuary of Lake Michigan. Elementa (Wash D C) 3:000064. https://doi.org/10.12952/journal.elementa.000064.

57. Lee S, Suits M, Wituszynski D, Winston R, Martin J, Lee J. 2020. Residential urban stormwater runoff: a comprehensive profile of microbiome and antibiotic resistance. Sci Total Environ 723:138033. https://doi.org/10.1016/j.scitotenv.2020.138033.

58. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. 2011. A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol Rev 75:14–49. https://doi.org/10.1128/MMBR.00028-10.

59. Jezbera J, Sharma AK, Brandt U, Doolittle WF, Hahn MW. 2009. "Candidatus Planktophila limnetica," an actinobacterium representing one of the most numerically important taxa in freshwater bacterioplankton. Int J Syst Evol Microbiol 59:2864–2869. https://doi.org/10.1099/ijs.0.010199-0.

60. Milwaukee Riverkeeper. 2017. 2017 Milwaukee River Basin report card. Milwaukee Riverkeeper, Milwaukee, WI.

61. Wisconsin Department of Natural Resources. 2016. WAV Level 2 Stream Monitoring Program. Wisconsin Department of Natural Resources, Madison, WI.

62. Templar HA, Dila DK, Bootsma MJ, Corsi SR, McLellan SL. 2016. Quantification of human-associated fecal indicators reveal sewage from urban watersheds as a source of pollution to Lake Michigan. Water Res 100:556–567. https://doi.org/10.1016/j.watres.2016.05.056.

63. Morrison HG, Grim SL, Vineis JH, Sogin M. 2013. 16S amplicon fusion primers and protocol for Illumina platform sequencing. Figshare https://doi.org/10.6084/m9.figshare.833944.v1. Retrieved 10 January 2018.

64. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.

65. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.

66. 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.

67. Rohwer RR, Hamilton JJ, Newton RJ, McMahon KD. 2018. TaxAss: leveraging a custom freshwater database achieves fine-scale taxonomic resolution. mSphere 3:e00327-18. https://doi.org/10.1128/mSphere.00327-18.

68. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The Human Microbiome Project. Nature 449:804–810. https://doi.org/10.1038/nature06244.

69. LaMartina EL, Mohaimani AA, Newton RJ. 2021. Urban wastewater bacterial communities assemble into seasonal steady states. Microbiome 9:1–13. https://doi.org/10.1186/s40168-021-01038-5.

70. De Cáceres M, Legendre P. 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90:3566–3574. https://doi.org/10.1890/08-1823.1.