Submitted by: James Clift MI Env. Council

Linking fecal bacteria in rivers to landscape, geochemical, and hydrologic factors and sources at the basin scale

Marc P. Verhougstraete[a,1], Sherry L. Martin[b], Anthony D. Kendall[b], David W. Hyndman[b], and Joan B. Rose[a]

[a]Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824; and [b]Department of Geological Sciences, Michigan State University, East Lansing, MI 48824

Edited* by Rita R. Colwell, University of Maryland, College Park, MD, and approved June 29, 2015 (received for review August 15, 2014)

Linking fecal indicator bacteria concentrations in large mixed-use watersheds back to diffuse human sources, such as septic systems, has met limited success. In this study, 64 rivers that drain 84% of Michigan's Lower Peninsula were sampled under baseflow conditions for *Escherichia coli*, *Bacteroides thetaiotaomicron* (a human source-tracking marker), landscape characteristics, and geochemical and hydrologic variables. *E. coli* and *B. thetaiotaomicron* were routinely detected in sampled rivers and an *E. coli* reference level was defined (1.4 $\log_{10}$ most probable number·100 mL^{-1}). Using classification and regression tree analysis and demographic estimates of wastewater treatments per watershed, septic systems seem to be the primary driver of fecal bacteria levels. In particular, watersheds with more than 1,621 septic systems exhibited significantly higher concentrations of *B. thetaiotaomicron*. This information is vital for evaluating water quality and health implications, determining the impacts of septic systems on watersheds, and improving management decisions for locating, constructing, and maintaining on-site wastewater treatment systems.

Escherichia coli | *Bacteroides thetaiotaomicron* | baseflow | reference conditions | septic system

Water quality degradation influenced by diffuse sources at large watershed scales has been difficult to describe. Human modifications of natural landscapes can permanently alter hydrologic cycles and affect water quality (1, 2). Deforestation (3) and increased impervious surface area (4) have been linked with decreased infiltration and thus increased surface runoff. Overland flows concentrate pollutants and rapidly transport them down gradient where they eventually enter surface water systems and affect water quality (5, 6). A number of models have been developed to calculate overland and surface water flows (7, 8) and nutrient/chemical transport (9), but few studies have focused on microbial movement from land to water, particularly nontraditional fecal indicator bacteria that can be used to track human sources of pollution.

Microbial contamination poses one of the greatest health risks to swimming areas, drinking water intakes, and fishing/shellfish harvesting zones where human exposures are highest (10–12). These highly visible areas often receive more attention than sources of contamination because identifying the origin of pollution in complex watersheds requires costly comprehensive investigation of environmental and hydrologic conditions across temporal and spatial scales (13). Grayson et al. (14) suggest using a "snapshot" approach that captures water quality characteristics at a single point in time across broad areas to provide information frequently missed during routine monitoring. Compared with long-term comprehensive investigations, the snapshot approach reduces the number of samples, cost, and personnel required to examine pollution sources.

Escherichia coli concentrations are commonly used to describe the relative human health risk during water quality monitoring in lieu of pathogen detection. Studies attempting to trace pollution in water back to a specific land use with *E. coli* have rarely produced definitive conclusions (15, 16). Using molecular approaches, specific source targets can be isolated in complex systems and have recently been used to investigate land use and water quality impairments (17). Furtula et al. (18) demonstrated ruminant, pig, and dog fecal contamination in an agriculturally dominated watershed (Canada) using *Bacteroides* markers. The *Bacteroides thetaiotaomicron* α-1–6 mannanase (*B. theta*) gene has a high human specificity (19–22), but no studies to date have linked its presence to land use patterns.

Reference conditions have been established for minimally disturbed environments based on measurements of macroinvertebrates, fish, and diatoms (23–25), but microbial reference conditions have not been adequately explored or defined. Based on 15 unimpaired California streams, microbial reference conditions for *E. coli* [1.0 $\log_{10}$ most probably number (MPN)·100 mL^{-1}] and enterococci (1.2 $\log_{10}$ MPN·100 mL^{-1}) were defined as being below state water quality thresholds (26). In the Great Lakes, a human health threshold of 2.37 $\log_{10}$ *E. coli* MPN·100 mL^{-1} (27), or a level equally protective of human health, has been adopted by all state governments. However, this health-associated reference level was derived from epidemiological studies undertaken at beaches throughout the United States (28, 29) with limited knowledge of local implications.

In response to water quality degradation from human stressors and the poorly understood microbial conditions in large-scale fresh water systems such as the Great Lakes basin, this paper aims to (*i*) examine the spatial distribution of *E. coli* and a human specific source marker (*B. theta*) in 64 river systems that drain most of the state's Lower Peninsula under baseflow conditions, (*ii*) identify baseflow reference levels of fecal contamination in rivers, and (*iii*) determine how key chemical, physical,

ENVIRONMENTAL SCIENCES

Significance

New microbial source-tracking tools can be used to elucidate important nonpoint sources of water quality degradation and potential human health risks at large scales. Pollution arising from septic system discharges is likely more important than previously realized. Identifying these sources and providing reference levels for water quality provides a basis to assess water quality trends and ultimately remediate degraded areas.

Author contributions: M.P.V., S.L.M., A.D.K., and D.W.H. designed research; M.P.V., S.L.M., and A.D.K. performed research; M.P.V. and A.D.K. contributed new reagents/analytic tools; M.P.V., S.L.M., and A.D.K. analyzed data; and M.P.V., S.L.M., A.D.K., D.W.H., and J.B.R. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: mverhougstraete@email.arizona.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1415836112/-/DCSupplemental.

environmental, hydrologic, and land use variables are linked to river water quality at large scales.

Results and Discussion

To address microbial water quality impairment, this study examined fecal bacteria source tracking across a large spatial scale with classification and regression tree (CART) statistical method to link fecal contamination in rivers to landscape, geochemical, and hydrologic factors as well as potential human fecal sources such as septic systems and sewage effluent at the basin scale. The *B. theta* results suggest human fecal contamination was affecting 100% of the studied river systems. These results have significant implications for water and environmental quality managers. Further details on hydrologic, geochemical, and land use characteristics, as well as a CART analysis of the reduced dataset, are described in *SI Materials and Methods*.

Microbial Water Quality and Reference Conditions. This project measured *E. coli* and *B. theta* concentrations in 64 rivers under baseflow conditions. Across all sites, *E. coli* concentrations ranged from 0.3 to 3.0 $\log_{10}$ MPN·100 mL^{-1} (geometric mean of 1.4 $\log_{10}$ MPN·100 mL^{-1}) and *B. theta* ranged from 4.2 to 5.9 $\log_{10}$ cell equivalents (CE)·100 mL^{-1} (geometric mean of 5.1 $\log_{10}$ CE·100 mL^{-1}). *E. coli* levels were below the detection limit (<1 MPN·100 mL^{-1}) in four rivers, whereas *B. theta* was detected in all samples (Fig. 1 and Table S1). Nine rivers (14% of sites) exceeded the US Environmental Protecion Agency (USEPA) suggested *E. coli* criterion for safe contact (2.37 $\log_{10}$ MPN·100 mL^{-1}), ranging in concentrations from 2.4 to 3.0 $\log_{10}$ MPN·100 mL^{-1}. In these same nine rivers, *B. theta* concentrations ranged from 4.6 to 5.6 $\log_{10}$ CE·100 mL^{-1}. These nine *E. coli* values were significantly different ($P < 0.001$) from those of the other 55 sites, which had a geometric mean of 1.3 $\log_{10}$ MPN·100 mL^{-1}. In contrast, there was no statistically significant difference ($P = 0.433$) between *B. theta* concentrations from these two sets of sites.

E. coli concentrations (geometric mean of 1.4 $\log_{10}$ MPN·100 mL^{-1}) were generally below USEPA recreational water quality criteria and consistent with previously measured ranges in Great Lakes tributary rivers (30–32). A comprehensive review (33) found that *E. coli* levels in freshwater below 2.23 $\log_{10}$ MPN·100 mL^{-1} were associated with low relative risks of gastrointestinal illness for swimmers compared with nonswimmers. Because the *E. coli* geometric mean concentration observed in this study was below the safety level reported by Wade et al. (33), we suggest a reference condition for *E. coli* of 1.4 $\log_{10}$ MPN·100 mL^{-1} for Michigan's Lower Peninsula rivers under baseflow conditions in the absence of recent storm runoff. Wade et al. (28) reported positive associations between occurrence of illness and molecularly detected *Bacteroides* at one Great Lakes beach with a geometric mean concentration of 3.08 $\log_{10}$ CE·100 mL^{-1}, while noting that the associations were statistically weak ($P < 0.1$). Yampara-Iquise et al. (19) reported *B. theta* levels ranged from 5.8 to 9.8 $\log_{10}$ copies·100 mL^{-1} in multiple urban, agricultural, and small-town creek systems that represented various levels of human impact. In the current study, *B. theta* concentrations (range = 4.2–5.9 $\log_{10}$ CE·100 mL^{-1}; geometric mean = 5.1 $\log_{10}$ CE·100 mL^{-1}) averaged 1.6 times higher than levels reported by Wade et al. (28) but slightly lower than those reported by Yampara-Iquise et al. (19). Establishing *B. theta* reference conditions for Michigan rivers under other flow conditions would require additional sample analysis and a greater understanding of the bacterial distributions because comparative *B. theta* datasets are relatively small relative to available *E. coli* data, a key aspect to defining reference levels (34). Reference levels are important for establishing acceptable levels of disturbances, defining long-term water quality changes, and supporting management decisions (34). Although the concept of a reference condition lies in the notion of minimal impact, it is recognized that few streams or rivers are truly unimpaired because most receive treated sewage effluent, and the current study supports this premise.

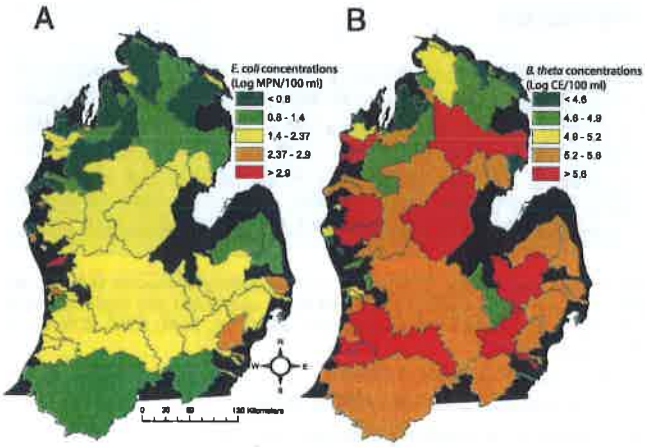

Fig. 1. (*A*) *E. coli* ($\log_{10}$ MPN·100 mL^{-1}) and (*B*) *B. theta* ($\log_{10}$ CE·100 mL^{-1}) concentrations measured in 64 rivers under baseflow conditions. Areas in black were not represented with samples.

CART Analysis of Microbial Water Quality. A primary goal of this study was to address diffuse pollution sources, historically a significant challenge in managing water quality. Major sources of nutrient loads from point and nonpoint sources of contamination were previously examined for Michigan's Lower Peninsula and shown to vary significantly between watersheds (35). The current study examines these drivers under baseflow conditions, where groundwater inputs dominate flows and wastewater effluent generally provides only a small fraction of total river discharges (Table S1). Effects of wastewater treatment plant (WWTP) effluent on microbial water quality were examined using multiple approaches (see *Supporting Information* for details), and it was ultimately determined that WWTP were not a driving factor of microbial water quality in the studied watersheds. Future analysis of the seasonal efficacy of WWTP could improve the understanding of wastewater impact on water quality by quantifying effluent discharge contributions in key urban areas.

The initial hypothesis of this research was that land use would best explain fecal bacterial concentrations in water. Instead, we found that land use characteristics such as septic systems and nutrients were the primary explanatory factors of microbial water quality. The influence of septic systems on microbial water quality, measured by *E.coli*, at a smaller watershed scale has also been reported in other regions (36, 37). In the current study, *E. coli* concentrations were linked primarily to total phosphorus and potassium. *B. theta* concentrations were primarily associated with the total number of septic systems in the watershed and within a 60-m buffer. Because WWTPs were not a driving factor of microbial water quality in the studied watersheds, these results indicate that under low flow conditions septic systems are a significant source of human fecal contamination to surface water in the studied watersheds.

CART analysis was used to evaluate the influence of the independent variables on *E. coli* and *B. theta*. Results from CART analyses for *E. coli* and *B. theta* concentrations at the full and reduced watersheds are summarized in Fig. 2 and Fig. S1, respectively. The CART outputs indicated complex causes of river water quality variability under baseflow conditions. For instance, *E. coli* concentrations at the full watershed scale were mainly related to total phosphorus (TP) concentrations, which is consistent with results by Carrillo et al. (38). TP concentrations accounted for 48% of *E. coli* variance with a threshold of 19.0 $\mu g \cdot L^{-1}$.

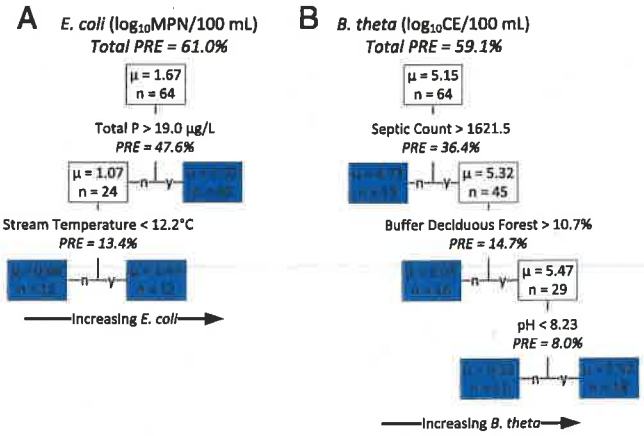

Fig. 2. CART analyses for (*A*) *E. coli* and (*B*) *B. theta* concentrations as dependent variables and land use, nutrient, chemical, hydrologic, and environmental parameters as independent variables in watersheds. PRE, proportion of reduction in error.

Although TP is essential for bacterial growth, the authors acknowledge that treated wastewater effluent includes high levels of both *E. coli* and TP. However, as stated above, WWTPs were not a driving factor of microbial water quality in the studied watersheds. Phosphorus, like *E. coli*, may be derived from sediments in the rivers, soil, plants, animal wastes, or manure and thus, unlike the *B. theta*, is not exclusive to fecal pollution.

The full watershed CART outputs and correlation analysis indicated *B. theta* concentrations were strongly associated with total numbers of septic systems in the watershed ($r = 0.364$, $P = 0.002$) and in the 60-m buffer ($r = 0.357$, $P = 0.004$). *B. theta* concentrations were not correlated with septic system density in the watershed ($P = 0.361$) or in the 60-m buffer ($P = 0.520$). Interestingly, the total number of septic systems in the watershed accounted for 36% of the *B. theta* concentration variance with a threshold count of 1,622 systems per watershed, as shown in Figs. 2*B* and 3. The snapshot sampling strategy used in this study focused on a spatial composite of the watersheds near the drainage point toward the Great lakes. Thus, the total number of people on septic tanks equates to the level of feces entering each watershed, and these levels are potentially dominated by failing septic systems contributing high concentrations of bacteria to nearby water systems. A Michigan health department reported a 26% on-site wastewater failure rate during time of sale or transfer inspections that discharged an estimated 65,000 gallons of untreated fecal waste each year to nearby water bodies (39). Future watershed-based studies should include analysis of total septic systems in the watershed and septic density, because it would be possible to overlook failing septic systems if the sample size were small or the focus were only on septic density. Additional efforts aimed at the condition of septic systems, their ability to remove bacteria, and microbial transport to nearby surface waters are required.

The direct and significant correlation between estimated number of septic systems and the human-specific marker *B. theta* in water (Fig. 3) illustrates a major issue for water quality of Michigan's streams and rivers, with an estimated 1.4 million on-site septic systems statewide (35, 40). In this study, the overall *B. theta* geometric mean was one $\log_{10}$ unit higher than secondary treated sewage effluent, whereas the highest measured concentrations were 1.5 logs higher than biologically treated septage effluent (20). Interestingly, when the CART analysis considered the entire upstream drainage area, including lakes, 2.5 times fewer septic systems were required to produce *B. theta* levels similar to when these drainage areas were restricted to downstream of the nearest lake, potentially indicating increased failure rates of septic systems surrounding lakes compared with rivers (see *Supporting Information* for details). Habteselassie et al. (41) identified that surface water and groundwater near failing on-site wastewater treatment systems contained higher concentrations of *E. coli* and enterococci than water surrounding properly functioning on-site wastewater treatment systems ($P < 0.001$). Combined, these results illustrate the importance and need for responsible development and septic system maintenance along lake and river riparian zones to protect water quality. Future analysis should include incremental spatial assessment of *B. theta* with respect to septic systems in watersheds to assess the fate and transport of bacteria from septic systems and define their acute/chronic impacts on water quality.

E. coli and *B. theta* Z-scores [(observed − mean)/SD] were compared using CART, as shown in Fig. 4, to identify the characteristics that could differentiate between *E. coli* and *B. theta* concentrations. Positive values of the Z-score differences occur when *E. coli* concentrations are higher, relative to their population mean, than *B. theta* concentrations. Negative values imply the opposite, with relatively higher *B. theta* concentrations. In catchments with discharge <0.66 $m^3 \cdot s^{-1}$ and with fewer than 294 septic systems in the 60-m buffer, *E. coli* concentrations were much higher than those of *B. theta*. In contrast, *B. theta* concentrations were much higher than those of *E. coli* in rivers with discharge >0.66 $m^3 \cdot s^{-1}$, particularly in catchments with dissolved organic carbon >5.4 $\mu g \cdot L^{-1}$. *E. coli*, which occurs in the feces of all warm-blooded mammals and birds, has been shown to persist and regrow in the environment under some conditions and has been associated with suspended particles that have low settling rates (42–45). Therefore, in watersheds with low discharge it is possible that *E. coli* can attach to particles and persist longer than *B. theta*, which is an anaerobic organism with a faster decay rate in rivers (46).

We compared the concentrations and loads of *E. coli* and *B. theta* across all sites (Fig. S2). No statistically significant relationship was identified between *E. coli* and *B. theta* concentrations ($r = 0.18$; $P = 0.16$). Bacterial entry to rivers during baseflow seems to be occurring from some of the same diffuse sources, including septic systems. The comparison of *E. coli* versus *B. theta* concentrations illustrated that each of these microorganisms was entering rivers from similar sources (i.e., diffuse sources such as septic systems) (Fig. 2). However, each organism was influenced by different environmental parameters as identified by the Z-score CART analysis (Fig. 4). *E. coli* was ubiquitous in most rivers and concentrations were primarily associated with TP and K levels. This study indicates that *B. theta* can be used as a source-tracking marker to investigate diffuse sources of human-derived contaminants from septic systems under baseflow hydrologic conditions at watershed scales.

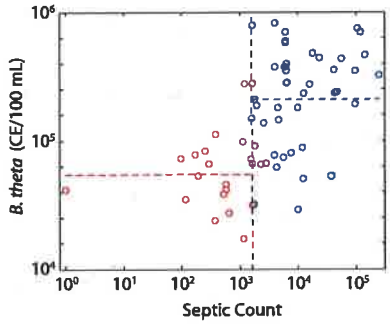

Fig. 3. *B. theta* versus septic systems illustrating the CART output from the first split of Fig. 2*B*.

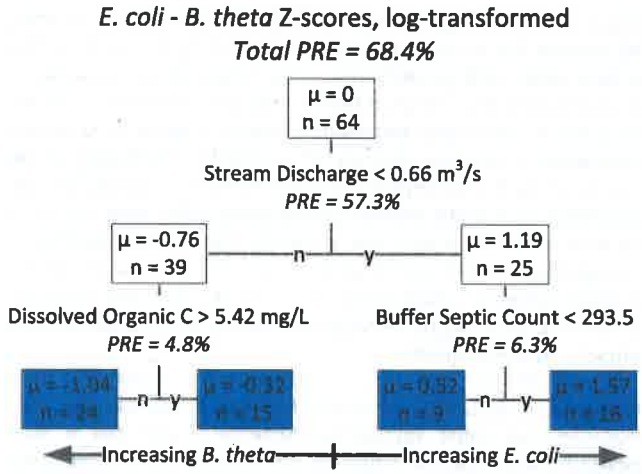

Fig. 4. CART of *E. coli* and *B. theta* Z-scores illustrating conditions associated with different concentrations between these two microbes. PRE, proportion of reduction in error.

Conclusions

To address impaired waters and restore them to designated uses, the process for total maximum daily loads (TMDLs) has been developed under the Clean Water Act. According to Stiles (47) there are currently 65,000 TMDLs and 43,000 listings that need to be addressed. Many stretches of water systems are impaired due to fecal pollution and *E. coli*, but there have been no established approaches or tools to identify nonpoint sources. This study provides a path forward to assess and ultimately improve water quality at large scales. More importantly, this study provides reference conditions for a large number of watersheds that, in the event of major landscape disturbance, could be used to measure remediation progress. Using a synoptic sampling approach for regional water quality assessment, this study found that human fecal contamination was prevalent under baseflow conditions. Baseflow in the study watersheds was generally dominated by groundwater and not by wastewater treatment effluent. Results suggest a regional *E. coli* reference condition below the current USEPA freshwater recreational criterion could be established. However, identifying specific sources of fecal contamination in rivers cannot be achieved using ubiquitous bacteria, such as *E. coli*. Assessing water quality using solely *E. coli* may mislead water quality managers and severely limit the ability to remediate impaired waterways. However, microbial source-tracking markers, such as the human-specific *B. theta* marker, can provide a more refined tool to identify the impacts of nonpoint sources of human fecal pollution, which could help prioritize restoration activities that should be implemented at watershed scales. The high variability of water quality measurements illustrates complex relationships between bacteria and landscape, geochemical, and hydrologic properties. The influence of septic systems in riparian zones also indicates that additional localized control measures, including septic system maintenance and construction, should be implemented to protect water quality and human health.

Materials and Methods

Study Area. This study investigated 64 watersheds draining Michigan's Lower Peninsula to the Great Lakes (Fig. S3). Watersheds were selected using the following criteria: (*i*) the 30 largest watersheds that represent >80% of Michigan's Lower Peninsula land area and (*ii*) 34 smaller watersheds randomly selected across the state from locations near their outlet to the lake. All sampling sites were located at bridge crossings and selected on the criteria that each was reasonably accessible, had adequate flow, river water dominated discharge, and the maximum amount of upstream land use was captured while meeting the above criteria.

Water Sample Collection. A synoptic sampling scheme was used to capture water quality characteristics under a single flow condition (i.e., baseflow) across broad spatial areas (14). Compared with long-term comprehensive investigations, this approach reduces the number of samples, cost, and personnel resources required to address pollution sources while providing essential information missed during routine monitoring.

Grab samples were collected from each river sampling site between October 1–13, 2010, which was chosen as a groundwater-dominated baseflow period based on historical hydrographs and antecedent precipitation. Groundwater-driven baseflow is critical to the preservation of water quality and quantity in the Great Lakes and provides year-round support for aquatic habitats. Before sampling each watershed, meteorological conditions were monitored to ensure that no significant precipitation had occurred within several days and hydrographs from nearby US Geological Survey (USGS) stream gauges were inspected to check that sampled rivers were at baseflow. October was chosen for the sampling period because the late growing season baseflow period is least likely to have large variability in water quality because flows are dominated by groundwater in the region. There is variability in water quality between baseflow periods (i.e., fall versus summer), but this variability is small relative to the variability between baseflow and other periods due to overland flow and dilution effects (48, 49). Water temperature (degrees Celcius), specific conductance (microsiemens per centimeter), and dissolved oxygen (milligrams per liter) were measured on-site using YSI 600R Sonde (YSI Incorporated). Field samples were placed on ice in coolers and transported to Michigan State University for other analyses, including bacterial testing (described below) within 24 h.

Water Analysis. Each sample was assayed for water chemistry as summarized in Table S2. The methods for assaying chemicals and nutrients are described in Table S3. *E. coli* analyses were performed within 24 h of collection using IDEXX Colilert Quanti-Tray 2000. Following incubation at 35 °C (±0.5 °C) for 24 h (±2 h), fluorescent wells were reported positive for *E. coli*, and reported as MPN per 100 mL. *E. coli* C-3000 (American Type Culture Collection 15597) was used as positive control for verification of media integrity. Sterile water was used for negative controls to verify method integrity. *E. coli* measurements below detection limits (1.0 MPN·100 mL^{-1}) were assigned the value of the detection limit.

Samples were analyzed for the human-specific marker *B. theta*, which has been shown to have a high sensitivity comparable to other human-associated markers in a multilaboratory evaluation (50). Compared with *B. theta*, HF183 and other source markers had greater false positive rates in animal feces collected in the same region as our study area (21). BacHum exhibited an even greater false positive rate than HF183 (51). Laboratories associated with our team and others have demonstrated that *B. theta* is a suitable human-specific marker and is related to human health outcomes (19–21, 52).

Analysis of the human-specific marker *B. theta* α-1–6 mannanase (5′CATCGTTCGTCAGCAGTAACA3′; 5′CCAAGAAAAAGGGACAGTGG3′) was performed according to Yampara-Iquise et al. (19), specifically by filtering 900 mL of water through a 0.45-μm hydrophilic mixed cellulose esters filter. Each filter was placed into a 50-mL centrifuge tube containing 20 mL of sterile phosphate-buffered water, vortexed, and centrifuged (30 min; 4,000 × *g*; 21 °C). Eighteen milliliters were decanted from the tube and the remaining eluent and pellet were stored at −80 °C. DNA was extracted from 200 μL of the thawed pellet via QIAamp DNA mini kit protocol. Quantitative PCR (qPCR) was performed on extracted DNA following Yampara-Iquise et al. (19) with a probe modification (20) using a Roche Light-Cycler 2.0 Instrument (Roche Applied Sciences). Each *B. theta* assay was carried out with 10 μL of LightCycler 480 Probe Mastermix (Roche Applied Sciences), 0.4 μL forward and reverse primers, 0.2 μL probe 62 (6FAM-ACCTGCTG-NFQ; Roche Applied Sciences Universal Probe Library), 1.0 μL BSA, 3.0 μL nuclease-free water, and 5.0 μL of extracted DNA and processed in triplicate. The qPCR analyses included a 15-min, 95 °C preincubation cycle, followed by 50 amplification cycles, and a 0.5-min 40 °C cooling cycle. A diluted plasmid standard was included during each qPCR run as a positive control and molecular-grade water was used in place of DNA template for negative controls. One copy of the targeted *B. theta* gene is assumed present per cell, and thus one gene copy number corresponded to one equivalent cell (19, 20). *B. theta* gene copies were converted to CE and reported as qPCR CE·100 mL^{-1}.

Climate and Hydrology. Hourly precipitation data were extracted from the Grand Rapids, Gaylord, and Detroit (Michigan) Next Generation Radar (NEXRAD) stations through the National Climate Data Center (www.ncdc.noaa.gov/nexradinv), with a base reflectivity of 0.50°, an elevation range of 124 nautical miles, and 16-km^2 cells. Hourly precipitation averages across each watershed were used to calculate total rainfall weighted by the

proportion of each NEXRAD cell within the sampled watershed. Precipitation was categorized into cumulative hourly totals (millimeters) before sample collection at intervals of 6, 12, 18, and 24 h and 2, 3, 4, 6, and 8 d, reported as millimeters per time before sample collection.

Real-time river discharge was measured at each site during sample collection using an Acoustic Doppler Current Profiler (53), colocated USGS stream gauges (waterwatch.usgs.gov), or current meter via wading following USGS protocol (54). River discharge is reported as cubic meters per second.

Land Use. Watersheds were delineated and then land use and septic system statistics were calculated for each watershed using Esri ArcMap GIS software (Table S4). The spatial analyst watershed tool was used to develop surface watersheds for each sampling point at 1 arc-second. Two watersheds were defined for each river site, referred to in this paper as full watersheds, which include the entire upstream drainage area ($n = 64$), and reduced watersheds, which only include drainage areas upstream of the sampling site to the nearest lake, reservoir, or pond ($n = 52$). The full watershed analysis ($n = 64$) included 12 sites that were at or near lake outlets, resulting in significantly smaller watersheds (average = 108 km^2) than the other 52 watersheds (average = 366 km^2). These 12 sites were removed in the reduced watershed analysis because it was originally hypothesized that longer retention time in the lentic water systems would likely reduce microbe concentrations owing to environmental decay. A digital map of land cover from 30-m resolution Landsat imagery and the National Land Cover Database (NLCD 2006; www.mrlc.gov/nlcd2006.php) was used to define land use in each watershed and buffer. Land use was categorized using the NLCD classification system with 16 categories and seven categories using the Anderson Level 1 Land Cover Classification System (55); Table S5 describes the Anderson classifications and equivalent NLCD categories. A 60-m riparian buffer was applied to streams in both full and reduced watersheds because land parcels are generally located adjacent to roads and require a buffer between surface waters and septic tanks. The average septic system setback from surface waters in Michigan is 15 m. Additionally, the 60-m riparian buffer ensured all riparian land uses were accounted for if the land use/river/septic system GIS layers were not completely matched under the 30-m resolution.

A map of households that likely use on-site septic systems to treat wastewater was previously developed for this study region (35). Briefly, septic system totals and locations were estimated following the cumulative examination of WWTP infrastructure, incorporated municipality areas, household location according to 2010 census blocks, 2006 NLCD and road layers, and residential drinking water well information. Estimated septic system numbers (per watershed) and densities (per square kilometer) in each watershed and 60-m-wide buffer around surface water bodies were calculated for the 64 river systems.

Estimates of total population and population relying on WWTPs for water treatment were performed for each watershed and 60-m buffer. The total population in each watershed was estimated by multiplying the number of households (based on 2010 census data, described above during septic system estimates) by the average household size in each census block. The number of people relying on WWTPs was estimated by overlaying census block information and wastewater treatment plant service area boundaries. Additionally, the USEPA Discharge Monitoring Report (DMR) Pollutant Loading Tool (cfpub.epa.gov/dmr/ez_search.cfm) was used to estimate the ratio of average annual WWTP effluent to measured baseflow. A full description of this method is provided in *Supporting Information*.

Statistical Analysis. A constant value of 1 was added to *E. coli* and *B. theta* concentrations before log transformation and analysis. Soil hydraulic conductivity values were $\log_{10}$-transformed before statistical analyses. Spearman correlation tests were used to examine relationships among physical, geochemical, and microbial measurements. Descriptive statistics were performed using IBM SPSS Statistics software (Version 19.0) with a significance threshold of (α) 0.01.

CART analysis was used to compare *E. coli* and *B. theta* (dependent variables) data to the independent geochemical, hydrologic, environmental, and land use variables. CART has been used to investigate pathogenic bacteria and parasite relationships with environmental and land use factors (56), to classify lakes based on chemistry and clarity (57), and to predict the occurrence of fecal indicator bacteria with respect to physiochemical variables (58). CART was selected because it allows for robust nonlinear model development using multiple potentially interacting predictor variables (59) that splits dependent variables into categories based on the influence of independent variables. Following previously published methods (56, 57), CART recursively split dependent variables using a recursive partitioning algorithm (rpart) and a 10-fold cross-validation criterion. The 10-fold cross-validation approach breaks all data into 10 subsets and calculates the split based on 9 of the 10 subsets. This method is used for each group until reaching a minimum stopping criterion of five observations per subgroup.

Fully developed CART outputs often required pruning to remove insignificant splits and ensure significant variable associations were not missed due to the splitting and stopping criteria (60). We first pruned CART outputs using the 1-SE rule (61–63), and, if needed, a subsequent pruning step was performed if splits did not reduce error by 5% or more. This rule minimized the cross-validated error of the model, which has been shown to produce optimal sized trees that are stable across replications (61, 64).

Detailed CART outputs were investigated to identify competitor and surrogate variables for each node. Competitor splits are ranked according to the reduction in model error from other potential splits, whereas surrogate splits are ranked according to how similar the resultant groups are relative to the primary split groups. Model accuracy was assessed by summing the proportional reduction of error from each split. All CART analyses were performed using the R software system (R Foundation for Statistical Computing).

To compare concentrations of the two organisms at each site relative to the average concentration of each organism, the Z-score of each sample was calculated. Z-scores [(observed − mean)/SD] for *E. coli* and *B. theta* were calculated in R using the "scale (dataset, center=TRUE, scale=TRUE)" command. This is defined as the sample concentration minus the mean of the population divided by the SD of the population. In this case, the Z-score of the log-transformed concentration was calculated. Positive Z-scores indicate samples with concentrations greater than the population mean, whereas negative Z-scores indicate the opposite. A CART analysis of the difference in Z-scores, calculated as *E. coli* − *B. theta*, was then performed using the same set of predictor variables in the single-organism models.

ACKNOWLEDGMENTS. We thank Steve Hamilton, Emily Luscz, Bobby Chrisman, Rebecca Ives, Sarah AcMoody, and Seth Hunt for vital support during this project and Drs. Shannon Briggs and Jon Bartholic for their technical support during the development of this manuscript. Partial funding for this project came from National Oceanic and Atmospheric Administration Great Lakes Environmental Research Laboratory Grant "Land Use Change and Agricultural Lands Indicators and Tipping Points." Partial support was also provided by the Environmental Protection Agency Grants 112013 and 118539.

1. Breuer L, et al. (2009) Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use. *Adv Water Resour* 32(2):129–146.
2. Vörösmarty C, Dork S (2000) Anthropogenic disturbance of the terrestrial water cycle. *Bioscience* 50(9):753–765.
3. Germer S, et al. (2009) Implications of long-term land-use change for the hydrology and solute budgets of small catchments in Amazonia. *J Hydrol (Amst)* 364(3-4): 349–363.
4. Arnold CL, Gibbons CJ (1996) Impervious surface coverage: The emergence of a key environmental indicator. *J Am Plann Assoc* 62(2):243–258.
5. Falkenmark M (2011) Water—A reflection of land use: Understanding of water pathways and quality genesis. *Int J Water Resour Dev* 27(1):13–32.
6. Evers DC, et al. (2011) Mercury in the Great Lakes region: Bioaccumulation, spatiotemporal patterns, ecological risks, and policy. *Ecotoxicology* 20(7):1487–1499.
7. Katz DM, Watts FJ, Burroughs ER, Daniel B, Associate MK (1995) Effects of surface roughness and rainfall impact on overland flow. *J Hydraul Eng* 121(7):546–553.
8. Ray DK, Duckles JM, Pijanowski BC (2010) The impact of future land use scenarios on runoff volumes in the Muskegon River Watershed. *Environ Manage* 46(3):351–366.
9. Cha Y, Stow CA, Reckhow KH, DeMarchi C, Johengen TH (2010) Phosphorus load estimation in the Saginaw River, MI using a Bayesian hierarchical/multilevel model. *Water Res* 44(10):3270–3282.
10. Kistemann T, et al. (2002) Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. *Appl Environ Microbiol* 68(5):2188–2197.
11. Wong M, et al. (2009) Evaluation of public health risks at recreational beaches in Lake Michigan via detection of enteric viruses and a human-specific bacteriological marker. *Water Res* 43(4):1137–1149.
12. Almeida C, Soares F (2012) Microbiological monitoring of bivalves from the Ria Formosa Lagoon (south coast of Portugal): A 20 years of sanitary survey. *Mar Pollut Bull* 64(2):252–262.
13. Soranno PA, et al. (2011) Quantifying regional reference conditions for freshwater ecosystem management: A comparison of approaches and future research needs. *Lake Reservior Manage* 27(2):138–148.
14. Grayson RB, Gippel CJ, Finlayson BL, Hart BT, Gippep CJ (1997) Catchment-wide impacts on water quality: The use of "snapshot" sampling during stable flow. *J Hydrol (Amst)* 199(1-2):121–134.
15. Vega M, Pardo R, Barrado E, Debán L (1998) Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res* 32(12): 3581–3592.
16. Wheeler E, Burke J, Hagan E, Alm EW (2006) Persistence and potential growth of the fecal indicator bacteria, *Escherichia coli*, in shoreline sand at Lake Huron. *J Great Lakes Res* 32(2):401–405.

17. Peed LA, et al. (2011) Combining land use information and small stream sampling with PCR-based methods for better characterization of diffuse sources of human fecal pollution. *Environ Sci Technol* 45(13):5652–5659.
18. Furtula V, et al. (2012) Inorganic nitrogen, sterols and bacterial source tracking as tools to characterize water quality and possible contamination sources in surface water. *Water Res* 46(4):1079–1092.
19. Yampara-Iquise H, Zheng G, Jones JE, Carson CA (2008) Use of a *Bacteroides thetaiotaomicron*-specific alpha-1-6, mannanase quantitative PCR to detect human faecal pollution in water. *J Appl Microbiol* 105(5):1686–1693.
20. Srinivasan S, Aslan A, Xagoraraki I, Alocilja E, Rose JB (2011) *Escherichia coli*, enterococci, and *Bacteroides thetaiotaomicron* qPCR signals through wastewater and septage treatment. *Water Res* 45(8):2561–2572.
21. Aslan A, Rose JB (2013) Evaluation of the host specificity of *Bacteroides thetaiotaomicron* alpha-1-6, mannanase gene as a sewage marker. *Lett Appl Microbiol* 56(1): 51–56.
22. Layton BA, et al. (2013) Performance of human fecal anaerobe-associated PCR-based assays in a multi-laboratory method evaluation study. *Water Res* 47(18):6897–6908.
23. Reynoldson TB, Norris RH, Resh VH, Day KE, Rosenberg DM (1997) The reference condition: A comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *J N Am Benthol Soc* 16(4): 833–852.
24. Davies SP, Jackson SK (2006) The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. *Ecol Appl* 16(4):1251–1266.
25. Carlisle DM, Hawkins CP, Meador MR, Potapova M, Falcone J (2008) Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. *J N Am Benthol Soc* 27(1):16–37.
26. Tiefenthaler LL, Stein ED, Lyon GS (2009) Fecal indicator bacteria (FIB) levels during dry weather from Southern California reference streams. *Environ Monit Assess* 155(1-4):477–492.
27. United States Environmental Protection Agency (2012) *Recreational Water Quality Criteria* (US Environmental Protection Agency, Washington, DC).
28. Wade TJ, et al. (2006) Rapidly measured indicators of recreational water quality are predictive of swimming-associated gastrointestinal illness. *Environ Health Perspect* 114(1):24–28.
29. Wymer LJ, Wade TJ, Dufour AP (2013) Equivalency of risk for a modified health endpoint: A case from recreational water epidemiology studies. *BMC Public Health* 13:459.
30. Byappanahalli M, Fowler M, Shively D, Whitman R (2003) Ubiquity and persistence of *Escherichia coli* in a Midwestern coastal stream. *Appl Environ Microbiol* 69(8): 4549–4555.
31. Byappanahalli MN, Whitman RL, Shively DA, Sadowsky MJ, Ishii S (2006) Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. *Environ Microbiol* 8(3):504–513.
32. Nevers MB, Whitman RL, Frick WE, Ge Z (2007) Interaction and influence of two creeks on *Escherichia coli* concentrations of nearby beaches: Exploration of predictability and mechanisms. *J Environ Qual* 36(5):1338–1345.
33. Wade TJ, Pai N, Eisenberg JNS, Colford JM, Jr (2003) Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. *Environ Health Perspect* 111(8): 1102–1109.
34. Russo RC (2002) Development of marine water quality criteria for the USA. *Mar Pollut Bull* 45(1-12):84–91.
35. Luscz EC, Kendall AD, Hyndman DW (2015) High resolution spatially explicit nutrient source models for the Lower Peninsula of Michigan. *J Great Lakes Res* 41(2):618–629.
36. Sowah R, Zhang H, Radcliffe D, Bauske E, Habteselassie MY (2014) Evaluating the influence of septic systems and watershed characteristics on stream faecal pollution in suburban watersheds in Georgia, USA. *J Appl Microbiol* 117(5):1500–1512.
37. Oliver CW, et al. (2014) Quantifying the contribution of on-site wastewater treatment systems to stream discharge using the SWAT model. *J Environ Qual* 43(2):539–548.
38. Carrillo M, Estrada E, Hazen TC (1985) Survival and enumeration of the fecal indicators Bifidobacterium adolescentis and *Escherichia coli* in a tropical rain forest watershed. *Appl Environ Microbiol* 50(2):468–476.
39. Barry-Eaton District Health Department (2011) Time of Sale or Transfer (TOST) Program: The first three years. Available at www.barryeatonhealth.org/Portals/9/EH/FIRST THREE YEARS OF TOST.pdf. Accessed June 8, 2015.
40. Michigan Department of Environmental Quality (2009) *Michigan's Nonpoint Source Program Plan* (Michigan Department of Environmental Quality, Lansing, MI).
41. Habteselassie MY, et al. (2011) Tracking microbial transport through four onsite wastewater treatment systems to receiving waters in eastern North Carolina. *J Appl Microbiol* 111(4):835–847.
42. Ishii S, Sadowsky MJ (2008) *Escherichia coli* in the Environment: Implications for Water Quality and Human Health. *Microbes Environ* 23(2):101–108.
43. Kobayashi H, Pohjanvirta T, Pelkonen S (2002) Prevalence and characteristics of intimin- and Shiga toxin-producing *Escherichia coli* from gulls, pigeons and broilers in Finland. *J Vet Med Sci* 64(11):1071–1073.
44. Fogarty LR, Haack SK, Wolcott MJ, Whitman RL (2003) Abundance and characteristics of the recreational water quality indicator bacteria *Escherichia coli* and enterococci in gull faeces. *J Appl Microbiol* 94(5):865–878.
45. Wilkinson J, Jenkins A, Wyer M, Kay D (1995) Modelling faecal coliform dynamics in streams and rivers. *Water Res* 29(3):847–855.
46. Ballesté E, Blanch AR (2010) Persistence of Bacteroides species populations in a river as measured by molecular and culture techniques. *Appl Environ Microbiol* 76(22): 7608–7616.
47. Stiles T (2014) Lightening a new candle: A new long-term vision for the Clean Water Act Section 303(d) program. *Water Resour IMPACT* 16(4):3–8.
48. Poor CJ, McDonnell JJ (2007) The effects of land use on stream nitrate dynamics. *J Hydrol (Amst)* 332(1-2):54–68.
49. Wayland KG, et al. (2003) Identifying relationships between baseflow geochemistry and land use with synoptic sampling and R-mode factor analysis. *J Environ Qual* 32(1): 180–190.
50. Boehm AB, et al. (2013) Performance of forty-one microbial source tracking methods: A twenty-seven lab evaluation study. *Water Res* 47(18):6812–6828.
51. Van De Werfhorst LC, Sercu B, Holden PA (2011) Comparison of the host specificities of two bacteroidales quantitative PCR assays used for tracking human fecal contamination. *Appl Environ Microbiol* 77(17):6258–6260.
52. 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):1–40.
53. Mueller DS, Wagner CR, Rehmel MS, Oberg KA, Rainville F (2013) Measuring discharge with acoustic Doppler current profilers from a moving boat. *Book 3: Applications of Hydraulics* (US Geological Survey, Reston, VA), Sect A, Chap 22, Version 2.0, p 95.
54. Jarrett RD (1991) Wading measurements of vertical velocity profiles. *Geomorphology* 4(3-4):243–247.
55. Anderson JR, Hardy EE, Roach JT, Witmer RE (1976) *A Land Use and Land Cover Classification System for Use with Remote Sensor Data* (US Government Printing Office, Washington, DC).
56. Wilkes G, et al. (2011) Associations among pathogenic bacteria, parasites, and environmental and land use factors in multiple mixed-use watersheds. *Water Res* 45(18): 5807–5825.
57. Martin SL, Soranno PA, Bremigan MT, Cheruvelil KS (2011) Comparing hydrogeomorphic approaches to lake classification. *Environ Manage* 48(5):957–974.
58. Bae HK, Olson BH, Hsu K-L, Sorooshian S (2010) Classification and regression tree (CART) analysis for indicator bacterial concentration prediction for a Californian coastal area. *Water Sci Technol* 61(2):545–553.
59. De'ath G (2002) Multivariate regression trees: A new technique for modeling species–environment relationships. *Ecology* 83(4):1105–1117.
60. Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W (2003) Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Ann Behav Med* 26(3):172–181.
61. Breiman L, Friedman J, Olshen R, Stone C (1984) *Classification and Regression Trees* (Chapman and Hall/CRC, New York), 1st Ed.
62. Venables WN, Ripley BD (2001) *Modern Applied Statistics With S-Plus* (Springer, New York).
63. De'ath G, Fabricius K (2000) Classification And Regression Trees: A powerful yet simple technique for ecological data analysis. *Ecology* 81(11):3178–3192.
64. Questier F, Put R, Coomans D, Walczak B, Vander Heyden Y (2005) The use of CART and multivariate regression trees for supervised and unsupervised feature selection. *Chemom Intell Lab Syst* 76(1):45–54.
65. Wetzel R, Likens G (2000) *Limnological Analyses* (Springer, New York), 3rd Ed.
66. Hamilton SK, Bruesewitz DA, Horst GP, Weed DB, Sarnelle O (2009) Biogenic calcite–phosphorus precipitation as a negative feedback to lake eutrophication. *Can J Fish Aquat Sci* 66(2):343–350.
67. Crumpton W, Isenhart T, Mitchell P (1992) Nitrate and organic N analyses with second-derivative spectroscopy. *Limnol Oceanogr* 37(4):907–913.
68. Clesceri LS, Greenberg AE, Eaton AD, eds (1998) *Standard Methods for the Examination of Water and Wastewater* (United Book, Baltimore), 20th Ed.