2011 Project Abstract
For the Period Ending June 30, 2014

PROJECT TITLE: Mississippi River Water Quality Assessment
PROJECT MANAGER: Michael Sadowsky
AFFILIATION: University of Minnesota
MAILING ADDRESS: 140 Gortner Lab, 1479 Gortner Ave
CITY/STATE/ZIP: Saint Paul, MN 55108
PHONE: (612) 626-0977
E-MAIL: Sadowsky@umn.edu
WEBSITE: http://www.cbs.umn.edu/main/news/inthefield/m3p.shtml
FUNDING SOURCE: Environment and Natural Resources Trust Fund
LEGAL CITATION: M.L. 2011, First Special Session, Chp. 2, Art.3, Sec. 2, Subd. 05c

APPROPRIATION AMOUNT: $ 557,000

Overall Project Outcome and Results

A metagenomics-based sequencing approach was utilized to characterize the bacterial community at sites along the Mississippi River in Minnesota to understand how these communities were influenced by or indicative of water quality. Results of this study revealed that the bacterial community throughout the river primarily consisted of a small number of highly abundant species that comprise a "core microbial community" that was stable both in terms of community membership and inferred functional traits. Variation in community membership and species abundances were primarily influenced by physicochemical parameters (*e.g.* pH and temperature) rather than spatial distance, and a reproducible community structure occurred annually toward the late summer. Furthermore, specific bacterial orders were related to chemical concentrations that co-varied with surrounding land use, suggesting that increases in abundance of these orders may be indicative of specific types of contamination throughout the river. Therefore, assessment of the total bacterial community provides more information about water quality and contamination sources than could be previously gleaned from traditional enumeration of indicator bacteria like *Escherichia coli*. In addition to these findings, construction of fosmid libraries to assess resistance of the bacterial community to antibiotics and heavy metals revealed that levels of resistance to both were low throughout the river. Municipal wastewater treatment was not associated with increased antibiotic resistance, but proximity to agricultural wastewater increased the frequency of resistance to the antibiotics kanamycin and ampicillin. Furthermore, the resistances to the heavy metals Cd and Cr were significantly elevated in primarily developed (urban) areas. These results indicate the influence of anthropogenic contaminants on the distribution of functional traits throughout the river. Results of this project as well as dissemination of these results are further discussed in an attached Final Report.

Project Results Use and Dissemination

Results of this study have been presented at national meetings of the American Society for Microbiology and submitted to peer-reviewed scientific journals for publication. In addition, exhibits have been prepared at the Bell Museum, the Science Museum of Minnesota, and Itasca State Park to inform the general community about the findings of this study. Summer workshops were also held in order to disseminate details of the methodology used in this study to high school teachers.

Environment and Natural Resources Trust Fund (ENRTF)
M.L. 2011 Work Plan Final Report

Date of Status Update: **9/3/14**

Final Report

Date of Work Plan Approval: 6/23/2011

Project Completion Date: 6/30/2014 **Is this an amendment request?** Yes _No_ X __

Project Title: Mississippi River Water Quality Assessment

Project Manager: Michael Sadowsky

Affiliation: U of MN

Address: 140 Gortner Lab, 1479 Gortner Ave

City: St Paul **State:** MN **Zipcode:** 55108

Telephone Number: (612) 626-0977

Email Address: sadowsky@umn.edu

Web Address: http://www.cbs.umn.edu/main/news/inthefield/m3p.shtml

Location:

Counties Impacted: Statewide

Ecological Section Impacted: Lake Agassiz Aspen Parklands (223N), Minnesota and Northeast Iowa Morainal (222M), North Central Glaciated Plains (251B), Northern Minnesota and Ontario Peatlands (212M), Northern Minnesota Drift and lake Plains (212N), Northern Superior Uplands (212L), Paleozoic Plateau (222L), Red River Valley (251A), Southern Superior Uplands (212J), Western Superior Uplands (212K)

Total ENRTF Project Budget:

ENRTF Appropriation:	$557,000
Amount Spent:	$499,801
Balance:	$ 57,199

Legal Citation: M.L. 2011, First Special Session, Chp. 2, Art.3, Sec. 2, Subd. 05c

Appropriation Language:
$278,000 the first year and $279,000 the second year are from the trust fund to the Board of Regents of the University of Minnesota to assess water quality in the Mississippi River using DNA sequencing approaches and chemical analyses. The assessments shall be incorporated into a Web-based educational tool for use in classrooms and public exhibits. This appropriation is available until June 30, 2014, by which time the project must be completed and final products delivered.

I. PROJECT TITLE: Mississippi River Water Quality Assessment

II. FINAL PROJECT STATEMENT:

A metagenomics-based sequencing approach was utilized to characterize the bacterial community at sites along the Mississippi River in Minnesota to understand how these communities were influenced by or indicative of water quality. Results of this study revealed that the bacterial community throughout the river primarily consisted of a small number of highly abundant species that comprise a "core microbial community" that was stable both in terms of community membership and inferred functional traits. Variation in community membership and species abundances were primarily influenced by physicochemical parameters (e.g. pH and temperature) rather than spatial distance, and a reproducible community structure occurred annually toward the late summer. Furthermore, specific bacterial orders were related to chemical concentrations that co-varied with surrounding land use, suggesting that increases in abundance of these orders may be indicative of specific types of contamination throughout the river. Therefore, assessment of the total bacterial community provides more information about water quality and contamination sources than could be previously gleaned from traditional enumeration of indicator bacteria like Escherichia coli. In addition to these findings, construction of fosmid libraries to assess resistance of the bacterial community to antibiotics and heavy metals revealed that levels of resistance to both were low throughout the river. Municipal wastewater treatment was not associated with increased antibiotic resistance, but proximity to agricultural wastewater increased the frequency of resistance to the antibiotics kanamycin and ampicillin. Furthermore, the resistances to the heavy metals Cd and Cr were significantly elevated in primarily developed (urban) areas. These results indicate the influence of anthropogenic contaminants on the distribution of functional traits throughout the river. Results of this project as well as dissemination of these results are further discussed in an attached Final Report.

Concurrent with the research on the microbial communities of the Mississippi River, the project also engaged high school teachers through two summer workshops and engaged the public through exhibits at the Science Museum of Minnesota, the Bell Museum of Natural History, and Itasca State Park. These exhibits will continue to be on display past the end of this funding.

III. PROJECT STATUS UPDATES:

Project Status as of December 31, 2011:

Work on this project began in earnest. Our first round of water samples from the Mississippi and Zumbro rivers were collected and processed in the lab during the summer of 2011. In addition to the 11 different sites we sampled, we also took water samples from two depths at one site (Hidden falls) every two weeks for DNA sequence analysis. This was done to access variability in sequence data over time and by depth. Water chemistry (nutrients, metals, and chemicals) and physical site data were also taken for all samples and are currently being analyzed. DNA sequencing and fosmid library production are currently underway and will be completed within 8 weeks. Sequence analysis will start right after this.

We have hired a laboratory technician (Trevor Gould) who is working on the data analysis and we have hired a post-doc who will begin work on the project in April 2012.

Work has begun on developing teacher workshops for August 2012 and 2013. Materials for the workshops are being developed. Shotgun gene sequencing of several samples is being used to produce sequence assemblies for teacher and student bioinformatics activities. Infrastructure and software for

hosting bioinformatics activities is in place and the interface is being developed for the teacher workshops. The workshop coordinators are creating modules for student use and identifying textbook materials for use by teachers in the classroom. Because this planning is not yet complete, the design for a website to host the materials has been postponed until spring semester, 2012.

Coordination of exhibits at the Science Museum of Minnesota, the Bell Museum of Natural History and at Itasca State Park has begun and all parties are excited to work together in developing this material for the public. We will be hosting a coordination meeting for all museum personnel in January 2012.

Project Status as of June 30, 2012:

Because of the State Shutdown last July, some of the work on this project has been delayed. However, despite this setback, we are well on track for completing the project on time. Our first round of water samples from the Mississippi and Zumbro rivers and the water samples from two depths at one site (Hidden falls) are well into processing, although the bills for these have not yet shown up on our University account. Comparative analysis of water chemistry (nutrients, metals, and chemicals), physical site data, and microbial populations is currently underway. Some of the data from our analyses can be viewed on the project website, http://www.cbs.umn.edu/m3p , and is being prepared for publication.

We have just hired a postdoctoral student (Christopher Staley) who is working on the data analysis and publications.

Work on the teacher workshops is progressing well. Materials for the workshops are being developed. Shotgun gene sequencing of several samples is being used to produce sequence assemblies for teacher and student bioinformatics activities. Infrastructure and software for hosting bioinformatics activities is in place and the interface is in final development for the teacher workshops. Two teacher co-leaders have been hired (Karen Casper from Saint Paul Public School district and Mary Raab from Minneapolis Public School district) and have been meeting with the course instructor, Brian Gibbens. The workshop is scheduled for July 30-August 3 on the St. Paul campus of the University of Minnesota. The costs of this workshop will show up on the December, 2012 update report.

Coordination of exhibits at the Science Museum of Minnesota, the Bell Museum of Natural History and at Itasca State Park continues to progress. The final contract with the Science Museum of Minnesota has just been signed and we already have a display up at Itasca State Park. Interestingly, our request for exhibits has been a catalyst for some exciting new ideas at all three venues that will highlight Minnesota waterways and their importance in our lives.

Project Status as of December 31, 2012:

The project is on track, and great progress was made on all sections.

We continue to analyze data obtained from the summers of 2011 and 2012 sampling of the Mississippi River. This analysis includes three parts: analysis of the microbial diversity from the ribosomal DNA sequencing, analysis of gene function through gene expression in the fosmid clones, and search for unexpressed genes through comparisons of data from shotgun sequencing with existing databases of DNA sequences. We have one paper submitted to the journal Applied and Environmental Microbiology and are awaiting results of its review (see attached).

Our teacher workshop held in August 2012 was successful with 14 enrolled teachers. The teachers were from 13 different high schools, representing 10 different school districts. More detailed information is provided in the Activity 2 section below.

Outreach to the general public is also progressing. The exhibit at the Mary Gibbs Mississippi Headwaters Center at Itasca State Park is in its final stages of design and production, with the goal to have it installed by March when the ground thaws. There also is a poster of the project on display at the Nature Center (Jacob V. Brower Visitor Center and Gift Shop) at the entrance of the park.

The exhibit at the Science Museum of Minnesota is also in development, with a launch on Science Buzz scheduled within the next month or so. The video exhibit at the Bell Museum has been modified to increase visibility and we have the final contract estimate.

Amendment Request (01/07/2013): Amendment approved by the LCCMR on January 14, 2013. We request movement of funds ($35,000) from the Professional/Technical/Contracts (Activity 1: Analysis of Microorganisms and Metagenomics) to the Equipment/Tools/Supplies to purchase additional lab materials for functional genomic analyses ($25,000, Activity 1: Analysis of Microorganisms and Metagenomics) and to enhance our public exhibit at the Bell Museum of Natural History ($10,000, Activity 3: Project Data Dissemination.) The functional genomic analyses have been more expensive than originally anticipated because of an expansion of the range and depth of analyses over what we thought we could accomplish. This has provided us with a lot more in depth information and this point was raised by one of the initial outside reviewers of this proposal. The video exhibit at the Bell Museum has been modified to increase visibility and we have the final contract estimate that is above the original budget request.

This shift in funds is now possible because the University of Minnesota is subsidizing sequencing costs - making this part of our project less expensive than originally budgeted. Consequently, despite this requested shift in budget items we will deliver more data, provide more visibility and enhance dissemination activities. These requested changes are also shown as a budget revision in Attachment A.

While there is a shift in funds, there are no changes in the budget items and no new budget items have been added. However the line in the Attachment A to which the functional analysis money will be spent currently lists “Laboratory supplies for filtering, cultures, genome preps of river samples, “ even though our project description clearly indicates that our proposal is to fund functional analyses (from Section IV. Project Activities and Outcomes. Activity 1: “Library clones will be picked into 384 well microplates using a Qbot colony picking robot and screened, by students, for functionally active genes involved in resistance to antibiotics and heavy metals, and those that encode for degradation of recalcitrant organic compounds. Functionally active fosmid clones will be sequenced at the Biomedical Genomics Center at the University of Minnesota, and this sequence data will be assembled into contigs and analyzed by Blast and IMG-ACT software and websites (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi). Based on these needs and our prior wording, we have reworded the line in Attachment A to read as follows: “Laboratory supplies for filtering, cultures, genome preps of river samples, including functional analyses of clones”, which was our original intention and goal. **This Amendment request was approved by the LCCMR on January 14, 2013.**

Project Status as of June 30, 2013

The project continues to stay on track, and progress was made on all tasks and activities. We continue to analyze data obtained from the summers of 2011 and 2012 sampling of the Mississippi River. We

have begun some sampling in 2013 to verify results and to increase the validity and confidence in the data. We are doing this to also increase replication and allow statistical analysis of results. This analysis includes three parts: analysis of the microbial diversity from rDNA sequencing, analysis of gene function through gene expression in the fosmid clones, and search for unexpressed genes through comparisons of data from shotgun sequencing with existing databases of DNA sequences. We have also received partial data from chemical analyses of the water samples obtained in 2011 and 2012 and anticipate having two years of chemical analyses completed by the end of July 2013.

Our teacher workshop will be held in early August 2013 now has 19 teacher participants signed up. However, since the workshop will not be held until August, this number is likely to increase slightly. The teachers currently are from 14 different high schools, representing 12 different school districts.

Outreach to the general public is also progressing very well. The exhibit at the Mary Gibbs Mississippi Headwaters Center at Itasca State Park is complete with the goal to have it installed by the July 4th weekend. This is slightly delayed from the original plan because of weather issues at the State Park. The cost of this exhibit will be invoiced in July so this will show up on the budget in the December 2013 report. As mentioned before, there also is a poster of the project on display at the Nature Center (Jacob V. Brower Visitor Center and Gift Shop) at the entrance of the park.

The exhibit at the Science Museum of Minnesota is complete. This exhibit can be seen at Science Buzz here: http://www.sciencebuzz.org/topics/m3p. We will make all teacher workshop materials web available at the end of this summer's teacher workshop. The video exhibit at the Bell Museum is almost complete and is in its final stages of video editing. We have reviewed it and are very excited about how well it explains the project and its impact to the general public.

Amendment Request (06/27/2013): We request movement of funds ($37,500) from Professional/ Technical Contracts (the genome sequencing budget) to Personnel (Wages and Benefits) in order to extend the contract of Technician Trevor Gould, until March 23, 2014. This will enable him to finish sequence analyses and the meta-analysis of climatic, edaphic, and chemical data needed to complete this project. The request to shift funds will not impact our ability to finish the sequencing itself, and is due in large part to continued cost reductions. We are now doing more sequencing than originally envisioned! Some of the extra sequence runs were originally requested by the reviewers of our proposal and are now present in our nearly accepted manuscript (see attached).
Amendment Request Approved by LCCMR on July 9, 2013.

Project Status as of December 31, 2013
The project is nearing completion in about 6 months. We are doing some final analyses of sequence data already acquired, have another paper that has been submitted for publication, continue to work on the functional analyses of the fosmid clones. In addition, we have some additional sequencing left to do on the 2011 and 2102 samples using new DNA primers, and have developed a plan to obtain the final information to strengthen the project's outcomes. This plan is given in more detail in Activity 1 below.

The teacher professional development program is completed. This past summer's workshop served 15 teacher participants from around the state. The post-workshop assessment indicated that the workshop was successful in introducing teachers to metagenomics, the importance of understanding the microbial populations in river systems, and the basic tools of studying these populations. Additional information is provided in Activity 2 below.

All the outreach activities are now installed and open to the public. The video presentation is in place at the Bell Museum of Natural History on the East Bank campus of the University; preliminary input from

viewers has been very positive and we invite all the members of the Commission to visit the Bell Museum. The exhibit is also installed at Itasca State Park at the Mary Gibbs Mississippi Headwaters Center. It is installed on the plaza directly outside the gift store and restaurant, an area with very high traffic volume. We are very pleased with the results of both of these final parts of our outreach work.

Amendment Request (01/08/2014): We request permission to use $1,500 of the personnel funds in Activity 2 to update the website. Activity 2 is where we budgeted the original cost of development of the website. We also request movement of the remaining personnel funds ($9,654) from Activity 2 to Activity 1 in order to extend the contract of Technician Trevor Gould to finish the functional analyses of fosmid clones. This analysis is taking more time than anticipated, but is an important part of reporting the microbial populations of antibiotic resistance organisms at the various sampling sites. The request to shift funds will not impact our ability to finish any other part of the project.

Amendment request approved by the LCCMR January 27, 2014

Related to this amendment request, and as requested of notification by the LCCMR, we will have approximately $35,000 remaining after all costs have been accounted for. These funds are available for the LCCMR.

Project Status as of June 30, 2014
The project is complete. Specifics on each Activity are given below. Because of continued savings on DNA sequencing costs, the amount of funding we will be returning to LCCMR is $57,199. Note that we were able to complete all facets of the project successfully within the amount spent.

IV. PROJECT ACTIVITIES AND OUTCOMES: More specific details of the project activities are provided in the attached addendum.

ACTIVITY 1: Analysis of Microorganisms and Metagenomics

Description:
This proposal will fund two and a half years of sampling and metagenome and chemical analysis of water samples from the Mississippi River at 11 critical junctures in Minnesota, from Lake Itasca to La Crescent, focusing on the headwaters and confluences with other major rivers. We are currently obtaining preliminary data from 10 of these 11 sites sampled this last summer, and are requesting funding here for in-depth studies of these 11 sites for two additional years. Surface water samples at each site will be analyzed twice per year and sediment samples at each site will be analyzed once per year. We also request funding for additional water sampling at one site to obtain information concerning the temporal and spatial variability of the microbial populations. At each sampling location we will also obtain information on other indicators of water quality, including industrial and agricultural chemicals and pharmaceuticals, inputs that impact bacterial diversity and at functionality.

The following information is abstracted from the attached addendum and the reader is directed there for more detailed information about each activity.

Metagenome Analysis

The Mississippi River will be sampled twice yearly at 11 sites from Lake Itasca to La Crescent during years 1 and 2 (see attached map), with sediment sampling done once yearly. We will also sample site #4

(Hidden Falls) 6 times per year (biweekly from May to August) at two sampling depths (0.3 and 1 meter below the surface). This will allow us to obtain information concerning the temporal and spatial variability of the microbial populations in the Minnesota River. The exact locations (latitude and longitude) of sampling sites at each location, when possible, will be the same as those used by MPCA and the Met Council to allow comparisons to existing data and those obtained in the future. At each site, two 1 L samples will be taken for water chemistry analysis (see below) and a 40L sample will be taken for metagenomic analyses. Total DNA will be extracted from cell pellets using Bio101 FP120 Fastprep instrument and MoBio Powersoil DNA extraction kits (Mo Bio Laboratories, Solana Beach, CA) as previously described (Ishii et al. 2006) and DNA corresponding to the V6 hypervariable region of the full-length 16S rDNA will be amplified by PCR using primers as described by Wang et al. 2007 and Lazarevic et al. 2009. The PCR primers will contain a unique sequence tag (Binladen et al. 2007 and the amplicons from each of the 11 samples will be pooled together and the multiplexed amplicons will be sequenced on a Illumina/Solexa Sequencer at the National Center for Genomic Research (NCGR) in Santa Fe, New Mexico. The 16S rDNA sequence data obtained in our studies will be compared to V6 reference databases as described by Dethlefsen et al. 2008 and Lazarevic et al, 2009. The taxonomic classification of 16S rDNA PCR products will be assigned using the GAST (Global Alignment for Sequence Taxonomy) taxonomic classification tool as described by Sogin et al. (2006), and by analyses done using reference database of V6 rDNA sequences (RefHVR_V6) from SILVA (Pruesse et al., 2007), the taxonomy from known cultured isolates, the Entrez Genome and the Ribosomal Database Project (Cole et al. 2009), Greengenes (DeSantis et al., 2006) and the software program ARB (Ludwig et al. 2004). The resulting phylogenetic relationships that are identified will be tested by maximum-likelihood bootstrap trees (with 1000 iterations) using distance-based subsampling and a minimum distance of 3% between sequences. Operational taxonomic units (OTUs) will be determined and compared by using the sortx subroutine of XplorSeq (Frank 2008). Comparisons of bacterial constituents of the river, between sites and sampling dates, will be determined by examining the numbers and types of phyla (or operational taxonomic units) at each sample site. We will also determine species diversity, species richness, and evenness using rarefaction analysis (Robertson et al. 2009).

In addition to phylogenetic information, our metagenomic analyses will also examine the functionality of the bacterial community of the Mississippi at each sampling site. To do this, we will send one half of the frozen cells from each site (as described above) to the Clemson University Genome Institute (http://www.genome.clemson.edu/) for the construction of functional gene libraries. This will be done only for the samples obtained once per year at each site. The libraries, consisting of randomly sheared metagenomic DNA, will be constructed in fosmid vector pEPIFOS-5 and transformed into *E. coli* DH10 as the host. Each fosmid will have an average insert size of ~39 kb (we have made libraries for year -1 already and have these data), enough to encode to about 20-40 bacterial genes. We will obtain ~10,000 clones (containing about 390,000 kb of DNA) from each sample. Library clones will be picked into 384 well microplates using a Qbot colony picking robot and screened, by students, for functionally active genes involved in resistance to antibiotics and heavy metals, and those that encode for degradation of recalcitrant organic compounds. Functionally active fosmid clones will be sequenced at the Biomedical Genomics Center at the University of Minnesota, and this sequence data will be assembled into contigs and analyzed by Blast and IMG-ACT software and websites (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi).

Water Analysis

We will also obtain data on the presence and concentration of chemical compounds known to influence human health and water quality. At each site we will examine for the presence of the following compounds and chemicals: **pharmaceuticals** – acetaminophen and caffeine; **antibiotics** - tylosin, erythromycin, and trimethoprim; **pesticides** - atrazine, acetochlor, and metolachlor; **personal care**

products - DEET, triclosan, and nonylphenol; and **endocrine disrupters** - trenbolone and estradiol. The data obtained from these analyses will be entered into the same relational database described above so that students and researchers can examine the possible relationship between the chemical constituents of the Mississippi River at each site and the presence and types of microbes recovered at each site. Water samples at each site and during each sampling event will also be analyzed for standard limnological parameters, including: water pH and temperature, turbidity, and nutrient (N, P, and C) concentrations, and inorganic trace elements via ICP analysis, and *E. coli* counts. All samples used for chemical analyses will be collected in glass bottles and will be analyzed by our collaborator Dr. William Koskinen at the USDA-ARS Soil and Water Management Unit at the University of Minnesota.

Depending on the compound under analysis, extracted chemicals will also be analyzed using a Waters Alliance high performance liquid chromatography/mass spectrometer with electrospray interface operating in positive-ion (LC/MS-ESI(+)) mode, or by using a Agilent 6890 gas chromatograph with capillary column coupled to a mass selective detector (GC/MS) operating in selected ion mode. Many of the chemicals to be analyzed in this study are also being examined in a current LCCMR-funded study of the Zumbro River led by Deborah L Swackhamer at the University of Minnesota. This will allow cross comparison of results obtained at different sites along this same waterway. Water chemical and physical data obtained from each site and sampling time will added to the relational database using to store metagenomic data, allowing for analysis of the potential correlative relationship between microbial community structure and the presence of chemical contaminants.

Summary Budget Information for Activity 1:

ENRTF Budget: **$ ~~407,291~~**
$ ~~397,291~~
$ 406,945
Amount Spent: **$ 381,495**
Balance: **$ 25,450**

Activity Completion Date:

Outcome	**Completion Date**	**Budget**
1. Sampling of the Mississippi River and analysis of samples for microbial species diversity and functionality at each sampling location.	12/31/2013	~~$397,736~~ ~~$387,736~~ $397,390
2. Correlations of structural (sequence of diversity) and functional metagenome data to physical and chemical data at each location.	12/31/2013	$3,822
3. Initial development of relational web database consisting of metagenomic and physical chemical data.	12/31/2011	$1,911
4. Uploading of metagenomic data into IMG-ACT for searching and retrieval by researchers, students, river managers, regulatory agencies, and the public to better understand how human activity influences the microbiology of the Mississippi River.	12/31/2013	$3,822

Activity Status as of December 31, 2011:

Twenty-two water samples and three sediment samples were taken from each of the 11 sample sites in August and we are in the process of analyzing these samples for phylogenetic analyses (costs incurred to 12/21/11 for this analysis: $2,451.29). Chemical analysis were completed for the 11 sites and analysis is underway (costs incurred to 12/21/11 for this analysis: $1,012.25). In addition, the salary and fringe for the laboratory technician who has been doing the sampling, sample preparation, and preliminary analysis is $15,197.60 and $6,276.71, respectively, to 12/21/11. The DNA samples are at the sequencing facility and fosmid cloning facility waiting to be processed.

The postdoctoral candidate has been selected and will start work in April so no funds have been spent on that yet.

Activity Status as of June 30, 2012:

Sample collection for the 2012 summer is underway, including all 11 sample sites. Sample processing of last summer's samples continued and this summer's sample processing has begun. Total costs of processing and analysis since grant inception is $11,205.81, but some large costs for DNA analysis and fosmid preparation have not yet been billed to our University account. In addition, the salary and fringe for the laboratory technician doing the sampling, sample preparation, and preliminary analysis now totals (since grant inception) $34,540.00 and $14,265.25 respectively. As requested in our letter of May 7, 2012 to you we were required by the University to provide a 2.5% increase to the laboratory technician. This increase is now reflected in the attached revised budget. We have just hired a postdoctoral student who is already contributing to the sample analysis and has begun working on scientific publications of the data we have obtained so far. Salary and fringe for this postdoctoral student as of is $3557.70 and $721,50. Finally, $210 in travel costs have been incurred for sample acquisition.

Activity Status as of December 31, 2012:

We continue to analyze data obtained from the summers of 2011 and 2012 sampling of the Mississippi River. This analysis includes three parts: analysis of the microbial diversity from the ribosomal DNA sequencing, analysis of gene function through gene expression in the fosmid clones, and the search for unexpressed genes through comparisons of data from shotgun sequencing with existing databases of DNA sequences. Interestingly, while the cost of the analysis of gene function is higher than originally anticipated, the cost of sequencing has become lower (in great part because the University of Minnesota is subsidizing the sequencing.) Thus, we would like to move funds that had been budgeted for sequencing to support the functional analysis work. This request is given in Section III and is shown as a budget revision in Attachment A.

We have one paper submitted to the journal Applied and Environmental Microbiology and are awaiting results of its review. The paper abstract is listed below and the paper is attached as a PDF for you here.

Title: Relationship between Land Use and Anthropogenic Factors Influencing Bacterial Community Structure in the Upper Mississippi River

Authors: Christopher Staley, Tatsuya Unno1, Trevor J. Gould, Bruce Jarvis, Jane Phillips, James B. Cotner, and Michael J. Sadowsky

ABSTRACT

A 16S metagenomics-based approach was used to examine the relationship between microbial community structure, land use, and anthropogenic activity at 10 sites along the Upper Mississippi River

in 2010. Microbiota were characterized using the V6 hypervariable region of the 16S rDNA gene and Illumina next-generation sequencing technology. A total of ~ 8.8 × 106 Illumina sequence reads were obtained and 2.8 × 105 per site were used for analysis after subsampling to control for site-specific differences in read number. Sample coverage was 99.2 ± 0.2% at each site and a total of 16,400 operational taxonomic units (OTUs) were observed, with a mean of 4,594 ± 824 OTUs per sample. Approximately 97% of OTUs were classified as Bacteria, 2.9% as Archaea, and 0.03% could not be assigned. Proteobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria, and Verrucomicrobia accounted for 93.6 ± 1.3% of all OTUs. As distance from the pristine headwaters increased, significant decreases in diversity and minority phyla were observed. Rainfall and pH were positively correlated with microbial diversity, but the relative numbers of the most abundant phyla did not vary significantly among sites. Microbial community structure analyses grouped sites by land use, suggesting that human activity likely has major impacts on the river's microbiota. To our knowledge, this is among the first studies to intensively characterize a freshwater riverine bacterial community using a next-generation sequencing approach and results of this study indicate that the Mississippi River microbiome is complex and is shaped, in large part, by anthropogenic activity.

Activity Status as of June 30, 2013:

We continue to analyze data obtained from the summers of 2011 and 2012 sampling of the Mississippi River. We will also generate and analyze additional data that will be generated from samples collected in 2013 to improve confidence in the results from 2011 and 2012 sampling and to increase replications. This analysis includes three parts: analysis of the microbial diversity from the ribosomal DNA sequencing, analysis of gene function through gene expression in the fosmid clones, and the search for unexpressed genes through comparisons of data from shotgun sequencing with existing databases of DNA sequences. We have also received partial data from chemical analyses of the water samples from 2011 and 2012 and anticipate having complete chemical analyses complete by the end of July.

Our paper, previously submitted to the Journal of Applied Microbiology, has been reviewed and is generally acceptable for publication after minor revision (see attached). In addition, data from the 2011 and 2012 samplings were presented as a poster at the General Meeting of the American Society for Microbiology in May (see below) and as an invited Divisional Lecture talk to. The abstracts for the paper and poster are listed below and the complete versions are attached for you here at the end of this report:

Paper Title: Variation in the Core Microbial Community of the Upper Mississippi River in Response to Upstream Environmental and Population Impacts

Authors: Christopher Staley, Tatsuya Unno, Trevor J. Gould, Bruce Jarvis, Jane Phillips, James B. Cotner, and Michael J. Sadowsky

ABSTRACT
Aims. A deep-sequencing approach was used to characterize the microbial community at ten sites along the Upper Mississippi River to evaluate shifts in the community resulting from upstream inputs and land use changes.
Methods and Results. Microbial structure and diversity in the river was determined by using Illumina next-generation sequencing technology and the V6 hypervariable region of 16S rDNA. A total of 16,400 operational taxonomic units (OTUs) were observed (4,594 ± 824 OTUs per sample). *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, and *Verrucomicrobia* accounted for 93.6 ± 1.3% of all sequence reads, and 90.5 ± 2.5% belonged to OTUs shared among all sites (n = 552). Among non-shared sequence reads at each site, 27-51% were introduced at the headwaters and 33-49%

were associated with potentially anthropogenic impacts upstream of the second sampling site. Alpha diversity decreased with distance from the pristine headwaters while rainfall and pH were positively correlated with diversity. Replication and smaller filter pore sizes minimally influenced the characterization of community structure.
Conclusions. A "core microbial community" is present throughout the Upper Mississippi River. Shifts in structure are related to changes in the relative abundance, rather than presence/absence, of OTUs, some of which are likely introduced as a result of anthropogenic impacts.
Significance and Impact of Study. This study is among the first to characterize a riverine microbial community using a deep-sequencing approach and demonstrates that upstream influences and potentially anthropogenic impacts can influence the presence and relative abundance of OTUs downstream resulting in significant variation in community structure.

Poster Title: Relationship Between Land Use and Anthropogenic Factors Influencing Bacterial Community Structure in the Upper Mississippi River

Authors: Christopher Staley, Tatsuya Unno, Trevor J. Gould, Bruce Jarvis, Jane Phillips, James B. Cotner, and Michael J. Sadowsky

ABSTRACT
A metagenomics-based approach targeting the V6 hypervariable region of 16S rDNA was used to examine the relationship between microbial community structure (MCS) and anthropogenic activity at ten sites along the Upper Mississippi River. Further study was performed at two sites to determine the effects of sampling depth, short-term temporal variation, sample volume, and filter size on MCS in the Mississippi River. A total of 16,400 operational taxonomic units (OTUs) were observed from all sampling sites (mean of 4,594 ± 824 OTUs per sample). The majority of OTUs (93.6 ± 1.3%) were classified as *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, and *Verrucomicrobia*. Significant decreases in diversity and minority phyla were observed as distance from the pristine headwaters at Lake Itasca increased. Analyses of MCS grouped sites by land use, suggesting major anthropogenic impacts on the river's microbiota. The MCS determined did not differ significantly as a result of depth or variation over a two-month period ($P \geq 0.117$). Samples collected during the third month varied significantly as a result of depth ($P < 0.001$), and MCS was also significantly different than the previous two months ($P < 0.001$) with an increase in the relative abundance of *Cyanobacteria* and a decline in *Elusimicrobia*. The MCS from sample volumes of ≤ 2 L were not significantly different from each other but differed from the 6 L sample ($P < 0.001$), with minority taxa having higher relative abundances in the larger sample. Filtration of flow-through from a 0.45 µm filter showed that the majority of OTUs belonged to the family *Microbacteriaceae*. The results of this study indicate that the Mississippi River microbiome is complex and is shaped extensively by anthropogenic activity. Furthermore, the MCS determined is less likely to be affected by sampling protocols than differences in the physical and chemical properties of the water column.

ASM Division Q Lecturer: MICHAEL J. SADOWSKY; Univ. of Minnesota, St. Paul, MN
Presentation Title: The Impact of Human Activity on the Microbial Metagenome of the Upper Mississippi River.
Session: 151 Role of Microbes in Environmental Sustainability
Session Date & Time: Monday, May 20 | 3:00 p.m. – 5:30 p.m., Denver, CO

Activity Status as of December 30, 2013
We are now doing final analyses of sequence data already acquired, have another paper that has been submitted for publication, and continue to work on the functional analyses of the fosmid clones. We also have some additional sequencing left to do on the 2011 and 2012 samples using new DNA primers, and

have developed a plan to obtain the final information to strengthen the project's outcomes. This new information will allow us to more accurately determine the taxonomy of microbes in the river, the contribution of sediments and sands to the river's microbiota and functional diversity of genes present in these microbes. We have a new paper submitted to the ISME journal and are awaiting results of its review. The abstract is listed below and the paper is attached at the end of this report:

Title: Species Sorting Dynamics and Sediment Resuspension Alter the Core Bacterial Community of the Upper Mississippi River

Authors: Christopher Staley, Trevor J. Gould, Ping Wang, Jane Phillips, James B. Cotner, and Michael J. Sadowsky

ABSTRACT

Bacterial community structure (BCS) in freshwater communities varies seasonally and due to physicochemical gradients. Variation in metacommunity structure along a major river, however, remains understudied. Here we characterize the BCS at 11 sites along the Mississippi River and contributing rivers in Minnesota during the summers of 2010 through 2012 using Illumina next-generation sequencing. Contributions from sediment to water microbial diversity were evaluated. Long-term variation in community membership was observed, and significant shifts in the relative abundance of major freshwater taxa, including *α-Proteobacteria*, *Burkholderiales*, and *Actinomycetales*, were observed due to temporal and spatial variation as well as depth. Abundances of all phyla identified were correlated with rainfall, temperature, or pH suggesting species sorting primarily shaped metacommunity structure. Biweekly changes in bacterial communities revealed a recurrent BCS associated with samples taken in late summer in 2011 and 2012, further suggesting that seasonal dynamics strongly influence community composition. Sediment communities differed from those in the water, but contributed up to 50% to community composition in the water column. Among water sampling sites, 34% showed significant variability in BCS of replicate samples indicating variability among riverine communities due to heterogeneity in the water column. Results of this study reveal how communities in environmental reservoirs impact waterborne BCS and highlight the need for a better understanding of spatial and temporal variation in riverine bacterial diversity. Moreover, techniques used in this study may prove useful to determine sources of sediments and soils to waterways, which will facilitate best management practices and total maximum daily loading studies.

Final Report Summary: June 30, 2014

The conclusion of this project leaves us with three years of data concerning the bacterial community in the Upper Mississippi River in Minnesota (including previous work performed in 2010 using Federal stimulus funds). Our results revealed that the bacterial community in the Upper Mississippi River is largely comprised of a taxonomically and functionally conserved core community that is made up of a small number of highly abundant species, in particular members of the *Proteobacteria, Actinobacteria*, and *Bacteroidetes*, which are known to comprise ubiquitous freshwater lineages. Furthermore, we have identified that variation in community membership in response to natural and anthropogenic impacts primarily influences the distribution of rare community members that account for ~ 90% of total species diversity.

We have shown that specific types of land cover, representing various sources of anthropogenic contamination, are associated with discrete changes in bacterial community structure. Monitoring of these shifts in the bacterial community may prove more useful to resource managers in evaluating

overall water quality than traditional indicator bacteria like *E. coli*. Natural temporal and spatial variation in community structure was also observed, but was less pronounced than variation due to presumed pollution. Of note, the community structure in the river appears to reform annually in the late summer, most likely in response to temperature. Furthermore, bacterial communities in sands and sediments may also be responsible for shifts in community structure when these reservoirs are disturbed by climatic and/or anthropogenic mechanisms.

Finally, despite variation in community structure, the functions performed by the bacterial community (*e.g.* metabolism and cell signaling) are highly conserved throughout Minnesota, although important variations in function can result from changes in land cover type, suggesting an adaptation of the community to various pollutants. Functional screening of fosmid libraries further supported this as phenotypic patterns of antibiotic and heavy metal resistances corresponded to changes in nutrient/chemical concentrations and major surrounding land cover.

Our study represents one of the largest t efforts, currently, to understand how bacterial communities in a major river change and adapt to natural and anthropogenic disturbances. Our findings, in addition to helping us understand the microbial ecology of this ecosystem, provide an essential basis to assist resource managers preserve and protect this vital ecosystem as well as public health.

ACTIVITY 2: Professional Development of Grade 7-12 Teachers

Description: In this result we will develop a hands-on professional development program for G7-12 teachers, offered both in the Twin Cities and in Northwest Minnesota (Itasca) to provide greater access to this opportunity statewide. This professional development program will focus on preparing teachers to include Mississippi metagenomics studies in their science curriculum in a way that meets state standards for science inquiry.

The following information is abstracted from the attached addendum and the reader is directed there for more detailed information about each activity.

Learning science requires building a foundation of skills and knowledge. However, science itself is essentially an inquiry-based endeavor. This is recognized nationally in the National Science Education Standards (National Research Council 1996; http://www.nap.edu/openbook.php?isbn=0309053269) and in the state of Minnesota's science standards for K-12 students (http://www.education.state.mn.us/MDE/Academic_Excellence /Academic_Standards/Science/index.html). In addition, the fast pace of the biological sciences requires constant attention to bring advances in biology to teachers, students, and the public so they are able to understand new discoveries and their social implications. Our program provides opportunities for G7-12 teachers and students and for the general public to become engaged in the Minnesota Mississippi Metagenomics project. Our project goals for the following individuals are outlined below:

G7-12 teachers and students:

Good inquiry-based science in the G7-12 classrooms is often complicated for schools due to cost - school districts often have difficulty finding adequate funding, difficulties that teachers have in obtaining commitments for professional development programs by scientists in the field, and the necessary "hook" to keep students engaged. The use of the data from the Minnesota Mississippi Metagenome project (M3P) in the classroom addresses each of these issues. Firstly, the M3P provides a platform for students to work on this project whether in a wet-lab setting, or by doing online genome

analyses requiring only a web browser. Thus, the costs can be scaled to the budget of the district. Secondly, our program will provide workshops for teachers that will allow them to interact with scientists and build a learning community with scientists and each other. Thirdly, the M3P provides a "hook" for engaging students by fostering the excitement of discovering something no human has known before. (How cool is that?). The other "hook" is that the Mississippi River is the largest and most historic river in the United States and, through this project, the students have an opportunity to contribute to knowledge about, and care of, this incredible environmental resource.

The professional development plan of this project is to engage 20 teachers per year, in 2012 and 2013, via workshops to bring Mississippi Metagenomics to their classrooms. Teachers participating in this project will receive graduate credit, a stipend, books for their own reference, teaching materials for their classrooms, and continuing support. Each cohort will begin with a one week, full time workshop. The 2012 workshop will be held at the University of Minnesota -Twin Cities Campus and will recruit districts and teachers in the metro area. The 2013 workshop will be held at the University of Minnesota Itasca Biological Station and Laboratories (http://www.cbs.umn.edu/itasca/) which has laboratories, housing, and dining facilities, as well as the same instrumentation (e.g., high throughput genomic and robotic facilities) as are present on campus. This will allow teachers at both locations to experience the real science of metagenomics. In addition, both the campus and Itasca sites have access to the Mississippi River for sampling activities.

In this result we will also develop dedicated and jointly administered websites making the metagenome diversity and functional and chemical data accessible to middle and high school students, undergraduate and graduate students, researchers, and the public. As discussed above, students and teachers participating in this activity will also be involved in annotating the functionally active fosmid clones that will be sequenced at the Biomedical Genomics Center at the University of Minnesota. They will also analyze Mississippi River metagenome data that reflects microbial diversity issues, the presence of pathogens, and the relationship between microbial data and chemical constituents in the river. School participants in this project will chiefly use the IMG-ACT software and websites (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi.) The IMG-ACT is a database of microbial genomes and metagenomic data that is maintained by the DOE. We will also utilize existing national web resources, such as Dolan DNA Learning Center (http://www.dnalc.org/) and Mothur databases (Schloss et al. 2009) to facilitate archiving, retrieval, and analysis of metagenome data. This will allow students to participate in the Mississippi Metagenome project, learn about bioinformatic and metagenomics, and help discover novel microbial genes that are related to growth and survival in the Mississippi River. It will also allow students throughout the state to gain an appreciation for how human activity influences the functioning of the river ecosystem. Since student access to the website and annotation of genes will be monitored by us, we will obtain near instantaneous metrics on how many students are participating in our educational and outreach programs.

The same inquiry-based activities that the teachers will be engaged in will be modified to be used in their classrooms. Teachers will be asked for their input on the development of curricular action plans for their students that will be worked on throughout the academic year. These curriculum plans will include mechanisms to meet state science standards using metagenomics as the focus. Three required follow-up meetings in the academic year will provide teachers with a learning community for support, continued access to the scientists in the project, and support to gauge additional needs for the incorporation of metagenomics in their classrooms. These follow-up meetings will also provide teachers with the increased and continued intellectual support identified by Huffman, et al (2003) as instrumental for changing teaching practice. Throughout this project we will also develop educational materials, such as webinars and PowerPoint presentations that can be used by teachers for instructional purposes.

Summary Budget Information for Activity 2:

ENRTF Budget: **~~$79,909~~ $70,255**
Amount Spent: **$ 48,243**
Balance: **$ 22,012**

Activity Completion Date:

Outcome	Completion Date	Budget
1. Provision of professional development workshops, in the summer 2012 and 2013.	9/1/2013	~~$78,495~~ ~~$68,841~~ $67,341
2. Production of a web accessible, searchable database with downloadable datasets for use in the 7-12 and undergraduate classrooms, as well as by researchers in Minnesota and elsewhere. This will occur via a partnership of researchers with G7-12, undergraduate, and graduate students and educators, and citizens working on this database.	6/30/2012	~~$1,414~~ $2,914
3. Production of curriculum packets, webinars, books, materials, presentations, and approaches that can be incorporated into G7-12 classrooms.	9/1/2013	*This outcome will be accomplished by the course instructor whose salary is included in 1 above (in the ~~$78,495~~).*
4. Annotation of gene identity and function in IMG-JGI website by G7-12 students.	12/31/2013	*No cost to grant, data input is by students in classrooms .*
5. Development of trained teachers that incorporate this cutting edge science into their classrooms and pedagogical materials for other teachers to use throughout the state (and nation).	12/31/2013	*This outcome will be accomplished by the course instructor whose salary is included in 1 above (in the ~~$78,495~~).*

Activity Status as of December 31, 2011:

We are beginning to organize materials for the Summer 2012 workshop, but no funds have yet been expended on this activity yet. Workshop development is being done by the PIs, technician and an HHMI fellow who has agreed to help us coordinate this activity.

Work is in progress to develop materials for Activity 2, including additional shotgun sequencing of full water metagenomes to produce sequence assemblies used by for student bioinformatics activities. Infrastructure and software for hosting bioinformatics activities is in place and the interface is being developed for the teacher workshops. Because the workshop and materials have yet to be fully developed, the design for the website has been postponed until spring semester 2012.

Activity Status as of June 30, 2012:

The organization of the teacher's workshop is in process. We have hired two teacher co-leaders (Karen Casper from Saint Paul Public School district and Mary Raab from Minneapolis Public School district) but they will not be paid until the end of the workshop. The two teacher co-leaders have been meeting with Brian Gibbens, the workshop instructor, to complete the plans. Textbooks have been chosen and the daily schedule is nearly complete. The workshop will be July 30-August 3 on the St. Paul campus. We currently have 16 teacher participants signed up for the workshop, but continue to recruit.

No funds have been spent in this section since the salaries, student fees, textbook and lab costs, etc. will all be incurred at the time of the workshop. Thus the costs for this year's workshop will show up on the December, 2012 report.

The website for the entire project is up and can be viewed at (http://www.cbs.umn.edu/m3p). The Teacher Resources section will be expanded as the workshop gets closer and will continue to have additions on a regular basis.

Activity Status as of December 31, 2012:

Our teacher workshop held in August 2012 was successful with 14 enrolled teachers. The teachers were from the following schools: St. Louis Park High School, Eagle Ridge Academy, Susan B. Anthony School, East Ridge High School, VOA SALT High School, Providence Academy, Nova Classical Academy, Shakopee Middle School, Champlin Park High School, Montevideo Senior High, Minnetonka High School, Owatonna ALC, and Jackson Middle School.

Representing the following school districts:
ISD 283, Eagle Ridge Academy, MPS – 1, South Washington County District 833, Minneapolis, Nova Classical Academy, Shakopee ISD 720, Anoka-Hennepin District #11, Montevideo Public Schools, Minnetonka 276

Workshop activities included

- Filtering and Microscopy Lab
- Finding/Reading Metageomics primary literature articles and discussing them as a group
- Poster Project: Hypothesize the effect physical factors (i.e. temperature, pH etc) would have on River microbes and test this hypothesis using Mothur. Create a poster describing the experiment, present it at the workshop, and bring it back home to hang up in their classrooms.
- Using the Mothur bioinformatics program
- Research specifics about a genus identified using mothur
- Genome annotation with IMG-ACT
- Discover novel enzymes: Use MetaBioMe and Blast to identify enzyme homologs in metagenomic data
- Functional Metagenomics Lab: Perform metagenomic screen for novel proteins that can break down casein or provide antibiotic resistance.

- Design individual lesson plans for bringing metagenomics into your classroom.
- Present Lesson Plans to the workshop group

A follow up survey was sent to the teachers and have just finished analyzing the results. In general individuals had planned on incorporating metagenomics into their classrooms in a variety of ways. Some teachers decided to lecture about metagenomics, some proposed lab experiments that involved sequencing microbes, and at least one participant substantially modified their syllabus to include many metagenomics-themed lectures and activities.

Although most of the participants do plan to incorporate metagenomics into their classrooms, only a few have done so as of this date. Those that did said that using websites like MetaBioME and using programs like Mothur made it relatively easy to incorporate metagenomics into their classes. Those that have not yet incorporated it into their classrooms stated that they were planning on doing it in the next semester after their students have gotten enough genetics background and after they've had more time to process the workshop and develop their own activities. When asked what the students thought of the new metagenomics activities one teacher participant said "Better on average than other trainings because there were so many options. The material was somewhat engaging".

When asked about their **future plans to incorporate metagenomics into their classrooms** and teacher-participants had a variety of responses. Some say, "I still am going to use the lesson plan that I designed in the workshop." A few stated that they felt metagenomics activities would fit best with advanced or AP classes. One teacher-participant said "I am hoping that in future years I will be able to design a class solely aimed at metagenomics-related topics, where students can do a more extended study of the ideas we did in the workshop - plating, reading, entry, synthesis/searches, and final reports. I would like students to be integrating information about genes and topics from current events as the basis for their research, i.e. cancer related genes."

The summer workshop in 2013 will be held on August 5th - 9th at the University of Minnesota Itasca Biological Station and Laboratories with Itasca State Park. We are currently organizing the workshop and beginning to identify teacher participants.

Activity Status as of June 30, 2013:

Our teacher workshop will be held in early August 2013 and now has 19 teacher participants signed- up. However, since the workshop will not be held until August, this number is likely to increase slightly. The teachers currently are from 14 different high schools, representing 12 different school districts. The workshop schedule is being modified based on our experience in the Summer 2012 workshop. Books and materials have been ordered so some of the costs of this workshop are already showing up in the expenses.

Activity Status as of December 30, 2013

The 2013 Metagenomics Teaching Workshop was held August 5-9 at Itasca Biological Research Station. This year's workshop had 15 enrolled teachers from the following schools: Hermantown High School, Brainerd High School, Quarry Hill Nature Center, Pine City Jr/Sr High School, Albany High School, Washington Technology Magnet, Arlington, Fishers Island School, Buffalo High School, West Jr. High, and Cloquet High School. School districts that were represented include: Hermantown School District 700, Brainerd Public School District ISD 181, Rochester Public Schools ISD 535, Pine City ISD 578, Albany Area Schools District 745, Saint Paul Public Schools, Rochester Public School ISD

535, Fishers Island Union Free School District, Buffalo, Hopkins 270, and Cloquet Public Schools ISD# 94.

While at the workshop, teachers were exposed to metagenomics through a variety of lecture and lab experiences. During the workshop, participants worked for ~40 hours on doing the pre-class reading, attending workshop presentations and labs, searching for articles, and developing a lesson plan presentation. Teacher participants used their newfound knowledge to create new teaching materials for use in their classrooms and they are required to give a brief presentation about how metagenomics will be used in their classrooms. Additionally teachers will be contacted after the workshop so we can learn what activities they have successfully implemented in their classrooms.
Workshop activities included:

- Filtering and Microscopy Lab
- Microbe Swabbing Lab
- Finding/Reading Metageomics primary literature articles and discussing them as a group
- Oral Microflora Lab
- Formulating a hypothesis about how physical factors affect Mississippi River microbes
- Using the Mothur bioinformatics program
- Genome annotation with IMG-ACT
- IMG and MetaBioMe Gene Discovery Activity
- Design individual lesson plans for bringing metagenomics into your classroom.
- Present Lesson Plans to the workshop group

Please find below a link and video describing more about the course: http://www.cbs.umn.edu/teachercourses/metagenomics

A survey administered at the end of the workshop indicated that participants really liked the PowerPoint presentations and activities, especially the background presentations about the basics of microbes and metagenomics. While participants generally liked the presentations and activities, some thought that the bioinformatics, comparative genomics, and gene annotation activities may have been a bit complex for their students. The participants unanimously agreed that the co-facilitators were knowledgeable and explained the material in an engaging manner. Nearly all of the participants also felt that the teacher co-leaders helped them integrate the workshop materials into their classrooms.

Here are some things that the teacher-participants said when asked about what they liked most about the workshop:

"The article assignment was the most useful. I also appreciated having a chance to culture some bacteria on plates; I had never done this before and have wanted to do it with my students, however I was nervous about it. I have an autoclave and am able to sterilize materials, so I am now ready to grow some cultures."

"The presentations were well done and engaging. I appreciate having the PowerPoint presentations available on Moodle."

"The beginning of the workshop was the best. I really appreciated the early lectures. Then finding articles and sharing them was really helpful. It gave me a good sense of the possible applications for metagenomics. Also, doing the swabbing and growing out the plates for the oral microflora lab was engaging and fun. I was glad we had three plates (and the freedom) to try different things."

"I feel like I have a good understanding of the use of metagenomics."

"I think the introductory microbial and metagenomial info was very well presented. I am a bigger fan of the wet labs, again, because of my digital deficits. But I see the critical importance of the computer-based work for students. As always the opportunity to interact with peers is a highlight. The staff knew their stuff and were very supportive and accommodating. [They] presented us with a menu of options to work with and adapt to our settings and inherent time constraints. Lovely setting! Nice accommodations and facilities. Good food. No complaints."

"I am excited about the topic of metagenomics and am glad that I was able to learn more about it. Again, I appreciate the time to reflect on incorporating it into our curriculum."

"As always, the opportunity to network with other teachers and bounce ideas off of each other is very beneficial. In addition, the exposure to the rapidly developing changes in the field allow me to stay on top of my content area."

"The first three days included slideshows that I would use in my class. I am going to grab some of the slides for my own slide show presentations in my microbiology course. The wet labs were also good for learning the technique needed for collecting samples and analyzing samples."

"I enjoyed the lecture and presentations. I also benefited from the active engagement with IMG."

"I like the real data that we can use."

"Everything was great. I thoroughly enjoyed stretching my mind and to begin thinking of what metagenomics could mean to the many disciplines of science I teach. It is obvious that this was well thought out and taught by impressive teachers that are well versed in many aspects of metagenomics. In addition everyone was so kind and truly created a welcoming environment for learning. I greatly appreciate the opportunity to attend. Thank you."

"The workshop was excellent! Real-world labs in a beautiful environment."

Finally, the website needs to be updated with new research results (Activity 1), additional information for teachers (Activity 2) and the public (Activity 3). Thus, we request that $1,500 of the savings in the personnel section of Activity 2 be earmarked to pay the website staff member to do these updates. This will not require moving any funds, we are just identifying them for this purpose. Also note that the website work for this entire project was originally budgeted in Activity 2, even though the materials in it spanned the entire project.

Final Report Summary: June 30, 2014

The intent of this activity was to engage G7-12 teachers and their students in the most current work in the field of environmental biology, metagenomics. To accomplish this, we offered two workshops for teachers, one in the summer of 2012 in the Twin Cities area and one in the summer of 2013 in Northwest Minnesota through our Biological Station within Itasca State Park. We had 14 and 15 teachers attending in each year, respectively. During the workshops, the teachers had opportunities to work with experts in the field, develop curriculum for their classrooms, and work collaboratively with other science teachers. Both workshops were well received. Some teachers have been able to incorporate what they learned into their classrooms, others found that this additional information was

helpful to them in their teaching, but the materials would not work within their curriculum. This is a typical outcome for teacher professional development activities. We also developed videos, classroom materials, and teacher resources that were provided to the teachers and are on our website for others to use. Recently, the University of Minnesota switched to a new web site format so we are just now completing the transition of our materials to this new format. Although the old website is still accessible, we will have the new and more accessible web site available by October 31, 2014.

ACTIVITY 3: Project Data Dissemination

Description:

Project data, teacher information, and research concerning the Minnesota Mississippi Metagenomics project will be disseminated via five main routes. Dissemination activity is paramount to the success of this project since the Minnesota Mississippi Metagenomics project has the potential to engage the public in the excitement of state-of-the art research, application of this research to real problems in our state, and discussions about the implications of policy decisions on our natural resources. We have chosen three venues through which to reach the general public: the Science Museum of Minnesota, the Bell Museum of Natural History on the University of Minnesota campus, and Itasca State Park. At the Science Museum, we will work with exhibit staff (via our collaborator Patrick Hamilton, Director, Environmental Sciences and Earth-System Science, Science Museum of Minnesota) to incorporate information and database access for this project both on their EarthBuzz website (http://www.sciencebuzz.org/buzz-tags/earth-buzz) and in kiosks in their exhibit about Minnesota and the environment. As part of the Minnesota Mississippi Metagenome Project, the SMM will build an Earth Buzz kiosk and install at public venue in the Twin Cities area and support the Earth Buzz project manager who will devote their time over the 2.5 year duration of the project. The SMM staff will coordinate the generation of stories and blogs about the Minnesota Mississippi Metagenomics project for the Museum's Earth Buzz network, mentor graduate students involved in the research project on how to write science blogs for general public audiences, and stay abreast of the research project in order to prepare relevant stories and blogs for Earth Buzz. At the Bell Museum, we will be working with the museum curators to add materials at existing aquatic dioramas to explain the use of metagenomics in measuring water quality. And finally, at Itasca State Park, we will work with the Park Naturalist to design and build two exhibits, one for the Nature Center at the East entrance to the Park, and one along the trail to the Mississippi headwaters, that will explain and engage the visitors in this research on the Mississippi. Together these venues have the ability to engage a large population of the public concerning metagenomics, the Mississippi River, the microbial constituents of the river, including pathogens, and how human activity influences the structure and function of this important waterway and ecosystem.

Results from this project will also be disseminated via reports made to the LCCMR, the generation of teaching materials, in periodic update reports made to cooperators, in seminars given throughout the state and nation, and in scientific publications in peer-reviewed scientific and teaching journals. Lastly, project data and approaches, including all teaching and learning activities will be disseminated via a dedicated web site that will be built specifically for the project.

Summary Budget Information for Activity 3:		
	ENRTF Budget:	**$ ~~69,800~~**
		$ 79,800
	Amount Spent:	**$ 70,062**
	Balance:	**$ 9,738**

Activity Completion Date:

Outcome	Completion Date	Budget
1. Development of Minnesota Mississippi Metagenome Website.	12/31/2011	*This outcome is also in activity #2 where it is budgeted.*
2. Production of public exhibits at SMM, Lake Itasca, and the Bell Museum.	6/30/2012	~~$69,800~~ $79,800
3. Production of curriculum packets, webinars, books, materials, and presentations for G7-12 students and teachers.	6/30/2013	*This outcome is also in activity #2 where it is budgeted.*
4. Dissemination of project data and results via webinars, seminars and workshops, and publications.	12/31/2013	*This outcome is also in activity #1 where it is budgeted. This will be accomplished by the postdoctoral student, the faculty PI, and others as a normal part of their scientific work.*

Activity Status as of December 31, 2011:

We met with Pat Hamilton and Liza Pryor of the Science Museum of Minnesota and with Barb Coffin and Gordon Murdock at the Bell Museum of Natural History to begin development of public exhibits. We have a meeting scheduled on January 11 at Itasca State Park to meet with Connie Cox, Park Naturalist, to begin planning the exhibit to be located on the trail leading to the headwaters. No funds have been spent on this Activity yet. The workshop coordinators are developing modules for student use in the classroom and textbooks are being chosen.
The development of the Minnesota Mississippi Metagenome website is in progress, as mentioned in Activity #2 and will be completed in Spring semester 2012.

Activity Status as of June 30, 2012:

We have just signed the contract with the Science Museum of Minnesota, and they are already working on incorporating this project information into their kiosk modules.

We have met with the Bell Museum staff in development of a public exhibit. The plan is to include this project's information in a larger exhibit about water use in Minnesota, with the Bell providing funding for the other sections, including some space remodeling, to make this an inviting and informative exhibit.

A poster describing this project is on display at the Nature Center at the entry of Itasca State Park. While waiting for construction of the more permanent (wood and metal) displays, we thought it best to get some material up as the summer vacation traffic began. The cost of this poster display was $516.04

The Minnesota Mississippi Metagenome website is up, as mentioned in Activity #2. http://www.cbs.umn.edu/m3p

Activity Status as of December 31, 2012:

The outreach to the general public also is progressing. The exhibit at the Mary Gibbs Mississippi Headwaters Center at Itasca State Park is in final stages of design and production, with the goal to have

it installed by March and invoiced at the time of completion. There also is a poster of the project on display at the Nature Center (Jacob V. Brower Visitor Center and Gift Shop) at the entrance of the park. The exhibit at the Science Museum of Minnesota (SMM) is also in development, with a launch on Science Buzz scheduled within the next month or so. The SMM will invoice us for the work when the site is launched.

The exhibit at the Bell Museum has been modified to include a series of short videos, seen via a kiosk, to increase visibility and interest in metagenomics. We now have the final contract estimate. This requires some increase in funding – and this project report includes a requested amendment to move $10,000 to fund this increase. This request is given in Section III and is shown as a budget revision in Attachment A.

Activity Status as of June 30, 2013:

Outreach to the general public is also progressing very well. The exhibit at the Mary Gibbs Mississippi Headwaters Center at Itasca State Park is complete with the goal to have it installed by the July 4 weekend. (This is slightly delayed from the original plan because of some weather issues at the State Park.) The plan is to have this exhibit up for at least two years. The cost of this exhibit will be invoiced in July so will show up on the budget in the December report. As mentioned before, there also is a poster of the project on display at the Nature Center (Jacob V. Brower Visitor Center and Gift Shop) at the entrance of the park.

The exhibit at the Science Museum of Minnesota is complete:
see http://www.sciencebuzz.org/topics/m3p. Our agreement with the Science Museum is that they will continue to update this website into the future as our project progresses.

The video exhibit at the Bell Museum is almost complete and is in its final stages of video editing. We have reviewed it and are very excited about how well it explains the project and its impact to the general public. Some of the costs of this project have already been invoiced, but there are some remaining costs that will show up on the December 30, 2013 report.

Activity Status as of December 30, 2013
All the outreach activities are now installed and open to the public. The video presentation is in place at the Bell Museum of Natural History on the East Bank campus of the University. Preliminary input from viewers has been very positive and we invite all the members of the Commission to visit the Bell Museum. The exhibit is also installed at Itasca State Park at the Mary Gibbs Mississippi Headwaters Center. The exhibit is installed on the plaza directly outside the gift store and restaurant, an area with very high traffic volume. We are very pleased with the results of both of these final parts of our outreach work.

This completes the work budgeted in Activity 3.

Final Report Summary: June 30, 2014

An important part of the process of science is the communication of results to others. In a project such as this, that involves a major river running through our state, we felt that this communication necessarily would involve not only communication to other scientists but also to the public. Thus, we used multiple avenues to reach both audiences. Our scientific communications included poster sessions at the American Society for Microbiology General Meeting (annual meeting), scientific papers (see attached papers), and seminars given on campus. Reports to LCCMR would reach legislative and citizen

members, and additional outreach to the public was also accomplished. We collaborated with the Science Museum of Minnesota to complete an exhibit that children and parents could access both at the Museum and at home (http://www.sciencebuzz.org/topics/m3p.) We worked with the Bell Museum of Natural History on the University of Minnesota's Twin Cities campus to prepare a video series that is currently available in a kiosk at the Bell Museum. We also worked with the Minnesota Department of Natural Resources to put up a permanent exhibit at Itasca State Park at the Mary Gibbs Mississippi Headwaters Center. The exhibit is installed on the plaza directly outside the gift store and restaurant, an area with very high traffic volume. We are very pleased with the results of our outreach work. Lastly, project data and approaches, including all teaching and learning activities are disseminated via a dedicated web site that is in final stages of upgrade to a new University website format and will be complete by October 31, 2014.

V. DISSEMINATION:

Description: NOTE THAT DISSEMINATION IS PART OF ACTIVITY 3 AND HAS BEEN FULLY DESCRIBED ABOVE.

Activity Status as of December 31, 2011:
We have met with Pat Hamilton and Liza Pryor of the Science Museum of Minnesota and with Barb Coffin and Gordon Murdock at the Bell Museum of Natural History to begin development of public exhibits. We have a meeting scheduled on January 11 at Itasca State Park to meet with Connie Cox, Park Naturalist, to begin planning the exhibit to be located on the trail leading to the headwaters. No funds have been spent on this Activity yet. The workshop coordinators are developing modules for student use in the classroom and textbooks are being chosen. The development of the Minnesota Mississippi Metagenome website is in progress, as mentioned in Activity #2 and will be completed in Spring semester 2012.

Activity Status as of June 30, 2012:

As mentioned in Activity 3 summary, we have just signed the contract with the Science Museum of Minnesota, and they are already working on incorporating this project information into their kiosk modules.

We have met with the Bell Museum staff in development of a public exhibit. The plan is to include this project's information in a larger exhibit about water use in Minnesota, with the Bell providing funding for the other sections, including some space remodeling, to make this an inviting and informative exhibit.

A poster describing this project is on display at the Nature Center at the entry of Itasca State Park. While waiting for construction of the more permanent (wood and metal) displays, we thought it best to get some material up as the summer vacation traffic began. The cost of this poster display was $516.04

The Minnesota Mississippi Metagenome website is up, as mentioned in Activity #2. http://www.cbs.umn.edu/m3p

As mentioned in Activity 1, the new postdoc has begun preparing the first paper for publication of the results of this research so far.

Activity Status as of December 31, 2012:

The outreach to the general public also is progressing. The exhibit at the Mary Gibbs Mississippi Headwaters Center at Itasca State Park is in final stages of design and production, with the goal to have it installed by March. There also is a poster of the project on display at the Nature Center (Jacob V. Brower Visitor Center and Gift Shop) at the entrance of the park. The exhibit at the Science Museum of Minnesota is also in development, with a launch on Science Buzz scheduled within the next month or so. The exhibit at the Bell Museum has been modified to include a series of short videos to increase visibility and interest in our project. We now have the final contract estimate. This requires some increase in funding so this project report includes an amendment to move funds from the Professional/Technical/Contracts (Sequencing) to the Equipment/Tools/Supplies to purchase additional lab materials for functional genomic analyses and to enhance our public exhibit at the Bell Museum of Natural History. One manuscript has already been submitted and another is being prepared now for publication in peer-reviewed journals (see attached). After some more analyses we will also submit a second manuscript for publication in a peer-reviewed journal.

Activity Status as of June 30, 2013:

Outreach to the general public is progressing very well. The exhibit at the Mary Gibbs Mississippi Headwaters Center at Itasca State Park is complete with the goal to have it installed by the July 4th weekend. The plan is to have this exhibit up for at least two years. The cost of this exhibit will be invoiced in July so will show up on the budget in the December report. As mentioned before, there also is a poster of the project on display at the Nature Center (Jacob V. Brower Visitor Center and Gift Shop) at the entrance of the park. The exhibit at the Science Museum of Minnesota is complete; see Science Buzz here: http://www.sciencebuzz.org/topics/m3p. Our agreement with the Science Museum is that they will continue to update this website into the future as our project progresses. The video exhibit at the Bell Museum is almost complete and is in its final stages of video editing. We have reviewed it and are very excited about how well it explains the project and its impact to the general public.

Our first paper that was submitted to the Journal of Applied Microbiology has been reviewed, and it is acceptable for publication after minor revisions (see attached). This is the first paper on the Mississippi Metagenome and presents some interesting data on bacterial taxa occupying the Mississippi River and how they are impacted by anthropogenic activity. In addition, data from the 2011 and 2012 samplings were presented as a poster and as a Divisional Lecture at the General Meeting of the American Society for Microbiology at Denver in May 2013. This lecture will also be presented in a conference at Seoul National University in July 2013.

Activity Status as of December 30, 2013

Final Report Summary: June 30, 2014

As mentioned in an earlier summary, an important part of the process of science is communication of results to others. In a project such as this that involves a major river running through our state, we felt that this communication necessarily would involve not only communication to other scientists but also to the public. Thus, we used multiple avenues to reach both audiences. Our scientific communications included poster sessions at the American Society for Microbiology General Meeting (annual meeting), scientific papers (see attached papers), and seminars given on campus. Reports to LCCMR would reach legislative and citizen members, and additional outreach to the public was also accomplished. We collaborated with the Science Museum of Minnesota to complete an exhibit that children and parents

could access both at the Museum and at home (http://www.sciencebuzz.org/topics/m3p.) We worked with the Bell Museum of Natural History on the University of Minnesota's Twin Cities campus to prepare a video series that is currently available in a kiosk at the Bell Museum. We also worked with the Minnesota Department of Natural Resources to put up a permanent exhibit at Itasca State Park at the Mary Gibbs Mississippi Headwaters Center. The exhibit is installed on the plaza directly outside the gift store and restaurant, an area with very high traffic volume. We are very pleased with the results of our outreach work. Lastly, project data and approaches, including all teaching and learning activities are disseminated via a dedicated web site that is in final stages of upgrade to a new University website format and will be complete by October 31, 2014.

VI. PROJECT BUDGET SUMMARY:

A. ENRTF Budget:

Budget Category	$ Amount	Explanation
Personnel:	$ 213,681	Postdoctoral student and lab technician to gather and process samples, website staff to assist with initial website setup for data uploads, instructor for teacher professional development program, and graduate student for assessment of the teacher professional development program
Professional/Technical Contracts:	$270,930	Two Teacher Co-leaders of workshop, $12,000; exhibit staff at the Science Museum of Minnesota, $54,800; Genome sequence analysis at NCGR, Chemical Analyses at USDA-ARS, and preparation of functional gene libraries at Clemson University. Genome preparation and all genomic, physical, and chemical analyses are done most cost effectively in specialty labs that charge by the sample, $204,130.
Service Contracts	$0	
Equipment/Tools/Supplies:	$45,097	Text books ($190 - 3-4 books per teacher) and information materials that cover DNA technology, metagenomics theory, and microbiology for the 40 teachers while in class, $7,600; lab materials such as filters, PCR materials, DNA sequencing, agar plates, tubes, etc. for 40 teachers in workshop, $15,400; filtering materials for samples, $7763; exhibit materials such as casework for dissemination activities, $15,000.
Travel Expenses in MN:	$15,452	Travel for taking river water samples, travel and room and board for teachers and instructors in teacher professional development program
Other:	$11,840	Administrative and student fees for teachers in professional development program; physical analysis of river water samples.
TOTAL ENRTF BUDGET:	**$557,000**	

Explanation of Use of Classified Staff: One technician (Lab Services Coordinator) hired specifically for this project will be paid on this funding to assist in river sampling and genomic analysis of the samples, as well as providing lab support to the teacher professional development programs and dissemination activities. The staff member identified has been working on this project for two years already and is highly skilled in the specific requirements of the job.

Explanation of Capital Expenditures Greater Than $3,500: NA

Number of Full-time Equivalent (FTE) funded with this ENRTF appropriation: 2.35 in year 1; 2.65 in year 2.

B. Other Funds:

Source of Funds	$ Amount Proposed	$ Amount Spent	Use of Other Funds
Non-state			
American Recovery and Reinvestment Act of 2009 (a.k.a. Federal Stimulus) funds	$16,670	$16,670	*ARRA Federal Stimulus funds paying the undergraduate course instructor and laboratory support personnel, for July and August, 2011.*
State			
University of Minnesota O&M funds	$4,533	$4,533	*Portion of Jane Phillips salary for management of teacher professional development programs; College match.*
TOTAL OTHER FUNDS:	**$21,203**	**$21,203**	

VII. PROJECT STRATEGY:

A. Project Partners: The project will be carried out under the direction of Drs. Michael Sadowsky (PI) and co-PI James Cotner. Funded project partners will include Pat Hamilton of the Science Museum of Minnesota, Itasca State Park, and the Bell Museum, Dr. William Koskinen (USDA-ARS) who will be involved in sample analysis for chemicals and the NCGR who will do DNA sequence analysis on a fee basis. We will also collaborate with the National Park Service at the SMM, Adam Birr at the Minnesota Department of Agriculture, and Barb Peichel at MPCA for dissemination activities.

B. Project Impact and Long-term Strategy: This request seeks funding for the first 2.5 years of this program. This will provide the basis for a long-term, continuing study of the health of the Mississippi River that will include all the states bordering the Mississippi and eventually all the states in the Mississippi watershed. Since the River starts in Minnesota at Itasca, this new in depth study and broad impact program begins in Minnesota. Additional funding for more long term and more extensive analyses (of the upper and lower Mississippi River) will be obtained from the National Science Foundation, other states, and other foundations. This National project will be organized similar to the MN project, but involve researchers, students, and the public all the way to New Orleans.

C. Spending History:

Funding Source	FY 2011
Federal Stimulus Funds	$383,300
(9/1/2009 – 6/20/2011)	

VIII. ACQUISITION/RESTORATION LIST: NA

IX. MAP: see attached

X. RESEARCH ADDENDUM: see attached

XI. REPORTING REQUIREMENTS:
Periodic work plan status update reports will be submitted not later than January 2012, July 2012, January 2013, July 2013, and January 2014. A final report and associated products will be submitted between June, 2014 and August 1, 2014.

Final Attachment A: Budget Detail for M.L. 2011 (FY 2012-13) Environment and Natural Resources Trust Fund Projects

Project Title: *Mississippi River Water Quality Assessment*
Legal Citation: *Fill in your project's legal citation from the appropriation language*
Project Manager: *Michael Sadowsky*
M.L. 2011 (FY 2012-13) ENRTF Appropriation: $ *557,000*
Project Length and Completion Date: 3 *years; expected completion June 30, 2014*
Date of Update: *July 25, 2014*

ENVIRONMENT AND NATURAL RESOURCES TRUST FUND BUDGET	Activity 1 Budget: Revised as of January 27, 2014	Amount Spent	Balance	Activity 2 Budget: Revised as of January 27, 2014	Amount Spent	Balance	Activity 3 Budget - Revised as of January 2013	Amount Spent	Balance	TOTAL BUDGET	TOTAL BALANCE
BUDGET ITEM				***Professional Development of Grade 7-12 Teachers***							
Personnel (Wages and Benefits)	238,254	239,842	-1,588	12,927	13,050	-123				251,181	-1,711
Chris Staley, Postdoctoral student, $95.200 (84% salary, 16% benefits) 100% time, 2 years											
Trevor Gould, Technician, $95,900 (73% salary, 27% benefits) 100% time, 2 years											
To be determined, Website staff, $1,414 (73% salary, 27% benefits), 2% time; approved in Jan. 27, 2014 amendment: $1500 for website update											
To be determined, one Instructor, $12,000 (76% salary, 24% fringe), 5% time, 2 years.											
To be determined, one Advanced Graduate Student from CEHD, $9,167 (77% salary, 23% benefits), 1 year - No tuition is required due to advanced status											
Professional/Technical Contracts -	131,630	108,383	23,247	12,000	12,000	0	54,800	54,000	800	198,430	24,047
Genome sequence analysis at NCGR, Chemical Analyses at USDA-ARS, and preparation of functional gene libraries at Clemson University. Genome preparation and all genomic, physical, and chemical analyses are done most cost effectively in specialty labs that charge by the sample.											
To be hired, Two Teacher Co-leaders of workshop, $12,000 (2 teachers per year X $3000/teacher X 2 years)											
One Exhibit staff member at the Science Museum of Minnesota, $54,800 (73% salary, 27% benefits), 2 years											
Equipment/Tools/Supplies *(list out general descriptions of item(s) or item type(s) and their purpose—one row per item/item type. Add rows as needed)*	32,097	31,437	660	23,000	12,714	10,286	25,000	16,062	8,938	80,097	19,883
Text and reference books, information materials for classrooms $7,600 ($190/teacher X 20 teachers/year X 2											
Laboratory supplies for teachers, $15,400 ($385 per teacher X 20 teachers/year X 2 years)											
Laboratory supplies for filtering, cultures, genome preps of river samples, including functional analyses of clones.											
Exhibit materials (metal/wood stands, glass frames, photographs), $15,000 (3 sites X $5000/site)											
Printing											
Travel expenses in Minnesota (*Specify types of travel expenses, e.g., mileage, lodging,*	3,684	1,101	2,583	11,768	6,841	4,928				15,452	7,511
In-State Travel for 10 samplings per year X 2 years @1800 mi *$0.50/mi		53									
Room & board for 4 people X 3 days/year X 2 years for sampling: $1664 for lodging; $1120 for food											
Participant travel (30 mi/dayX 5 day*20 teachers/yr *2year * 0.50/mi)											
Participant room and board for Itasca workshop held 2012 (2011 workshop held in the Twin Cities for Metro area teachers so no room and board needed): ($362.50 each X 20 teachers)											
Instructor/co-leader travel to Itasca in 2012 (2 cars: 430 miles round trip * $0.50/mile)											
Room and board for instructor and co-teachers during teacher professional development program in 2012 (2011 workshop held in Twin Cities so no room and board needed): 1 week at Itasca ($362.50 per person) for instructor and 2 co-teachers											
Other	1,280	732	548	10,560	3,638	6,922				11,840	7,470
Fees for participants: Tuition is waived for the project, but there will be an administrative fee of $100 per registrant X 20 teachers X 2 years = $4000. Plus we are estimating fees for 2012 and 2013 based on current fees(see http://www.cce.umn.edu/Summer-Term/Costs/ and http://cce.umn.edu/documents/DCP/Summer-Tuition-2011.pdf): (University fees of $65/credit X 2 credit) + (CCE College and Technology fee of $25 for students taking less than 6 credits) = $155 for teachers in both 2012 and 2013. Plus the teachers who take the oncampus course in 2012 will be assessed a transportation fee of $18. So $155 X 20 teachers X 2 years = $6200; $18 X 20 teachers X 1 year = $360. Total for administrative fee and University, College and Technology fee, and transportation fee = $4000 + $6200 + $360 = $10560.											
~~Stipend ($1500 per teacher for 1 week workshop/followup work X 20 teachers X 2 years) = $60,000~~ [We have redistributed the $60,000 for additiional sampling consistent with the purposes of our project to achieve the purposes of the project.]- see workplan											
Water samples physical analysis @ $20/sample X 64 samples = $1280.											
COLUMN TOTAL	**$406,945**	**$381,495**	**$25,450**	**$70,255**	**$48,243**	**$22,012**	**$79,800**	**$70,062**	**$9,738**	**$557,000**	**$57,199**

Published Manuscripts

- 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.

Abstract: Aims: A next-generation, Illumina-based sequencing approach was used to characterize the bacterial community at ten sites along the Upper Mississippi River to evaluate shifts in the community potentially resulting from upstream inputs and land use changes. Furthermore, methodological parameters including filter size, sample volume and sample reproducibility were evaluated to determine the best sampling practices for community characterization.
Methods and Results: Community structure and diversity in the river was determined using Illumina next-generation sequencing technology and the V6 hypervariable region of 16S rDNA. A total of 16 400 operational taxonomic units (OTUs) were observed (4594 ± 824 OTUs per sample). *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria* and *Verrucomicrobia* accounted for 93.6 ± 1.3% of all sequence reads, and 90.5 ± 2.5% belonged to OTUs shared among all sites (n = 552). Among nonshared sequence reads at each site, 33‐49% were associated with potentially anthropogenic impacts upstream of the second sampling site. Alpha diversity decreased with distance from the pristine headwaters, while rainfall and pH were positively correlated with diversity. Replication and smaller filter pore sizes minimally influenced the characterization of community structure.
Conclusions: Shifts in community structure are related to changes in the relative abundance, rather than presence/absence of OTUs, suggesting a 'core bacterial community' is present throughout the Upper Mississippi River. Significance and Impact of the Study: This study is among the first to characterize a large riverine bacterial community using a next-generation sequencing approach and demonstrates that upstream influences and potentially anthropogenic impacts can influence the presence and relative abundance of OTUs downstream resulting in significant variation in community structure.

- Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. 2014. Core functional traits of bacterial communities in the Upper Mississippi River show limited variation in response to land cover. Frontiers Microbiol. 5: 414.

Abstract: Taxonomic characterization of environmental microbial communities via high-throughput DNA sequencing has revealed that patterns in microbial biogeography affect community structure. However, shifts in functional diversity related to variation in taxonomic composition are poorly understood. To overcome limitations due to the prohibitive cost of high-depth metagenomic sequencing, tools to infer functional diversity based on phylogenetic distributions of functional traits have been developed. In this study we characterized functional microbial diversity at 11 sites along the Mississippi River in Minnesota using both metagenomic sequencing and functional-inference-based (PICRUSt) approaches. This allowed us to determine how distance and variation in land cover throughout the river influenced the distribution of functional traits, as well as to validate PICRUSt inferences. The distribution and abundance of functional traits, by metagenomic analysis, were similar among sites, with a median standard deviation of 0.0002% among tier 3 functions in KEGG. Overall inferred functional variation was

significantly different ($P \leq 0.035$) between two water basins surrounded by agricultural vs. developed land cover, and abundances of bacterial orders that correlated with functional traits by metagenomic analysis were greater where abundances of the trait were inferred to be higher. PICRUSt inferences were significantly correlated ($r = 0.147$, $P = 1.80 \times 10^{-30}$) with metagenomic annotations. Discrepancies between metagenomic and PICRUSt taxonomic-functional relationships, however, suggested potential functional redundancy among abundant and rare taxa that impeded the ability to accurately assess unique functional traits among rare taxa at this sequencing depth. Results of this study suggest that a suite of "core functional traits" is conserved throughout the river and distributions of functional traits, rather than specific taxa, may shift in response to environmental heterogeneity.

Accepted Manuscripts

- Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. 2014. Bacterial community structure is indicative of chemical inputs in the Upper Mississippi River. Frontiers Microbiol.

Abstract: Local and regional associations between bacterial communities and nutrient and chemical concentrations were assessed in the Upper Mississippi River in Minnesota to determine if community structure was associated with discrete types of chemical inputs associated with different land cover. Bacterial communities were characterized by Illumina sequencing of the V6 region of 16S rDNA and compared to > 40 chemical and nutrient concentrations. Local bacterial community structure was shaped primarily by associations among bacterial orders. However, order abundances were correlated regionally with nutrient and chemical concentrations, and were also related to major land coverage types. Total organic carbon and total dissolved solids were among the primary abiotic factors associated with local community composition and co-varied with land cover. *Escherichia coli* concentration was poorly related to community composition or nutrient concentrations. Abundances of fourteen bacterial orders were related to land coverage type, and seven showed significant differences in abundance ($P \leq 0.046$) between forested or anthropogenically-impacted sites. This study identifies specific bacterial orders that were associated with chemicals and nutrients derived from specific land cover types and may be useful in assessing water quality. Results of this study reveal the need to investigate community dynamics at both the local and regional scales and to identify shifts in taxonomic community structure that may be useful in determining sources of pollution in the Upper Mississippi River.

Submitted Manuscripts

- Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. Species sorting dynamics promote community resilience in response to natural and anthropogenic disturbance in the Upper Mississippi River. Sci. Total Environ. *Note: Accepted pending revisions.*

Abstract: Bacterial community structure (BCS) in freshwater ecosystems varies seasonally and due to physicochemical gradients, but metacommunity structure of a major river remains understudied. Here we characterize the BCS along the Mississippi River and contributing rivers in Minnesota over three years using Illumina next-generation sequencing, to determine how changes in environmental conditions as well as inputs from surrounding land and confluences impacted community structure. Contributions of sediment to water microbial diversity were also

evaluated. Long-term variation in community membership was observed, and significant shifts in relative abundances of major freshwater taxa, including *α-Proteobacteria*, *Burkholderiales*, and *Actinomycetales*, were observed due to temporal and spatial variation. Taxa abundances were correlated primarily with temperature and rainfall, but also nutrient concentrations, suggesting that species sorting played a predominant role in shaping BCS. Furthermore, an annually-recurrent BCS was observed in late summer, further suggesting that seasonal dynamics strongly influence community composition. Sediment communities differed from those in the water, but contributed up to 50% to community composition in the water column. Among water sampling sites, 34% showed significant variability in BCS of replicate samples indicating variability among riverine communities due to heterogeneity in the water column. Results of this study highlight the need for a better understanding of spatial and temporal variation in riverine bacterial diversity associated with physicochemical gradients and reveal how communities in sediments, and potentially other environmental reservoirs, impact waterborne BCS. Techniques used in this study may prove useful to determine sources of microbes from sediments and soils to waterways, which will facilitate best management practices and total maximum daily load determinations.

- Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. High-throughput functional screening reveals low frequencies of antibiotic resistance genes in DNA recovered from the Upper Mississippi River. J. Water Health.

Abstract: In this study, we determined the frequency of antibiotic resistance genes (ARGs) in the Upper Mississippi River using a high throughput, functional, metagenomic screening procedure. Fosmid libraries containing ~10,000 clones were screened for resistance to ampicillin, cephalothin, kanamycin, and tetracycline. We hypothesized that nutrient concentrations, land cover type, and taxonomic community composition may select for ARGs. Resistance to ampicillin, cephalothin, and kanamycin was low (< 1.00%), and no resistance to tetracycline was detected. Ammonium and total dissolved solids (TDS) concentrations were correlated with kanamycin and cephalothin resistances ($r = 0.617$ and -0.449, $P = 0.002$ and 0.036, respectively). Cephalothin resistance was also positively correlated with the percentage of forested land cover ($r = 0.444$, $P = 0.039$). Only the candidate division OD1, among 35 phyla identified, was correlated with ampicillin resistance ($r = 0.456$, $P = 0.033$), suggesting that minority members of the community may be responsible for dissemination of ARGs in this ecosystem. Results of this study suggest that ammonium and TDS may be involved in a complex selection process for ARGs. Furthermore, we suggest that minority species, potentially contributed in low numbers from sediment and biofilm reservoirs, may be the primary carriers of ARGs in this riverine system.

Manuscripts in Preparation/Planned

- Staley C, Johnson D, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. Frequencies of heavy metal resistance are associated with land cover type in the Upper Mississippi River.

Abstract: Taxonomic compositions of freshwater bacterial communities have been well-characterized via metagenomic-based approaches, especially next-generation sequencing; however, functional diversity of these communities remains less well-studied. Various anthropogenic sources are known to impact the bacterial community composition in freshwater river systems and potentially alter functional diversity. In this study, high-throughput functional screening of large (~10,000 clones) fosmid libraries representing communities in the Upper Mississippi River revealed low frequencies of resistance to heavy metals in the following order: $Mn^{2+} > Cr^{3+} > Zn^{2+} > Cd^{2+} > Hg^{2+}$, and no resistance to Cu^{2+} was detected. Significant correlations were observed between resistance frequencies of Cd and Cr with developed land cover ($r = 0.296$, $P = 0.016$ and $r = 0.257$, $P = 0.037$, respectively). Discriminant function analysis further supported these associations while redundancy analysis further indicated associations with forested land cover and greater resistance to Hg and Zn. Nutrient and metal ion concentrations and abundances of bacterial orders were poorly correlated with heavy metal resistance frequencies, except for an association of *Pseudomonadales* abundance and resistance to Hg and Zn. Taken together, results of this study suggest that allochthonous bacteria contributed from specific land cover types influence the patterns of metal resistance throughout this river.

- Summary of methodological experiments including 1) effects of volume and replication on community characterization, 2) influence of different DNA extraction kits on community characterization, and 3) effect of sequence read length on community characterization and taxonomic resolution.

- More rigorous evaluation of the contribution of sand and sediment communities to waterborne microbial community structure.

Posters Presented

- C. Staley, T. Unno, T.J. Gould, B. Jarvis, J. Phillips, J.B. Cotner, and M.J. Sadowsky. Relationship Between Land Use and Anthropogenic Factors Influencing Bacterial Community Structure in the Upper Mississippi River. American Society for Microbiology General Meeting, May 18-21, 2013, Denver, CO.

- C. Staley, T.J. Gould, P. Wang, J. Phillips, J.B. Cotner, and M.J. Sadowsky. Species Sorting Dynamics in the Bacterial Community of the Upper Mississippi River are Influenced by Land Use and Sediment Resuspension. American Society for Microbiology General Meeting, May 17-21, 2014, Boston, MA. (*also presented at the University of Minnesota Microbial and Plant Genomics Institute symposium on Aug 27, 2014.*)

High-throughput functional screening reveals low frequencies of antibiotic resistance genes in DNA recovered from the Upper Mississippi River

Christopher Staley[1], Trevor J. Gould[1,2], Ping Wang[1], Jane Phillips[2], James B. Cotner[3], and Michael J. Sadowsky[1,4,*]

[1]BioTechnology Institute, [2]Biology Program, [3]Department of Ecology, Evolution, and Behavior, and [4]Department of Soil, Water and Climate, University of Minnesota, St. Paul, MN

Running title: Antibiotic resistance genes in Mississippi River bacterial communities

[*]Corresponding Author: Michael J. Sadowsky, BioTechnology Institute, University of Minnesota, 140 Gortner Lab, 1479 Gortner Ave, St.Paul, MN 55108; Phone: (612)-624-2706, Email: sadowsky@umn.edu

ABSTRACT

In this study, we determined the frequency of antibiotic resistance genes (ARGs) in the Upper Mississippi River using a high throughput, functional, metagenomic screening procedure. Fosmid libraries containing ~10,000 clones were screened for resistance to ampicillin, cephalothin, kanamycin, and tetracycline. We hypothesized that nutrient concentrations, land cover type, and taxonomic community composition may select for ARGs. Resistance to ampicillin, cephalothin, and kanamycin was low (< 1.00%), and no resistance to tetracycline was detected. Ammonium and total dissolved solids (TDS) concentrations were correlated with kanamycin and cephalothin resistances (r = 0.617 and -0.449, P = 0.002 and 0.036, respectively). Cephalothin resistance was also positively correlated with the percentage of forested land cover (r = 0.444, P = 0.039). Only the candidate division OD1, among 35 phyla identified, was correlated with ampicillin resistance (r = 0.456, P = 0.033), suggesting that minority members of the community may be responsible for dissemination of ARGs in this ecosystem. Results of this study suggest that ammonium and TDS may be involved in a complex selection process for ARGs. Furthermore, we suggest that minority species, potentially contributed in low numbers from sediment and biofilm reservoirs, may be the primary carriers of ARGs in this riverine system.

Keywords: antibiotic resistance / functional metagenomics / microbial diversity / microbial ecology / Mississippi River / next-generation sequencing

INTRODUCTION

Antibiotic-resistant bacteria were first reported in the 1940s and have become an increasing public health concern due to the diversity of resistant species, as well as an increasing number of antibiotics against which resistance has been observed (Levy & Marshall 2004; Capita & Alonso-Calleja 2013). Acquisition of antibiotic resistance is primarily attributed to the misuse and overuse of antibiotics in medicine and animal husbandry (Finch 2004; Bywater 2005). However, it is also well known that environmental bacteria produce antibiotics and, naturally, possess intrinsic resistance mechanisms, as well (Martínez 2008). While the role of antibiotics in inhibiting growth of microbial competitors is well understood, several hypotheses suggest that low concentrations of antibiotics may additionally serve as signaling molecules (Fajardo & Martínez 2008) or in some cases, as nutrient sources (Dantas *et al.* 2008). Furthermore, genes that confer resistance to elevated concentrations of antibiotics may also have drastically different metabolic or ecological functions in the environment (Martinez *et al.* 2009).

Some, if not all, antibiotic resistance genes (ARGs) have evolved chromosomally to either confer resistance to an antimicrobial compound or to perform a separate function which may confer antibiotic resistance in another organism (D'Costa *et al.* 2006). However, due to anthropogenic impacts, including the reintroduction of antibiotics into the environment, many of these genes have also been incorporated into mobile genetic platforms (*e.g.* plasmids and transposons) making their spread to pathogens more efficient via horizontal gene transfer (HGT) (Alonso *et al.* 2001; LaPara *et al*. 2011). The spread of ARGs among pathogenic bacteria as well as in wastewater communities has received considerable attention, but only in the last decade or so

have studies focused on environmental reservoirs, especially surface waters, as important reservoirs for ARGs (Baquero *et al*. 2008; LaPara *et al*. 2011). Recently, bacterial communities in water environments including rivers, streams, and lakes have been implicated as important vehicles for the retention and transfer of ARGs to human pathogens (Baquero *et al.* 2008; Lupo *et al.* 2012).

Although river water samples have been shown to harbor lower concentrations of antibiotics than associated sediment samples (Kim & Carlson 2007), sub-inhibitory concentrations of antibiotics have been shown to support the development of resistant phenotypes (Gullberg *et al.* 2011). Maintenance of antibiotic resistance generally confers a fitness cost to the organism. This suggests, however, that a lack of selective pressure would select for reversal of resistance (*i.e*. sensitivity). However, compensatory evolution and/or mitigation of fitness cost through community-level interactions has been shown to drastically slow reversal among bacterial communities (Andersson & Hughes 2010). Due to their persistence and potential for HGT between pathogens and non-pathogens, ARGs themselves, separate from antibiotics, have been suggested to be emerging contaminants in surface waters (Pruden *et al.* 2006; Martinez 2009).

Aquatic systems, in particular rivers, have recently been suggested as drivers for the spread of ARGs due to constant mixing of the bacterial community as well as a multitude of anthropogenic impacts (Baquero *et al.* 2008; Taylor *et al.* 2011). The spread of ARGs as a result of this mixing may result in the transfer of novel resistance genes from indigenous bacteria to potential human pathogens that, in turn, transfer the newly acquired resistance to the clinical setting (Poirel *et al.*

2002, 2005; Wright 2010). Furthermore, selective pressures from the contribution of antibiotics from agricultural, industrial, and municipal runoff, as well as wastewater effluent, may promote the emergence of resistant phenotypes, or ARGs may be indirectly selected for in response to adaptations to other nutrient or chemical contamination from anthropogenic sources (Martinez 2009).

One of the major difficulties in evaluating the resistome of environmental communities – the suite of ARGs present in the community (D'Costa *et al*. 2006) – is the lack of culturability of > 99% of environmental species (Amann *et al.* 1995). To circumvent this obstacle, recent studies assessing the prevalence and distribution of ARGs in environmental samples have relied on culture-independent qPCR assays (Huerta *et al.* 2013; Marti *et al.* 2013). However, functional metagenomic screening of large clone libraries has been recently suggested to characterize antimicrobial resistance frequencies of environmental samples (Martínez & Osburne 2013). This approach has been taken using bacterial artificial chromosome (BAC) libraries to assess antibiotic resistance in soil communities (Riesenfeld *et al.* 2004), and more recently fosmid libraries were constructed and screened to assess levels of antibiotic resistance in river sediments (Amos *et al.* 2014). The later system is very amenable to manipulation of genetic material as there is less a requirement for isolation of very large DNA fragments from environmental samples.

In this study, we used high-throughput fosmid library screening method to assess the frequencies of antibiotic resistance to a β-lactam (ampicillin), a cephalosporin (cephalothin), an

aminoglycoside (kanamycin), and a tetracycline (tetracycline) in bacterial communities throughout the Mississippi River in Minnesota. Fosmid libraries consisting of approximately 10,000 clones were constructed from each of 11 sites sampled in the summers of 2011 and 2012. Physicochemical, nutrient, land use, and bacterial community taxonomic data were also collected to determine how these factors influenced antibiotic resistance frequencies. We initially hypothesized that chemical parameters would have a greater impact on antibiotic resistance than community composition, in large part due to the persistence of a taxonomically- and functionally-conserved, core microbial community throughout the study area (Staley *et al.* 2013, 2014). Results of this study reveal how anthropogenic chemical inputs and community structure presumably influence the distribution of ARGs in a major river ecosystem.

METHODS

Water sampling and metadata collection

Eight sampling sites were selected along the main branch of the Mississippi River in Minnesota from the headwaters at Lake Itasca to near the southern border at La Crescent (Figure 1). In addition the Minnesota, St. Croix, and Zumbro Rivers were sampled. The 11 sites were selected to cover the length of the Mississippi River throughout the state and characterize communities in pristine (forested), agricultural, and urban-developed areas as well as in the major confluent rivers. Each site was sampled once between May and July in both 2011 and 2012. At each site, 40 L of water was collected from the surface, approximately 1.8 m from the shoreline in two 20 L carboys and transported back to the laboratory. Water temperature and pH were also recorded

at the time of sampling, and rainfall up to three days prior to sampling was obtained from [http://www.wunderground.com].

Additional 1 L samples were also collected for nutrient analysis in sterile amber bottles. Determination of the concentrations of ammonium, colorimetric nitrite/nitrate (NO_2/NO_3), orthophosphate, total phosphorus, total dissolved solids (TDS), and total organic carbon (TOC) was performed at the Research Analytical Lab at the University of Minnesota (Saint Paul) via standard methods [http://ral.cfans.umn.edu/types-of-analysis-offered/water/]. Land cover data was extrapolated from the 2006 National Land Cover Database (Fry *et al.* 2011) by overlaying a map of hydrologic unit code (HUC) boundaries at a scale of 1:250,000 using ArcGIS (Esri, Redlands, CA). Maps were obtained from the United States Geological Survey [http://water.usgs.gov/ maps.html]. Major land cover types (forested, developed, or agricultural) were assigned based on percentage area within the HUC boundary (Table 1).

Sample processing

Samples were either processed immediately or stored at 15 °C for < 24 h before filtration. Water was filtered as previously described (Staley *et al.* 2013). Briefly, water was pre-filtered through 90 mm diameter P5 filters (Whatman Inc., Piscataway, NJ) and microorganisms were concentrated on 142 mm diameter, 0.45 µm polyethane-sulfonate filters (Pall Co., Port Washington, NY) followed by elutriation in pyrophosphate Buffer (0.1% sodium pyrophosphate buffer, pH 7.0, 0.2% Tween 20). Cell pellets (six per sample, each representing 6-7 L of water) from cell suspensions were stored at -80 °C.

Construction of fosmid libraries

One cell pellet per sample was shipped on dry ice to the Clemson University Genomics Institute (CUGI) [http://www.genome.clemson.edu/] for fosmid construction. DNA from each of the samples was extracted using the Metagenomic DNA Isolation Kit for Water (Epicentre Biotechnologies, Madison, WI) followed by end-repair/phosphorylation. DNA fragments between 35-50 kb were size selected by pulsed-field gel electrophoresis and ligated into pCC2FOS (Epicentre Biotechnologies, Madison, WI). Ligated fosmids were transduced into *E. coli* DH10B by λ phage at CUGI. Fosmid libraries for each site contained a minimum of 50,000 clones and were shipped back to the laboratory on dry ice as glycerol stocks.

Fosmid libraries were diluted to 2.5 CFU μl^{-1} and 1 ml aliquots were plated on 20 × 20 cm Luria Bertani (LB) agar plates containing 12.5 $\mu g\ ml^{-1}$ chloramphenicol (CAM). Colonies (approximately 10,000 per library per site per year, Table 1) were transferred to 384-well plates containing Hogness modified freezing media (HMFM) (Yan *et al.* 2007) with 12.5 $\mu g\ ml^{-1}$ CAM using the QBot colony picking robot (Genetix, Sunnyvale, CA). Fosmid libraries were stored at -80 °C.

Antibiotic resistance screening

Fosmid libraries were thawed at room temperature immediately prior to functional screening. Functional screening was performed by plating libraries on 20 × 20 cm Müller-Hinton plates

(Himedia, Mumbai, India) amended with 7 μg ml^{-1} CAM and the experimentally-determined MIC of antibiotic – 20 μg ml^{-1} ampicillin (AMP), 35 μg ml^{-1} cephalothin (CET), 15 μg ml^{-1} kanamycin (KAN), or 10 μg ml^{-1} tetracycline (TET; see below). Plating was performed using a flame-sterilized 384-well replicator (Boekel Scientific, Feasterville, PA) and up to 6 384-well plates were stamped on each 20 × 20 cm plate. For each set of plates, a negative control (*E. coli* DH10B containing a fosmid without insert) was also streaked. Plates were incubated overnight (16-18 h) at 37 °C. Resistant isolates were determined as those that formed opaque colonies at least 1 mm in diameter and reported as a percentage of the total clone library for each sample.

Minimal inhibitory concentrations (MICs) of clones were determined as the lowest concentrations of antibiotic that reproducibly (triplicate cultures) prevented growth of a control strain (*E. coli* DH10B containing pCC2FOS without insert) grown overnight at 37 °C in 5 ml LB broth with agitation at 250 rpm. Concentrations of antibiotic were adjusted in 5 μg ml^{-1} increments until MICs were established and all plates were amended with 7 μg ml^{-1} CAM. All antibiotics were obtained from Sigma-Aldrich (St. Louis, MO).

High-throughput sequencing

DNA was extracted from two separate cell pellets using the Metagenomic DNA Isolation Kit for Water (Epicentre Biotechnologies). The V6 hypervariable region of the 16S rRNA gene was amplified using barcoded 967F/1046R primers as described previously (Sogin *et al.* 2006; Staley

et al. 2013) and purified using the QiaQuick® Gel Extraction Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Purified amplicons were pooled in equal amounts for sequencing. Amplicons originating from one cell pellet were sequenced on the Illumina MiSeq platform (2 × 150 read length) at the University of Minnesota Genomics Center (UMGC, Minneapolis). For replication and to enable greater sample multiplexing, DNA from the second cell pellet was sequenced in duplicate by UMGC on the HiSeq2000 (2 × 100 read length), as cross-platform variation in data has been previously reported (Caporaso *et al.* 2012). All sequences were deposited in the National Center for Biotechnology Information Sequence Read Archive under accession number SRP018728.

Sequence processing

All sequence processing was performed using Mothur ver. 1.29.2 (Schloss *et al.* 2009). Sequences were trimmed to 100 bp and paired-end aligned using fastq-join (Aronesty 2013). Sequences were quality trimmed using a window of 50 bp and an average quality score of 35. In addition, singleton sequences, those containing an ambiguous base, homopolymers > 8 bp, and sequences that did not have 100% identity to primer and barcode sequences were removed. Samples were aligned against the SILVA reference database ver. 102 (Pruesse *et al.* 2007) and subjected to a 2% precluster (Huse *et al.* 2010; Kunin *et al.* 2010). Chimeric sequences were removed using UCHIME (Edgar *et al.* 2011), samples were normalized to 25,717 sequence reads per sample by random subsampling, and operational taxonomic units were assigned at 97% similarity using the furthest-neighbor algorithm. Taxonomic classification was also performed against the Ribosomal Database Project database ver. 9 (Cole *et al.* 2009).

194

195 **Statistical analyses**

196 To compare taxonomic data with antibiotic resistance frequencies, the abundances of phyla were
197 averaged among triplicates. Analysis of variance (ANOVA) and Spearman rank correlations
198 relating physicochemical data, land cover, taxonomic data, and antibiotic resistance frequencies
199 were performed using SPSS software ver. 19 (IBM, Armonk, NY). Redundancy analysis was
200 performed using XLSTAT (Addinsoft, Belmont, MA). All statistics were evaluated at $\alpha = 0.05$.

201

202 **RESULTS**

203 The frequency of antibiotic resistance to ampicillin, cephalothin, and kanamycin was observed
204 throughout the study area during both years. However, frequencies were extremely low, never
205 with frequencies of $\geq 1.0\%$ observed among all fosmid libraries (Table 1). Resistance to
206 tetracycline was not observed among any of the fosmid libraries screened. Annual differences
207 were observed in the frequency of resistance to both cephalothin and kanamycin, but not
208 ampicillin. In 2011, the frequency of cephalothin resistance was significantly higher compared
209 to 2012 ($P = 0.019$), but resistance to kanamycin was lower ($P = 0.008$). However, throughout
210 the study, the frequencies of cephalothin and kanamycin resistance were positively correlated (r
211 $= 0.426$, $P = 0.048$).

212

213 Relationships between resistance frequencies and climatic and physicochemical parameters were
214 investigated to determine if these parameters were associated with increased resistance

frequencies. Few relationships were significant, although the ammonium concentration was significantly correlated with kanamycin resistance frequency ($r = 0.617$, $P = 0.002$) and the concentration of TDS was negatively correlated with cephalothin resistance frequency ($r = -0.449$, $P = 0.036$). These were among the only parameters, in addition to temperature and total phosphorus concentration, which differed significantly ($P < 0.05$) between years (Table 2).

The surrounding land cover types of the basins under study were further interrogated to determine if specific land cover types (developed, agricultural, or forested cover) were related to increases in resistances to specific antibiotics. The frequency of resistance to cephalothin was significantly positively correlated with the percentage of surrounding forested cover ($r = 0.444$, $P = 0.039$) and inversely correlated with agriculturally-associated cover ($r = -0.435$, $P = 0.043$). Forested area also co-varied significantly with TDS concentration ($r = -0.518$, $P = 0.014$). No co-variation was observed between other parameters that were significantly associated with antibiotic resistance frequency and surrounding land cover.

Since runoff from a variety of non-point sources might be associated with the increase in the frequency of antibiotic resistance genes, due potentially to introduction of non-indigenous, resistant taxa, the bacterial community was characterized via 16S rRNA sequencing of the V6 hypervariable region. The community was found to be comprised primarily of α-, β, and γ-*proteobacteria*, *Actinobacteria*, and *Bacteroidetes,* among 35 phyla identified (Table 2). No differences in the relative abundance of these groups were observed between years of study at $\alpha = 0.05$. Among all the phyla identified, only the relative abundances of *Planctomycetes* ($r = -$

237 0.524, $P = 0.012$) and candidate division OD1 ($r = 0.456$, $P = 0.033$) were significantly
238 correlated with the frequency of ampicillin resistance. Resistance to cephalothin or kanamycin
239 was not associated with any of the phyla identified.

240

241 Lastly, redundancy analysis was performed incorporating all of the parameters investigated by
242 traditional correlation analysis except rainfall, to simplify the model (Figure 2). The results of
243 the redundancy analysis generally corroborated the results of the traditional analyses, showing
244 weak to moderate, positive associations between ammonium and forested land cover with
245 resistances of kanamycin and cephalothin, respectively. Similarly, TDS concentration was
246 shown to be negatively associated with the frequency of cephalothin resistance. Interestingly,
247 however, there was an apparent weak, but positive, association between the relative abundance
248 of *γ-proteobacteria* and ampicillin resistance frequency, which was not captured by traditional
249 analysis.

250

251 **DISCUSSION**

252 Antibiotic resistance genes are now recognized as emerging contaminants of surface waters and
253 pose a risk of conferring antibiotic resistance to human pathogens (Martinez 2009), and recent
254 research has suggested that resistant bacteria are able to enter drinking water supplies indicating
255 a further risk of these bacteria spreading to the human food chain (Walsh *et al.* 2011). The
256 Mississippi River is used as a source of drinking water for more than 50 cities, affecting millions
257 of people, so it is critical to protect this ecosystem from the introduction of ARGs, among other
258 pollutants.

In this study, functional metagenomic screening of fosmid libraries revealed extremely low frequencies of ARGs in the Upper Mississippi River. The frequencies of antibiotic resistance reported here are similar to those found in a recent study of river sediments using functional metagenomic screening (Amos *et al.* 2014). However, our results are in contrast to a previous study that employed direct plating of water collected from the Mississippi River near Minneapolis, MN (Ash *et al.* 2002). In the prior study, a mean concentration of 3.09 $\log_{10}$ colony forming units (CFU) ml^{-1} was observed for two water samples analyzed, and 19.7 and 23.7% of isolates were resistant to ampicillin. Furthermore, among ampicillin resistant isolates (n = 72), 38% were resistant to cephalothin and 28% were resistant to amoxicillin. The discrepancy in the frequency of antibiotic resistance is likely due to differences in methodology, where direct plating may have selected for rare, antibiotic-resistant species. Conversely, since the community in the Mississippi River is comprised primarily of a small number of highly abundant species (Staley *et al.* 2013), it is likely that these species were over-represented in the fosmid libraries and did not possess ARGs. It is also possible, based on the nature of fosmid library construction, that ARGs, which are less abundant versus genes encoding metabolic enzymes, were diluted out of the fosmid libraries in favor of more abundant housekeeping genes. In addition, some genes incorporated into the fosmid libraries may simply have not been expressed by the *E. coli* host strain used due potentially to orientation in the fosmid vector.

To our knowledge, this study is among the first to examine the relationships between nutrient concentrations, which are potentially related to anthropogenic impacts, and antibiotic resistance

281 frequency. While most parameters measured were poorly related to resistance frequency,
282 ammonium and TDS concentrations were found to be significantly associated with resistance to
283 cephalothin and kanamycin. While cephalothin resistance was observed to be higher in 2011, so
284 were TDS concentrations, which were negatively correlated with this resistance frequency.
285 Similarly in 2012, kanamycin resistance was higher, but so were ammonium concentrations, and
286 the two were negatively correlated. However, concentrations of cephalothin and kanamycin
287 resistance were also positively correlated, suggesting that if nutrient concentrations play a role in
288 selecting for ARGs, the dynamics are likely complex and require further study. In addition to
289 responses directly related to nutrient concentrations, it is also possible that biofilms potentially
290 present on TDS particles contribute to increased frequencies of antibiotic resistance as these
291 communities are thought to harbor high densities of antibiotic resistant phenotypes (Marti *et al.*
292 2013).

293

294 Interestingly, both traditional correlation and redundancy analyses revealed that forested land
295 cover was significantly associated with antibiotic resistant phenotypes among fosmid libraries,
296 while agricultural cover was negatively correlated with cephalothin resistance and no significant
297 relationships were observed for developed land cover. Only two sites (Itasca and La Crescent)
298 were classified as forested areas, and the Itasca site is relatively shallow and potentially more
299 highly influenced by soil communities with intrinsic resistance mechanisms (D'Costa *et al.*
300 2006). Factors such as canopy cover that would reduce exposure to UV light may also serve to
301 protect resistant species, but conclusive determination of factors affecting this finding remain to
302 be studied. This result is in contrast to a prior study conducted on the South Platte River Basin,
303 where a positive correlation between the capacities of upstream wastewater treatment plants

304 (WWTPs) and animal feeding operations to the frequency of sulfonamide resistance was
305 observed using qPCR targeting *sul1* (Pruden *et al.* 2012). Furthermore, this study found that
306 distribution of tetracycline resistance, targeting *tet*(W), was independent of land use. Another
307 study investigating tertiary-treated wastewater and surface waters in the Duluth-Superior Harbor
308 identified approximately 20-fold higher concentrations of genes encoding tetracycline resistance
309 [*tet*(A), *tet*(W), and *tet*(X)] and a gene encoding the integrase of class 1 integrons (*intI*1) by
310 qPCR in wastewater compared to surface waters (LaPara *et al.* 2011). The lack of detection of
311 tetracycline in this study is surprising and may potentially be a result of the method used.
312 However, it has been shown (De Francesco *et al.* 2010) that qPCR may overestimate the
313 frequency of antibiotic resistance due, at least in part, to detection of heteroresistant organisms.

314

315 A prior study implicated the phyla *Actinobacteria* and *Firmicutes* as responsible for the transport
316 and dissemination of ARGs in Mediterranean water reservoirs (Huerta *et al.* 2013). Similarly, in
317 a study of wastewater, the phyla *Bacteroidetes* and *Firmicutes*, as well as the *β-proteobacteria*,
318 were positively correlated with antibiotic resistant populations (Novo *et al.* 2013). In the present
319 study, the abundance of only one candidate division, OD1, was positively correlated with
320 resistance to ampicillin, and this division has been reported to be present at relatively low
321 abundance in the Mississippi River (Staley *et al.* 2013). Given the relatively low frequency of
322 resistance detection and the potential restriction of the fosmid libraries to more abundant species,
323 it is possible to suggest that minority members of the community in the water column are
324 responsible for ARG dissemination, and their detection may have been limited in this study.
325 Such minority members of the community may be harbored in greater densities in sediments and
326 biofilms where ARGs may be more abundant (Marti *et al.* 2013), and antibiotic resistant

communities in these reservoirs have been suggested to protect planktonic bacteria from perturbation due to antibiotic release (Baquero *et al.* 2008). The interchange of ARGs and resistant phenotypes between the water column and sediment or biofilm communities will require further study.

CONCLUSION

In this study a functional metagenomic screening strategy revealed that the frequency of resistance to ampicillin, cephalothin, kanamycin, and tetracycline were low-to-non-detectable. Evaluation of nutrient concentrations, land cover, and taxonomic composition of the river community suggest that ammonium and TDS may be involved in a complex selection process for ARGs and that much of the resistance observed may be of natural origin. Furthermore, we suggest that minority species, potentially contributed in low numbers from sediment and biofilm reservoirs, may be the primary carriers of ARGs in this riverine system.

ACKNOWLEDGEMENTS

Funding for this project was provided by the American Recovery and Reinvestment Act of 2009 (ARRA) and the Minnesota Environment and Natural Resources Trust Fund as recommended by the Legislative-Citizen Commission on Minnesota Resources (LCCMR). This work was carried out in part using computing resources at the University of Minnesota Supercomputing Institute.

REFERENCES

Alonso, A., Sanchez, P., Martinez, J.L., 2001 Environmental selection of antibiotic resistance genes. *Environ. Microbiol.* **3**, 1–9. doi:10.1046/j.1462-2920.2001.00161.x

Amann, R.I., Ludwig, W., Schleifer, K.H., 1995 Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev*. **59**, 143–69.

Amos, G.C. a, Zhang, L., Hawkey, P.M., Gaze, W.H., Wellington, E.M., 2014 Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. *Vet. Microbiol.* **171**, 441–7. doi:10.1016/j.vetmic.2014.02.017

Andersson, D.I., Hughes, D., 2010 Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat. Rev. Microbiol.* **8**, 260–71. doi:10.1038/nrmicro2319

Aronesty, E., 2013 Comparison of sequencing utility programs. *Open Bioinforma. J.* **7**, 1–8. doi:10.2174/1875036201307010001

Ash, R.J., Mauck, B., Morgan, M., 2002 Antibiotic resistance of Gram-negative bacteria in rivers, United States. *Emerg. Infect. Dis.* **8**, 713–716. doi:10.3201/eid0807.010264

Baquero, F., Martínez, J.-L., Cantón, R., 2008 Antibiotics and antibiotic resistance in water environments. *Curr. Opin. Biotechnol.* **19**, 260–5. doi:10.1016/j.copbio.2008.05.006

Bywater, R.J., 2005 Identification and surveillance of antimicrobial resistance dissemination in animal production. *Poult. Sci.* **84**, 644–648.

Capita, R., Alonso-Calleja, C., 2013 Antibiotic-resistant bacteria: a challenge for the food industry. *Crit. Rev. Food Sci. Nutr.* **53**, 11–48.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012 Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M., Tiedje, J.M., 2009 The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37**, D141–D145.

D'Costa, V.M., McGrann, K.M., Hughes, D.W., Wright, G.D., 2006 Sampling the antibiotic resistome. *Science* **311**, 374–7. doi:10.1126/science.1120800

Dantas, G., Sommer, M.O.A., Oluwasegun, R.D., Church, G.M., 2008 Bacteria subsisting on antibiotics. *Science* **320**, 100–3. doi:10.1126/science.1155157

De Francesco, V., Zullo, A., Ierardi, E., Giorgio, F., Perna, F., Hassan, C., Morini, S., Panella, C., Vaira, D., 2010 Phenotypic and genotypic Helicobacter pylori clarithromycin resistance and therapeutic outcome: benefits and limits. *J. Antimicrob. Chemother.* **65**, 327–32. doi:10.1093/jac/dkp445

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011 UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200.

Fajardo, A., Martínez, J.L., 2008 Antibiotics as signals that trigger specific bacterial responses. *Curr. Opin. Microbiol.* **11**, 161–7. doi:10.1016/j.mib.2008.02.006

Finch, R.G., 2004 Antibiotic resistance: a view from the prescriber. *Nat. Rev. Microbiol.* **2**, 989–994.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., Wickham, J., 2011 Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogramm. Eng. Remote Sens.* **77**, 858–864.

Gullberg, E., Cao, S., Berg, O.G., Ilbäck, C., Sandegren, L., Hughes, D., Andersson, D.I., 2011 Selection of resistant bacteria at very low antibiotic concentrations. *PLoS Pathog.* **7**, e1002158. doi:10.1371/journal.ppat.1002158

Huerta, B., Marti, E., Gros, M., López, P., Pompêo, M., Armengol, J., Barceló, D., Balcázar, J.L., Rodríguez-Mozaz, S., Marcé, R., 2013 Exploring the links between antibiotic occurrence, antibiotic resistance, and bacterial communities in water supply reservoirs. *Sci. Total Environ.* **456**, 161–70. doi:10.1016/j.scitotenv.2013.03.071

Huse, S.M., Welch, D.M., Morrison, H.G., Sogin, M.L., 2010 Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* **12**, 1889–98. doi:10.1111/j.1462-2920.2010.02193.x

Kim, S.-C., Carlson, K., 2007 Temporal and spatial trends in the occurrence of human and veterinary antibiotics in aqueous and river sediment matrices. *Environ. Sci. Technol.* **41**, 50–57. doi:10.1021/es060737+

Kunin, V., Engelbrektson, A., Ochman, H., Hugenholtz, P., 2010 Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* **12**, 118–23. doi:10.1111/j.1462-2920.2009.02051.x

LaPara, T.M., Burch, T.R., McNamara, P.J., Tan, D.T., Yan, M., Eichmiller, J.J., 2011 Tertiary-treated municipal wastewater is a significant point source of antibiotic resistance genes into Duluth-Superior Harbor. *Environ. Sci. Technol.* **45**, 9543–9549.

Levy, S.B., Marshall, B., 2004 Antibacterial resistance worldwide: causes, challenges and responses. *Nat. Med.* **10**, S122–S129.

412 Lupo, A., Coyne, S., Berendonk, T.U., 2012 Origin and evolution of antibiotic resistance: the
413 common mechanisms of emergence and spread in water bodies. *Front. Microbiol.* **3**, 18.
414 doi:10.3389/fmicb.2012.00018

415 Marti, E., Jofre, J., Balcazar, J.L., 2013 Prevalence of antibiotic resistance genes and bacterial
416 community composition in a river influenced by a wastewater treatment plant. *PLoS One* **8**,
417 e78906. doi:10.1371/journal.pone.0078906

418 Martínez, A., Osburne, M.S., 2013 Preparation of fosmid libraries and functional metagenomic
419 analysis of microbial community DNA. *Methods Enzymol.* **531**, 123–42. doi:10.1016/B978-
420 0-12-407863-5.00007-1

421 Martinez, J.L., 2009 Environmental pollution by antibiotics and by antibiotic resistance
422 determinants. *Environ. Pollut.* **157**, 2893–2902.

423 Martínez, J.L., 2008 Antibiotics and antibiotic resistance genes in natural environments. *Science*
424 **321**, 365–7. doi:10.1126/science.1159483

425 Martinez, J.L., Fajardo, A., Garmendia, L., Hernandez, A., Linares, J.F., Martínez-Solano, L.,
426 Sánchez, M.B., 2009 A global view of antibiotic resistance. *FEMS Microbiol. Rev.* **33**, 44–
427 65. doi:10.1111/j.1574-6976.2008.00142.x

428 Novo, A., André, S., Viana, P., Nunes, O.C., Manaia, C.M., 2013 Antibiotic resistance,
429 antimicrobial residues and bacterial community composition in urban wastewater. *Water*
430 *Res.* **47**, 1875–87. doi:10.1016/j.watres.2013.01.010

431 Poirel, L., Kämpfer, P., Nordmann, P., 2002 Chromosome-encoded Ambler class A beta-
432 lactamase of Kluyvera georgiana, a probable progenitor of a subgroup of CTX-M extended-
433 spectrum beta-lactamases. *Antimicrob. Agents Chemother.* **46**, 4038–40.

434 Poirel, L., Rodriguez-Martinez, J.-M., Mammeri, H., Liard, A., Nordmann, P., 2005 Origin of
435 plasmid-mediated quinolone resistance determinant QnrA. *Antimicrob. Agents Chemother.*
436 **49**, 3523–5. doi:10.1128/AAC.49.8.3523-3525.2005

437 Pruden, A., Arabi, M., Storteboom, H.N., 2012 Correlation between upstream human activities
438 and riverine antibiotic resistance genes. *Environ. Sci. Technol.* **46**, 11541–11549.

439 Pruden, A., Pei, R., Storteboom, H., Carlson, K.H., 2006 Antibiotic resistance genes as emerging
440 contaminants: studies in northern Colorado. *Environ. Sci. Technol.* **40**, 7445–7450.
441 doi:10.1021/es0604131

442 Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W.G., Peplies, J., Glockner, F.O., 2007
443 SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA
444 sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196.

Riesenfeld, C.S., Goodman, R.M., Handelsman, J., 2004 Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ. Microbiol.* **6**, 981–9. doi:10.1111/j.1462-2920.2004.00664.x

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009 Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541. doi:AEM.01541-09 [pii] 10.1128/AEM.01541-09

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J., 2006 Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U. S. A.* **103**, 12115–12120.

Staley, C., Gould, T.J., Wang, P., Phillips, J., Cotner, J.B., Sadowsky, M.J., 2014 Core functional traits of bacterial communities in the Upper Mississippi River show limited variation in response to land cover. *Front. Microbiol.* **5**, 414. doi:10.3389/fmicb.2014.00414

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

Taylor, N.G.H., Verner-Jeffreys, D.W., Baker-Austin, C., 2011 Aquatic systems: maintaining, mixing and mobilising antimicrobial resistance? *Trends Ecol. Evol.* **26**, 278–284.

Walsh, T.R., Weeks, J., Livermore, D.M., Toleman, M.A., 2011 Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: an environmental point prevalence study. *Lancet Infect. Dis.* **11**, 355–62. doi:10.1016/S1473-3099(11)70059-7

Wright, G.D., 2010 Antibiotic resistance in the environment: a link to the clinic? *Curr. Opin. Microbiology* **13**, 589–94. doi:10.1016/j.mib.2010.08.005

Yan, T., Hamilton, M.J., Sadowsky, M.J., 2007 High-throughput and quantitative procedure for determining sources of Escherichia coli in waterways by using host-specific DNA marker genes. *Appl. Environ. Microbiol.* **73**, 890–6. doi:10.1128/AEM.01395-06

476 **Table 1**. Percentages of the fosmid libraries showing resistance to antibiotics tested in 2011

477 (top) and 2012 (bottom). The limit of detection for all screens was 0.01%.

Site	Land Cover*	Library Size	% of Resistant Clones			
			Amp	Cet	Kan	Tet
Itasca	Forest	10368	0.16	0.18	0.46	ND†
		9984	ND	0.70	0.02	ND
St. Cloud	Agriculture	9984	ND	0.06	ND	ND
		9984	0.01	0.16	ND	ND
Clearwater	Agriculture	9984	0.02	0.01	0.01	ND
		9984	ND	0.28	0.02	ND
Twin Cities	Developed	9600	ND	ND	ND	ND
		9984	ND	0.50	0.51	ND
Minnesota River	Agriculture	9984	0.01	0.04	ND	ND
		9984	0.01	0.09	ND	ND
Confluence	Developed	9600	ND	0.01	ND	ND
		9984	ND	0.71	ND	ND
Hastings	Developed	9984	ND	0.29	0.02	ND
		9984	ND	0.09	0.02	ND
St. Croix River	Agriculture	9216	ND	0.20	ND	ND
		9600	ND	0.49	ND	ND
Red Wing	Agriculture	9984	0.02	0.10	0.07	ND
		9984	ND	0.06	ND	ND
La Crescent	Forest	9984	ND	0.61	0.86	ND
		9984	ND	0.17	ND	ND
Zumbro River	Agriculture	9984	0.02	0.15	0.08	ND
		9600	ND	0.01	ND	ND

478 *Predominant surrounding land cover of the water basin in which the site is located.

479 †ND: not detected.

480 **Table 2**. Mean and standard deviations (in parentheses) of major bacterial phyla (classes of *Proteobacteria*), antecedent rainfall, and

481 physicochemical data collected in 2011 and 2012.

Year	Bacteria (% sequence reads)					Rainfall (mm)				Physicochemical parameters (mg L^{-1})							
	β-proteobacteria	*γ-proteobacteria*	*α-proteobacteria*	*Actinobacteria*	*Bacteroidetes*	72 h	48 h	24 h	Cumulative	Temp (°C)	pH	NH_4^+	NO_2/NO_3	Orthophosphate	Total phosphorus	Total organic carbon	TDS*
2011	58.0 (7.6)	7.4 (13.8)	1.8 (1.1)	1.8 (1.2)	0.8 (0.5)	1.8 (2.8)	3.0 (5.0)	2.6 (5.9)	7.4 (7.8)	18.2[a] (2.4)	7.7 (0.3)	0.1[a] (< 0.1)	2.5 (2.4)	0.2 (0.1)	0.1[a] (< 0.1)	6.2 (2.0)	79.9[a] (33.4)
2012	61.4 (7.9)	0.7 (0.6)	0.5 (0.2)	0.5 (0.2)	0.2 (0.1)	4.7 (8.6)	3.5 (11.6)	5.2 (6.9)	13.4 (15.3)	21.5[b] (1.8)	7.6 (0.3)	0.05[b] (< 0.1)	2.0 (2.0)	0.1 (< 0.1)	0.1[b] (< 0.1)	8.7 (3.7)	44.2[b] (19.1)

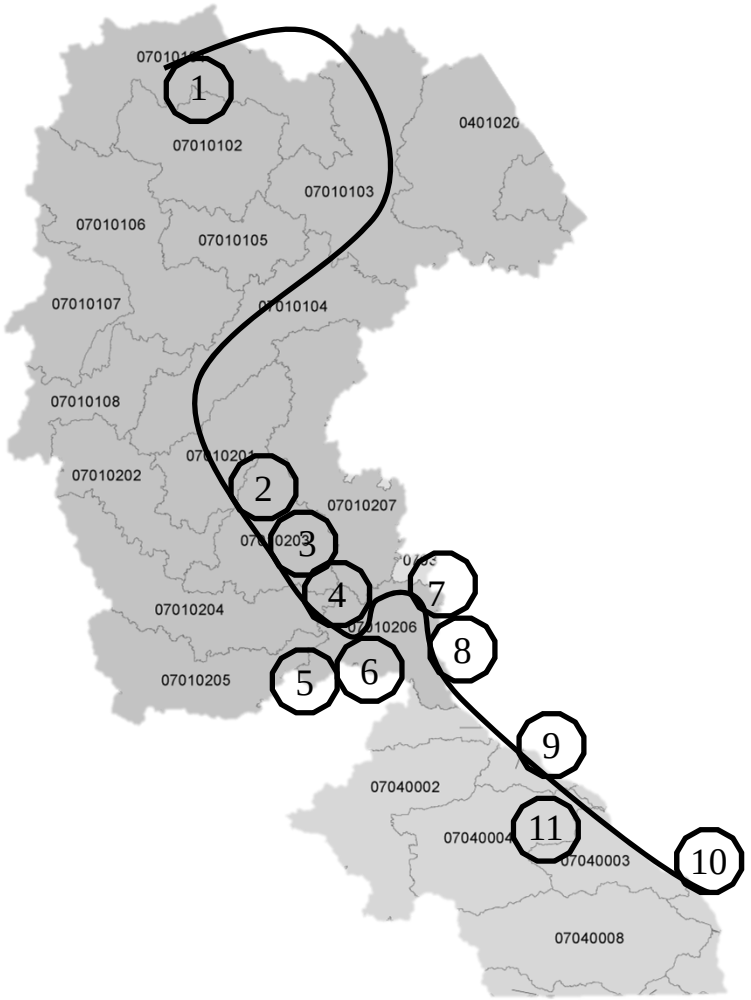
482 *Total dissolved solids.

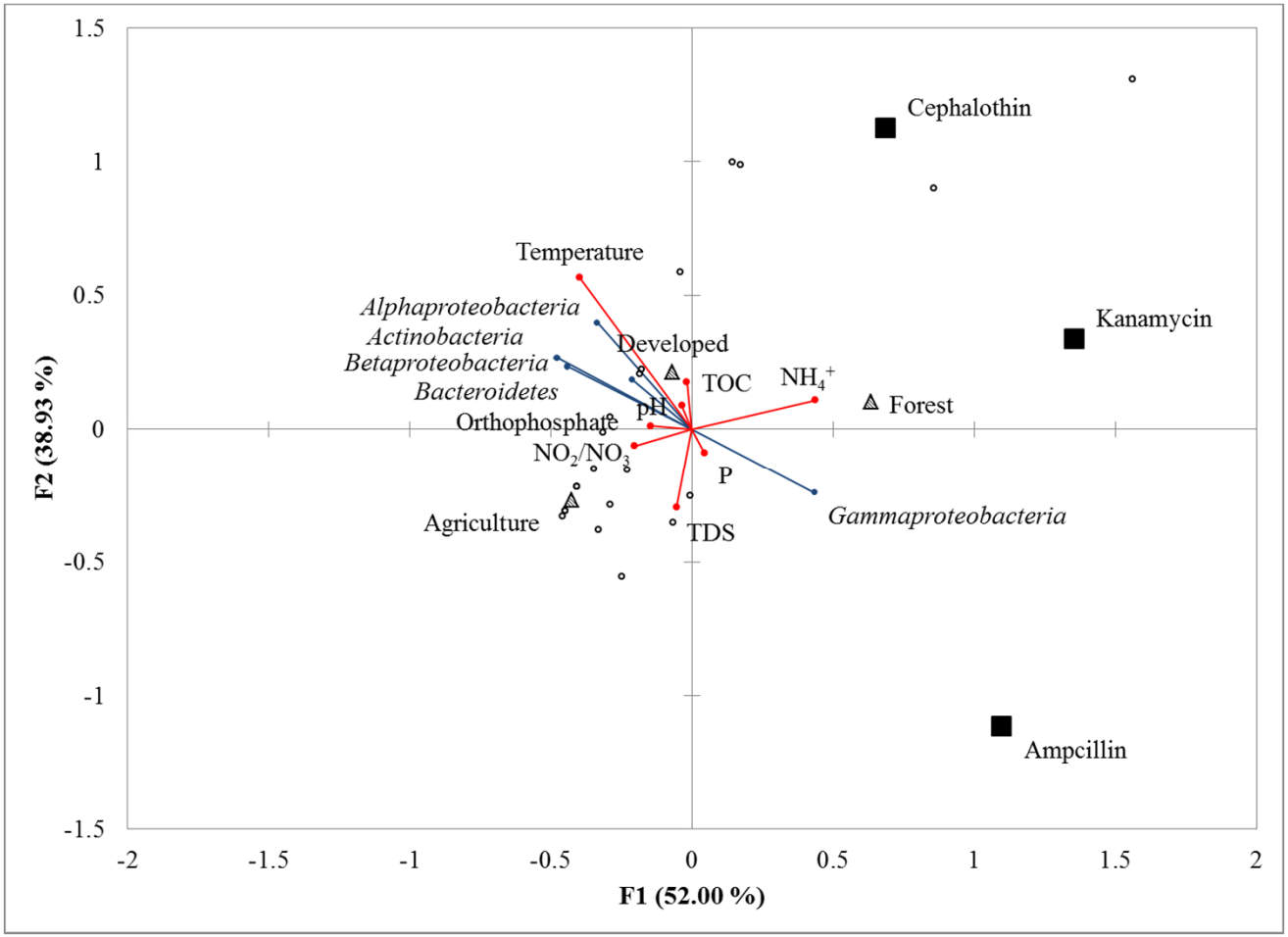
483 [a,b]Where indicated, data was significantly different between years (ANOVA, $\alpha = 0.05$)

FIGURE LEGENDS

Figure 1. Approximate location of sampling sites with basin boundaries. The darker gray area (top) represents the Upper Mississippi River drainage area, while the light gray area represents that of the Lower Mississippi River. Sites represented include 1: Itasca, 2: St. Cloud, 3: Clearwater, 4: Twin Cities, 5: Minnesota River, 6: Confluence, 7: Hastings, 8: St. Croix River, 9: Red Wing, 10: La Crescent, and 11: Zumbro River. This map was modified from Minnesota Pollution Control Agency [http://www.pca.state.mn.us/index.php/water/water-types-and-programs/surface-water/basins/basins-and-watersheds-in-minnesota.html] and numbers indicate 8-digit basin unit codes.

Figure 2. Redundancy analysis relating major bacterial groups, physicochemical data, and land cover to frequencies of antibiotic resistance observed. Antibiotics are shown as black squares, bacterial groups are shown as blue lines, physicochemical data are shown as red lines, and land cover type is shown as shaded triangles. Open, black circles represent sampling points.


1
07010102
07010103
07010106
07010105
07010107
07010104
07010108
07010202
2
07010207
3
07010204
4
7
8
07010205
5
6
9
07040002
11
10
07040003
07040008


1.5
1
0.5
0
-0.5
-1
-1.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
F2 (38.93 %)
F1 (52.00 %)
Cephalothin
Kanamycin
Ampcillin
Temperature
Alphaproteobacteria
Actinobacteria
Betaproteobacteria
Bacteroidetes
Developed
TOC
NH4+
Forest
pH
Orthophosphate
NO2/NO3
P
Gammaproteobacteria
Agriculture
TDS

Species sorting dynamics promote community resilience in response to natural and anthropogenic disturbance in the Upper Mississippi River

Christopher Staley[1], Trevor J. Gould[1,2], Ping Wang[1], Jane Phillips[2], James B. Cotner[3], and Michael J. Sadowsky[1,4,*]

[1]BioTechnology Institute, [2]Biology Program, [3]Department of Ecology, Evolution and Behavior, and [4]Department of Soil, Water and Climate, University of Minnesota, St. Paul, MN

Running title: Bacterial community dynamics in Mississippi River

[*]Corresponding Author: Michael J. Sadowsky, BioTechnology Institute, University of Minnesota, 140 Gortner Lab, 1479 Gortner Ave, Saint Paul, MN 55108; Phone: (612)-624-2706, Email: sadowsky@umn.edu

Abstract

Bacterial community structure (BCS) in freshwater ecosystems varies seasonally and due to physicochemical gradients, but metacommunity structure of a major river remains understudied. Here we characterize the BCS along the Mississippi River and contributing rivers in Minnesota over three years using Illumina next-generation sequencing, to determine how changes in environmental conditions as well as inputs from surrounding land and confluences impacted community structure. Contributions of sediment to water microbial diversity were also evaluated. Long-term variation in community membership was observed, and significant shifts in relative abundances of major freshwater taxa, including *α-Proteobacteria*, *Burkholderiales*, and *Actinomycetales*, were observed due to temporal and spatial variation. Taxa abundances were correlated primarily with temperature and rainfall, but also nutrient concentrations, suggesting that species sorting played a predominant role in shaping BCS. Furthermore, an annually-recurrent BCS was observed in late summer, further suggesting that seasonal dynamics strongly influence community composition. Sediment communities differed from those in the water, but contributed up to 50% to community composition in the water column. Among water sampling sites, 34% showed significant variability in BCS of replicate samples indicating variability among riverine communities due to heterogeneity in the water column. Results of this study highlight the need for a better understanding of spatial and temporal variation in riverine bacterial diversity associated with physicochemical gradients and reveal how communities in sediments, and potentially other environmental reservoirs, impact waterborne BCS. Techniques used in this study may prove useful to determine sources of microbes from sediments and soils to waterways, which will facilitate best management practices and total maximum daily load determinations.

24 Keywords: bacterial community structure/microbial ecology/high-throughput
25 sequencing/metacommunity theory/Mississippi River/recreational water

1. Introduction

The emergence of high-throughput, next-generation sequencing technology has allowed for a better characterization of bacterial communities from a variety of environments, including marine and fresh waters (Sogin et al., 2006; Gilbert et al., 2009; Fortunato et al., 2012; Staley et al., 2013), soils (Jones et al., 2009), wastewater (Sanapareddy et al., 2009), and the human microbiome (Peterson et al., 2009). The most common focus of these studies has been on the ecological biogeography of bacterial communities – how community structure varies within and between habitats in response to biotic interactions and shifts in abiotic environmental parameters (*e.g.* pH, temperature, etc.) (Horner-Devine et al., 2003; Pernthaler 2005; Gilbert et al., 2009, 2012; Lindström and Langenheder 2012). Bacterial communities in aquatic systems have been shown to fluctuate in response to physicochemical factors including temperature, day length, and nutrient concentrations (Gilbert et al., 2009, 2012), and these fluctuations have been demonstrated to be seasonally reproducible in both marine and freshwater ecosystems (Crump and Hobbie 2005; Shade et al., 2007; Fortunato et al., 2012; Gilbert et al., 2012). Deep sequencing of a sample from the English Channel, however, has revealed that a persistent microbial seed bank may exist for particular ecosystems, *e.g*. marine waters (Caporaso et al., 2012b). This finding has generally been interpreted to support the Baas-Becking hypothesis that 'everything is everywhere, but the environment selects' (Baas-Becking 1934), suggesting that due to the small size of microorganisms, their rapid rates of reproduction, and high dispersal ability, among other factors, the extinction of particular taxa is unlikely (Lindström and Langenheder 2012).

A metacommunity framework has been proposed for exploring the role of environmental variation as well as population dispersal dynamics in shaping bacterial community structure

(Leibold et al., 2004). In this framework, a metacommunity consists of local communities of potentially interacting taxa that are linked by dispersal. Four perspectives have been presented to describe environmental and spatial influences on local community structure (Leibold et al., 2004), and these include: (1) the patch-dynamics perspective which assumes that colonization and extinction of environmentally indistinct patches influence the local distribution of particular taxa, and diversity is limited by species dispersal; (2) the species sorting perspective, which is equivalent to the Baas-Becking hypothesis, where resource gradients and environmental conditions primarily drive community composition; (3) the mass-effect perspective which describes the influence of environmental gradients and dispersal dynamics among local communities on local community composition; and (4) the neutral perspective that assumes all populations have equivalent competitive ability, and community variation occurs due to stochastic processes of dispersal and extinction.

The species sorting and mass-effect perspectives are generally favored in most environmental studies of microbial biogeography, as the patch-dynamics and neutral theories are thought to be poorly applicable to environmental studies conducted over broad spatial scales (Lindström and Langenheder 2012; Winter et al., 2013). Abundant evidence of bacterial biogeography structured along numerous gradients of environmental parameters suggests that the patch dynamics perspective can be ruled out as patches show apparent variation in physicochemical parameters and/or nutrient availability (Martiny et al., 2006). Furthermore, several studies have demonstrated that the habitat-specific distribution of bacterial populations suggests that the neutral perspective of the metacommunity framework can likely be excluded (Zwart et al., 2002; Nemergut et al., 2011).

71 The influences of environmental heterogeneity and dispersal effects on contemporary
72 local community structures have been reviewed numerous times (Martiny et al., 2006; Hanson et
73 al., 2012; Lindström and Langenheder 2012). However, influences on community structure
74 resulting strictly from variation in environmental parameters vs. that from spatial distance have
75 been difficult to discern, in part due to the co-variation of physicochemical parameters with
76 spatial distance in many recent studies (Lindström and Langenheder 2012; Winter et al., 2013).

77 Temperature and salinity have been reported to be among the most influential physical
78 and chemical parameters affecting the composition of bacterial communities in aquatic habitats
79 (Tamames et al., 2010). However, many recent studies examining the effect of spatial distance
80 on bacterial community composition have been confounded by co-variation of distance with
81 temperature and/or salinity (Fortunato et al., 2012; Winter et al., 2013). Although all riverine
82 systems will likely exhibit spatial and temporal heterogeneity in physical and chemical
83 parameters (Winemiller et al., 2010), restriction of a study area and sampling period to a region
84 with relatively consistent temperature and salinity may allow for better characterization of spatial
85 variation within this system. Furthermore, while co-variation of environmental parameters, such
86 as nutrient availability with distance, may occur as a result of a riverine continuum (Vannote et
87 al., 1980), such factors will likely be offset by more significant terrestrial inputs such as
88 agricultural runoff or wastewater effluent discharge (Pereira and Hostettler 1993; Drury et al.,
89 2013). These inputs may result in predictable variation in physicochemical parameters in local
90 environments and bacterial communities may be shaped by anthropogenic impacts, as has been
91 previously suggested (Staley et al., 2013). These local communities nevertheless remain
92 interconnected by dispersal of upstream populations and may be influenced by communities in
93 sediments, as well.

In order to evaluate effects of environmental parameters (*i.e.* rainfall, temperature, pH, and nutrient concentrations) and spatial distances on community structure in a riverine system, bacterial community structures were characterized annually from 2010 to 2012 at 11 sampling sites along the Upper Mississippi River using Illumina next-generation sequencing. Sampling was conducted during the early summer following annual flushing due to snowmelt in mid-late spring to avoid confounding effects of this temporally inconsistent event. We hypothesized that physicochemical parameters and nutrient concentrations, as well as stochastic inputs from runoff, would account for greater variability in community structure than spatial distance, as has been previously demonstrated in a marine system (Winter et al., 2013). Furthermore, seasonally-associated variations of environmental parameters were anticipated to result in an annually reproducible community structure, as was observed in the English Channel (Gilbert et al., 2012).

To evaluate these seasonal dynamics, samples were collected biweekly over a 12-week period in 2011 and 2012 to observe short-term seasonal progressions in community structure. In addition, since environmental reservoirs (*i.e.* sediments) and depth were suspected to be among variables that impacted the bacterial community structure, the influences of these parameters were evaluated in the later years of study. Results of this study will help elucidate processes controlling and shaping bacterial community structure in a riverine ecosystem over a relatively large spatial area (> 400 km), and will reveal potentially significant influences of environmental reservoirs influencing community composition.

2. Materials and methods

2.1. Sample collection

Water samples were collected in the early summer of 2010 to 2012 from 8 sites along the Upper Mississippi River in Minnesota, one site near the confluence of the Minnesota River with the Mississippi, and one site on the St. Croix River near its confluence with the Mississippi River, as described previously (Staley et al., 2013). In 2011 and 2012, an 11th sampling site was added on the Zumbro River [44.314, -91.996]. Sampling sites were selected to span the entirety of the Mississippi River in Minnesota and to capture unimpacted communities as well as potential variations due to major agricultural and urban inputs and variation at confluences. Briefly, 40 L of water was collected at each site from the water's surface, transported back to the lab in sterile 20 L carboys, and either processed immediately or stored < 24 h at 15° C. In 2011 and 2012, samples (20 L) were collected from both the surface and at a depth of approximately 1.5 m using a canoe pump at the Twin Cities site, on six dates each year, two weeks apart. In 2012, sediment samples (approximately 5 cm depth) were also collected using an ethanol-sterilized, stainless steel soil auger, approximately 0.5 m from the river bank at the same time of water sample collection. Water temperature and pH were measured at the surface at each sampling date in the field. Rainfall data, up to three days prior to sampling, was also obtained from Weather Underground [http://www.wunderground.com]. Site locations and physical data are summarized in Supplementary Table S1.

In 2011 and 2012, at the time of collection of the 40 L sample, additional 1 L water samples were collected at all sites in a sterile amber bottles and stored at 4° C for nutrient analyses. From these samples, total carbon, colorimetric nitrite/nitrate, total phosphorus, and total dissolved solids (TDS) were measured by the Research Analytical Laboratories at the University of Minnesota (St. Paul), and these data are shown in Supplementary Table S1.

139 *2.2. Sample processing*

140 Sample processing was performed as previously described (Staley et al., 2013). Briefly,
141 water samples were passed through four layers of sterile cheesecloth and sequentially filtered
142 through P5 pre-filters (Whatman Inc., Piscataway, NJ) followed by bacterial capture on 0.45-µm
143 pore-sized filters (Pall Co., Port Washington, NY). Filters were changed as needed due to
144 clogging, and six to eight 0.45-µm-pore-size filters were required to filter the total 40 L sample
145 volume. We have previously evaluated the effect of filter size on bacterial community
146 characterization and found this filter pore size best accommodated the filtration of large volumes
147 of water with minimal bias in OTUs identified (Staley et al., 2013). Cells were elutriated as
148 previously described by vortexing in pyrophosphate buffer (Staley et al., 2013). Cell suspensions
149 were pelleted in 1.7 ml microcentrifuge tubes at 16,000 × *g*. A total of six cell pellets were
150 obtained for each 40 L water sample and thus, each pellet represented approximately a 6 L water
151 sample. For surface/depth samples, three pellets were obtained per sample so that all pellets
152 represented an equivalent sample volume. Cell pellets were stored frozen at -80° C.

153 DNA was extracted from two cell pellets per sampling site using the DNA Isolation Kit
154 for Water (Epicentre, Madison, WI). Each pellet was resuspended in 600 µl of TE buffer and 300
155 µl was removed to a separate 1.7 ml tube. DNA was extracted from both half-pellet suspensions
156 according to the manufacturer's instructions, and the final eluates from both half-pellet
157 suspensions were combined. DNA extracts from three separate cell pellets were not used because
158 three pellets were not available for many of the sites at the time of the second DNA extraction
159 due to use in other experiments not described here. DNA was extracted from individual 0.25 g
160 sediment samples from each site using the MoBio PowerSoil® (Carlsbad, CA) kit, according to

161 the manufacturer's instructions. Additional sediment samples were not available for replicate
162 DNA extraction at the time of the second DNA extraction.

163

164 *2.3. PCR and sequencing*

165 The V6 hypervariable region of the 16S rDNA was amplified using the 967F/1046R
166 primer set containing unique 6 nt multiplexing barcode sequences to identify each sample (Sogin
167 et al., 2006). Amplicons were gel purified using the QiaQuick® Gel Extraction Kit (Qiagen,
168 Valencia, CA), according to the manufacturer's instructions. Purified amplicons were pooled in
169 equal amounts for sequencing. Amplicons from one set of DNA extracts (one pellet per water
170 sample) and sediment samples were paired-end sequenced on the Illumina MiSeq platform at a
171 read length of 2 × 150 bp. To accommodate sample pools composed of larger numbers of
172 samples, and to reduce sequencing cost, duplicate amplicons from the second set of DNA
173 extracts, from water samples only, were paired-end sequenced (duplicates were separated into
174 different pools) on the Illumina HiSeq2000 platform at a read length of 2 × 100 bp. All
175 sequencing was performed at the University of Minnesota Genomics Center (Minneapolis).
176 Conclusions drawn from sequences generated from MiSeq and HiSeq2000 have been previously
177 demonstrated to be reproducible across platforms (Caporaso et al., 2012a). Sequencing results
178 were obtained as .fastq files and were submitted to the National Center for Biotechnology
179 Information Sequence Read Archive under BioProject accession number SRP018728.

180

181 *2.4. Sequence processing and analysis*

All sequence processing and analysis was performed using mothur software ver. 1.29.2 (Schloss et al., 2009). Sequence lengths were adjusted to 100 bp to account for differences in read length. Trimmed sequences were paired-end aligned using fastq-join (Aronesty 2013) and screened for quality using the following parameters: quality score ≥ 35 over a 50 nt window, no ambiguous bases, homopolymers ≤ 8 nt, and primer and barcode matching with 100% identity. Samples were processed as described previously (Schloss and Westcott 2011), with some modifications. Sequences with an abundance < 2 were removed, chimeras were removed using UCHIME software (Edgar et al. 2011), and sequences were aligned against the SILVA-alignment database ver. 102 (Pruesse et al. 2007). Operational taxonomic units (OTUs) were assigned at 97% similarity using the furthest neighbor algorithm and classified against the Ribosomal Database Project dataset ver. 9 (Cole et al. 2009). For all comparisons, sequence numbers per sample were normalized, by random subsampling, to that of the sample (single replicate) with fewest reads. This was 25,717 reads per sample for main (40 L) sample and sediment analyses and 14,591 sequence reads per sample for depth (20 L) samples.

To determine the influence of sediment communities on those of the water column, all sediment samples collected in 2012 were treated as a source and compared against a sink community of water samples also collected in 2012. The percentage contribution was determined using the SourceTracker subroutine (Knights et al., 2011) using default parameters.

2.5. Statistical analyses

Diversity indices [number of OTUs (S_{obs}), Shannon, non-parametric Shannon, and Simpson], sequencing coverage estimation, UniFrac analysis (Lozupone and Knight 2005),

ANOSIM analysis (Clarke 1993), principal coordinate analysis (PCoA), Mantel tests (Sokal and Rohlf 1995), Kruskal-Wallis analysis (Acar and Sun 2013), and analysis of molecular variance (AMOVA) (Excoffier et al. 1992) were performed using mothur ver. 1.32.1 (Schloss et al. 2009). For statistical tests in mothur, replicates were grouped via use of .design files, but were maintained as separate groups. Bray-Curtis distance matrices were calculated for community comparisons (Bray and Curtis 1957). One-way analysis of variance (ANOVA) followed by Tukey's *post-hoc* test was performed using SPSS Statistics software v. 19.0 (IBM, Armonk, NY). Spearman rank correlation analyses and multiple linear regression for variance partitioning were also conducted using SPSS software. All statistical analyses were conducted at $\alpha = 0.05$.

3. Results

3.1. Diversity of water and sediment samples

Among all water samples, 13,582, 13,848, and 14,588 OTUs were identified in 2010, 2011, and 2012 samples, respectively. In addition, 22,004 OTUs were identified in the 2012 sediment samples. Individual water samples (represented by DNA from one cell pellet) had a mean of 1,481 ± 252 OTUs (Table 1), while sediment samples had about 3-times that amount, 4,728 ± 523 OTUs. Diversity was similar among water samples over all three years (mean Shannon indices of 4.37, 3.93, and 4.12 for 2010, 2011, and 2012, respectively), but the lower diversity observed for 2011 samples was statistically significant ($P \leq 0.034$). Diversity was greater in sediment samples compared to water samples (mean Shannon index of 7.03 ± 0.23, $P < 0.001$).

225 Physicochemical parameters, nutrient concentrations, and distance from the headwaters
226 were also individually correlated with diversity (Supplementary Tables S2 and S3), but
227 responses were different depending on the study year. In 2010 and 2011, water temperature was
228 significantly and positively correlated with Shannon diversity indices ($r = 0.423$-0.482, $P \leq$
229 0.020). In 2012, however, a negative correlation between temperature was seen between
230 temperature and diversity measured by the inverse Simpson index ($r = -0.380$, $P = 0.047$). While
231 cumulative rainfall was correlated with diversity indices in 2010 ($r = 0.466$-0.504, $P \leq 0.009$), in
232 2011 and 2012, negative correlations between 48 h antecedent rainfall and diversity indices were
233 observed ($r = -0.426$-0.543, $P \leq 0.013$). There were no significant relationships between nutrient
234 concentrations and diversity measures.

235 Distance from the headwaters also showed variable effects on diversity. Distance was not
236 correlated with diversity indices in 2010 or 2012 ($P \geq 0.140$). However, distance was
237 significantly positively correlated with both Shannon and inverse Simpson indices in 2011 ($r =$
238 0.428-0.636, $P \leq 0.037$). ANOVA analyses of differences in physicochemical and nutrient
239 parameters as well as intercorrelations of distance, physicochemical parameters, and nutrient
240 concentrations are discussed in Supplementary Results.

241

242 *3.2. Annual variability in bacterial community structure*

243 Over all three years, the bacterial community structure of water samples was dominated
244 by the *Proteobacteria* (*α*, *β*, and *γ*) and *Actinobacteria* (Supplementary Table S4). Forty-six
245 OTUs, primarily classified to the *β-Proteobacteria*, were found in all samples, and accounted for
246 a mean 61.8 ± 15.8% of sequence reads among individual replicates. Higher percentages of

247 shared sequence reads were observed in samples collected in 2011 (75.4 ± 13.4%) compared to
248 2010 (69.5 ± 13.4%) and 2012 (67.0 ± 14.1%). Both unweighted and weighted UniFrac analysis,
249 however, revealed significant differences in the presence and abundance of OTUs, respectively,
250 among all three years (Sites 1-10 only, $P < 0.001$ for all comparisons). Significant differences in
251 community structure between years were also observed by ANOSIM ($P \leq 0.008$). Despite
252 differences in community structures, however, Mantel tests revealed that the bacterial
253 community compositions of the ten sites sampled each year were related, even between the 2010
254 and 2012 communities ($P < 0.001$).

255 Among all samples grouped by year, 1,479 OTUs were found to differ significantly in
256 relative abundance between years by using the Kruskal-Wallis test (Fig. 1). Among these OTUs,
257 and when classified to the order level, *Burkholderiales* (specifically *Polynucleobacter* spp.)
258 accounted for the greatest percentage of sequence reads in all three years. *Pseudomonadales*
259 showed greater prevalence in 2011 and was also found to be one of the predominant orders
260 among variable OTUs in 2012. This, however, is likely due to spikes in relative abundance of
261 this order at both the Itasca and St. Cloud sites in 2011 and the Twin Cities and Minnesota River
262 sites in 2012 (Figs 3 and 4 and Supplementary Table S4).

263 Shifts in the relative abundance of phyla were correlated with variation in
264 physicochemical and nutrient parameters (Supplementary Tables S2 and S3). Among
265 physicochemical parameters, several of these associations showed annual variation in
266 significance and direction. Discrepancies in correlations between sampling years are likely due
267 in part to co-variation of the physicochemical parameters measured. Some differences, such as
268 the negative correlation between *Alphaproteobacteria* abundance and 48 h antecedent rainfall,
269 were observed during multiple years of study and likely reflect consistent influences of these

parameters on particular taxa. Among nutrient concentrations measured (Table S3), nearly all of the relationships observed over both study years were conserved in at least one of the sampling years (*e.g.* positive correlations were observed over both years and in an individual year).

Replicate water samples did not differ significantly in the OTUs identified (unweighted UniFrac $P = 1.000$ for all pairwise comparisons); however, in 34.4% of replicate sets (n = 32), differences in community structure, determined as the relative abundances of OTUs, in at least one replicate were significant ($P \leq 0.048$) when analyzed using weighted UniFrac. Furthermore, ANOSIM analysis revealed significant differences in beta diversity among all replicate pairs ($P < 0.001$). No pattern in variation among replicates was observed in relationship to the sequencing platform or the pellet used for DNA extraction (*e.g.*, communities characterized from extracts from different pellets did not consistently differ significantly).

3.3. Spatial variability in bacterial community structure

Bacterial communities at several sampling sites showed variation in the relative abundance of OTUs, but not in OTU presence/absence or overall community structure within the same year. Almost all pairwise analyses, performed using weighted UniFrac analyses, revealed significant differences in the abundances of OTUs among sampling sites ($\alpha = 0.05$), but no apparent trends were observed in variability during any year of study. Unweighted UniFrac analysis, however, showed that variation in the OTUs present at each site was not significant during any single year ($P \geq 0.189$). Analysis by ANOSIM also revealed no significant pairwise differences in beta diversity among sampling sites during any single year ($P \geq 0.080$).

Kruskal-Wallis analysis of the OTUs showing significant variation in relative abundance among sampling sites revealed that the majority of these OTUs belonged to the same orders that showed annual variability in abundance (Figs 1-4). Annual temporal variation in OTU abundances accounted for fewer sequence reads than did spatial variation within a year, where OTUs that varied in abundance accounted for approximately 35% of sequence reads for a single year, but up 45-80% of total sequence reads at a single sampling site. The majority of OTUs that showed variation in abundance among sampling sites were classified to *Burkholderiales* in all three years. Site-specific increases in the relative abundance of certain orders were observed, for example *Verrucomicrobiales* and *Caulobacterales* were abundant in the 2010 Hastings sample (Fig. 2), but these increases in abundance were not maintained at downstream sites. Distance from the headwaters, however, was significantly correlated with the relative abundance of many of the phyla identified in river water samples (Supplementary Table S2).

3.4. Evaluation of species sorting dynamics

In order to partition variation in community structure between environmental variables and spatial distances, multiple linear regression analyses were performed to determine which parameters significantly affected community diversity as well as the relative abundance of major phyla (Table 2). Over both years of study, only temperature was significantly related to diversity measured by the Simpson index ($\beta = -2.363$, $P = 0.024$). Temperature and rainfall also had among the largest significant effects on the relative abundances of the most abundant phyla, although concentrations of nutrients, TDS, and pH also had significant relationships with phylum abundance. Distance was only significantly related to the abundances of *Verrucomicrobia* and

313 *Cyanobacteria*. Of note was the lack of association of any parameter with the abundances of
314 *Actinobacteria* or *Bacteroidetes* over both years or during individual study years.

315 Among individual years of study, drastically different relationships between
316 environmental parameters, distance, and community structure were observed. In 2011, nearly
317 every parameter modeled had a significant effect on diversity, phylum abundances, or both.
318 Conversely, in 2012, only temperature and rainfall had significant effects on community
319 structure and distance could not be incorporated into the model.

320

321 *3.5. Seasonal spatiotemporal variation in bacterial community structure*

322 In 2011 and 2012, 6 sets of samples were collected biweekly from the water's surface and
323 at a depth of approximately 1.5 m at the Twin Cities sampling site (Table 3). Diversity was
324 higher in 2012 than 2011 by all indices, except the Simpson index ($P < 0.001$). In 2011, diversity
325 was also significantly higher in surface samples than those taken at depth, by all indices except
326 the Shannon index ($P \leq 0.037$). No difference in diversity due to depth was observed in 2012.
327 Among samples collected within a given year, pairwise differences in bacterial community
328 composition were not significantly different by unweighted UniFrac analyses ($P \geq 0.425$).
329 However, nearly all pairwise comparisons revealed significant differences in community
330 structure by weighted UniFrac ($P \leq 0.044$). In 2011, samples collected during the first and
331 second sampling dates, from the surface, did not differ significantly from the third and fourth
332 samples collected at depth ($P \geq 0.054$), and in 2012 the first depth and second surface samples
333 did not differ significantly by weighted UniFrac ($P = 0.057$).

Significant differences ($P \leq 0.046$) in community structure were also seen between samples collected during different years (at both depths), at different depths (both years), and between samples collected early vs. later in the summer (all samples) as tested by both UniFrac and ANOSIM analyses. To assess potentially reproducible seasonal dynamics, replicate samples were merged together and treated as a single sample, for simplicity, for principal coordinate analysis (Fig. 5). AMOVA analysis revealed no significant grouping of merged samples by year or depth ($P \geq 0.090$). However, samples collected later in the study (mid-July to August) did show significant clustering, apart from samples collected earlier regardless of depth or year collected (AMOVA, $P < 0.001$), and this grouping was supported when replicates were analyzed separately (*i.e.* not merged, $P = 0.002$). Sample position along both axes was significantly positively correlated with water temperature ($r = 0.629$ and 0.502, $P = 0.001$ and 0.012 for axes 1 and 2, respectively), but not pH ($P \geq 0.106$), and nutrient analysis was not performed on biweekly samples.

3.6. Bacterial communities in sediments

Bacterial communities in sediments consisted predominantly of the α-, *β*-, and *γ*-*Proteobacteria*, *Bacteroidetes*, *Acidobacteria*, and *Actinobacteria* (Fig. 6). Communities in sediments were significantly different from those in the water during any year, as determined by both weighted and unweighted UniFrac and ANOSIM analyses ($P \leq 0.003$). Analysis by ANOSIM revealed significant differences between all sediment samples ($P < 0.001$), but these differences were generally not supported by UniFrac analyses. Unweighted UniFrac analysis did not show significant differences among OTUs present in sediment samples at all sites ($P = 1.000$). Sediment samples were tested as a source of OTUs against water samples using

SourceTracker and were found to contribute 9.0-57.3% to total sequence reads in water samples, with the highest contributions observed in the Minnesota River, the Twin Cities, and near confluences of the Mississippi River with the Minnesota and St. Croix Rivers (Table 4).

4. Discussion

In this study, the bacterial communities in water samples from 11 sites spanning the length of the Mississippi River in Minnesota and major contributing, confluent rivers were characterized over three years via Illumina high-throughput sequencing. Temperature and rainfall were among the predominant factors found to influence community structure, but associations between phylum abundances and nutrient concentrations were also observed suggesting that species sorting dynamics played a role in shaping the bacterial community structure. Associations of some taxa with distance and greater shifts in OTU abundance within-year spatial variability versus annual variability via Kruskal-Wallis test, however, suggest that dispersal and input dynamics may also play a role. Less variation among all sampling sites annually, and lack of significant changes in community membership in a given study year, may indicate that species sorting dynamics, rather than altering community structure, are promoting community resilience in response to inconsistent runoff and other terrestrial inputs as has been previously suggested (García-Armisen et al., 2014). Furthermore, community structure was observed to be similar at the Twin Cities site in the late summer of 2011 and 2012 suggesting that seasonal parameters, specifically temperature, may result in annually reproducible bacterial assemblages.

Based on a higher percentage of shared sequence reads among sites sampled in the same year, as well as unweighted UniFrac and ANOSIM analyses, spatial differences in bacterial communities were likely due to shifts in the relative abundance of OTUs rather than variation in presence/absence. The majority of spatial variation was observed in taxa generally regarded as members of a ubiquitous freshwater community (Zwart et al., 2002). However, sporadic increases in the relative abundance of certain OTUs were observed, most notably in relative abundance of OTUs classified to *Pseudomonadales.* This variation in OTU abundances may be due to terrestrial and anthropogenic impacts, as damming, for example, has been shown to result in greater relative abundances of *α-Proteobacteria* and *Actinobacteria* (Ruiz-González et al., 2013), and effluent outfall has been shown to contribute to biotic homogenization (Drury et al., 2013). Conversely, a *Vibrio* bloom observed in the English Channel was suggested to be primarily associated with eukaryotic activity indicating that biotic interactions may also account for dramatic variation in the relative abundance of taxa (Gilbert et al., 2012). The mechanism(s) responsible for the increase in *Pseudomonadales* observed here require further investigation.

Consistent correlations of certain phyla, such as *Verrucomicrobia* and *Cyanobacteria,* with distance from the headwaters suggest that a geographic gradient exists, as has been previously, but inconsistently, shown for *Cyanobacteria* (Schultz et al., 2013; Kolmakova et al., 2014). However, results of linear regression indicated that temperature and rainfall were more strongly associated with community variation than was distance alone. In addition to physicochemical variation, shifts in OTU abundances may also result from either consistent contribution from terrestrial sources, competitive advantage of certain indigenous OTUs, or a corresponding biological gradient. It is also probable that biological gradients have a significant influence in community variation than physicochemical parameters in this system (Fortunato and

401 Crump 2011), although this cannot be concluded based on data presented here. The influence of
402 physicochemical and spatial gradients is difficult to distinguish in the current study as distance
403 was found to be a significant covariant throughout the study. However, consistent relationships
404 between nutrient concentrations and phylum abundances in 2011 and 2012 suggest that either
405 gradients of these nutrients or input of these nutrients (and other bacteria) from terrestrial sources
406 are primarily driving community structure.

407 Analysis of biweekly samples revealed that the seasonal period of sampling more
408 significantly affected community similarity than did the depth or year during which the samples
409 were collected. This result is consistent with a previous study which found that sample date and
410 site contributed more to community variability than did depth in nearshore freshwater lake
411 samples (Mueller-Spitz et al., 2009). Of particular note, however, was the observation that
412 samples collected from mid-July onward clustered together regardless of depth or year of
413 sample. Recurrent annual seasonal dynamics have been well-documented in the English Channel
414 (Gilbert et al., 2012) and have been suggested in temperate rivers (Crump and Hobbie 2005), and
415 we hypothesize that the close-relatedness of the later river samples from both years is due to a
416 similar seasonal phenomenon and may continue through the late Fall.

417 There was no apparent association among bacterial communities in river water samples
418 collected from May to June over the two year study, suggesting that the river community may
419 stabilize during the late summer and show recurrent fluctuations throughout the rest of the year.
420 The lack of recurrent community structure as well as the lack of a clear association of early
421 samples with a particular study year may be a result of more highly variable community structure
422 in response to flushing in the late spring and early summer (Thurman et al., 1991), as well as
423 variable temperature effects. A recent study has also suggested that decreased *in situ* activity

424 results in less resilient bacterial communities during colder seasons (García-Armisen et al.,
425 2014), so similarity of communities during the late summer may also be due to greater bacterial
426 activity allowing faster rebound from terrestrial perturbation.

427 As has been shown in a previous 6-year study of the English Channel (Caporaso et al.,
428 2012b), however, the observed fluctuation in presence/absence of taxa in the current study may
429 have resulted from the relatively shallow sequencing effort presented here. The percentage of
430 reads shared was much lower over three years than observed in our initial 2010 study (Staley et
431 al., 2013), and alpha diversity was also found to change significantly between years suggesting
432 that community membership may indeed shift in this system annually. However, the finding that
433 the majority of sequence reads in each sample were shared over samples collected from three
434 years suggests the presence of a "core" Upper Mississippi River bacterial community. The
435 success of this core community may result from better competitive ability in the river biome.
436 Superior resistance of this group to stresses imposed by annual freezing, which has been shown
437 to reduce microbial biomass, productivity, and inputs from terrestrial surroundings in lakes
438 (Bertilsson et al., 2013), may also contribute to the success of particular taxa. In the present
439 study, samples were only collected during summer months, so this is only hypothesized at this
440 time but merits further investigation, particularly in flowing freshwater ecosystems.

441 The bacterial community in sediments was found to have an often considerable influence
442 on the bacterial community structure in the water column. Stark differences in the bacterial
443 communities of the water versus sediment samples, as was seen here using a high-throughput
444 sequencing approach, have been well-established in the literature (Gløckner et al., 2000; Zwart et
445 al., 2002). Nevertheless, this study offers novel information regarding the contribution of the
446 sediment community to that in the water column using an OTU-based analysis that was

447 previously employed to detect and quantify sources of fecal contamination (Knights et al., 2011;
448 Newton et al., 2013). Interestingly, the Minnesota River was found to have > 50% of sequence
449 reads attributed to sediment, and this watershed has been reported to contribute as much as 90%
450 of sediment to Lake Pepin (Engstrom et al., 2009). Consequently, it may be possible to use a
451 source tracking approach to determine sources of sediments and soils to waterways using an
452 OTU-based approach. A previous *in vitro* study has shown that taxonomic history of source
453 communities has a greater influence on community structure than environmental conditions
454 (Langenheder et al., 2006), suggesting that bacterial communities from suspended sediments
455 may significantly influence the bacterial community in the water column more so than
456 physicochemical or nutrient parameters.

457 Variability in the relative abundance of OTUs in replicate/pseudo-replicate samples was,
458 interestingly, more frequent between the pseudo-replicates sequenced on HiSeq2000 and was
459 less attributable to cross-platform variation. We have previously demonstrated that 1 L replicate
460 river water samples differ in bacterial community structure (Staley et al., 2013); however, in the
461 present study, cell pellets were archived from the same 40 L (2 × 20 L) water sample. Recent
462 attention has been paid to issues of PCR/sequencing bias and OTU clustering on diversity
463 estimation (Patin et al., 2013). OTU clustering has been demonstrated to reduce diversity
464 estimation (Patin et al., 2013) and likely would contribute minimally to community differences
465 among replicates. Sequencing reproducibility between HiSeq2000 and MiSeq platforms has also
466 been demonstrated for soil and host-associated samples (Caporaso et al., 2012a). Differences in
467 replicate samples observed here may result from biases due to sequencing platform or sample
468 pool size, but may also reflect inherent heterogeneity in the water sample. The reason for the
469 difference among replicate samples remains to be further explored; however, these findings

highlight the need for replicate sampling in high-throughput sequencing studies, as previously suggested (Prosser 2010).

The results of this study suggest that a species sorting dynamic contributes to community resilience and potentially shapes the bacterial community structure of the Upper Mississippi River ecosystem. This was previously indicated for stream biofilms and in marine waters (Besemer et al., 2009; Winter et al., 2013), and supports recent findings in a Belgian river (García-Armisen et al., 2014). The current study is among the first to evaluate these dynamics over multiple years of study with inputs from multiple non-point sources in a large river biome. Furthermore, this is among the first reports that a recurrent community structure appears to occur seasonally and is associated with water temperature in a major river using next-generation sequencing (Crump and Hobbie 2005; Fortunato et al., 2012). As was suggested in our earlier work, a 'core microbial community' appears to show relatively long-term persistence throughout the study area, accounting for > 50% of the community based on sequence reads (Staley et al., 2013). Importantly, though, this study offers novel insight into the contribution of sediment bacterial communities to those of the overlying water column. Specific input and/or reservoir sources such as sediment may play an important role on shaping non-core bacterial community membership, persistence, and abundance, separate from that resulting from gradients and seasonal dynamics alone in this riverine system. The ability to identify these and other impacts on waterborne bacterial diversity using a high-throughput sequencing approach may improve future best management practices and enhance total maximum daily loading calculations.

Acknowledgements

This project was funded, in part, by the American Recovery and Reinvestment Act of 2009 (ARRA) and the Minnesota Environment and Natural Resources Trust Fund, as recommended by the Legislative-Citizen Commission on Minnesota Resources (LCCMR). This work was carried out in part using computing resources at the University of Minnesota Supercomputing Institute.

Conflict of interest

The authors declare no conflict of interest.

References

Acar EF, Sun L. A generalized Kruskal-Wallis test incorporating group uncertainty with application to genetic association studies. Biometrics 2013;69(2):427–35.

Aronesty E. Comparison of sequencing utility programs. Open Bioinforma J 2013 Jan 31;7(1):1–8.

Baas-Becking LGM. Geobiologie of inleiding tot de milieukunde. The Hague, the Netherlands: W. P. van Stockum & Zoon; 1934.

Bertilsson S, Burgin A, Carey CC, Fey SB, Grossart HP, Grubisic LM, et al. The under-ice microbiome of seasonally frozen lakes. Limnol Oceanogr 2013;58:1998–2012.

Besemer K, Singer G, Hodl I, Battin TJ. Bacterial community composition of stream biofilms in spatially variable-flow environments. Appl Environ Microbiol 2009;75(22):7189–95.

Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr 1957;27:325–49.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 2012a;6(8):1621–4.

Caporaso JG, Paszkiewicz K, Field D, Knight R, Gilbert JA. The Western English Channel contains a persistent microbial seed bank. ISME J 2012b;6(6):1089–93.

Clarke KR. Non-parametric multivariate analyses of changes in community structure. Aust J Ecol 1993;18(1):117–43.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 2009;37:D141–D145.

Crump BC, Hobbie JE. Synchrony and seasonality in bacterioplankton communities of two temperate rivers. Limnol Oceanogr 2005;50(6):1718–29.

Drury B, Rosi-Marshall E, Kelly JJ. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. Appl Environ Microbiol 2013;79(6):1897–905.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 2011;27(16):2194–200.

Engstrom DR, Almendinger JE, Wolin JA. Historical changes in sediment and phosphorus loading to the upper Mississippi River: mass-balance reconstructions from the sediments of Lake Pepin. J Paleolimnol 2009;41(4):563–88.

Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial DNA restriction data. Genetics 1992;131(2):479–91.

Fortunato CS, Crump BC. Bacterioplankton community variation across river to ocean environmental gradients. Microb Ecol 2011;62(2):374–82.

Fortunato CS, Herfort L, Zuber P, Baptista AM, Crump BC. Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. ISME J 2012;6(3):554–63.

García-Armisen T, İnceoğlu Ö, Ouattara NK, Anzil A, Verbanck M a, Brion N, et al. Seasonal variations and resilience of bacterial communities in a sewage polluted urban river. PLoS One 2014;9(3):e92579.

Gilbert JA, Field D, Swift P, Newbold L, Oliver A, Smyth T, et al. The seasonal structure of microbial communities in the Western English Channel. Environ Microbiol 2009;11(12):3132–9.

Gilbert JA, Steele JA, Caporaso JG, Steinbrueck L, Reeder J, Temperton B, et al. Defining seasonal marine microbial community dynamics. ISME J 2012;6(2):298–308.

Gløckner FO, Zaichikov E, Belkova N, Denissova L, Pernthaler J, Pernthaler A, et al. Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of *Actinobacteria*. Appl Environ Microbiol 2000;66(11):5053–65.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol 2012;10(7):497–506.

Horner-Devine MC, Leibold MA, Smith VH, Bohannan BJM. Bacterial diversity patterns along a gradient of primary productivity. Ecol Lett 2003;6(7):613–22.

Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J 2009;3(4):442–53.

Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, et al. Bayesian community-wide culture-independent microbial source tracking. Nat Methods 2011;8(9):761–U107.

Kolmakova O V, Gladyshev MI, Rozanov AS, Peltek SE, Trusova MY. Spatial biodiversity of bacteria along the largest Arctic river determined by next-generation sequencing. FEMS Microbiol Ecol 2014 May;in press.

Langenheder S, Lindstrom ES, Tranvik LJ. Structure and function of bacterial communities emerging from different sources under identical conditions. Appl Environ Microbiol 2006;72(1):212–20.

Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, et al. The metacommunity concept: a framework for multi-scale community ecology. Ecol Lett 2004;7(7):601–13.

Lindström ES, Langenheder S. Local and regional factors influencing bacterial community assembly. Env Microbiol Rep 2012;4(1):1–9.

Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005;71(12):8228–35.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 2006;4(2):102–12.

Mueller-Spitz SR, Goetz GW, McLellan SL. Temporal and spatial variability in nearshore bacterioplankton communities of Lake Michigan. FEMS Microbiol Ecol 2009;67(3):511–22.

Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, et al. Global patterns in the biogeography of bacterial taxa. Environ Microbiol 2011;13(1):135–44.

Newton RJ, Bootsma MJ, Morrison HG, Sogin ML, McLellan SL. A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. Microb Ecol 2013;65(4):1011–23.

Patin N V, Kunin V, Lidstrom U, Ashby MN. Effects of OTU clustering and PCR artifacts on microbial diversity estimates. Microb Ecol 2013;65(3):709–19.

Pereira WE, Hostettler FD. Nonpoint-source contamination of the Mississippi River and its tributaries by herbicides. Environ Sci Technol 1993;27(8):1542–52.

Pernthaler J. Predation on prokaryotes in the water column and its ecological implications. Nat Rev Microbiol 2005;3(7):537–46.

Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, et al. The NIH Human Microbiome Project. Genome Res 2009;19(12):2317–23.

Prosser JI. Replicate or lie. Environ Microbiol 2010;12(7):1806–10.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 2007;35(21):7188–96.

Ruiz-González C, Proia L, Ferrera I, Gasol JM, Sabater S. Effects of large river dam regulation on bacterioplankton community structure. FEMS Microbiol Ecol 2013;84(2):316–31.

Sanapareddy N, Hamp TJ, Gonzalez LC, Hilger HA, Fodor AA, Clinton SM. Molecular diversity of a North Carolina wastewater treatment plant as revealed by pyrosequencing. Appl Environ Microbiol 2009;75(6):1688–96.

Schloss PD, Westcott SL. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl Environ Microbiol 2011;77(10):3219–26.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009;75(23):7537–41.

Schultz GE, Kovatch JJ, Anneken EM. Bacterial diversity in a large, temperate, heavily modified river, as determined by pyrosequencing. Aquat Microb Ecol 2013;70(2):169–79.

Shade A, Kent AD, Jones SE, Newton RJ, Triplett EW, McMahon KD. Interannual dynamics and phenology of bacterial communities in a eutrophic lake. Limnol Oceanogr 2007;52(2):487–94.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc Natl Acad Sci U S A 2006;103(32):12115–20.

Sokal RR, Rohlf FJ. Biometry. 3rd ed. New York: Freeman; 1995.

Staley C, Unno T, Gould TJ, Jarvis B, Phillips J, Cotner JB, et al. Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. J Appl Microbiol 2013;115:1147–58.

Tamames J, Abellan JJ, Pignatelli M, Camacho A, Moya A. Environmental distribution of prokaryotic taxa. BMC Microbiol 2010;10:85.

Thurman EM, Goolsby DA, Meyer MT, Kolpin DW. Herbicides in surface waters of the Midwestern United States - the effect of spring flush. Environ Sci Technol 1991;25(10):1794–6.

Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. River continuum concept. Can J Fish Aquat Sci 1980;37(1):130–7.

Winemiller KO, Flecker AS, Hoeinghaus DJ. Patch dynamics and environmental heterogeneity in lotic ecosystems. J North Am Benthol Soc 2010;29(1):84–99.

Winter C, Matthews B, Suttle CA. Effects of environmental variation and spatial distance on Bacteria, Archaea and viruses in sub-polar and arctic waters. ISME J 2013;7(8):1507–18.

Zwart G, Crump BC, Agterveld MPK V, Hagen F, Han SK. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquat Microb Ecol 2002;28(2):141–55.

Fig Legends

Fig 1 Distribution of taxonomic orders of OTUs found to vary significantly by year. The OTUs were found to vary significantly by Kruskal-Wallis test.

Fig 2 Distribution of taxonomic orders of OTUs found to differ significantly by site in 2010. The OTUs were found to vary significantly by using the Kruskal-Wallis test.

Fig 3 Distribution of taxonomic orders of OTUs found to differ significantly by site in 2011. The OTUs were found to vary significantly by using the Kruskal-Wallis test.

Fig 4 Distribution of taxonomic orders of OTUs found to differ significantly by site in 2012. The OTUs were found to vary significantly by using the Kruskal-Wallis test.

Fig 5 Principal coordinate analysis of surface and depth samples (all replicates merged for ordination). The r^2 value relating ordination to the distance matrix for both axes is 0.87. Symbols represent samples collected in 2011 (○), 2012 (□), from the surface (shaded) and at depth (filled). Numbers refer to the order in which samples were collected.

Fig 6 Distribution of abundant phyla found in sediment samples collected during 2012.

Table 1. Mean and standard deviation of coverage and diversity indices in the water column for all three years of study.

Year	Site	Coverage	S_{obs}[a]	Shannon	NP Shannon[†]	Simpson
2010	Itasca	0.96 ± 0.00[a,b,c]	1457 ± 108[a,c]	3.98 ± 0.14[a,d]	4.12 ± 0.13[a,d]	0.051 ± 0.011[a]
	St. Cloud	0.96 ± 0.[00a,b,c]	1618 ± 101[a,b,c]	4.41 ± 0.03[b,c,d]	4.55 ± 0.03[b,c,d]	0.036 ± 0.001[a]
	Clearwater	0.96 ± 0.[00a,b]	1683 ± 99[a,b]	4.58 ± 0.14[b]	4.71 ± 0.12[b]	0.032 ± 0.007[a]
	Twin Cities	0.96 ± 0.00[a,b,c]	1646 ± 79[a,b,c]	4.62 ± 0.05[b,c]	4.75 ± 0.04[b,c]	0.034 ± 0.001[a]
	MN River	0.96 ± 0.00[a,b,c]	1578 ± 46[a,b,c]	4.36 ± 0.14[b,c,d]	4.49 ± 0.12[b,c,d]	0.042 ± 0.001[a]
	Confluence	0.97 ± 0.00[a,b,c]	1329 ± 105[a,c]	4.19 ± 0.12[a,b,d]	4.30 ± 0.10[a,b,d]	0.048 ± 0.008[a]
	Hastings	0.96 ± 0.00[a,b,c]	1585 ± 194[a,b,c]	4.62 ± 0.20[b,c]	4.74 ± 0.21[b,c]	0.038 ± 0.016[a]
	St. Croix River	0.97 ± 0.00[a,b,c]	1462 ± 168[a,c]	4.61 ± 0.02[b,c]	4.71 ± 0.00[b,c]	0.031 ± 0.002[a]
	Red Wing	0.97 ± 0.00[a,c]	1166 ± 125[a,c]	4.25 ± 0.07[a,b,d]	4.34 ± 0.06[a,b,d]	0.037 ± 0.002[a]
	La Crescent	0.97 ± 0.01[a,c]	1287 ± 187[a,c]	4.11 ± 0.02[a,b,d]	4.23 ± 0.03[a,b,d]	0.050 ± 0.008[a]
2011	Itasca	0.97 ± 0.00[a,b,c,e,f]	1147 ± 168[a,b,e,g,h]	2.79 ± 0.40[a]	2.94 ± 0.41[a]	0.231 ± 0.113[a]
	St. Cloud	0.97 ± 0.01[a,b,f]	1036 ± 218[a,b,g,h]	2.94 ± 0.43[a]	3.07 ± 0.45[a]	0.213 ± 0.109[a]
	Clearwater	0.96 ± 0.00[a,c,d,e]	1551 ± 106[c,d,e,f,g,h]	4.14 ± 0.03[b,c,d]	4.28 ± 0.02[b,c,d]	0.052 ± 0.004[b]
	Twin Cities	0.95 ± 0.00[c,d]	1901 ± 132[c,d,f]	4.52 ± 0.05[b,c]	4.67 ± 0.03[b,c]	0.043 ± 0.002[b]

	MN River	$0.96 \pm 0.00^{a,c,e,f}$	$1511 \pm 20^{a,c,e,f,g,h}$	$4.10 \pm 0.24^{b,c,d}$	$4.24 \pm 0.22^{b,c,d}$	0.055 ± 0.010^{b}
	Confluence	$0.96 \pm 0.00^{a,c,d,e}$	$1638 \pm 138^{c,d,e,f,g}$	$4.24 \pm 0.22^{b,c,d}$	$4.38 \pm 0.23^{b,c,d}$	0.055 ± 0.002^{b}
	Hastings	$0.97 \pm 0.00^{a,b,c,e,f}$	$1350 \pm 87^{a,b,c,e,f,g,h}$	$3.91 \pm 0.03^{b,c,d}$	$4.04 \pm 0.01^{b,c,d}$	0.068 ± 0.003^{b}
	St. Croix River	$0.97 \pm 0.00^{a,b,e,f}$	$1236 \pm 134^{a,b,c,e,g,h}$	$4.30 \pm 0.13^{b,c,d}$	$4.40 \pm 0.11^{b,c,d}$	0.041 ± 0.006^{b}
	Red Wing	$0.96 \pm 0.00^{a,c,e,f}$	$1502 \pm 156^{a,c,e,f,g,h}$	$4.23 \pm 0.06^{b,c,d}$	$4.36 \pm 0.04^{b,c,d}$	0.056 ± 0.001^{b}
	La Crescent	$0.96 \pm 0.00^{a,c,d,e}$	$1599 \pm 22^{c,d,e,f,g,h}$	$4.45 \pm 0.26^{b,c,d}$	$4.58 \pm 0.25^{b,c,d}$	0.047 ± 0.014^{b}
	Zumbro River	$0.97 \pm 0.00^{a,b,c,e,f}$	$1304 \pm 15^{a,b,c,e,f,g,h}$	$3.63 \pm 0.22^{b,d}$	$3.77 \pm 0.21^{b,d}$	$0.100 \pm 0.028^{a,b}$
2012	Itasca	$0.97 \pm 0.00^{a,b,c}$	$1233 \pm 44^{a,b}$	$3.66 \pm 0.17^{a,b,d,f}$	$3.79 \pm 0.17^{a,b,d,f}$	$0.081 \pm 0.014^{a,b,c}$
	St. Cloud	$0.96 \pm 0.00^{a,b,c}$	$1617 \pm 133^{a,b,c}$	$4.21 \pm 0.32^{a,b,c,d,e,f}$	$4.36 \pm 0.32^{a,b,c,d,e,f}$	$0.050 \pm 0.015^{a,b,d}$
	Clearwater	$0.96 \pm 0.00^{a,b,c}$	$1823 \pm 50^{a,b,c}$	$4.52 \pm 0.21^{b,c,e}$	$4.67 \pm 0.20^{b,c,e}$	$0.039 \pm 0.009^{a,b,d}$
	Twin Cities	$0.97 \pm 0.00^{a,b,c}$	$1479 \pm 17^{a,b,c}$	$3.82 \pm 0.17^{a,b,d,e,f}$	$3.97 \pm 0.17^{a,b,d,e,f}$	$0.116 \pm 0.043^{a,c}$
	MN River	$0.97 \pm 0.00^{a,b,c}$	$1363 \pm 129^{a,b,c}$	$3.97 \pm 0.12^{a,b,c,d,e,f}$	$4.10 \pm 0.13^{a,b,c,d,e,f}$	$0.073 \pm 0.018^{a,b,c}$
	Confluence	$0.95 \pm 0.01^{a,b}$	$1972 \pm 199^{b,c}$	$4.64 \pm 0.27^{b,c,e}$	$4.79 \pm 0.27^{b,c,e}$	$0.040 \pm 0.002^{a,b,d}$
	Hastings	$0.96 \pm 0.00^{a,b,c}$	$1685 \pm 83^{a,b}$	$4.22 \pm 0.18^{a,b,c,d,e}$	$4.37 \pm 0.18^{a,b,c,d,e}$	$0.058 \pm 0.002^{a,b,d}$
	St. Croix River	$0.97 \pm 0.00^{a,b,c}$	$1244 \pm 186^{a,b,c}$	$4.30 \pm 0.04^{b,c,d,e}$	$4.40 \pm 0.06^{b,c,d,e}$	$0.041 \pm 0.006^{a,b,d}$
	Red Wing	$0.96 \pm 0.00^{a,b,c}$	$1840 \pm 47^{a,b,c}$	$4.48 \pm 0.17^{a,b,c,d,e,f}$	$4.63 \pm 0.15^{a,b,c,d,e,f}$	$0.047 \pm 0.003^{a,b,c}$

La Crescent	$0.97 \pm 0.00^{a,c}$	$1379 \pm 114^{a,b}$	$3.96 \pm 0.05^{a,b,d,f}$	$4.09 \pm 0.06^{a,b,d,f}$	$0.066 \pm 0.014^{a,b,c}$
Zumbro River	0.97 ± 0.00^{d}	1183 ± 63^{d}	3.58 ± 0.39^{g}	3.70 ± 0.39^{g}	$0.095 \pm 0.037^{b,d}$

[*]Number of OTUs observed. Sobs was significantly ($P = 0.041$) higher among water samples collected in 2012 compared to 2011 by post-hoc test, and diversity as measured by all indices was significantly lower in 2011 compared to other years ($P \leq 0.034$).

[†]Non-parametric Shannon index.

[a-g]Indices sharing the same superscript were not significantly different among other samples collected that year by *post-hoc* analysis. Sediment samples were not replicated and could not be analyzed statistically.

Table 2. Multiple linear regression standardized β coefficients for environmental parameters and distance describing changes in diversity and relative abundance of major phyla. *P* values are shown in parentheses and significant values are bolded.

Year	Descriptor	Shannon	Simpson	*Betaproteobacteria*	*Gammaproteobacteria*	*Alphaproteobacteria*	*Actinobacteria*[a]	*Bacteroidetes*[a]	*Verrucomicrobia*	*Cyanobacteria*
All	72 h rainfall	0.134	0.880	-0.736	1.330	**-2.551**	-0.844	0.073	0.642	1.827
		(0.646)	(0.384)	(0.466)	(0.192)	**(0.015)**	(0.404)	(0.942)	(0.525)	(0.076)
	48 h rainfall	-1.694	1.343	**-4.366**	**4.001**	0.539	0.107	-1.574	**-2.295**	-1.903
		(0.099)	(0.187)	**(< 0.001)**	**(< 0.001)**	(0.593)	(0.915)	(0.124)	**(0.027)**	(0.065)
	24 h rainfall	1.068	-1.426	0.723	-1.973	**4.115**	1.036	0.203	-1.681	-1.992
		(0.292)	(0.162)	(0.474)	(0.056)	**(< 0.001)**	(0.307)	(0.841)	(0.101)	(0.054)
	Cumulative rainfall	NA[b]	NA	NA	NA	NA	NA	NA	NA	NA
	Temperature	1.437	**-2.363**	**3.492**	**-4.082**	**2.760**	1.225	0.723	**-2.116**	**-2.440**
		(0.159)	**(0.024)**	**(0.001)**	**(< 0.001)**	**(0.009)**	(0.228)	(0.474)	**(0.041)**	**(0.020)**

	pH	-0.287	1.188	-1.923	**2.581**	**-2.668**	-1.184	-0.370	1.993	**2.410**
		(0.776)	(0.242)	(0.062)	**(0.014)**	**(0.011)**	(0.244)	(0.713)	(0.054)	**(0.021)**
	TDS	1.240	-0.744	1.541	-1.234	-1.015	-0.173	0.921	0.202	0.453
		(0.223)	(0.462)	(0.132)	(0.225)	(0.317)	(0.863)	(0.363)	(0.841)	(0.653)
	Carbon	0.211	1.048	**-2.909**	**3.044**	-1.600	-1.347	-0.434	**2.447**	1.422
		(0.834)	(0.302)	**(0.006)**	**(0.004)**	(0.118)	(0.186)	(0.666)	**(0.019)**	(0.163)
	Nitrite/nitrate	-0.624	0.025	-1.688	1.012	**2.143**	0.344	-0.878	0.255	-1.078
		(0.536)	(0.980)	(0.100)	(0.318)	**(0.039)**	(0.733)	(0.386)	(0.800)	(0.288)
	Phosphorus	-0.126	-0.363	-0.464	-0.205	**2.118**	0.598	-0.453	-1.507	-1.380
		(0.900)	(0.718)	(0.645)	(0.839)	**(0.041)**	(0.553)	(0.653)	(0.140)	(0.176)
	Distance	1.183	-0.107	0.132	-0.026	-0.617	-0.392	0.221	**2.559**	**3.480**
		(0.244)	(0.916)	(0.896)	(0.979)	(0.541)	(0.698)	(0.827)	**(0.015)**	**(0.001)**
2011	72 h rainfall	**-2.969**	**2.175**	**-3.053**	**3.416**	**-4.359**	-0.071	-0.167	2.924	0.019
		(0.009)	**(0.045)**	**(0.008)**	**(0.004)**	**(< 0.001)**	(0.944)	(0.892)	(0.010)	(0.985)
	48 h rainfall	**2.326**	-1.959	**2.923**	**-3.270**	**4.221**	0.245	0.247	**-4.600**	-0.682
		(0.034)	(0.068)	**(0.010)**	**(0.005)**	**(0.001)**	(0.810)	(0.808)	**(< 0.001)**	(0.505)

	24 h rainfall	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Cumulative	**4.141**	**-2.699**	**4.103**	**-4.706**	**6.553**	0.359	0.677	**-3.745**	-0.091
	rainfall	**(0.001)**	**(0.016)**	**(0.001)**	**(< 0.001)**	**(< 0.001)**	(0.724)	(0.508)	**(0.002)**	(0.929)
	Temperature	2.086	-1.989	**2.955**	**-2.997**	1.496	0.110	-0.246	**-7.197**	-2.107
		(0.053)	(0.064)	**(0.009)**	**(0.009)**	(0.154)	(0.914)	(0.809)	**(< 0.001)**	(0.051)
	pH	NA	NA	NA	NA	NA	NA	NA	NA	NA
	TDS	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Carbon	**-3.512**	**2.499**	**-3.744**	**4.204**	**-5.448**	-0.255	-0.385	**3.040**	0.043
		(0.003)	**(0.024)**	**(0.002)**	**(0.001)**	**(< 0.001)**	(0.802)	(0.705)	**(0.008)**	(0.966)
	Nitrite/nitrate	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Phosphorus	**-3.520**	1.682	**-3.742**	**4.450**	**-7.361**	-0.898	-2.800	**-7.701**	-1.407
		(0.003)	(0.112)	**(0.002)**	**(< 0.001)**	**(< 0.001)**	(0.383)	(0.013)	**(< 0.001)**	(0.179)
	Distance	**5.881**	**-3.324**	**4.930**	**-6.061**	**11.543**	0.618	1.578	**2.235**	**2.240**
		(< 0.001)	**(0.004)**	**(< 0.001)**	**(< 0.001)**	**(< 0.001)**	(0.545)	(0.134)	**(0.040)**	**(0.040)**
2012	72 h rainfall	-1.171	1.156	**3.395**	-1.927	**-3.877**	-0.405	-0.418	**-4.256**	-0.875
		(0.259)	(0.265)	**(0.004)**	(0.072)	**(0.001)**	(0.691)	(0.681)	**(0.001)**	(0.395)

48 h rainfall	**-2.189**	1.863	**-2.246**	**5.070**	**-2.482**	-0.209	-0.696	**-2.116**	-1.036
	(0.044)	(0.081)	**(0.039)**	**(< 0.001)**	**(0.025)**	(0.837)	(0.496)	**(0.046)**	(0.316)
24 h rainfall	**2.839**	-1.629	**-4.591**	1.699	**5.784**	-0.230	-0.270	**2.860**	0.511
	(0.012)	(0.123)	**(< 0.001)**	(0.109)	**(< 0.001)**	(0.821)	(0.791)	**(0.011)**	(0.617)
Cumulative rainfall	NA	NA	NA	NA	NA	NA	NA	NA	NA
Temperature	**3.503**	-1.828	**-3.059**	1.260	**3.429**	-0.033	0.380	**3.655**	2.035
	(0.003)	(0.086)	**(0.007)**	(0.226)	**(0.003)**	(0.974)	(0.709)	**(0.002)**	(0.059)
pH	0.600	-0.401	-0.702	0.207	1.225	0.251	0.603	0.535	1.750
	(0.557)	(0.694)	(0.493)	(0.838)	(0.238)	(0.805)	(0.555)	(0.600)	(0.099)
TDS	NA	NA	NA	NA	NA	NA	NA	NA	NA
Carbon	NA	NA	NA	NA	NA	NA	NA	NA	NA
Nitrite/nitrate	0.110	-0.502	-1.257	1.278	0.858	0.430	0.382	1.508	0.903
	(0.914)	(0.622)	(0.227)	(0.220)	(0.403)	(0.673)	(0.708)	(0.151)	(0.380)
Phosphorus	-0.181	-0.505	-0.439	0.585	0.007	0.037	0.648	0.329	-0.810
	(0.859)	(0.621)	(0.667)	(0.566)	(0.995)	(0.971)	(0.526)	(0.746)	(0.430)

Distance	NA	NA	NA	NA	NA	NA	NA	NA	NA

[a]Multiple regression models were not significant at $\alpha = 0.05$.

[b]Parameter was not included in the model.

Table 3. Date and mean diversity indices of samples taken for depth and seasonal variation analyses.

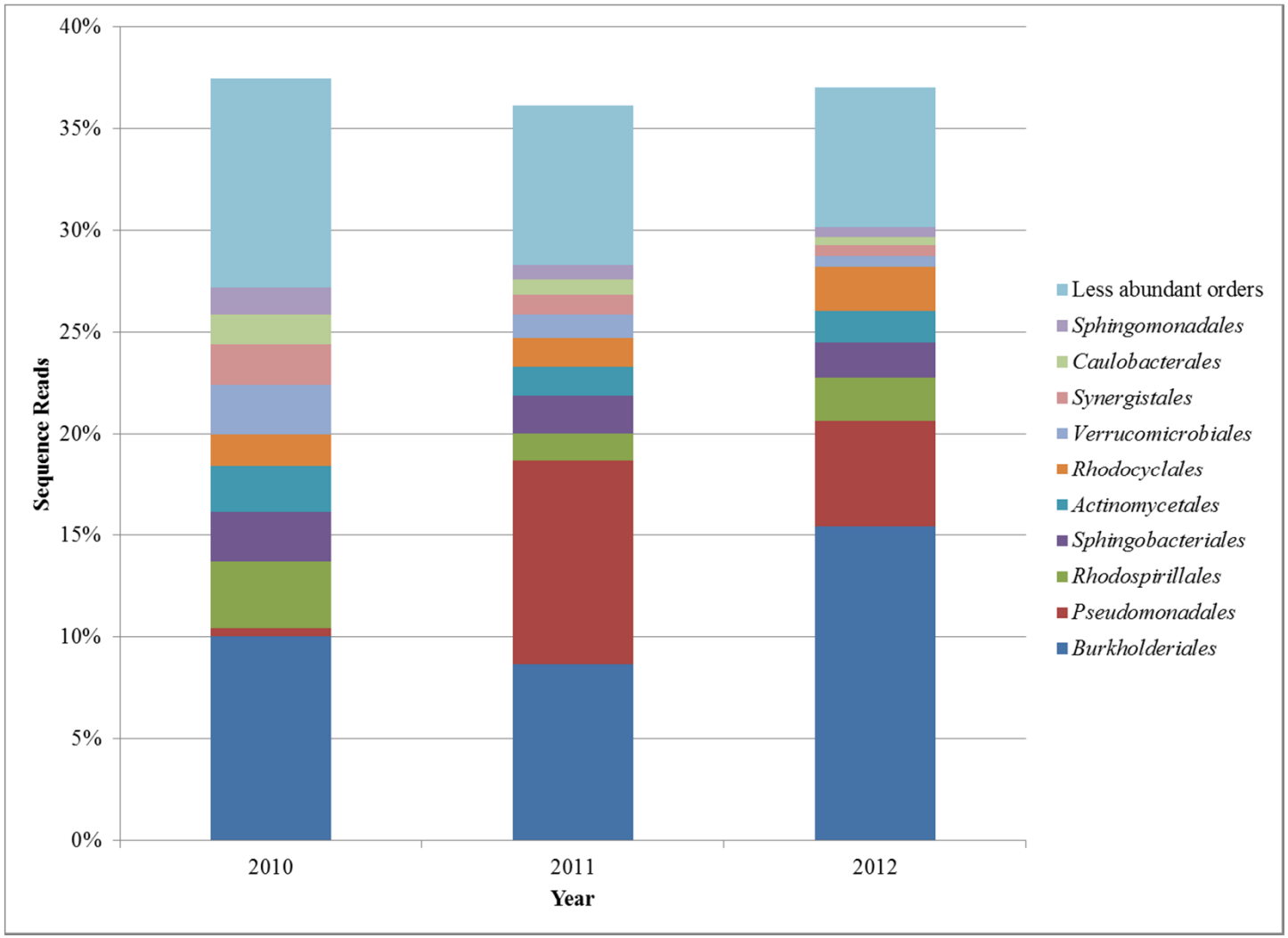
Year (Depth)	Date	Coverage	S_{obs}	Shannon	NP_Shannon	Simpson
2011 (surface)	18-May	0.96 ± 0.00	1022 ± 43	3.92 ± 0.14	4.08 ± 0.13	0.061 ± 0.007
	1-Jun	0.95 ± 0.00	1109 ± 1	4.08 ± 0.19	4.25 ± 0.18	0.053 ± 0.007
	15-Jun	0.95 ± 0.01	1121 ± 97	4.21 ± 0.28	4.37 ± 0.28	0.047 ± 0.005
	29-Jun	0.95 ± 0.00	1253 ± 66	4.31 ± 0.18	4.49 ± 0.18	0.048 ± 0.003
	13-Jul	0.94 ± 0.00	1284 ±71	4.46 ± 0.13	4.64 ± 0.14	0.041 ± 0.004
	4-Aug	0.95 ± 0.00	1213 ±96	4.45 ± 0.21	4.61 ± 0.21	0.036 ± 0.003
2011 (1.5 m)	18-May	0.96 ± 0.00	1012 ± 30	4.01 ± 0.15	4.16 ± 0.13	0.057 ± 0.004
	1-Jun	0.96 ± 0.00	1048 ± 82	3.92 ± 0.28	4.07 ± 0.28	0.096 ± 0.024
	15-Jun	0.96 ± 0.00	953 ± 26	3.75 ± 0.10	3.89 ± 0.10	0.100 ± 0.035
	29-Jun	0.96 ± 0.01	980 ± 92	4.03 ± 0.30	4.17 ± 0.30	0.062 ± 0.019
	13-Jul	0.95 ± 0.00	1173 ± 70	4.28 ± 0.14	4.44 ± 0.15	0.046 ± 0.002
	4-Aug	0.95 ± 0.00	1084 ± 104	4.38 ± 0.22	4.52 ± 0.22	0.038 ± 0.001
2012 (surface)	4-Jun	0.94 ± 0.00	1394 ± 45	4.43 ± 0.24	4.63 ± 0.23	0.045 ± 0.011
	20-Jun	0.95 ± 0.00	1054 ± 35	3.84 ± 0.22	4.01 ± 0.21	0.097 ± 0.063

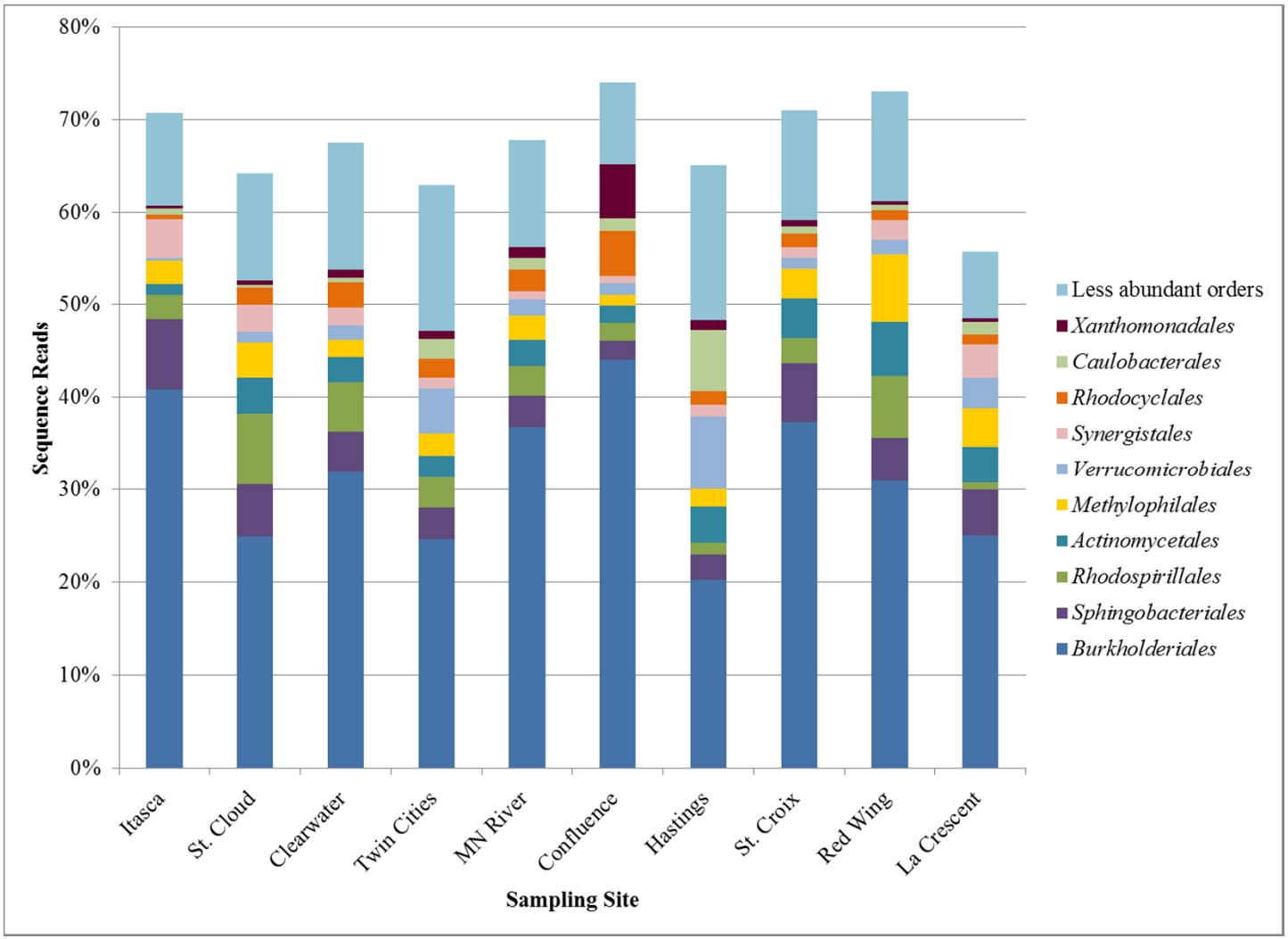
	3-Jul	0.95 ± 0.00	1217 ± 29	4.40 ± 0.20	4.57 ± 0.19	0.044 ± 0.007
	18-Jul	0.94 ± 0.00	1349 ± 25	4.70 ± 0.14	4.87 ± 0.13	0.037 ± 0.008
	1-Aug	0.95 ± 0.00	1295 ±23	4.60 ± 0.13	4.75 ± 0.12	0.044 ± 0.006
	16-Aug	0.95 ± 0.01	1180 ±96	4.62 ± 0.12	4.75 ± 0.13	0.042 ± 0.009
2012 (1.5 m)	4-Jun	0.94 ± 0.01	1294 ± 172	4.21 ± 0.45	4.40 ± 0.45	0.055 ± 0.018
	20-Jun	0.94 ± 0.00	1345 ± 20	4.47 ± 0.12	4.65 ± 0.11	0.044 ± 0.002
	3-Jul	0.95 ± 0.00	1138 ±17	4.26 ± 0.15	4.41 ± 0.14	0.045 ± 0.006
	18-Jul	0.94 ± 0.00	1403 ±21	4.64 ± 0.16	4.81 ± 0.15	0.041 ± 0.009
	1-Aug	0.95 ± 0.00	1321 ±29	4.62 ± 0.10	4.78 ± 0.10	0.043 ± 0.004
	16-Aug	0.94 ± 0.00	1418 ±79	4.64 ± 0.04	4.81 ± 0.04	0.048 ± 0.004

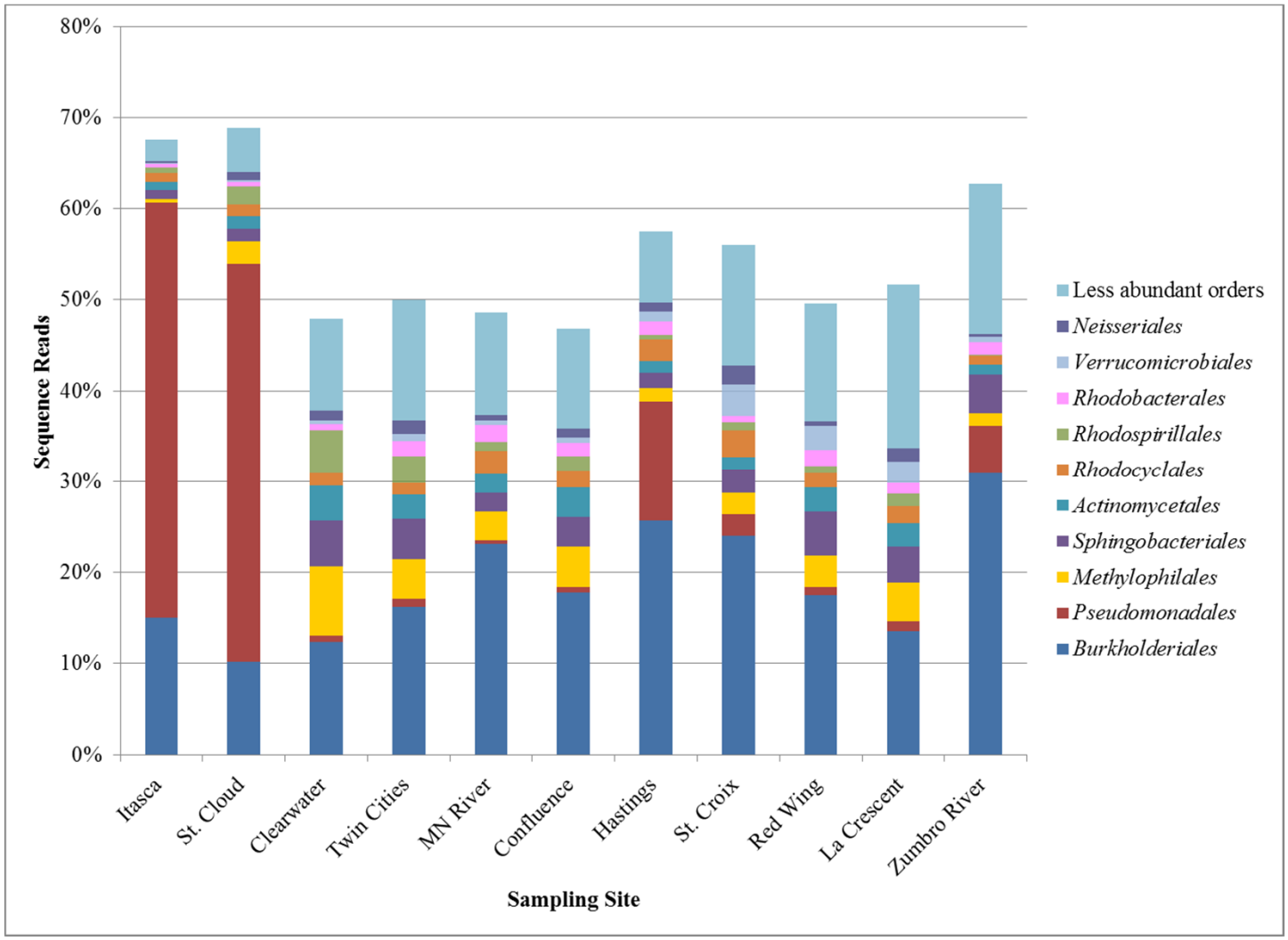
Table 4. Contribution of sediment to OTUs identified in the water column for samples collected in 2012.

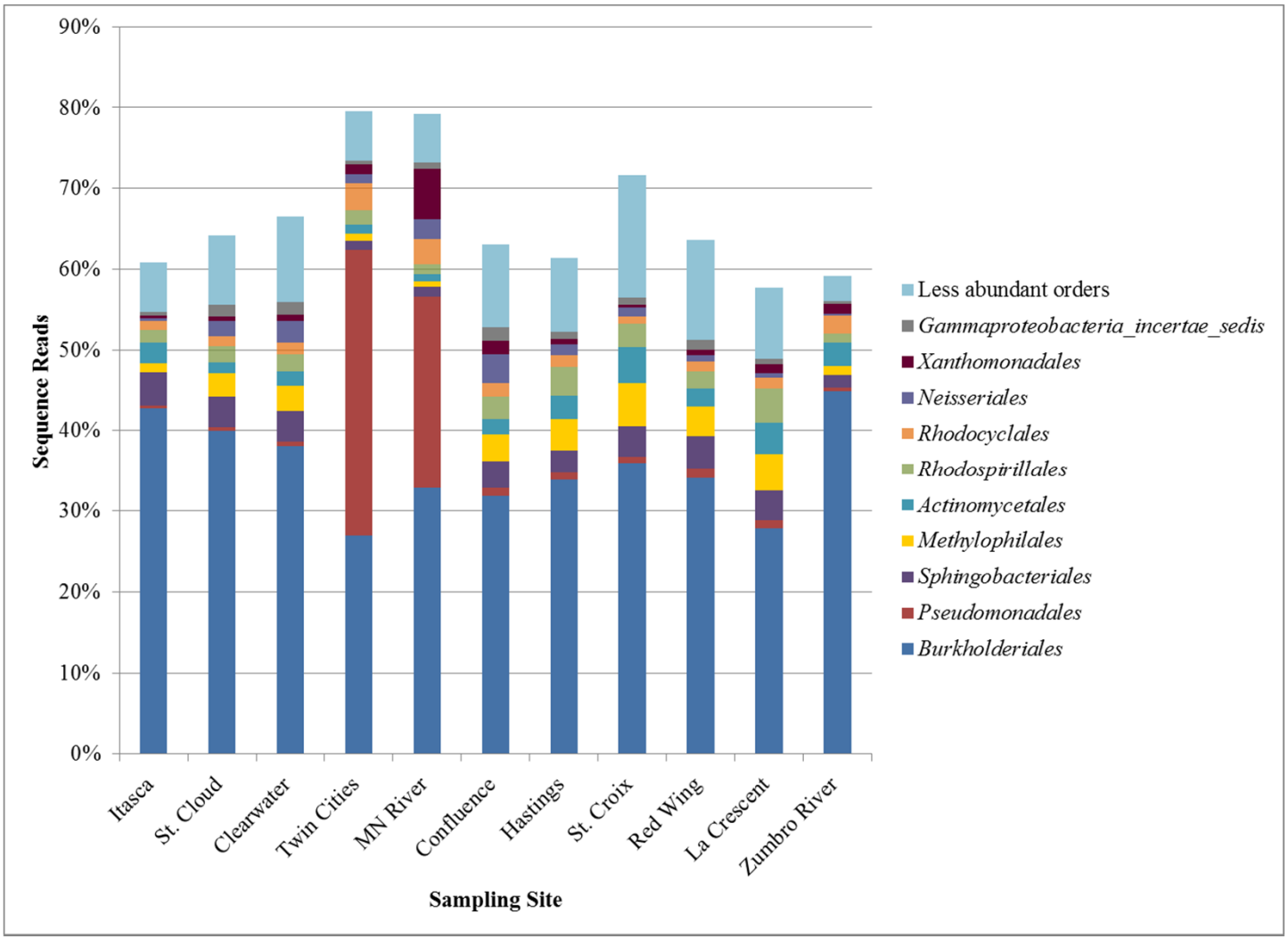
Site	% Contribution
Itasca	9.0 ± 3.0[a]
St. Cloud	13.3 ± 0.6
Clearwater	18.7 ± 0.6
Twin Cities	23.0 ± 1.7
MN River	57.3 ± 9.5
Confluence	20.0 ± 1.0
Hastings	18.7 ± 1.5
St. Croix River	14.3 ± 3.1
Red Wing	21.0 ± 1.7
La Crescent	11.0 ± 2.6
Zumbro River	13.7 ± 3.8

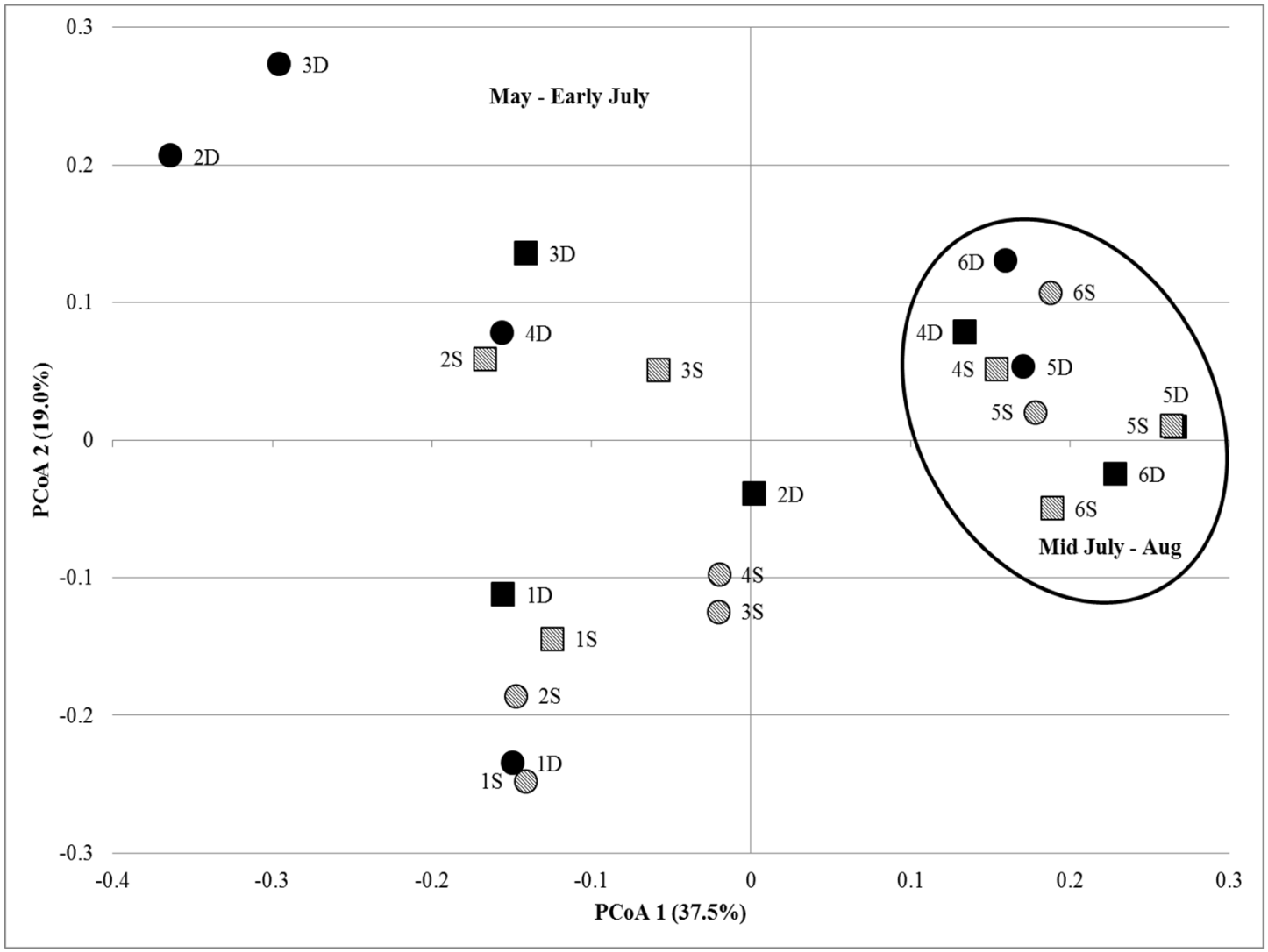
[a]Mean percentages of total sequence reads and standard deviation among replicate samples are shown.

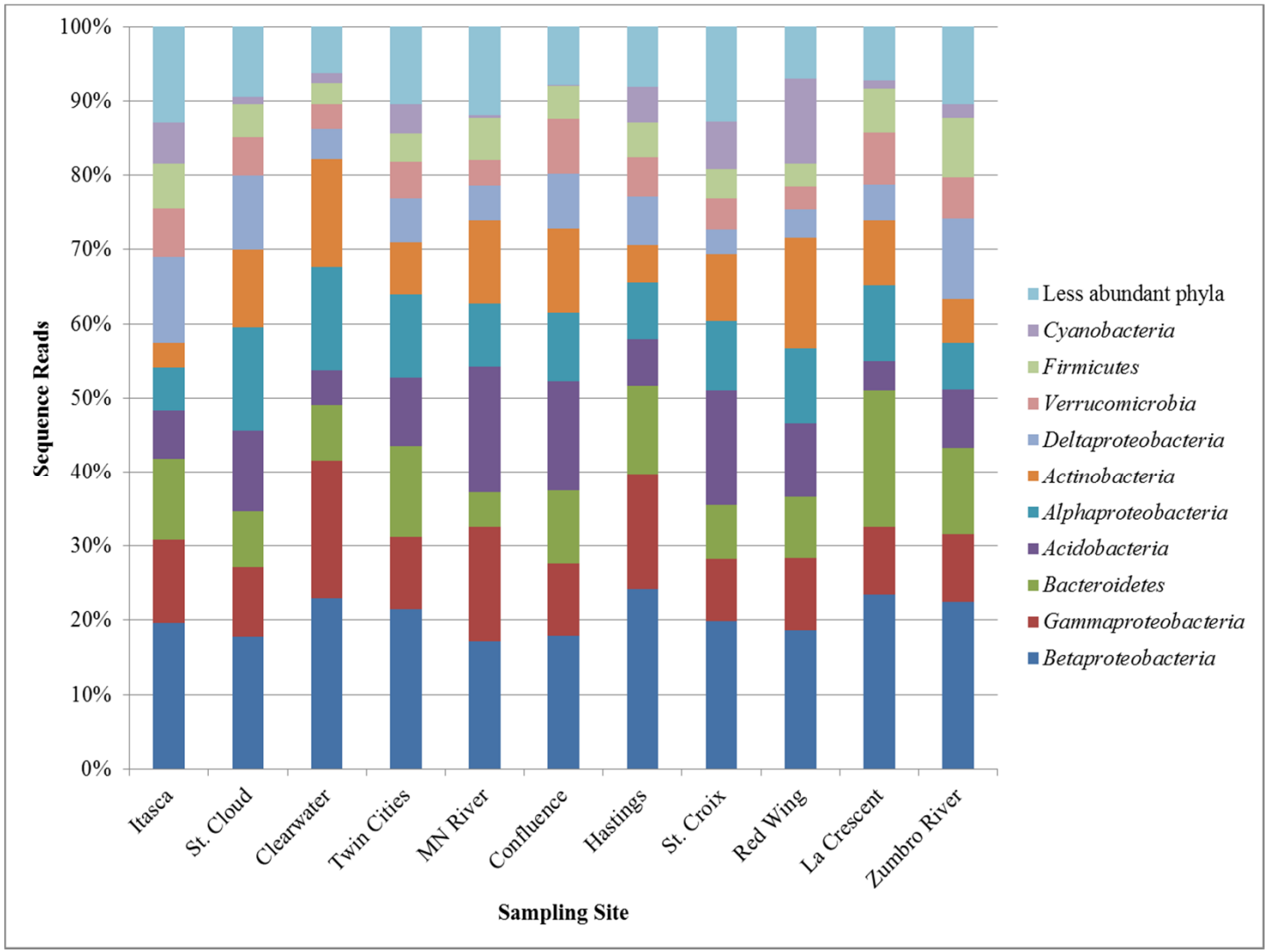

40%
35%
30%
25%
20%
15%
10%
5%
0%
Sequence Reads
2010
2011
2012
Year
Less abundant orders
Sphingomonadales
Caulobacterales
Synergistales
Verrucomicrobiales
Rhodocyclales
Actinomycetales
Sphingobacteriales
Rhodospirillales
Pseudomonadales
Burkholderiales


80%
70%
60%
50%
40%
30%
20%
10%
0%
Sequence Reads
Itasca
St. Cloud
Clearwater
Twin Cities
MN River
Confluence
Hastings
St. Croix
Red Wing
La Crescent
Sampling Site
Less abundant orders
Xanthomonadales
Caulobacterales
Rhodocyclales
Synergistales
Verrucomicrobiales
Methylophilales
Actinomycetales
Rhodospirillales
Sphingobacteriales
Burkholderiales


80%
70%
60%
50%
40%
30%
20%
10%
0%
Sequence Reads
Itasca
St. Cloud
Clearwater
Twin Cities
MN River
Confluence
Hastings
St. Croix
Red Wing
La Crescent
Zumbro River
Sampling Site
Less abundant orders
Neisseriales
Verrucomicrobiales
Rhodobacterales
Rhodospirillales
Rhodocyclales
Actinomycetales
Sphingobacteriales
Methylophilales
Pseudomonadales
Burkholderiales


90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Sequence Reads
Itasca
St. Cloud
Clearwater
Twin Cities
MN River
Confluence
Hastings
St. Croix
Red Wing
La Crescent
Zumbro River
Sampling Site
Less abundant orders
Gammaproteobacteria_incertae_sedis
Xanthomonadales
Neisseriales
Rhodocyclales
Rhodospirillales
Actinomycetales
Methylophilales
Sphingobacteriales
Pseudomonadales
Burkholderiales


0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
PCoA 2 (19.0%)
PCoA 1 (37.5%)
May - Early July
Mid July - Aug
3D
2D
3D
4D
2S
3S
2D
4S
3S
1D
1S
2S
1D
1S
6D
6S
4D
4S
5D
5S
5D
5S
6D
6S


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Sequence Reads
Itasca
St. Cloud
Clearwater
Twin Cities
MN River
Confluence
Hastings
St. Croix
Red Wing
La Crescent
Zumbro River
Sampling Site
Less abundant phyla
Cyanobacteria
Firmicutes
Verrucomicrobia
Deltaproteobacteria
Actinobacteria
Alphaproteobacteria
Acidobacteria
Bacteroidetes
Gammaproteobacteria
Betaproteobacteria

Science of the Total Environment

Species sorting dynamics promote community resilience in response to natural and anthropogenic disturbance in the Upper Mississippi River

Christopher Staley[1], Trevor J. Gould[1,2], Ping Wang[1], Jane Phillips[2], James B. Cotner[3], and

Michael J. Sadowsky[1,4,*]

[1]BioTechnology Institute, [2]Biology Program, [3]Department of Ecology, Evolution and Behavior, and [4]Department of Soil, Water and Climate, University of Minnesota, St. Paul, MN

[*]Corresponding Author: Michael J. Sadowsky, BioTechnology Institute, University of Minnesota, 140 Gortner Lab, 1479 Gortner Ave, Saint Paul, MN 55108; Phone: (612)-624-2706, Email: sadowsky@umn.edu

Supplemental Materials

Supplementary Results

Variation in physicochemical and nutrient parameters

Physicochemical parameters and nutrient concentrations showed significant differences between sample years and several co-varied significantly, depending on sampling year. Rainfall (72 h, 48 h, and cumulative) was significantly greater in 2010 compared to 2011, but not 2012 ($P \leq 0.08$), water temperature was higher in 2012 than both preceding years ($P < 0.001$), and pH was greater in 2010 compared to 2012 ($P = 0.002$). Concentrations of TDS and phosphorus were higher in 2011 than in 2012 ($P < 0.001$), while higher concentrations of carbon were observed in 2012 ($P = 0.001$).

Correlations between distance from the headwaters, rainfall, temperature, and pH were intercorrelated during all three years of study, but correlations were not consistent in significance or direction. In 2011, distance was positively correlated with temperature, pH, and the concentrations to TDS and nitrite/nitrate (r = 0.618, 0.610, 0.595, and 0.762, $P = 0.001$, 0.002, 0.002, and < 0.001, respectively), and carbon concentration decreased farther from the headwaters (r = -0.667, $P < 0.001$). Rainfall 48 h, 24 h, and 3-days cumulative to sampling were negatively correlated with nitrite/nitrate (r = -0.454 to -0.592, $P \leq 0.015$), and cumulative rainfall was also negatively correlated with TDS concentration (r = -0.349, $P = 0.047$) but inversely correlated with carbon concentration (r = 0.495, $P = 0.003$). pH and the concentrations of nitrite/nitrate were positively intercorrelated (r = 0.731 – 0.855, $P < 0.001$) and negatively correlated with carbon concentration (r = -0.658 to -0.918, $P < 0.001$).

Similar to 2011, concentrations of TDS and nitrite/nitrate were higher farther downstream (r = 0.667 and 0.619, $P < 0.001$ and 0.001, respectively), and phosphorus concentrations were lower (r = -0.835, $P < 0.001$). pH, TDS, nitrite/nitrate concentrations were again positively intercorrelated (r = 0.542 – 0.776, $P \leq 0.001$), as were carbon and phosphorus (r = 0.506, $P = 0.003$), and these two groups of intercorrelated variables were negatively correlated with each other (r = -0.577 to -0.784, $P \leq 0.023$).

Supplementary Table S1. Sampling site locations and dates as well as physicochemical parameters measured.

GPS	Site Number	Sampling Site	Distance (km)[a]	Sampling Date	72 h Rainfall (mm)	48 h Rainfall (mm)	24 h Rainfall (mm)	Cumulative Rainfall (mm)	Temperature (°C)	pH	TDS (mg L^{-1})[b]	Carbon (mg L^{-1})	Nitrite/Nitrate (mg L^{-1})	Phosphorus (mg L^{-1})
47.347, -95.182	1	Itasca	0	5/19/2010	0.00	0.00	0.00	0.00	17.0	8.0	NA	NA	NA	NA
				6/22/2011	0.00	11.18	10.67	21.85	12.0	7.4	66.44	6.82	0.05	0.16
				7/10/2012	3.30	0.25	0.00	3.55	20.6	7.9	43.68	4.82	0.55	0.12
45.548, -94.145	2	St. Cloud	263	6/15/2010	0.00	3.05	2.54	5.59	18.0	7.8	NA	NA	NA	NA
				6/19/2011	1.52	10.67	0.00	12.19	18.0	7.7	58.03	7.62	0.29	0.26
				6/28/2012	0.00	0.00	0.00	0.00	22.4	7.3	20.07	14.47	0.17	0.13
45.419, -94.042	3	Clearwater	271	6/15/2010	0.00	3.05	2.54	5.59	18.0	7.9	NA	NA	NA	NA
				6/19/2011	1.52	10.67	0.00	12.19	18.0	7.3	54.47	7.35	0.39	0.12
				6/28/2012	0.00	0.00	0.00	0.00	23.1	7.4	23.78	13.18	0.16	0.12

44.904, -93.190	4	Twin Cities	311	6/8/2010	0.00	0.25	21.34	21.59	21.0	8.3	NA	NA	NA	NA
				6/14/2011	0.00	0.00	17.53	17.53	19.0	7.9	65.20	7.60	0.71	0.20
				6/20/2012	2.03	38.61	3.05	43.69	23.5	7.9	40.86	9.72	0.51	0.10
44.886, -93.174	5	Minnesota River	NA[c]	6/8/2010	0.00	0.25	21.34	21.59	21.0	8.1	NA	NA	NA	NA
				6/12/2011	7.37	0.51	0.00	7.88	18.0	7.9	139.42	4.40	5.01	0.23
				6/26/2012	21.84	0.00	7.87	29.71	23.8	7.9	71.13	6.41	6.19	0.11
44.918, -93.129	6	Confluence[d]	313	6/8/2010	0.00	0.25	21.34	21.59	20.0	8.1	NA	NA	NA	NA
				6/12/2011	7.37	0.51	0.00	7.88	19.0	7.8	125.94	5.07	3.94	0.28
				6/26/2012	21.84	0.00	7.87	29.71	23.3	7.6	51.65	8.21	3.82	0.10
44.746, -92.850	7	Hastings	330	6/6/2010	24.38	0.00	4.06	28.44	21.0	8.2	NA	NA	NA	NA
				6/5/2011	0.76	0.00	0.00	0.76	21.0	8.1	92.80	6.16	3.66	0.22
				6/14/2012	0.00	0.00	17.78	17.78	20.5	7.7	54.59	9.58	2.17	0.11
44.749, -92.806	8	St. Croix River	NA	6/6/2010	24.38	0.00	4.06	28.44	19.0	7.0	NA	NA	NA	NA
				6/5/2011	0.76	0.00	0.00	0.76	19.0	7.2	21.87	8.99	0.31	0.16
				6/14/2012	0.00	0.00	17.78	17.78	19.9	6.9	13.35	12.43	0.18	0.11
44.567,	9	Red Wing	362[e]	6/6/2010	24.38	0.00	4.06	28.44	21.0	7.1	NA	NA	NA	NA

-92.538[¶]				6/5/2011	0.25	0.00	0.00	0.25	21.0	7.8	85.10	6.51	3.24	0.08
				6/7/2012	0.00	0.00	0.00	0.00	21.3	7.6	46.24	7.50	2.25	0.10
43.857, -91.303	10	La Crescent	401	6/2/2010	5.33	0.00	0.00	5.33	15.8	9.1	NA	NA	NA	NA
				6/1/2011	0.00	0.00	0.00	0.00	18.0	7.8	70.61	6.13	2.31	0.23
				5/30/2012	0.51	0.00	3.30	3.81	20.1	7.5	48.56	7.50	1.29	0.09
44.314, -91.996	11	Zumbro River	NA	6/1/2011	0.00	0.00	0.00	0.00	17.0	8.0	98.77	1.59	7.57	0.17
				5/30/2012	1.78	0.00	0.00	1.78	18.2	7.9	72.13	2.06	4.40	0.09

[a]Approximate distance from the headwaters at Lake Itasca.

[b]Total dissolved solids.

[c]Site not located on the Mississippi River.

[d]Confluence of the Mississippi and Minnesota Rivers.

[e]Sampling site was moved in 2011 to these coordinates; in 2010, the site was sampled at 44.745, -92.800, approximately 30 km upstream.

Supplementary Table S2. Spearman correlation coefficients relating phyla abundance with physicochemical parameters. Only significant correlations are shown.

Phylum	Antecedent Rainfall*				Temperature	pH	Distance†
	72 h	48 h	24 h	Cumulative	(°C)		
Betaproteobacteria	-0.225 (0.028)[a] -0.529 (0.003)[b]			-0.400 (< 0.001)[a] -0.689 (< 0.001)[b] -0.448 (0.009)[d]	-0.489 (0.006)[b]	0.381 (0.038)[b] -0.441 (0.010)[c]	
Gammaproteobacteria		0.296 (0.003)[a] 0.621 (< 0.001)[b]	0.595 (0.001)[b]	0.303 (0.003)[a] 0.521 (0.002)[d]	0.482 (0.007)[b]	0.443 (0.010)[d]	
Alphaproteobacteria	0.223 (0.022)[a] 0.720 (< 0.001)[b]	-0.517 (< 0.001)[a] -0.401 (0.028)[b] -0.436 (0.011)[c] -0.561 (0.001)[d]		0.478 (0.008)[b]		-0.288 (0.004)[a] -0.407 (0.025)[b] 0.450 (0.009)[c]	0.330 (0.005)[a] 0.610 (0.002)[c]
Actinobacteria	0.392 (0.032)[b]				-0.344 (0.050)		
Bacteroidetes	-0.383 (0.028)[d]	-0.257 (0.012)[a] -0.347 (0.048)[c]	-0.563 (0.001)[b]	-0.272 (0.007)[a] -0.490 (0.004)[d]	-0.228 (0.025)[a] -0.492 (0.006)[b] -0.549 (0.001)[d]		
Verrucomicrobia	0.534 (0.002)[b] -0.521 (0.002)[d]	-0.325 (0.001)[a] -0.710 (< 0.001)[c] -0.487 (0.004)[d]		0.442 (0.014)[b] -0.483 (0.004)[c]	0.413 (0.023)[b] 0.648 (< 0.001)[c] -0.419 (0.015)[d]	0.370 (0.044)[b] -0.356 (0.042)[d]	0.475 (< 0.001)[a] 0.594 (0.002)[b] 0.904 (< 0.001)[c]

	-0.400 (0.028)[b]	0.373 (0.043)[b]		-0.539 (0.002)[b]	-0.300 (0.003)[a]	0.618 (< 0.001)[a]	0.249 (0.035)[a]
Cyanobacteria		-0.604 (< 0.001)[c]		-0.636 (< 0.001)[c]		0.653 (< 0.001)[b]	0.809 (< 0.001)[c]
						0.570 (0.001)[c]	
	0.364 (0.048)[b]	-0.327 (0.001)[a]		0.447 (0.013)[b]		-0.280 (0.006)[a]	
Deltaproteobacteria		-0.502 (0.003)[d]				-0.431 (0.017)[b]	
	0.409 (0.025)[b]	-0.268 (0.008)[a]		0.463 (0.010)[b]	0.420 (0.021)[b]		0.729 (< 0.001)[c]
Firmicutes		-0.407 (0.019)[c]		-0.364 (0.037)[c]	0.522 (0.002)[c]		
		-0.381 (0.029)[d]					
Acidobacteria	-0.365 (0.037)[d]	-0.428 (0.013)[d]			0.265 (0.009)[a]	-0.216 (0.034)[a]	
	-0.434 (0.012)[d]		-0.912 (< 0.001)[b]	-0.208 (0.042)[a]	-0.407 (< 0.001)[a]	-0.378 (0.030)[d]	
Synergistetes				-0.609 (< 0.001)[b]	-0.690 (< 0.001)[b]		
				-0.559 (0.001)[d]			
	0.654 (< 0.001)[b]	-0.376 (0.031)[c]		0.612 (< 0.001)[b]	0.415 (0.023)[b]		0.473 (< 0.001)[a]
Planctomycetes					0.384 (0.028)[c]		0.564 (0.004)[c]
							0.619 (0.001)[d]
	-0.354 (0.043)[d]	-0.374 (0.032)[d]		0.537 (0.002)[b]	0.589 (0.001)[b]	0.376 (0.031)[c]	0.703 (< 0.001)[c]
Chloroflexi					-0.443 (0.010)[d]		
		-0.465 (0.006)[d]	-0.385 (0.036)[b]	-0.229 (0.025)[a]		-0.318 (0.002)[a]	-0.533 (0.007)[b]
Nitrospira				-0.464 (0.007)[d]		-0.356 (0.042)[d]	

Epsilonproteobacteria		-0.230 (0.024)[a] -0.453 (0.008)[d]			0.386 (0.035)[b]		0.460 (0.024)[c]
Gemmatimonadetes	-0.490 (0.004)[d]	-0.271 (0.008)[a] -0.470 (0.006)[c] -0.430 (0.012)[d]	0.233 (0.022)[a]	0.423 (0.020)[b]	0.275 (0.007)[a] 0.461 (0.010)[b] 0.422 (0.014)[c]		0.259 (0.028)[a] 0.540 (0.007)[c]
Deinococcus-Thermus			-0.346 (0.049)[d]	-0.358 (0.041)[d]		-0.378 (0.030)[d]	
Armatimonadetes	0.498 (0.005)[b] -0.427 (0.013)[d]	-0.277 (0.006)[a] -0.413 (0.017)[c]		-0.223 (0.029)[a] -0.472 (0.006)[d]		-0.643 (< 0.001)[b] -0.359 (0.040)[d]	
Chlamydiae	0.337 (0.001)[a] 0.492 (0.006)[b]	-0.397 (< 0.001)[a] -0.443 (0.014)[b] -0.385 (0.027)[c]	0.235 (0.021)[a] 0.590 (0.001)[b] 0.476 (0.005)[d]	0.282 (0.005)[a] 0.730 (< 0.001)[b] -0.367 (0.036)[c]	0.383 (< 0.001)[a] 0.777 (< 0.001)[b] 0.484 (0.004)[c]		0.391 (0.001)[a] 0.575 (0.003)[b] 0.756 (< 0.001)[c]
Spirochaetes		0.373 (0.043)[b]			0.472 (0.006)[c]	0.454 (< 0.001)[a] 0.610 (< 0.001)[b]	0.525 (0.008)[c]
Euryarchaeota	-0.229 (0.025)[a] -0.381 (0.029)[d]	0.590 (0.001)[b] -0.396 (0.023)[d]	0.451 (0.012)[b]	-0.509 (0.002)[d]	0.424 (0.020)[b]	-0.247 (0.015)[a] -0.432 (0.012)[d]	
Elusimicrobia	-0.286 (0.005)[a] -0.605 (< 0.001)[b]	0.854 (< 0.001)[b] -0.409 (0.018)[c]			0.461 (0.007)[c]	0.536 (0.001)[c] -0.394 (0.023)[d]	0.487 (0.016)[c]
Chlorobi	0.372 (0.033)[c]				-0.384 (0.027)[d]		
WS3		0.388 (0.034)[b]					

		-0.348 (0.047)[d]					
Fusobacteria	-0.495 (< 0.001)[b]	0.721 (< 0.001)[b]					-0.348 (0.003) -0.553 (0.005)[b]
Deferribacteres		-0.396 (0.023)[d]					
Thermotogae		-0.358 (0.041)[d]	0.361 (0.039)[d]		-0.349 (0.046)[d]		
Lentisphaerae				-0.266 (0.009)[a] -0.373 (0.042)[b]	-0.215 (0.035)[a] -0.424 (0.014)[d]	0.397 (0.022)[c]	0.423 (0.039)[c]
Aquificae		-0.204 (0.046)[a] -0.474 (0.008)[b]					
Crenarchaeota			-0.617 (< 0.001)[b]	0.356 (0.042)[c]	-0.243 (0.017)[a] -0.417 (0.022)[b]		-0.248 (0.035)
OD1					-0.394 (0.023)[d]		
TM7			-0.402 (0.028)[b]	-0.205 (0.045)[a]			-0.435 (0.034)[b]
Chrysiogenetes			0.315 (0.002)[a] 0.523 (0.003)[b]	0.323 (0.001)[a] 0.544 (0.002)[b]	0.494 (0.005)[b]		
Zetaproteobacteria							0.401 (< 0.001)[a] 0.479 (0.018)[b] 0.523 (0.009)[d]
Caldiserica			0.488 (0.004)[c]	0.396 (0.023)[c]			
SR1					0.407 (0.019)[c]		

Thermodesulfobacteria		0.374 (0.032)[c]			-0.517 (0.002)[d]		
Tenericutes	-0.372 (0.033)[c]	0.580 (0.001)[b]			-0.213 (0.037)[a] -0.434 (0.012)[d]	0.411 (0.017)[d]	-0.423 (0.039)[b]
unclassified			0.235 (0.021)[a] 0.407 (0.026)[b]		0.370 (0.044)[b]	0.419 (< 0.001)[a] 0.424 (0.020)[b] 0.571 (0.001)[d]	
Fibrobacteres	-0.202 (0.048)[a]		0.396 (0.023)[c]	-0.467 (0.009)[b]	-0.228 (0.026)[a] -0.435 (0.011)[d]		-0.302 (0.010)[a] -0.546 (0.006)[b] -0.547 (0.006)[d]
OP11		0.399 (0.021)[d]	-0.380 (0.029)[d]				-0.282 (0.016)[a]
BRC1					-0.499 (0.003)[d]	-0.359 (0.040)[d]	

[*]Rainfall that occurred 72, 48, or 24 h prior to sampling or cumulative rainfall over the three-day period.

†Distance from the headwaters at Lake Itasca.

[a]The *P* values are shown in parentheses and only statistically significant values are shown. Phyla and classes of *Proteobacteria* are shown in order of decreasing abundance among all samples. Correlations were significant among all samples collected from 2010-2012.

[b]Correlations were significant among samples collected during 2010.

[c]Correlations were significant among samples collected during 2011.

[d]Correlations were significant among samples collected during 2012.

Supplementary Table S3. Spearman correlation coefficients relating phyla abundance with TDS and nutrient concentrations for 2011 and 2012 data. Only significant correlations are shown.

Phylum	TDS	Carbon	Nitrite/Nitrate	Phosphorus
Betaproteobacteria		-0.368 (0.035)[b]	0.574 (< 0.001)[b]	
Gammaproteobacteria	0.310 (0.011)[a]		-0.421 (0.015)[b]	
Alphaproteobacteria	0.389 (0.025)[c]		0.383 (0.028)[c]	-0.261 (0.034)[a]
Actinobacteria				
Bacteroidetes				
Verrucomicrobia				
Cyanobacteria	0.463 (< 0.001)[a] 0.508 (0.003)[b]	-0.458 (< 0.001)[a] -0.647 (< 0.001)[b]	0.604 (< 0.001)[a] 0.736 (< 0.001)[b] 0.353 (0.044)[c]	-0.397 (0.022)[c]
Deltaproteobacteria	-0.267 (0.030)[a]	0.342 (0.005)[a]		
Firmicutes	0.374 (0.032)[b]		0.367 (0.002)[a] 0.434 (0.012)[b]	
Acidobacteria	-0.272 (0.027)[a]	0.304 (0.013)[a]		-0.324 (0.008)[a]
Synergistetes	-0.382 (0.028)[c]	-0.251 (0.042)[a]		0.394 (0.023)[c]
Planctomycetes			0.367 (0.036)[c]	-0.284 (0.021)[a] -0.464 (0.006)[c]
Chloroflexi	0.470 (0.006)[b]	-0.444 (0.010)[b]	0.502 (0.003)[b]	
Nitrospira	-0.255 (0.039)[a]			-0.313 (0.011)[a]

Epsilonproteobacteria				
Gemmatimonadetes	-0.283 (0.021)[a]	0.375 (0.002)[a]		-0.260 (0.035)[a]
Deinococcus-Thermus	-0.452 (0.008)[c]			
Armatimonadetes	-0.398 (0.022)[c]			
Chlamydiae			0.276 (0.025)[a] 0.433 (0.012)[c]	
Spirochaetes	0.266 (0.031)[a] 0.361 (0.039)[b]	-0.243 (0.049)[a]	0.323 (0.0008)[a] 0.370 (0.034)[b]	
Euryarchaeota	-0.396 (0.001)[a] -0.444 (0.010)[b] -0.407 (0.019)[c]	0.355 (0.003)[a] 0.352 (0.045)[b] 0.366 (0.036)[c]	-0.288 (0.019)[a]	
Elusimicrobia	-0.365 (0.037)[c]	0.376 (0.031)[c]	0.451 (0.008)[b]	
Chlorobi	0.395 (0.023)[b]		0.344 (0.050)[b]	-0.361 (0.039)[c]
WS3				
Fusobacteria				
Deferribacteres				
Thermotogae			0.431 (0.012)[c]	-0.373 (0.002)[a] -0.497 (0.003)[c]
Lentisphaerae	0.353 (0.044)[b]	-0.276 (0.025)[a] -0.525 (0.002)[b]	0.318 (0.009)[a] 0.530 (0.002)[b]	
Aquificae				
Crenarchaeota				
OD1	0.493 (0.004)[b]	-0.278 (0.024)[a]	0.306 (0.012)[a]	

		-0.516 (0.002)[b]	0.407 (0.019)[b]	
TM7				
Chrysiogenetes				
Zetaproteobacteria				-0.323 (0.008)[a] -0.566 (0.001)[c]
Caldiserica				
SR1				
Thermodesulfobacteria				-0.290 (0.018)[a] -0.372 (0.033)[c]
Tenericutes		-0.302 (0.014)[a] -0.445 (0.009)[c]		-0.250 (0.043)[a] -0.404 (0.020)[c]
unclassified	0.251 (0.042)[a]	-0.264 (0.032)[a] -0.372 (0.033)[c]	0.395 (0.001)[a] 0.348 (0.047)[b] 0.431 (0.012)[c]	
Fibrobacteres				
OP11	-0.322 (0.008)[a]			
BRC1		-0.327 (0.007)[a] -0.501 (0.003)[c]	0.258 (0.037)[a] 0.400 (0.021)[c]	-0.311 (0.011)[a] -0.460 (0.007)[c]

[a]The *P* values are shown in parentheses and only statistically significant values are shown. Phyla and classes of *Proteobacteria* are shown in order of decreasing abundance among all samples. Correlations were significant among all samples collected from 2011 and 2012.

[b]Correlations were significant among samples collected during 2011.

[c]Correlations were significant among samples collected during 2012.

Supplementary Table S4. Mean and standard deviation of most abundant phyla and *Proteobacteria* classes identified in river water samples during the study period. Other phyla had < 3% mean abundance among all samples tested

Year	Site	*Betaproteobacteria*	*Gammaproteobacteria*	*Alphaproteobacteria*	*Actinobacteria*	*Bacteroidetes*	*Verrucomicrobia*	*Cyanobacteria*
2010	**Itasca**	80.1 ± 2.7	7.3 ± 0.7	10.4 ± 3.5	6.7 ± 9.4	12.5 ± 8.8	0.6 ± 0.0	3.0 ± 0.3
	St. Cloud	76.7 ± 1.3	10.2 ± 0.3	10.8 ± 1.1	13.2 ± 15.3	7.2 ± 2.0	3.1 ± 0.1	5.4 ± 0.6
	Clearwater	75.7 ± 0.5	12.2 ± 0.4	9.9 ± 0.4	9.2 ± 11.2	5.5 ± 2.3	3.6 ± 0.3	5.9 ± 0.4
	Twin Cities	74.8 ± 0.9	12.6 ± 0.8	9.8 ± 0.3	9.6 ± 12.7	4.3 ± 1.7	7.2 ± 1.0	11.2 ± 1.2
	MN River	78.0 ± 0.3	10.7 ± 0.2	9.3 ± 0.2	11.0 ± 14.3	4.5 ± 2.4	3.0 ± 0.6	2.5 ± 0.5
	Confluence	79.6 ± 0.2	11.6 ± 0.1	7.3 ± 0.2	7.8 ± 10.3	2.5 ± 1.3	2.1 ± 0.3	1.6 ± 0.2
	Hastings	71.6 ± 5.7	11.4 ± 1.7	14.6 ± 4.0	16.0 ± 20.6	4.1 ± 1.9	9.8 ± 1.1	4.0 ± 0.2
	St. Croix River	70.4 ± 1.4	9.7 ± 0.2	16.7 ± 1.0	12.4 ± 13.7	7.3 ± 2.3	3.4 ± 0.1	0.4 ± 0.0

	Red Wing	73.2 ± 1.7	9.2 ± 0.2	15.2 ± 1.6	15.0 ± 15.8	5.2 ± 2.1	4.3 ± 0.1	0.9 ± 0.1
	La Crescent	80.0 ± 1.1	8.5 ± 0.2	10.0 ± 0.9	14.9 ± 18.5	6.6 ± 2.8	4.5 ± 0.6	7.4 ± 0.4
2011	**Itasca**	47.4 ± 12.9	47.0 ± 12.8	4.3 ± 0.0	2.8 ± 3.2	1.8 ± 0.9	0.2 ± 0.0	0.2 ± 0.1
	St. Cloud	46.0 ± 10.7	47.2 ± 11.9	5.5 ± 1.0	5.2 ± 6.5	1.7 ± 0.4	0.6 ± 0.1	0.4 ± 0.1
	Clearwater	78.0 ± 0.2	8.5 ± 1.0	11.0 ± 1.0	15.6 ± 19.9	6.3 ± 2.7	1.3 ± 0.1	1.5 ± 0.1
	Twin Cities	76.2 ± 1.4	9.6 ± 0.9	11.2 ± 0.1	11.3 ± 14.7	5.6 ± 2.3	1.9 ± 0.0	3.0 ± 0.4
	MN River	77.1 ± 1.3	11.2 ± 0.6	9.7 ± 0.3	9.7 ± 13.6	2.6 ± 1.9	1.3 ± 0.4	2.4 ± 0.8
	Confluence	76.5 ± 2.1	11.0 ± 0.9	10.3 ± 0.4	14.5 ± 20.0	4.2 ± 2.3	1.7 ± 0.7	2.4 ± 1.3
	Hastings	70.2 ± 5.1	20.5 ± 5.7	7.8 ± 0.6	5.8 ± 7.6	2.1 ± 1.1	1.5 ± 0.2	1.4 ± 0.3
	St. Croix River	74.8 ± 0.8	12.5 ± 1.2	10.9 ± 0.5	8.5 ± 12.4	3.3 ± 1.8	5.3 ± 0.9	0.8 ± 0.1
	Red Wing	76.3 ± 0.1	11.5 ± 0.3	10.2 ± 0.7	12.5 ± 17.0	6.2 ± 3.2	3.6 ± 0.1	3.3 ± 0.5
	La Crescent	72.7 ± 0.4	12.5 ± 0.3	12.4 ± 0.2	9.3 ± 11.6	5.2 ± 3.4	4.3 ± 0.1	6.5 ± 2.9
	Zumbro River	80.5 ± 3.9	10.7 ± 0.9	7.1 ± 2.6	4.3 ± 5.4	7.3 ± 6.6	0.8 ± 0.1	10.8 ± 1.7
2012	**Itasca**	86.7 ± 1.7	5.3 ± 1.0	6.7 ± 0.5	13.7 ± 19.4	5.4 ± 3.9	0.4 ± 0.1	0.6 ± 0.2

St. Cloud	78.8 ± 1.7	8.1 ± 1.1	10.2 ± 0.1	7.9 ± 11.7	4.9 ± 3.4	1.9 ± 0.8	0.3 ± 0.1
Clearwater	75.4 ± 1.1	8.7 ± 0.7	12.6 ± 0.2	8.4 ± 11.8	4.9 ± 2.5	2.5 ± 0.6	0.8 ± 0.2
Twin Cities	53.1 ± 6.5	37.6 ± 6.7	7.6 ± 0.3	4.3 ± 5.4	1.6 ± 1.0	0.8 ± 0.1	1.0 ± 0.2
MN River	55.5 ± 3.7	30.0 ± 3.9	12.6 ± 0.3	4.3 ± 5.6	1.7 ± 0.6	0.6 ± 0.1	0.9 ± 0.1
Confluence	76.1 ± 2.1	9.1 ± 1.7	11.7 ± 0.4	12.1 ± 18.1	4.5 ± 2.7	1.7 ± 0.9	1.6 ± 0.9
Hastings	80.4 ± 1.6	7.7 ± 1.0	9.8 ± 0.3	14.9 ± 21.9	3.9 ± 2.3	1.9 ± 1.2	1.6 ± 1.6
St. Croix River	67.5 ± 2.8	12.1 ± 0.2	17.9 ± 2.6	15.5 ± 19.6	4.5 ± 1.9	2.6 ± 0.2	0.9 ± 0.0
Red Wing	75.7 ± 0.5	11.1 ± 0.5	10.7 ± 0.5	11.8 ± 17.1	5.4 ± 3.0	3.6 ± 0.2	2.1 ± 0.6
La Crescent	79.7 ± 0.7	8.6 ± 0.5	10.1 ± 0.3	18.3 ± 26.2	4.6 ± 1.8	1.8 ± 0.1	1.0 ± 0.2
Zumbro River	86.4 ± 5.9	5.1 ± 1.4	7.1 ± 4.1	12.8 ± 19.6	2.5 ± 1.9	0.5 ± 0.1	0.4 ± 0.3

Frequencies of heavy metal resistance are associated with land cover type in the Upper Mississippi River

Christopher Staley[1], Dylan Johnson[2], Trevor J. Gould[1,2], Ping Wang[1], Jane Phillips[2], James B. Cotner[3], and Michael J. Sadowsky[1,4,*]

[1]BioTechnology Institute, [2]Biology Program, [3]Department of Ecology, Evolution and Behavior, and [4]Department of Soil, Water and Climate, University of Minnesota, St. Paul, MN

[*]Corresponding Author: Michael J. Sadowsky, BioTechnology Institute, University of Minnesota, 140 Gortner Lab, 1479 Gortner Ave, Saint Paul, MN 55108; Phone: (612)-624-2706, Email: 1lemson1@umn.edu

Abstract

Taxonomic compositions of freshwater bacterial communities have been well-characterized via metagenomic-based approaches, especially next-generation sequencing; however, functional diversity of these communities remains less well-studied. Various anthropogenic sources are known to impact the bacterial community composition in freshwater river systems and potentially alter functional diversity. In this study, high-throughput functional screening of large (~10,000 clones) fosmid libraries representing communities in the Upper Mississippi River revealed low frequencies of resistance to heavy metals in the following order: $Mn^{2+} > Cr^{3+} > Zn^{2+} > Cd^{2+} > Hg^{2+}$, and no resistance to Cu^{2+} was detected. Significant correlations were observed between resistance frequencies of Cd and Cr with developed land cover ($r = 0.296$, $P = 0.016$ and $r = 0.257$, $P = 0.037$, respectively). Discriminant function analysis further supported these associations while redundancy analysis further indicated associations with forested land cover and greater resistance to Hg and Zn. Nutrient and metal ion concentrations and abundances of bacterial orders were poorly correlated with heavy metal resistance frequencies, except for an association of *Pseudomonadales* abundance and resistance to Hg and Zn. Taken together, results of this study suggest that allochthonous bacteria contributed from specific land cover types influence the patterns of metal resistance throughout this river.

Keywords: bacterial community / fosmid library / functional metagenomics / heavy metal resistance / Mississippi River

1. Introduction

Over the last several decades, many studies have assessed the sensitivity and tolerance of environmental microbial communities to heavy metals in water, soils, and the rhizosphere (Hassen et al. 1998; Abou-Shanab et al. 2007). It has been well-established that high concentrations of heavy metal pollutants in the environment are associated with declines in bacterial diversity, particularly among rare taxa (Gans et al. 2005; Ancion et al. 2010; Hemme et al. 2010). However, not all metals are equally toxic, and several (*e.g.* Cr and Zn) are known to be critical for cellular functions (Seiler and Berendonk 2012), while others (*e.g.* Cd and Hg) are known to form highly toxic complexes (Nies 1999). Nevertheless, even biologically important trace elements can be toxic at elevated concentrations. Furthermore, environmental conditions, such as pH, that affect the valence state of metal ions are also important factors affecting metal toxicity (Nies 1999; Seiler and Berendonk 2012).

Taxonomic structures in bacterial communities in lotic systems (rivers and streams) in both the water and sediment have been shown to be influenced by surrounding land cover type, resulting from both the introduction of non-indigenous bacteria as well as other pollutants that cause variations in community structure (Wang et al. 2011; Gibbons et al. 2014; Staley et al. 2014a). Heavy metals are among the contaminants contributed to the environment resulting from a variety of anthropogenic practices including agriculture (Han et al. 2000), aquaculture (Burridge et al. 2010), and discharge of industrial and municipal effluent (Ahluwalia and Goyal 2007). A recent study has also indicated that bacterial biofilms can capture and retain these metals and

potentially transfer them to higher trophic levels, representing a concern for human health (Ancion et al. 2010).

In general, mechanisms of heavy metal resistance fall into one of three types. Heavy metal ions may be incorporated into complexes for sequestration, toxicity of intracellular ions may be lessened by reduction of metal ions to less toxic valence states, and/or toxic ions may be removed from the cell via efflux pumps (Nies 2003; Seiler and Berendonk 2012). Several bacteria, such as *Ralstonia metallidurans*, have naturally adapted mechanisms for heavy metal resistance chromosomally to survive in highly metal-rich habitats (Mergeay et al. 2003), and the functions and distribution of these genes have been well reviewed (Silver and Phung 1996; Nies 2003). However, anthropogenic pollution, in the form of increased heavy metal ion concentrations or other pollutants, imposes a selective pressure in favor of these resistance mechanisms and has resulted in their incorporation onto mobile genetic elements (*e.g.* plasmids), enabling horizontal gene transfer (Silver and Phung 1996; Nies 2003). Of particular concern is the spread of resistance mechanisms that also confer resistance to other antimicrobial compounds including antibiotics (Chapman 2003; Nies 2003).

Monitoring of the microbial community has been proposed over the last several decades as a biological indicator of soil health (Pankhurst et al. 1995; Yakovchenko et al. 1996), with some suggesting that bacterial community responses may be detectable in advance of detectable changes in abiotic, edaphic parameters (Pankhurst et al. 1995). Similarly, in a recent study we have shown that changes in the structure of the microbial community in the Upper Mississippi River may be indicative of more subtle variations in chemical contaminants as a result of anthropogenic practices (Staley et al. 2014a).

However, a number of studies have suggested that bacterial functional responses, rather than taxonomic community structure, are better indicators of perturbation (Comte and del Giorgio 2009; Burke et al. 2011; Steffen et al. 2012). In response to specifically metal contamination, for example, an earlier study indicated that copper contamination resulted in a decrease in photosynthetic potential among a phototrophic community (Massieux et al. 2004). More recently, in the Upper Mississippi River, we have found that while the distribution of functional traits is highly conserved, slight but significant variations in the distribution of functional traits are also linked with surrounding land cover using inference-based and whole genome shotgun approaches (Staley et al. 2014b).

Culture-dependent methods have been suggested to better represent the physiological state of bacteria than culture-independent methods (Ellis et al. 2003), due to discrepancies between genes present versus those expressed in the environment. Functional screening of large, metagenomic fosmid libraries has been recently proposed as a method by which to characterize functional traits in a microbial community (Martínez and Osburne 2013), while at the same time circumventing the lack of culturability of > 99% of bacterial species in the laboratory (Amann et al. 1995). This method has been recently used to characterize antibiotic resistance frequencies in river sediments (Amos et al. 2014).

In the current study, we performed functional metagenomic screening of fosmid libraries constructed from water samples collected throughout the Upper Mississippi River in 2011 and 2012 to determine frequencies of resistance to the metals Cd^{2+}, Cr^{3+}, Cu^{2+}, Hg^{2+}, Mn^{2+}, and Zn^{2+}. We have previously taxonomically characterized the bacterial community at these sites (Staley et al. 2014a), and we hypothesize that factors

91 influencing taxonomic community structure, such as total carbon and nitrite/nitrate
92 concentrations as well as major land cover types, will also be associated with resistances
93 to specific metals. Results of this study elucidate the interrelationships between
94 physicochemical parameters, taxonomic variation, and resistances of environmental
95 communities to heavy metals in a major riverine ecosystem.
96
97 **2. Materials and methods**
98 *2.1. Sample collection and processing*
99 Surface water samples (40 L) were collected from the shore in sterile, 20 L
100 carboys from 11 sites along the Upper Mississippi River in Minnesota and major
101 contributing rivers from near the headwaters at Lake Itasca to the southern border near La
102 Crescent, MN, as previously described (Staley et al. 2014a). Water samples were
103 transported back to the lab and either processed immediately or stored at 15 °C overnight
104 and processed the following day. As described previously, samples were strained through
105 sterile cheesecloth and filtered through a P5 pre-filter (Whatman Inc., Piscataway, NJ) to
106 remove aggregate bacteria prior to concentrating larger, planktonic bacterial cells on a
107 0.45-µm-pore-size polyethane-sulfonate filter. We have previously reported that this pore
108 size was necessary to efficiently filter this large volume of water (Staley et al. 2013).
109 Cells were elutriated from filters by vortexing in pyrophosphate buffer and cell pellets,
110 six per site, representing approximately 6-7 L of water, were stored at -80 °C until used.
111
112 *2.2. Fosmid library preparation*

One cell pellet per site, per year was shipped on dry ice to the Clemson University Genomics Institute (CUGI) [http://www.genome.clemson.edu/] where DNA extraction and fosmid library construction was performed. Briefly, DNA from each of the pellets was extracted using the Metagenomic DNA Isolation Kit for Water (Epicentre Biotechnologies, Madison, WI) followed by end-repair/phosphorylation according to the manufacturer's instructions. DNA fragments between 35-50 kb were size selected by pulsed-field gel electrophoresis and were subsequently ligated into pCC2FOS (Epicentre Biotechnologies). Ligated fosmids were transduced into *Escherichia coli* DH10B by λ phage. Fosmid libraries for each site contained a minimum of 50,000 clones and were shipped back to the laboratory on dry ice as glycerol stocks.

Stock fosmid libraries were diluted to 2.5 CFU μl^{-1} and aliquots of 1 ml were plated on 20 × 20 cm Luria Bertani (LB) agar plates containing 12.5 $\mu g\ ml^{-1}$ chloramphenicol. Colonies (approximately 10,000 per library, Table 1) were transferred to 384-well plates containing Hogness modified freezing media (HMFM) (Yan et al. 2007) with 12.5 $\mu g\ ml^{-1}$ chloramphenicol using the Qbot colony-picking robot (Genetix, Sunnyvale, CA). Fosmid libraries were stored at -80 °C.

2.3. Heavy metal resistance screening

Six metals (Cd^{2+}, Cr^{3+}, Cu^{2+}, Hg^{2+}, Mn^{2+}, and Zn^{2+}) were selected for metal resistance screening based on a wide range of expected toxicities and suspected presence in environmental samples. Chloride salts were used to make stock solutions at the following concentrations: 0.1 M $CdCl_2$, 1 M $CrCl_3$, 1 M $CuCl_2$, 0.1 M $HgCl_2$, 2 M $MnCl_2$, and 1 M $ZnCl_2$. Stock solutions were made fresh, weekly and filter-sterilized

136 through a 0.22 μm pore-size filter. Fosmid libraries were screened using a flame-
137 sterilized replicator to stamp plates onto 20 × 20 cm nutrient agar plates containing 7 μg
138 ml^{-1} chloramphenicol and the inhibitory concentration of a single metal (see below).
139 Nutrient agar has been previously utilized as a solid medium for heavy metal resistance
140 screening of environmental samples (Hassen et al. 1998). Six 384-well plates were
141 stamped onto each 20 × 20 cm plate. Plates were incubated for approximately 18 h at 37
142 °C. For each set of plates made, a control strain (*E. coli* DH10B containing a fosmid
143 without insert) was also streaked to verify that the media contained an inhibitory metal
144 concentration. Resistant colonies were reported as those that produce opaque colonies at
145 least 1 mm in diameter.
146 Inhibitory metal concentrations were experimentally determined by plating a
147 control strain of *E. coli* DH10B containing a fosmid without insert, grown overnight in
148 LB supplemented with 7 μg ml^{-1} chloramphenicol, in triplicate over a range of
149 concentrations from 0.050 mM to 20 mM (0.05 mM increments). The inhibitory
150 concentration was selected as the lowest concentration which completely inhibited
151 growth of all three replicates. Nutrient agar plates containing 7 μg ml^{-1} chloramphenicol
152 were used for all plating, and inhibitory concentrations were determined as 0.25 mM Cd,
153 2 mM Cr, 1 mM Cu, 0.075 mM Hg, 20 mM Mn, and 0.50 mM Zn.
154
155 *2.4. Bacterial community structure*
156 The bacterial community structure of water samples was determined and is
157 reported elsewhere (Staley et al. 2014a). Briefly, DNA was extracted from cell pellets
158 using the Metagenomic DNA Isolation Kit for Water (Epicentre Biotechnologies). The

159 V6 hypervariable region of the 16S rDNA was amplified using the 967F/1046 primer set
160 (Sogin et al. 2006), and amplicons were paired-end sequenced using the MiSeq and
161 HiSeq2000 platforms by the University of Minnesota Genomics Center (Minneapolis).
162 Sequence processing and OTU calling (97% similarity, furthest-neighbor algorithm) were
163 performed as described elsewhere (Staley et al. 2014a), using the SILVA database (ver.
164 102) for alignment and the RDP database (ver. 9) for taxonomic assignment (Pruesse et
165 al. 2007; Cole et al. 2009). Order abundances evaluated in the current work represented
166 the most abundant orders, representing an average of at least 1% of the community
167 among all samples, as well as those that were found to be significantly associated with
168 land cover type by discriminant function analysis in our prior study (Staley et al. 2014a).
169
170 *2.5. Metadata*
171 The collection of metadata is also described elsewhere (Staley et al. 2014a).
172 Briefly, land cover types were determined by overlaying the 2006 National Land Cover
173 Database onto a map of hydrologic basin boundaries at a 1:250,000 scale (Fry et al.
174 2011). Major land cover types (agricultural, developed, or forested) for sampling sites
175 were assigned based on the highest percentage of aggregate land cover pixels for that site
176 within the basin in which the site was located. Nutrient (ammonium, carbon,
177 nitrite/nitrate, and orthophosphate) and metal ion (Al, Mn, and Cu) concentrations were
178 determined by the Research Analytical Labs at the University of Minnesota (St. Paul)
179 using standard methods [http://ral.cfans.umn.edu/types-of-analysis-offered/water/]. Water
180 for these analyses (1 L) was collected in addition to the main, 40 L sample in sterile
181 amber bottles. The analytes evaluated here were those that were previously directionally

182 associated with community structure by Bayesian analysis (Staley et al. 2014a), and,
183 therefore, were expected to have potential associations with community function.
184
185 *2.6. Statistical analyses*
186 Spearman rank correlations, ANOVA analyses, and discriminant function analysis
187 were performed using SPSS Statistics software ver.19 (Chicago, Illinois). Redundancy
188 analysis was performed using XLSTAT (Addinsoft, Belmont, MA). All statistics were
189 evaluated at $\alpha = 0.05$.
190
191 **3. Results and discussion**
192 Frequencies of heavy metal resistances were generally low for all metals tested
193 (Table 1). The highest frequencies of resistance were observed for Cr, Mn, and Zn, lower
194 resistances were observed for Cd and Hg, and resistance to Cu was not observed. In 2011,
195 Cr, Mn, and Zn resistance frequencies observed were greater than those observed in
196 2012. However, only the frequencies of resistance to Cd and Cr were significantly
197 different between years ($P \leq 0.001$). The frequency of Cd resistance was greater in 2012
198 versus 2011, while the inverse was true for Cr (Table 1). Frequencies of metal resistances
199 observed also varied by sampling site during both years (standard deviations between
200 0.01 to 3.57% of resistant clones; Table 1). In 2011, clones from Itasca had the greatest
201 observed frequencies of resistance to Cr, Cd, Hg, and Zn while the greatest frequency of
202 Mn resistance was observed at Hastings. This trend was not observed in 2012, and only
203 the frequency of Zn resistance was greatest at Itasca.

204 Resistance frequencies detected here are considerably less than those previously
205 reported from environmental waters using culture-based methods (Sabry et al. 1997;
206 Hassen et al. 1998). While resistance to Cd was still low (1%), resistances to Cu and Hg
207 were considerably greater than those observed here (22 and 9%, respectively) among
208 isolates from marine water by agar dilution method, with concentrations ranging from
209 0.005 – 80 mM (Sabry et al. 1997). Similarly, bacteria from various environments
210 including freshwater and wastewater showed resistance to Cd, Cr, and Hg by agar
211 dilution method when greater concentrations of these metals were used than those applied
212 in the current study (Hassen et al. 1998). Our lack of detection of resistance to Cu in this
213 study is particularly surprising; however, Cu concentration in the water was undetectable
214 in all but five of the samples collected (n = 22) and was generally low (0.01 to 0.57 mg L^{-1}
215) suggesting a lack of selective pressure to maintain resistance to this metal.
216 Furthermore, the frequency of heavy metal resistance observed in this study is similar to
217 frequencies of antibiotic resistance observed in river sediments via functional
218 metagenomic screening (Amos et al. 2014).
219 The low frequencies of resistance observed here compared to previous studies are
220 thought to result from methodological differences. Direct plating of environmental
221 samples in previous studies may have selected for resistant isolates that are minority
222 members of the microbial community. We have previously reported that a large
223 percentage of the bacterial community in this ecosystem (~90%) is comprised of only a
224 small number of species (Staley et al. 2013), and these species are likely to be over-
225 represented in the fosmid libraries. Resultantly, metal resistance among minority
226 members of the community are unlikely to be detected by the method used here despite a

large number of clones screened. Similarly, housekeeping genes, which are present in greater abundance than those conferring heavy metal resistance, are probably also over-represented in the fosmid libraries resulting in low frequencies of resistance detected. Finally, environmental DNA may not have been incorporated into fosmids in the correct orientation, so while genes conferring resistance could be present, these genes may not be being expressed in the *E. coli* host. Nevertheless, given the large numbers of clones screened, the resistance frequencies reported here potentially better represent actual resistance patterns in the total riverine bacterial community than would be determined using methods which select for resistant phenotypes.

In the current study, multiple resistances to heavy metals were also infrequently observed (mean of 0.34 ± 0.01% of all clones screened among all sites during both years). Resistances to multiple metals were more frequent in 2011 (0.67 ± 1.09%) compared to 2012 (0.33 ± 0.22%) among all sampling sites, but this difference was not statistically significant ($P = 0.319$). Resistances to Cr and Mn were most commonly observed among clones showing multiple resistances. Not surprisingly, the frequency of resistances to multiple metals was greatest in 2011 at the Itasca sampling site (3.84%). In 2012, the greatest frequency of multiple resistances was observed in the St. Cloud sample (0.86%), which also had among the greatest frequencies of resistance to single metals (Table 1). Multiple resistances observed among 81 aerobic heterotrophic bacterial isolates, from marine water, have previously been described to exhibit high frequencies of multiple metal resistance (95.1%), and the most commonly observed resistance pattern was pentametal resistance (25.9%) among eight metals screened (Sabry et al. 1997). A more recent study of soil and rhizosphere bacterial cultures found multiple metal

250 resistance in all isolates with heptametal resistance as the predominant pattern (28.8% of
251 isolates) (Abou-Shanab et al. 2007). Differences in multiple metal resistance patterns
252 observed here are likely due to the screening method used, as described above, but may
253 also reflect differences in this riverine bacterial community from communities in marine
254 waters or terrestrial habitats.
255
256 *3.1. Association of land coverage category with metal resistance frequency*
257 Among data collected from both years of study, a discriminant function analysis
258 (DFA) revealed that one function relating heavy metal resistance frequencies to major
259 land coverage types was significant and explained 93.9% of the variance within the
260 model among all samples (Wilks' $\lambda = 0.131$, $P < 0.001$; Figure 1). The second
261 discriminant function was also statistically significant and explained the remaining 6.1%
262 of variance in this model (Wilks' $\lambda = 0.763$, $P = 0.002$). Clustering of resistance
263 frequencies with land coverage types by DFA analysis was similar to results obtained via
264 Spearman correlation analysis. Cd and Cr resistance frequencies were significantly
265 positively correlated with the percentage of developed land cover ($r = 0.296$, $P = 0.016$
266 and $r = 0.257$, $P = 0.037$, respectively). Cd resistance frequency was also negatively
267 correlated with forested land cover ($r = -0.387$, $P = 0.001$) while Cr resistance frequency
268 was negatively correlated with agricultural land cover ($r = -0.246$, $P = 0.047$). Similarly,
269 resistance frequencies over both years were higher at sites surrounded by the land cover
270 type with which they clustered by DFA (Table 2). Frequencies of resistance to Hg and Zn
271 were not significantly correlated with land cover ($P \geq 0.065$) but both were significantly
272 greater at forested sites than developed or agricultural sites (Table 2).

273 Several studies have indicated that specific anthropogenic practices generally
274 contribute to higher metal concentrations in the environment (Han et al. 2000; Ahluwalia
275 and Goyal 2007; Burridge et al. 2010). However, there is presently a paucity of studies
276 evaluating associations between land cover type and the frequency of specific heavy
277 metal resistances in bacteria. However, a recent study investigated heavy metal
278 concentrations in topsoil and found that, among the metals also tested in this study, land
279 use was associated with large-scale (11 km) variation in Cd and Hg (Zhao et al. 2010).
280 This study further reported that anthropogenic activities have dramatic influences on
281 concentrations of Cd, Cu, Hg, and Zn. Concentrations of Cu and Zn were higher in
282 forested areas, and Cd and Hg concentrations were generally higher in areas dominated
283 by vegetable fields and were presumed to result from industrial emissions and
284 agrochemical usage (Zhao et al. 2010). The findings in the previous study generally
285 corroborate the association of Cd with developed land cover and the high frequency of Zn
286 resistance at forested sites found here. Conversely, Hg concentration was greater at
287 forested sites in the current study.
288
289 *3.2. Relationship of water chemistry on metal resistance frequency*
290 Nutrient and chemical concentrations measured are reported elsewhere (Staley et
291 al. 2014a) and those that were significantly correlated to heavy metal resistance
292 frequencies and were previously found to be significantly related to bacterial community
293 structure by Bayesian analysis are summarized in (Table 3). The concentration of
294 ammonium was higher in 2011 while concentrations of carbon and Al were higher in
295 2012 ($P \leq 0.001$). No other analyte concentrations varied significantly by year ($P \geq 0.05$).

296 Based on our previous study (Staley et al. 2014a), several intercorrelations were
297 observed between land cover type/analyte concentration and land cover type/metal
298 resistance frequency (Table 3). These intercorrelations were investigated to determine if
299 particular nutrients associated with land cover types may also be associated with an
300 increased metal resistance frequency, perhaps by providing a competitive advantage to
301 autochthonous, resistant taxa. Conversely, metal resistance frequencies may be
302 independently related to land cover due to contributions of resistant, allochthonous taxa.
303 For the purpose of evaluating these intercorrelations, Hg and Zn resistance frequencies
304 were presumed to be associated with forested land cover on the basis of the ANOVA
305 result (Table 2). Redundancy analysis relating nutrient and metal ion concentrations, land
306 cover, and metal resistance was also performed (Figure 2).
307 Redundancy analysis revealed few strong trends among nutrients and metal ions
308 with metal resistance frequencies. Ammonium, total carbon, and Mn concentrations
309 appeared to be very weakly associated with Cd and Mn resistance frequencies and
310 developed land cover. Of these, only Mn concentration was significantly positively
311 correlated with developed land cover in our prior study (Staley et al. 2014a), where Cd
312 was also related with this land cover type here. A prior study has suggested that
313 treatment of soils with industrial wastewater contributed to increased concentrations of
314 Cd and promoted an increase in the biosorption capability of bacteria to Cd (Ansari and
315 Malik 2007). Therefore, it is possible to suggest that wastewater discharges in developed
316 areas may also be contributing to increased frequencies of Cd resistance observed here,
317 perhaps by co-selecting for Mn resistance in response to greater concentrations of Mn
318 ions. Wastewater is also known to contain high levels of other metals promoting greater

319 bacterial resistance (Leung et al. 2000), so it is surprising that only Cd, and very weakly
320 Mn, resistances were associated with developed land.

321 As expected based on ANOVA results (Table 2), redundancy analysis indicated a
322 weak relationship between forested land cover and resistance to Hg and Zn. All nutrient
323 and metal concentrations that were significantly correlated with Hg resistance here (Table
324 3), were generally also significantly negatively correlated with forested land cover in our
325 previous study (Staley et al. 2014a). These results suggest that anthropogenic pollutants
326 are not contributing to elevated frequencies of resistance to Hg and Zn, although a
327 previous study of topsoils indicated that pesticide use was a large contributor of Hg (Zhao
328 et al. 2010). Atmospheric transport of herbicides potentially containing these metals has
329 been reported (Goolsby et al. 1997), suggesting a possibility that these metals were
330 deposited in soils removed from anthropogenic contamination and may still contribute to
331 a reservoir of resistance genes. This possibility will require more thorough investigation
332 of soil communities in forested regions.

333 Redundancy analysis revealed that forested land cover was more closely
334 associated with Cr resistance, although DFA suggested an association between developed
335 land cover and Cr resistance. The reasons for this discrepancy are unclear but may result
336 from co-variation of developed land cover with orthophosphate and nitrite/nitrate
337 concentrations. The primary consequence of increased nitrogen and phosphate in surface
338 waters is eutrophication, which has been shown to increase bacterial biomass (Smith and
339 Schindler 2009). This increase in biomass could potentially result in an increase in non-
340 resistant species impeding our ability to detect Cr resistance, which is more clearly
341 illustrated by redundancy analysis than more simplistic DFA.

342

343 *3.3. Taxonomic relationship to metal resistance frequency*

344 The taxonomic distributions of bacterial communities, derived from 16S rDNA
345 sequence data, have been reported previously, and fourteen orders were found to be
346 significantly associated with major land cover types by DFA (Staley et al. 2014a). Those
347 orders that were significantly correlated with land cover type and heavy metal resistance
348 are shown in Table 4. These orders were also incorporated into a redundancy analysis
349 with land use and nutrient concentrations, but all except *Pseudomonadales* were poorly
350 correlated with metal resistance frequencies and are not shown for simplicity. Only
351 *Pseudomonadales* abundance was strongly associated with forested land use in the
352 redundancy analysis, and resultantly, Hg and Zn resistance frequencies. This association
353 is consistent with prior studies of environmental heavy metal resistance where
354 *Pseudomonas* isolates were among the most frequently classified (Hassen et al. 1998).
355 With the exception of the most abundant order (*Burkholderiales*), all of the orders
356 considered were significantly correlated with the resistance frequency of at least one
357 metal (Table 4); however, the majority of relationships were inconsistent in significance
358 and/or direction between years. The lack of clear associations between orders and
359 resistance frequencies here is not surprising as order-level classification is very broad and
360 can include a large number of functional traits. However, more specific classification
361 using short sequence reads has been shown to result in inaccurate classification (Mizrahi-
362 Man et al. 2013). Taken together, these results suggest that environmental heavy metal
363 resistance in this ecosystem is most likely explained by inputs of allochthonous bacteria
364 associated with specific land cover as neither nutrient concentrations nor abundances of

365 specific bacterial taxa were strongly associated with resistance patterns observed. Further
366 investigation will be necessary to more discretely evaluate these associations and
367 potentially other processes driving patterns of heavy metal resistance in this ecosystem.
368
369 **4. Conclusions**
370 This study is among the first to apply a metagenomic functional screening
371 approach to assess the frequency of heavy metal resistance throughout a large riverine
372 ecosystem. Low resistance frequencies throughout the river were quantified in the
373 following order: $Mn^{2+} > Cr^{3+} > Zn^{2+} > Cd^{2+} > Hg^{2+}$, nearly inverse to the order of
374 toxicity reported elsewhere (Duxbury 1981; Hassen et al. 1998), and surprisingly no
375 resistance to Cu was observed. The frequencies of resistance to specific metals could be
376 related to land cover where Cd and Mn, and potentially Cr, were generally greater in
377 developed areas and Hg and Zn resistances were higher in forested areas. Nutrient
378 concentrations and specific orders of bacteria were poorly related to the frequency of
379 heavy metal resistance, with the exception that the abundance of *Pseudomonadales*
380 showed an association with Hg and Zn resistances. These findings suggest that in this
381 lotic ecosystem, heavy metal resistance is likely due, in part, to land-cover-associated
382 inputs of allochthonous bacteria.
383
384 **Acknowledgements**
385 Funding for this study was provided, in part, by the Minnesota Environment and Natural
386 Resources Trust Fund, as recommended by the Legislative-Citizen Commission on
387 Minnesota Resources (LCCMR). This work was completed using resources at the

388 University of Minnesota Supercomputing institute. We would also like to thank

389 Cassandra Barrett and Derek Schultz for helping to screen fosmid libraries.

References

Abou-Shanab R, van Berkum P, Angle JS. Heavy metal resistance and genotypic analysis of metal resistance genes in gram-positive and gram-negative bacteria present in Ni-rich serpentine soil and in the rhizosphere of Alyssum murale. Chemosphere 2007;68(2):360–7.

Ahluwalia SS, Goyal D. Microbial and plant derived biomass for removal of heavy metals from wastewater. Bioresour Technol 2007;98(12):2243–57.

Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev. 1995;59(1):143–69.

Amos GC a, Zhang L, Hawkey PM, Gaze WH, Wellington EM. Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. Vet Microbiol 2014;171(3-4):441–7.

Ancion P-Y, Lear G, Lewis GD. Three common metal contaminants of urban runoff (Zn, Cu & Pb) accumulate in freshwater biofilm and modify embedded bacterial communities. Environ Pollut 2010;158(8):2738–45.

Ansari MI, Malik A. Biosorption of nickel and cadmium by metal resistant bacterial isolates from agricultural soil irrigated with industrial wastewater. Bioresour Technol 2007;98(16):3149–53.

Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T. Bacterial community assembly based on functional genes rather than species. Proc Natl Acad Sci U S A 2011;108(34):14288–93.

Burridge L, Weis JS, Cabello F, Pizarro J, Bostick K. Chemical use in salmon aquaculture: a review of current practices and possible environmental effects. Aquaculture 2010;306(1-4):7–23.

Chapman JS. Disinfectant resistance mechanisms, cross-resistance, and co-resistance. Int Biodeterior Biodegradation 2003;51(4):271–6.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 2009;37:D141–D145.

Comte J, del Giorgio PA. Links between resources, C metabolism and the major components of bacterioplankton community structure across a range of freshwater ecosystems. Environ Microbiol 2009;11(7):1704–16.

Duxbury T. Toxicity of heavy metals to soil bacteria. FEMS Microbiol Lett 1981;11(2-3):217–20.

Ellis RJ, Morgan P, Weightman AJ, Fry JC. Cultivation-dependent and -independent approaches for determining bacterial diversity in heavy-metal-contaminated soil. Appl Environ Microbiol 2003;69(6):3223–30.

Fry J, Xian G, Jin S, Dewitz J, Homer C, Yang L, et al. Completion of the 2006 National Land Cover Database for the conterminous United States. Photogramm Eng Remote Sens 2011;77(9):858–64.

Gans J, Wolinsky M, Dunbar J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science 2005:1387–90.

Gibbons SM, Jones E, Bearquiver A, Blackwolf F, Roundstone W, Scott N, et al. Human and environmental impacts on river sediment microbial communities. PLoS One 2014;9(5):e97435.

Goolsby DA, Thurman EM, Pomes ML, Meyer MT, Battaglin WA. Herbicides and their metabolites in rainfall: Origin, transport, and deposition patterns across the midwestern and northeastern United States, 1990-1991. Environ Sci Technol 1997;31(5):1325–33.

Han FX, Kingery WL, Selim HM, Gerard PD. Accumulation of heavy metals in a long-term poultry waste-amended soil. Soil Sci 2000;165(3):260–8.

Hassen A, Saidi N, Cherif M, Boudabous A. Resistance of environmental bacteria to heavy metals. Bioresour Technol 1998;64:7–15.

Hemme CL, Deng Y, Gentry TJ, Fields MW, Wu L, Barua S, et al. Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. ISME J 2010;4(5):660–72.

Leung W, Wong M, Chua H, Lo W, Yu P, Leung C. Removal and recovery of heavy metals by bacteria isolated from activatedsludge treating industrial effluents and municipal wastewater. Water Sci Technol 2000;41:233–40.

Martínez A, Osburne MS. Preparation of fosmid libraries and functional metagenomic analysis of microbial community DNA. Methods Enzymol 2013;531:123–42.

Massieux B, Boivin MEY, Van Den Ende FP, Langenskiöld J, Marvan P, Barranguet C, et al. Analysis of structural and physiological profiles to assess the effects of Cu on biofilm microbial communities. Appl Environ Microbiol 2004;70(8):4512–21.

Mergeay M, Monchy S, Vallaeys T, Auquier V, Benotmane A, Bertin P, et al. Ralstonia metallidurans, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes. FEMS Microbiol Rev 2003;27(2-3):385–410.

Mizrahi-Man O, Davenport ER, Gilad Y. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. PLoS One 2013;8(1):e53608.

Nies DH. Microbial heavy-metal resistance. Appl Microbiol Biotechnol 1999;51(6):730–50.

Nies DH. Efflux-mediated heavy metal resistance in prokaryotes. FEMS Microbiol Rev 2003;27(2-3):313–39.

Pankhurst C, Hawke B, McDonald H, Kirkby C, Buckerfield J, Michelsen P, et al. Evaluation of soil biological properties as potential bioindicators of soil health. Aust J Exp Agric 1995;35(7):1015.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 2007;35(21):7188–96.

Sabry SAA, Ghozlan HAA, Abou-Zeid D-MM. Metal tolerance and antibiotic resistance patterns of a bacterial population isolated from sea water. J Appl Microbiol 1997;82(2):245–52.

Seiler C, Berendonk TU. Heavy metal driven co-selection of antibiotic resistance in soil and water bodies impacted by agriculture and aquaculture. Front Microbiol 2012;3:399.

Silver S, Phung L. Bacterial heavy metal resistance: new surprises. Annu Rev Microbiol 1996;50:753–89.

Smith VH, Schindler DW. Eutrophication science: where do we go from here? Trends Ecol Evol 2009;24(4):201–7.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere."Proc Natl Acad Sci U S A 2006;103(32):12115–20.

Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. Bacterial community structure is indicative of chemical inputs in the Upper Mississippi River. Front Microbiol. 2014a;in press.

Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. Core functional traits of bacterial communities in the Upper Mississippi River show limited variation in response to land cover. Front Microbiol 2014b;5:414.

Staley C, Unno T, Gould TJ, Jarvis B, Phillips J, Cotner JB, et al. Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. J Appl Microbiol 2013;115:1147–58.

Steffen MM, Li Z, Effler TC, Hauser LJ, Boyer GL, Wilhelm SW. Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. PLoS One 2012;7(8):e44002.

Wang S-Y, Sudduth EB, Wallenstein MD, Wright JP, Bernhardt ES. Watershed urbanization alters the composition and function of stream bacterial communities. PLoS One 2011;6(8):e22972.

Yakovchenko V, Sikora LJ, Kaufman DD. A biologically based indicator of soil quality. Biol Fertil Soils 1996;21(4):245–51.

Yan T, Hamilton MJ, Sadowsky MJ. High-throughput and quantitative procedure for determining sources of Escherichia coli in waterways by using host-specific DNA marker genes. Appl Environ Microbiol 2007;73(3):890–6.

Zhao Y, Wang Z, Sun W, Huang B, Shi X, Ji J. Spatial interrelations and multi-scale sources of soil heavy metal variability in a typical urban-rural transition area in Yangtze River Delta region of China. Geoderma 2010;156:216–27.

Table 1. Percentages of resistances to heavy metals observed in fosmid clone libraries.

Year	Site	No. Clones	Cd	Cr	Cu	Hg	Mn	Zn
2011	Itasca	10,368	0.32%	10.92%	0.00%	0.13%	9.39%	2.45%
2011	St. Cloud	9,984	0.00%	7.19%	0.00%	0.00%	1.36%	0.07%
2011	Clearwater	9,984	0.00%	3.55%	0.01%	0.02%	0.75%	0.02%
2011	Twin Cities	9,600	0.15%	5.70%	0.00%	0.04%	2.19%	0.01%
2011	MN River	9,984	0.10%	3.32%	0.00%	0.00%	3.53%	0.07%
2011	Confluence	9,600	0.00%	4.95%	0.00%	0.00%	7.18%	0.08%
2011	Hastings	9,984	0.21%	0.99%	0.00%	0.00%	14.92%	0.08%
2011	St. Croix River	9,216	0.05%	0.18%	0.00%	0.00%	8.16%	0.12%
2011	Red Wing	9,984	0.08%	0.03%	0.00%	0.01%	4.71%	0.18%
2011	La Crescent	9,984	0.03%	0.23%	0.00%	0.00%	7.70%	0.06%
2011	Zumbro River	9,984	0.12%	0.00%	0.00%	0.01%	9.44%	0.30%
	Avg.	9,879	0.10%	3.37%	0.00%	0.02%	6.30%	0.31%
	St. Dev.	302	0.10%	3.57%	0.00%	0.04%	4.27%	0.71%

Year	Site		Cd	Cr	Cu	Hg	Mn	Zn
2012	Itasca	9,984	0.03%	0.22%	0.00%	0.00%	5.49%	0.29%
2012	St. Cloud	9,984	0.52%	1.82%	0.00%	0.04%	5.45%	0.06%
2012	Clearwater	9,984	0.06%	0.17%	0.00%	0.01%	13.27%	0.09%
2012	Twin Cities	9,984	0.37%	0.66%	0.00%	0.00%	8.21%	0.08%
2012	MN River	9,984	0.27%	0.37%	0.00%	0.00%	7.02%	0.12%
2012	Confluence	9,984	0.45%	0.64%	0.00%	0.00%	5.83%	0.02%
2012	Hastings	9,984	0.73%	0.88%	0.00%	0.01%	1.16%	0.04%
2012	St. Croix River	9,600	0.01%	0.79%	0.00%	0.00%	4.15%	0.01%
2012	Red Wing	9,984	0.07%	1.64%	0.00%	0.00%	2.25%	0.09%
2012	La Crescent	9,984	0.01%	0.61%	0.00%	0.00%	2.53%	0.04%
2012	Zumbro River	9,600	0.19%	0.28%	0.00%	0.00%	3.08%	0.02%
	Avg	9,914	0.25%	0.74%	0.00%	0.01%	5.31%	0.08%
	St. Dev.	155	0.24%	0.55%	0.00%	0.01%	3.40%	0.08%

Table 2. Mean frequency of metal resistance at sites grouped by major land cover type.

Land Cover	**Cd**	**Cr**	**Hg**	**Mn**	**Zn**
Agricultural	$0.12 \pm 0.15^{a,b}$	1.61 ± 2.15^{a}	0.01 ± 0.01^{a}	5.26 ± 3.66^{a}	0.10 ± 0.08^{a}
Forested	0.14 ± 0.14^{b}	2.91 ± 5.34^{a}	0.03 ± 0.06^{b}	6.42 ± 2.74^{a}	0.71 ± 1.17^{b}
Developed	0.32 ± 0.26^{a}	2.30 ± 2.35^{a}	0.01 ± 0.02^{a}	6.58 ± 4.94^{a}	0.05 ± 0.03^{a}

[a,b]Resistance frequencies sharing the same superscript are not statistically different via ANOVA analysis ($\alpha = 0.05$).

Table 3. Spearman correlation coefficients relating nutrient/chemical concentrations to the frequency of heavy metal resistance observed. *P* values are shown in the parentheses and significant correlations are bolded.

Analyte	Year	Cd	Cr	Hg	Mn	Zn
Ammonium	2011*	0.123 (0.494)	**-0.460 (0.007)**†	0.103 (0.569)	**0.536 (0.001)**†	**0.358 (0.041)**†
	2012	0.277 (0.119)	0.147 (0.415)	0.091 (0.616)	-0.178 (0.665)	-0.030 (0.870)
Carbon	2011	-0.034 (0.850)	**0.384 (0.028)**†	0.100 (0.580)	**-0.346 (0.048)**†	**-0.368 (0.035)**§
	2012*	0.226 (0.206)	**0.442 (0.010)**†	**0.669 (< 0.001)**‡	0.251 (0.160)	-0.188 (0.295)
Nitrite/nitrate	2011	-0.005 (0.978)	**-0.562 (0.001)**§	**-0.500 (0.003)**‡	0.219 (0.221)	0.104 (0.563)
	2012	0.214 (0.231)	-0.091 (0.615)	**-0.534 (0.001)**‡	-0.282 (0.112)	-0.050 (0.781)
Orthophosphate	2011	0.018 (0.921)	-0.149 (0.407)	-0.211 (0.239)	0.106 (0.558)	0.073 (0.685)
	2012	0.090 (0.619)	-0.184 (0.306)	**-0.558 (0.001)**‡	-0.207 (0.248)	-0.185 (0.302)
Al	2011	**-0.619 (< 0.001)**†	-0.227 (0.240)	**-0.452 (0.008)**‡	**-0.400 (0.021)**†	-0.002 (0.989)
	2012*	**0.460 (0.007)**†	-0.191 (0.287)	0.035 (0.847)	**0.345 (0.049)**†	-0.064 (0.723)
Cu	2011	**-0.568 (0.001)**‡	0.066 (0.714)	-0.232 (0.193)	**-0.360 (0.040)**†	-0.335 (0.057)
	2012	-0.301 (0.089)	**-0.400 (0.021)**‡	-0.191 (0.286)	0.100 (0.580)	**0.503 (0.003)**‡
Mn	2011	-0.211 (0.239)	0.331 (0.060)	-0.105 (0.560)	**-0.431 (0.012)**†	-0.314 (0.076)
	2012	0.201 (0.261)	-0.106 (0.556)	0.248 (0.165)	**0.577 (< 0.001)**†	0.012 (0.949)

*Higher analyte concentrations were observed in the underlined year of study via ANOVA ($\alpha = 0.05$). Where no year is underlined, the difference in concentration was not significant.

†Unknown or insignificant intercorrelation between analyte concentration and land cover type as well as heavy metal resistance frequency and land cover type.

‡Positive intercorrelation between analyte concentration and land cover type as well as heavy metal resistance frequency and land cover type.

§Inverse intercorrelation between analyte concentration and land cover type as well as heavy metal resistance frequency and land cover type.

Table 4. Spearman correlations of the relative abundance of orders with metal resistances by year. *P* values are shown in parentheses and significant values are shown in bold.

Order	Year	Cd	Cr	Hg	Mn	Zn
Actinomycetales	2011	**-0.360 (0.040)**	0.020 (0.912)	-0.277 (0.119)	**-0.389 (0.025)**	-0.277 (0.119)
	2012	-0.316 (0.073)	0.034 (0.850)	-0.137 (0.448)	-0.296 (0.095)	-0.185 (0.303)
Aeromondales	2011	-0.185 (0.302)	-0.154 (0.393)	-0.255 (0.152)	-0.265 (0.136)	-0.255 (0.152)
	2012	0.081 (0.654)	**0.361 (0.039)**	-0.095 (0.598)	0.056 (0.0757)	-0.063 (0.729)
Bacillales	2011	-0.044 (0.806)	**-0.467 (0.006)**	**-0.404 (0.020)**	0.066 (0.716)	**-0.404 (0.020)**
	2012	0.235 (0.188)	0.112 (0.535)	-0.113 (0.532)	-0.007 (0.971)	0.181 (0.314)
Bacteroidales	2011	**-0.411 (0.017)**	-0.170 (0.344)	-0.261 (0.143)	**-0.412 (0.017)**	-0.261 (0.143)
	2012	0.155 (0.390)	0.297 (0.093)	**0.509 (0.002)**	-0.182 (0.311)	0.127 (0.480)
Burkholderiales	2011	0.146 (0.419)	-0.271 (0.127)	-0.221 (0.216)	0.218 (0.223)	-0.221 (0.216)
	2012	-0.023 (0.900)	-0.105 (0.560)	0.213 (0.235)	-0.108 (0.551)	0.045 (0.805)
Chromatiales	2011	**-0.516 (0.002)**	0.097 (0.590)	-0.299 (0.091)	**-0.524 (0.002)**	-0.299 (0.091)
	2012	0.026 (0.886)	0.194 (0.280)	0.261 (0.143)	-0.048 (0.789)	-0.098 (0.587)
Cyanobacteria (unclassified)	2011	-0.091 (0.615)	**-0.661 (< 0.001)**	**-0.498 (0.003)**	0.212 (0.237)	**-0.498 (0.003)**
	2012	-0.192 (0.285)	0.331 (0.060)	**-0.400 (0.021)**	-0.220 (0.219)	-0.171 (0.343)
Gammaproteobacteria[*]	2011	-0.038 (0.833)	-0.118 (0.511)	**-0.498 (0.003)**	-0.203 (0.258)	**-0.498 (0.003)**
	2012	0.275 (0.121)	**0.432 (0.012)**	**0.507 (0.003)**	0.028 (0.876)	-0.075 (0.676)
Legionellales	2011	-0.075 (0.679)	**-0.464 (0.007)**	-0.116 (0.519)	0.150 (0.405)	-0.116 (0.519)
	2012	-0.249 (0.163)	**0.351 (0.045)**	-0.149 (0.407)	**-0.367 (0.036)**	-0.147 (0.414)
Methylophilales	2011	**-0.690 (< 0.001)**	-0.037 (0.840)	**-0.498 (0.003)**	**-0.678 (< 0.001)**	**-0.498 (0.003)**
	2012	-0.122 (0.499)	**0.508 (0.003)**	0.230 (0.198)	**-0.501 (0.003)**	**-0.443 (0.010)**
Neisseriales	2011	**-0.439 (0.011)**	-0.162 (0.369)	**-0.465 (0.006)**	-0.260 (0.144)	**-0.465 (0.006)**
	2012	**0.458 (0.007)**	0.080 (0.659)	**0.369 (0.035)**	**0.504 (0.003)**	0.056 (0.755)
Nitrosomonadales	2011	-0.243 (0.173)	-0.021 (0.908)	-0.211 (0.238)	-0.130 (0.472)	-0.211 (0.238)
	2012	**0.391 (0.025)**	**0.446 (0.009)**	0.284 (0.110)	**-0.376 (0.031)**	-0.197 (0.271)
Prolixibacter	2011	-0.005 (0.979)	0.079 (0.661)	0.199 (0.266)	-0.239 (0.179)	0.199 (0.266)
	2012	**-0.448 (0.009)**	0.244 (0.171)	-0.060 (0.740)	**-0.418 (0.015)**	0.110 (0.544)

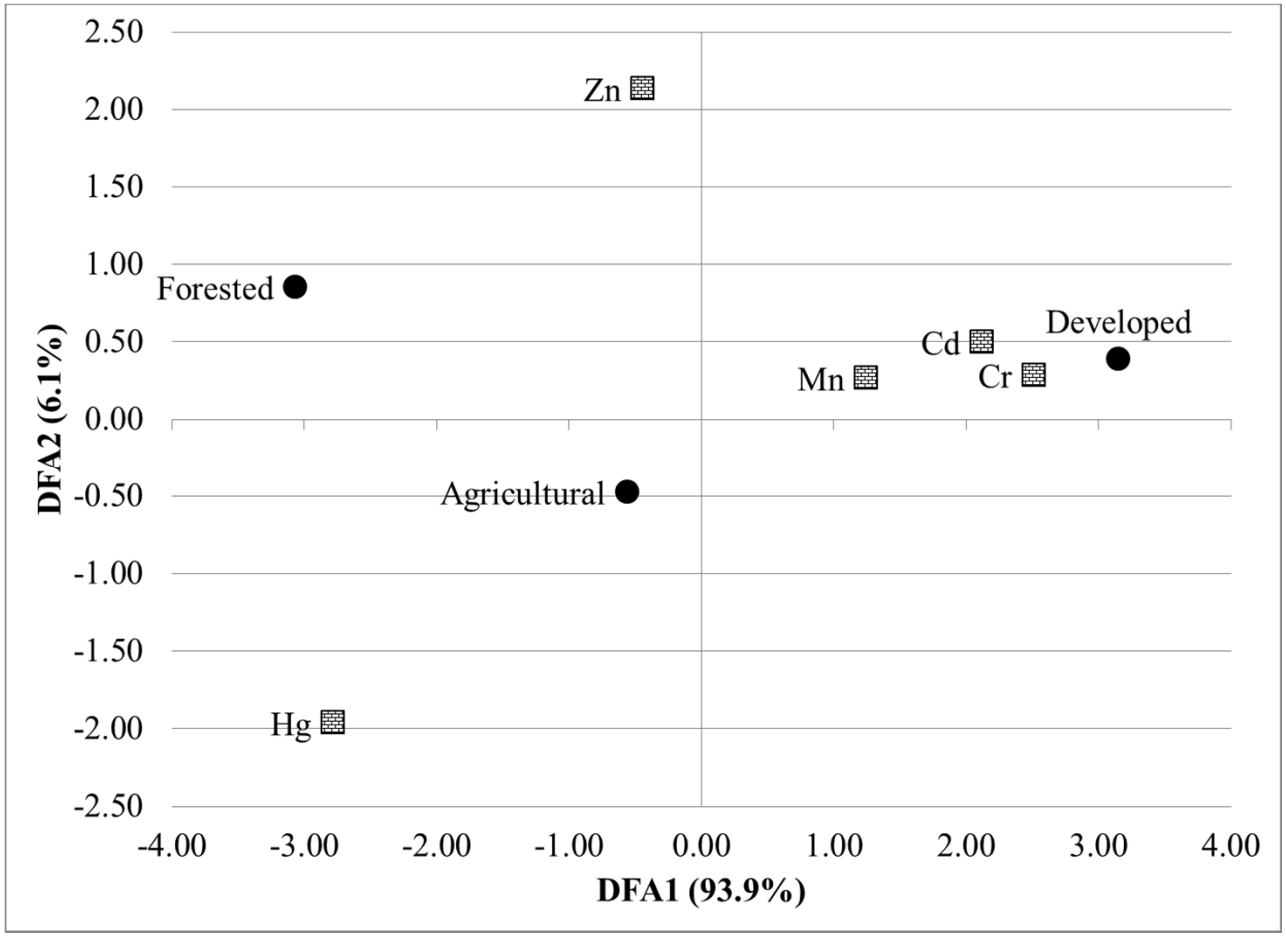
Pseudomonadales	2011	**0.437 (0.011)**	0.238 (0.183)	**0.476 (0.005)**	**0.463 (0.007)**	**0.476 (0.005)**
	2012	0.172 (0.337)	0.142 (0.431)	**-0.469 (0.006)**	0.048 (0.789)	0.026 (0.886)
Rhizobiales	2011	-0.321 (0.069)	**-0.408 (0.019)**	**-0.471 (0.006)**	-0.294 (0.096)	**-0.471 (0.006)**
	2012	0.093 (0.607)	**0.496 (0.003)**	**0.441 (0.010)**	0.031 (0.865)	-0.301 (0.089)
Rhodobacterales	2011	0.140 (0.436)	-0.279 (0.116)	**-0.443 (0.010)**	-0.078 (0.667)	**-0.443 (0.010)**
	2012	0.138 (0.445)	0.157 (0.383)	**0.368 (0.035)**	0.154 (0.392)	**0.480 (0.005)**
Rhodocyclales	2011	.028 (0.879)	**-0.358 (0.041)**	-0.332 (0.059)	0.123 (0.497)	-0.332 (0.059)
	2012	**0.365 (0.036)**	-0.263 (0.140)	-0.139 (0.439)	0.295 (0.096)	0.002 (0.991)
Rhodospirillales	2011	**-0.707 (0.000)**	**0.370 (0.034)**	-0.299 (0.091)	**-0.845 (0.000)**	-0.299 (0.091)
	2012	0.047 (0.795)	**0.509 (0.002)**	0.215 (0.231)	**-0.376 (0.031)**	**-0.478 (0.005)**
Shingobaccteriales	2011	**-0.377 (0.031)**	**-0.397 (0.022)**	**-0.489 (0.003)**	-0.326 (0.064)	**-0.498 (0.003)**
	2012	**-0.373 (0.032)**	0.288 (0.104)	0.140 (0.437)	-0.220 (0.218)	-0.045 (0.803)
Synergistales	2011	**-0.448 (0.009)**	-0.152 (0.397)	-0.299 (0.091)	**-0.531 (0.001)**	-0.299 (0.091)
	2012	**-0.621 (< 0.001)**	0.129 (0.474)	0.042 (0.817)	-0.332 (0.059)	0.191 (0.288)
Verrucomicrobia Subdivision 3[*]	2011	**-0.523 (0.002)**	0.141 (0.434)	**-0.432 (0.012)**	**-0.581 (< 0.001)**	**-0.432 (0.012)**
	2012	0.212 (0.235)	**0.480 (0.005)**	**0.525 (0.002)**	-0.084 (0.642)	-0.325 (0.065)
Xanthomonadales	2011	0.171 (0.342)	0.027 (0.881)	**-0.354 (0.043)**	0.036 (0.841)	**-0.354 (0.043)**
	2012	**0.394 (0.023)**	-0.051 (0.778)	-0.196 (0.274)	0.260 (0.144)	0.020 (0.913)

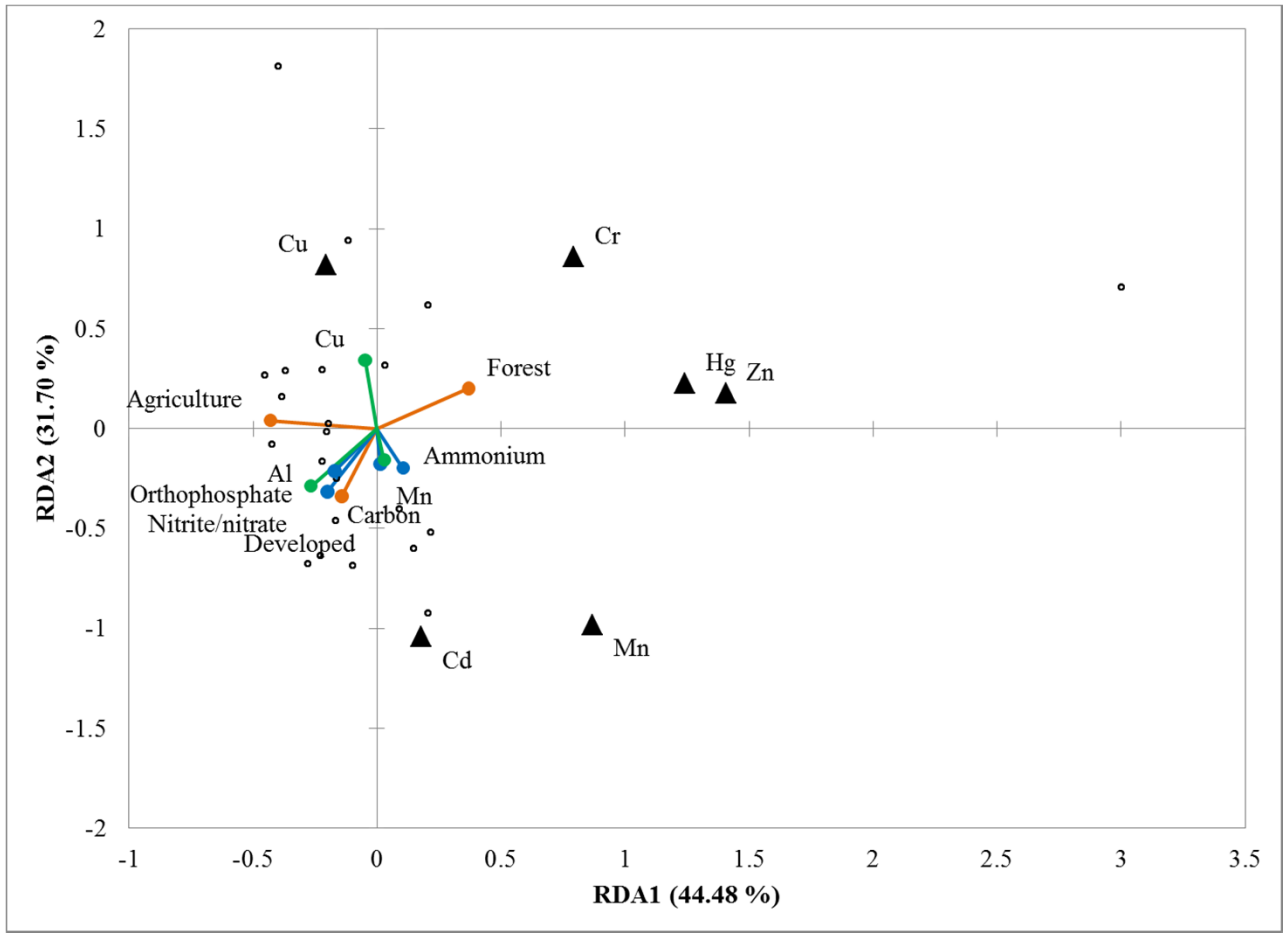
[*]Orders are designated *incertae sedis*.

Figure legends

Figure 1. Discriminant function analysis plot relating heavy metal resistance frequencies to major land cover types.

Figure 2. Redundancy analysis relating nutrient (blue) and metal ion (green) concentrations, percentage of basin land cover (orange), and frequencies of resistance to heavy metals (triangles). Observations are shown as open circles.


2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
DFA2 (6.1%)
-4.00
-3.00
-2.00
-1.00
0.00
1.00
2.00
3.00
4.00
DFA1 (93.9%)
Zn
Forested
Developed
Cd
Mn
Cr
Agricultural
Hg


RDA2 (31.70 %)
RDA1 (44.48 %)
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
Cu
Cr
Cu
Forest
Hg
Zn
Agriculture
Ammonium
Al
Orthophosphate
Mn
Carbon
Nitrite/nitrate
Developed
Cd
Mn

Bacterial Community Structure is Indicative of Chemical Inputs in the Upper Mississippi River

Christopher Staley[1], Trevor J. Gould[1,2], Ping Wang[1], Jane Phillips[2], James B. Cotner[3], and Michael J. Sadowsky[1,4,#]

[1]BioTechnology Institute, [2]Biology Program, [3]Department of Ecology, Evolution, and Behavior, and [4]Department of Soil, Water and Climate, University of Minnesota, St. Paul, MN

Running title: Modeling of the Mississippi River bacterial community

[#]Corresponding Author: Michael J. Sadowsky, BioTechnology Institute, University of Minnesota, 140 Gortner Lab, 1479 Gortner Ave, St.Paul, MN 55108; Phone: (612)-624-2706, Email: sadowsky@umn.edu

Abstract

Local and regional associations between bacterial communities and nutrient and chemical concentrations were assessed in the Upper Mississippi River in Minnesota to determine if community structure was associated with discrete types of chemical inputs associated with different land cover. Bacterial communities were characterized by Illumina sequencing of the V6 region of 16S rDNA and compared to > 40 chemical and nutrient concentrations. Local bacterial community structure was shaped primarily by associations among bacterial orders. However, order abundances were correlated regionally with nutrient and chemical concentrations, and were also related to major land coverage types. Total organic carbon and total dissolved solids were among the primary abiotic factors associated with local community composition and co-varied with land cover. *Escherichia coli* concentration was poorly related to community composition or nutrient concentrations. Abundances of fourteen bacterial orders were related to land coverage type, and seven showed significant differences in abundance ($P \leq 0.046$) between forested or anthropogenically-impacted sites. This study identifies specific bacterial orders that were associated with chemicals and nutrients derived from specific land cover types and may be useful in assessing water quality. Results of this study reveal the need to investigate community dynamics at both the local and regional scales and to identify shifts in taxonomic community structure that may be useful in determining sources of pollution in the Upper Mississippi River.

Keywords: diversity / ecology / environmental/recreational water / metagenomics / water quality

Introduction

The Mississippi River is an important natural resource that is used as a source for drinking water by many cities, as well as for recreational, agricultural, and industrial purposes. The Upper Mississippi River Basin (UMRB), however, has a well-documented history of contamination by heavy metals, including mercury and lead that peaked in the 1960s (Balogh et al., 1999, 2009; Wiener and Sandheinrich, 2010). The UMRB is also impacted by chemicals, including the pesticide DDT, herbicides, and polychlorinated biphenyls (Pereira and Hostettler, 1993; Wiener and Sandheinrich, 2010). In addition, nutrient loading, especially nitrogen and phosphorus, primarily from agricultural runoff (Schilling et al., 2010), has been a persistent concern, with estimates that the UMRB contributed as much as 43 and 26% of total nitrogen and phosphorus, respectively, to the northern Gulf of Mexico from 2001 to 2005 (United States Environmental Protection Agency, 2007). Furthermore, effluent from urban and suburban wastewater treatment plants has been shown to increase organic and inorganic nutrient content downstream in temperate river systems and to decrease bacterial diversity (Drury et al., 2013). While the input of many pollutants to the UMRB has declined since passage of the Clean Water Act and other regulatory measures in the 1970s (Balogh et al., 2009; Wiener and Sandheinrich, 2010), non-point sources of pollution remain a contemporary concern (Wiener and Sandheinrich, 2010).

In addition to long-recognized nutrient contamination of the UMRB, several other pollutants, including fecal contamination, traditionally measured by using fecal indicator bacteria (*Escherichia coli* and *Enterococcus*) (United States Environmental Protection Agency, 2012, 2002), antimicrobials, other pharmaceuticals, and personal care product pollutants have potential negative impacts on the river's water quality. Recently, reaches of the Mississippi River between the Coon Rapids Dam and Minneapolis, as well as a section near Saint Paul, MN have been deemed impaired due to elevated concentrations of *E. coli* (Russell and Weller, 2012). Furthermore, pharmaceuticals such as acetaminophen and caffeine have also been detected in the Mississippi River and are likely contributed from wastewater discharge and runoff (Ellis, 2006; Zhang et al., 2007). While elevated nutrient and bacterial concentrations, and the presence of potentially toxic xenobiotic compounds, suggest negative impacts on ecosystem processes, they provide little indication as to the source(s) of contamination, and little is known about resultant effects on bacterial community structure and ecosystem functioning as a result of their presence in the UMRB.

Many recent studies evaluating variation in microbial community structure in response to environmental gradients have relied on indirect measures and the use of statistical methods, such as non-metric multidimensional scaling, Mantel tests, and Spearman rank correlations, among others (Fortunato et al., 2012; Brandsma et al., 2013; Portillo et al., 2012; Staley et al., 2013). While these methods provide valuable information, such as how bacterial communities vary in response to changes in salinity and depth on a regional scale (Fortunato et al., 2012), local and temporal differences in microbial assemblages can be difficult to determine due to considerable data reduction using these methods such that an entire study period or study area is typically analyzed and relationships on a local scale may be overlooked. Furthermore, spatial alteration in community structure is difficult to disentangle from changes due to environmental gradients,

especially when multiple sources of variation co-vary (Fortunato and Crump, 2011; Fortunato et al., 2012). Local similarity analysis (LSA) has been proposed as a method to explain complex, non-linear relationships between microbial assemblages and co-varying environmental parameters on a local scale that can be displayed in a graphical format (Ruan et al., 2006). Extended LSA (eLSA) is a recent expansion of this analysis to incorporate replicate sample data (Xia et al., 2011), and this method has been subsequently used to characterize seasonal relationships within and between microbial taxa and environmental factors in the English Channel (Gilbert et al., 2012).

In addition to being able to better evaluate complex inter- and intra-community responses to environmental variables in high-throughput DNA sequencing datasets, studies of river water have suggested that the microbial communities may change predictably in response to specific types of anthropogenic inputs or contaminant sources (Unno et al., 2010; Staley et al., 2013). Recent development of several analytical software packages has allowed for relatively rapid assignment of operational taxonomic units (OTUs) to specific sources (Knights et al., 2011; Unno et al., 2012), and these have been used in the study of surface water to identify human-specific fecal contamination (Unno et al., 2012; Newton et al., 2013). These studies suggest that evaluation of the entire bacterial community composition may be a useful water quality monitoring tool, potentially providing accurate assessment of the magnitude and distribution of contamination from a variety of sources. In addition to inputs of specific non-indigenous taxa from allochthonous sources, shifts in bacterial community structure may also be associated with the introduction or increase in concentration of specific chemical contaminants, including pharmaceuticals or heavy metals.

In the current study, we characterized concentrations of nutrients and xenobiotic compounds present in the water column of the UMRB in Minnesota during the summers of 2011 and 2012 in order to determine how chemical inputs to the river influenced bacterial community structure. We hypothesized that associations of bacterial orders with nutrient and chemical concentrations would vary both locally and regionally in response to different sources of contamination. To evaluate local associations, all environmental parameters were modeled into an association network with abundances of bacterial orders, and a predictive Bayesian network was developed to display significant, directional associations over a regional scale. Regional bacterial community structure was further hypothesized to be related to specific land-coverage types that are likely to be influencing chemical gradients throughout the river. While previous studies have identified specific taxa associated with human fecal contamination and wastewater discharge (Newton et al., 2013; Drury et al., 2013), the wide variety of analytes measured here, as well as the large study area, offered a more thorough evaluation of how land-use practices and nutrient concentrations might co-vary, and how bacterial communities vary in response to anthropogenically-impacted chemical contributions both locally and regionally. Furthermore, the ability to associate specific bacterial taxa with discrete types of contamination will prove useful in future water quality monitoring, the adoption of best management practices, and development of total maximum daily load (TMDLs).

Materials and methods

Sampling collection and processing

Samples were collected during early summer (May through July) in 2011 and 2012 from 11 sites along the Upper Mississippi River and major contributing rivers near their confluences with the Mississippi River as previously described (Staley et al., 2013). Locations of sampling sites are listed beginning at the headwaters near Lake Itasca to the southern border of Minnesota (near La Crescent; Table 1). Water samples (40 L) were collected in sterile 20 L carboys, transported to the lab, and either processed immediately or stored at 15° C overnight and processed the following day (Staley et al., 2013). Briefly, samples were strained through sterile cheesecloth and sequentially filtered (5 – 10 µm) to remove debris and aggregate bacteria. A subset of planktonic bacterial cells were subsequently captured on a 0.45-µm-pore-sized polyethane sulfonate filters. The effects of filtration and filter pore-size on bacterial community characterization have been previously explored in our laboratory (Staley et al., 2013), and the 0.45-µm pore size was selected to allow the most efficient filtration of large volumes of water. However, while the 0.45 µm pore size allows efficient filtration of large volumes of water, only larger, free-living planktonic bacteria can be characterized by this method. Cells were elutriated by vortexing filters in pyrophosphate buffer, pH 7.0, and six cell pellets representing approximately 6 L each of water were stored at -80° C.

DNA extraction and sequencing

DNA was extracted from two replicate cell pellets per site using the DNA Isolation Kit for Water (Epicentre, Madison, WI). The V6 hypervariable region of the 16S rDNA was amplified using the 967F/1046R barcoded primer set (Sogin et al., 2006), and amplicons were gel-purified using the QiaQuick® Gel Extraction Kit (Qiagen, Valencia, CA). Replicate sequence data was generated by paired-end sequencing of purified amplicon pools on the Illumina MiSeq (DNA from one cell pellet, 2 × 150 bp read length) and HiSeq2000 (DNA from the second cell pellet, amplicons sequenced in duplicate at 2 × 100 bp). Duplicate sequencing was performed because a third cell pellet was not available for additional DNA extraction resulting from use of pellets in other experiments not described here. Interpretations of sequence data have been previously shown to be reproducible across platforms (Caporaso et al., 2012), and use of the HiSeq2000 enabled more efficient multiplexing of samples. Sequencing was performed by the University of Minnesota Genomics Center. Sequences were deposited to the GenBank Sequence Read Archive under accession number SRP018728.

Metadata collection

Physicochemical parameters, including temperature, pH, and rainfall up to 72 h prior to sampling was collected at the time of sampling. In addition, two additional 1 L water samples were collected in sterile amber bottles and stored at 4° C for further chemical analyses. *Escherichia coli* concentration was determined by membrane filtration and plating on mTEC agar using standard methods (United States Environmental Protection Agency, 2002) and data are expressed as colony-forming units (CFU) per 100 ml. Concentrations of ammonium (NH_4), nitrite/nitrate

(NO_2/NO_3), total phosphorus (total-P), orthophosphate, total organic carbon (carbon), and total dissolved solids (TDS) were measured by the Research Analytical Laboratories (RAL; University of Minnesota, St. Paul, MN). In addition, inductively coupled plasma mass spectrometry (ICP) analysis was used to quantify ion concentrations, and quantification of concentrations of various xenobiotic compounds, classified as antibiotics, endocrine disruptors, pharmaceuticals, personal care products, and pesticides was performed by the RAL using liquid chromatography-mass spectrometry. Analytes were exhaustively selected based on the capability of the RAL, and xenobiotic compounds measured were determined by RAL capabilities. All physical parameters and analytes measured are shown in Table 2.

Land cover data was extrapolated from the National Land Cover Database (NLCD 2006) (Fry et al., 2011) by overlaying a map of hydrologic unit code (HUC) boundaries (1:250,000 scale) using ArcGIS (Esri, Redlands, CA) and expressed as a percentage of the total area of the HUC boundary (Table 1). Maps were obtained from the United States Geological Survey [http://water.usgs.gov/ maps.html]. NLCD codes for similar land cover types (*e.g.* 'developed, low' and 'developed, med') were summed in order to evaluate the influence of major land cover types. Major land cover categories investigated were "developed" (urban anthropogenic impacts), "forested" (unimpacted by anthropogenic activity), and "agricultural" (agricultural anthropogenic impacts). Agricultural land throughout this manuscript refers to the sum of pastureland and cultivated (crop) land, while "pasture" specifically references pastureland alone.

Sequence processing

All sequence processing was performed using mothur software ver. 1.29.2 and 1.32.0 (Schloss et al., 2009). Sequences were trimmed to 100 nt, paired-end aligned using fastq-join (Aronesty, 2013), and screened for quality. Sequences that had a quality score < 35 over a window of 50 nt, had a mismatch to a primer or barcode sequence, had homopolymers > 8 nt, or had an ambiguous base (N) were excluded from analysis. Singleton sequences were removed in mothur and chimeras were removed using UCHIME (Edgar et al., 2011). The number of sequence reads in each sample was normalized by random subsampling to 25,703 sequence reads per sample. Sequences were aligned against the SILVA database ver. 102 (Pruesse et al., 2007), OTUs were clustered using the furthest-neighbor algorithm at 97% similarity, and OTUs were classified against the Ribosomal Database Project ver. 9 database (Cole et al., 2009).

Statistical analyses

All diversity calculations, ordination plots, and community comparisons were performed using mothur (Schloss et al., 2009) and Bray-Curtis dissimilarity matrices (Bray and Curtis, 1957). For all analyses, unless otherwise stated, replicates were maintained as separate samples and grouped by using various .design files. Three analyses were used to evaluate differences in community structure among sampling sites: i) Beta-diversity differences between sites were determined using UniFrac metrics (Lozupone and Knight, 2005), which take into account raw phylogenetic differences between sets of taxa (unweighted) or abundance-weighted phylogenetic differences

(weighted); ii) analysis of similarity (ANOSIM) (Clarke, 1993), in which rank order differences in community structure are evaluated from a dissimilarity matrix was performed, and iii) analysis of molecular variance (AMOVA) (Excoffier et al., 1992), which is similar to a non-parametric analysis of variance (ANOVA) utilizing dissimilarity matrices was also performed. All community-level analyses were also done using mothur and statistics were calculated using 1000 iterations. Spearman correlations, ANOVA, generalized linear modeling, and discriminant function analysis (DFA) were performed using SPSS Statistics software ver. 19.0 (IBM, Armonk, NY). All statistics were evaluated at $\alpha = 0.05$.

Network modeling

For network modeling, bacteria were classified to orders, as this level of taxonomic resolution has been used previously for interrogating associations among bacterial communities (Gilbert et al., 2012), and the accuracy of more specific classification (*e.g.* to family or genus) may be unreliable with short sequence reads used here (Mizrahi-Man et al., 2013). Local similarity analyses were performed using the eLSA software package (Xia et al., 2011) with no time-delay [-d 0], normalization to both percentile and Z-score [-n percentileZ] and replicates were averaged (default setting). A total of 191 parameters were included in the LSA model; parameters that were not detected during both years of study were excluded. LSA results were visualized using Cytoscape ver 3.0.2 (Shannon et al., 2003).

For Bayesian network inference, environmental parameters were normalized as described previously (Larsen et al., 2012) using the equation $Env\_norm_i^j = \frac{\left(MAX(Env^j) - Env_i^j\right)}{\left(MAX(Env^j) - MIN(Env^j)\right)} \times 99 + 1$. Environmental parameters were incorporated into a single input matrix with the 15 most abundant bacterial orders, and for simplicity, reported as the mean percentages of relative abundance among triplicates (a number between 0 and 100). Relationships were inferred using the Bayesian Network Inference with Java Objects (BANJO) ver. 2.2.0 (Smith et al., 2006), and settings used were similar to those previously described (Larsen et al., 2012). Networks were considered using simulated annealing and the All Local Moves proposer, with a maximum of five parents. Only environmental parameters were considered as parents. Only significant associations (at $\alpha = 0.05$) were incorporated into the final consensus network of highest-scoring networks inferred and were visualized using Cytoscape software.

Results

Bacterial community composition

A mean of 1,450 ± 266 OTUs were identified among all samples and could be classified to 153 orders. An average of 0.02 ± 0.02% of sequence reads could not be classified to an order among all replicate samples. Differences in alpha diversity were not significantly different between years (mean Shannon index 4.02 ± 0.50, $P = 0.128$). The bacterial community composition of all samples was predominantly comprised of the orders were *Burkholderiales* (54.0% mean

abundance), *Actinomycetales* (10.1%), *Pseudomonadales* (8.3%), *Sphingobacteriales* (3.4%), *Methylophilales* (3.1%), *Rhodocyclales* (2.4%), and *Rhodospirillales* (2.0%; Figure 1). All other orders accounted for a mean of < 2.0% of sequence reads. At two sites sampled in either year (Itasca and St. Cloud in 2011 and Twin Cities and Minnesota River in 2012), a majority of sequence reads classified to *Pseudomonadales,* and this result was consistent among replicate sequence data at these sites. While the dominant orders were generally consistent among samples, bacterial communities differed significantly between samples collected in 2011 and 2012 (unweighted and weighted UniFrac $P < 0.001$, ANOSIM $P = 0.007$, AMOVA $P = 0.007$; Figure 2).

Local similarity analysis

Local similarity analysis revealed that the relative abundances of bacterial orders were generally more significantly intercorrelated amongst themselves, than with nutrient or chemical concentrations (Figure 3). Importantly, *E. coli* concentration was not significantly associated with other analyte concentrations or abundances of specific orders. Specific and stronger associations (-0.7 > Spearman's r > 0.7) among all parameters and bacterial orders included in LSA are shown in Supplementary Figures S1 and S2.

Among the nutrient concentrations measured, only the amounts of organic carbon and TDS were significantly associated with order abundances. Organic carbon concentration was associated with the abundances of *Acidobacteria* Group 2, *Gemmatimonadales*, and *Sneathiellales* (Spearman's r = 0.640, 0.631, and 0.792), while TDS concentration was inversely correlated with abundance of *Acidobacteria* Group 2 (r = -0.709). Among ions measured, the concentration of potassium was significantly associated with abundances of *Puniceicoccales*, *Thiotrichales* and *Verrucomicrobia* Subdivision 5 (r = 0.624, 0.584 and 0.761).

Developed land use was also associated with potassium concentration (r = 0.730) as well as *Thiotrichales* and *Verrucomicrobia* Subdivision 5 abundances (r = 0.734 and 0.715). Magnesium concentration was negatively correlated with abundance of *Anaerolineales* (r = -0.556), and boron concentration was positively correlated with abundances of *Rhodobacterales* (r = 0.534) and several rare orders for which correlation coefficients could not be obtained. Several of these rare orders were also negatively associated with iron and zinc concentrations. Among the xenobiotic chemicals tested, only the concentration of acetochlor was significantly associated with *Acidobacteria* Group 1 (r = 0.766).

Association of physical and chemical parameters with community composition

Correlations among nutrient and chemical concentrations with community diversity and order abundances were also observed regionally via traditional correlation analyses. Similar to LSA, however, significant intercorrelations were also observed between analyte concentrations,

making the biological importance of associations with community composition difficult to interpret.

Temperature and cumulative, 3-day antecedent rainfall were significantly higher in 2011 than 2012 ($P = 0.041$ and < 0.001, respectively; Table 2). Among all the nutrients measured, organic carbon, ammonium, total phosphorous, and TDS were significantly greater in 2011 vs. 2012 ($P \leq 0.001$). Over both years of the study, temperature and organic carbon concentrations were positively correlated with diversity, as measured by the Shannon index (r = 0.292 and 0.369, $P = 0.018$ and 0.002, respectively) while pH, 48 h antecedent rainfall, and TDS concentration were inversely correlated (r = -0.325, -0.462, and -0.246, $P = 0.008$, <0.001, and 0.046, respectively). A generalized linear model was constructed to relate pH and nutrient concentrations to Shannon diversity. Temperature and rainfall were excluded as these were not associated with land cover. Concentrations of organic carbon and nitrite/nitrate were found to have significant main effects on Shannon diversity ($P = 0.009$ and 0.045, respectively), although the abundances of most of the orders were significantly correlated with the concentration of at least one of the nutrients measured ($P < 0.05$). Supplementary Figures S3 and S4 show a subset of these orders that also had relative abundances related to land coverage.

The concentrations of the majority of ions also differed significantly ($P < 0.05$) between years, except for Cr, Cu, K, and Mn. The concentrations of Al, Mn, and K were significantly correlated with richness measured as the number of OTUs identified (r = 0.294 – 0.310, $P \leq 0.016$), while the concentrations of Ca were negatively correlated to the Shannon diversity index (r = -0.246, $P = 0.047$). However, the concentration of N,N-Diethyl-meta-toluamide (the insecticide DEET) significantly increased richness (r = 0.283, $P = 0.021$), while the concentration of carbaryl was inversely correlated (r = -0.288, $P = 0.019$) to richness by traditional analysis.

Intercorrelation of water quality parameters

Among parameters traditionally used to evaluate water quality, traditional bivariate correlation analysis indicated that the concentration of *E. coli* was only significantly correlated with total phosphorus and TDS concentrations (r = 0.527 and 0.328, $P < 0.001$ and 0.007, respectively) and inversely correlated with organic carbon concentrations (r = -0.363, $P = 0.003$). Nitrite/nitrate, orthophosphate, and TDS concentrations were all positively correlated with each other (r = 0.665 – 0.850, $P < 0.001$), and negatively correlated with organic carbon concentration (r = -0.546 to -0.804, $P < 0.001$). Ammonium, total phosphorus, and TDS concentrations were also significantly positively correlated with each other (r = 0.261 – 0.353, $P \leq 0.035$). Negative correlations between ammonium or phosphorus concentrations and organic carbon concentrations were not significant at $\alpha = 0.05$.

Association of nutrient and chemical concentrations with land cover

To simplify the interpretation of intercorrelations among analytes and land cover, analyte concentrations were related to major land coverage categories observed (developed, forested, or pasture) by traditional bivariate correlation analysis (Table 3). Pastureland alone was evaluated because cultivated land was poorly correlated with the relative abundances of nearly every order, and a negative correlation between pastureland and *E. coli* concentration was significant ($r = -0.303$, $P = 0.013$). Generally, a greater percentage of developed land area was associated with higher pH and increased nitrite/nitrate, orthophosphate, and TDS concentrations, as well as concentrations of several ions. In contrast, the concentrations of these parameters tended to decrease with greater percentages of forested or pasture land within the HUC boundary (Table 3). A greater percentage of pastureland within a hydrologic unit was also well correlated with increased organic carbon concentration ($r = 0.695$, $P < 0.001$). Concentrations of some pesticides, endocrine disrupters, and personal care products were also negatively correlated with the percentage of forested area within a hydrologic unit ($r = -0.279$ to -0.375, $P \leq 0.033$), while acetochlor concentration was positively correlated with developed area ($r = 0.421$, $P < 0.001$).

Association of land coverage with community structure

Sites were grouped by land coverage category (developed, forested, or combined agricultural) based on the greatest percentage of land coverage within the HUC boundary in which the site was located (Table 1). Changes in community membership (differences in phylogenetic branching) and relative abundance of taxa were significantly different ($P \leq 0.008$) among land coverage categories (both years combined), as evaluated by comparing unweighted and weighted UniFrac metrics, respectively. Similarly, these differences were found to be significant ($P \leq 0.037$) in individual years, except when using unweighted UniFrac metrics of differences between the developed and forested categories in 2012 ($P = 0.066$). Within a land coverage category, however, differences in phylogenetic structure did not differ significantly among sampling sites from either year ($P \geq 0.200$), but the majority of pairwise comparisons showed significant differences when relative abundance of taxa was considered at $\alpha = 0.05$.

However, despite phylogenetic differences found by evaluating UniFrac values, β-diversity was not significantly different ($P = 0.345$) among land coverage categories when evaluated using ANOSIM over both years or during a single year ($P = 0.359$ and 0.237 for 2011 and 2012, respectively). Furthermore, clustering of sites by primary surrounding land-coverage category was not significant in 2011 or 2012 (AMOVA $P = 0.205$ and 0.101, respectively).

Richness, determined as the number of OTUs, was significantly higher at sites surrounded by developed land than those surrounded by forest or agriculture ($P = 0.001$). Shannon diversity was significantly higher at developed sites compared to forested sites ($P = 0.017$), but the difference in diversity between developed and agricultural sites was not significant by *post-hoc* test ($P = 0.371$). The percentage of total developed and pasture land coverage, but not total agricultural land coverage, within an HUC boundary were significantly

correlated with richness (number of OTUs observed, r = 0.583, $P < 0.001$; r = 0.261, $P = 0.034$; respective to coverage type) and Shannon indices (r = 0.334, $P = 0.006$; r = 0.463, $P < 0.001$) by traditional bivariate correlation analysis. Taxonomic orders that were ≥ 0.05% of sequence reads (on average) among all samples (n = 51) were evaluated by traditional analysis to determine associations with the predominant surrounding land coverage. Thirty-seven orders were correlated with the percentage of at least one of these land use types (Supplementary Figures S3 and S4).

Bayesian modeling of community variation

Modeling by Bayesian inference was performed to elucidate potentially biologically important parameters influencing community composition among analytes and land cover measured. Bayesian inference revealed that nine analytes and the percentage of pasture coverage were significantly and directionally associated with the relative abundance of 12 of 15 (80%) of the most abundant taxonomic orders (Figure 4). Interestingly, the concentration of *E. coli* at sites was not related to the relative abundance of any of these orders. However, among the nutrients examined, the concentrations of ammonium, organic carbon, nitrite/nitrate, and orthophosphate were all associated with the relative abundance of at least one order. Furthermore, while directional relationships among specific orders could be inferred, they did not include the two most abundant orders identified – *Burkholderiales* and *Actinomycetales*.

Discriminant function analysis (DFA) of land cover

The DFA revealed two discriminant functions (*i.e.* linear functions in which order abundances were weighted such that land coverage types are maximally dispersed) to explain variance in major land coverage types ($P < 0.001$ for both functions; DFA1 canonical coefficient = 0.994, Wilks' $\lambda = 0.027$; DFA 2 canonical coefficient = 0.871, Wilks' $\lambda = 0.242$). Fourteen orders were identified that were best related to the primary surrounding land coverage of samples collected during both years (orders and coefficients, the absolute values of which indicate the effect size, are shown in Table 4). Of these orders, 7 of 14 (50%) showed differences in relative abundance over both years of study between the three major land-coverage types, via Tukey's *post-hoc* test (Table 4). Among the orders that were significantly more abundant at sites with primarily developed land coverage, only *Aeromonadales* and *Nitrosomonadales* were significantly more abundant compared to primarily agricultural sites ($P \leq 0.005$). Among those more abundant at forested sites, only the *Gammaproteobacteria* order was also greater when compared against agricultural sites ($P = 0.019$).

To better resolve members of the community that might indicate inputs from specific land cover types, the sequences of representative OTUs from each order that showed a significant association with land cover were compared manually against the GenBank non-redundant (nr/nt) database via blastn search (BLAST). Representative OTUs were the most abundant OTUs in each order and generally had 10- to 100-fold more sequence reads than all other OTUs classified to that order over the entire dataset. The isolation sources of top-matches

were taken from the submitters' annotations within GenBank metadata. In general, top BLAST matches for orders that were more abundant at forested sites tended to be strictly from freshwater habitats while those more abundant at developed sites were associated with wastewater treatment or oil contamination (Table 5).

Discussion

In this study, we evaluated the relationships between environmental and chemical parameters and the abundances of bacterial orders and observed that throughout the river at large, the abundances of bacterial orders were associated with regional variation in nutrient and chemical concentrations that are intercorrelated with land coverage patterns. The relationship between land cover and nutrient concentrations has been frequently reported (Gilbert et al., 2009; Fortunato and Crump, 2011; Fortunato et al., 2012). However, specific community compositions associated with specific land cover patterns or nutrient concentrations remains poorly explored in riverine communities, particularly in the water column.

At the local scale, biotic interactions have been reported to be primary drivers of variation in bacterial community structure (Fortunato and Crump, 2011), and this result is similar to that reported here in which associations were primarily observed among bacterial orders rather than nutrient and chemical concentrations by LSA. The finding that the majority of significant associations were observed among bacterial orders in the Mississippi River is also similar to observations of microbial communities in the English Channel (Gilbert et al., 2012), although the scale of these two ecosystems differ considerably in ecosystem type, study area, and nutritional availability. The local associations identified here may be particularly important when assessing community relationships in a lotic ecosystem prone to continual changes in nutrient availability and bacterial taxa as well as variation in residence time associated with seasonal variation and, potentially, land cover. A recent study of microbial communities in river sediments found that while primarily environmental drivers shaped community structure (Gibbons et al., 2014), dispersal dynamics and stochastic forces may play a minor role. When residence time is short, as might be expected in the late Spring and early Summer in this system, biotic interactions may play a stronger role in defining community dynamics than during later months when residence time increases. This possibility, as well as potential time-delayed effects on community structure in this ecosystem will require further study.

Despite the highly intercorrelated relationships among orders and seeming independence from local environmental and chemical parameters in the association network, significant directional relationships were inferred between both water chemistry, especially organic carbon and TDS concentrations, and order abundances as well as among bacterial orders alone at the regional scale via Bayesian analysis. These results suggest that physicochemical and nutrient concentrations may influence only a small number of taxa locally, but these effects may be propagated throughout the entire community, potentially as a result of interactions among taxa throughout this ecosystem. For example, the negative Bayesian correlation of temperature with

E. coli concentration may indicate increases in abundance of other community members that out-compete this species, and *E. coli* is known to be a minority member of the community (Byappanahalli and Fujioka, 2004). Elucidation of these potential interactive effects have been previously utilized to create a predictive artificial neural network for microbial communities (Larsen et al., 2012). However, development of such a network was beyond the scope of the current study as few time points were analyzed.

Discrepancies between associations observed from local versus regional analyses suggest that the spatial scale at which data are analyzed affects the conclusions drawn (Ruan et al., 2006; Xia et al., 2011; Gilbert et al., 2012). Future work will be required to validate whether the associations observed in this study represent actual community-level bacterial interactions, as this could not be determined here. A previous study indicated that responses to variation in nutrient concentration affect the distribution of functional traits in a community while taxonomic composition is habitat-specific (Comte and del Giorgio, 2009). Similarly, a study of geographically disparate freshwater cyanobacterial blooms found that functional traits were conserved among ecosystems despite variation in taxonomic community composition (Steffen et al., 2012). Based on these previous findings, community composition here may not reflect community adaptation to changes in nutrient concentrations but may instead be indicative of the bacteria contributed from natural runoff and anthropogenic impacts.

Furthermore, this study is restricted by the removal of aggregate bacteria, which have been shown to be phylogenetically and functionally diverse (Grossart, 2010), as well as the limitation to only bacterial orders that are represented in taxonomic classification databases. Complex trophic interactions are known to influence the bacterial community (Verreydt et al., 2012), and the lack of these associations between domains here may result in an over-simplified model of inter-specific associations among the total community. Finally, assessing bacterial associations when taxa are classified to orders may also reduce model complexity as bacterial orders can show significant diversity in functional traits, functional overlap, and differential survival ability. However, based on the shortness of sequence reads, more specific classification was not performed as taxonomic assignment to more specific levels have been shown to have poor accuracy in classification (Mizrahi-Man et al., 2013).

We hypothesized here that variation in major land coverage types would result in consistent directional shifts in the relative abundance of bacterial orders due to alteration of nutrient and chemical concentrations, as well as inputs of source-specific bacterial groups. Alteration of nutrient concentrations as a result of land coverage has been previously reported (Fisher et al., 2000; Miller et al., 2011), and these reports are generally consistent with the results presented here. Similarly, water quality has also been shown to be influenced by land cover in a predominantly urban setting (Tu, 2011). A previous study of stream sediments has further reported that variation in bacterial communities was significantly associated with impervious land cover associated with urbanization, although taxonomic assignments of variable OTUs was not addressed (Wang et al., 2011). Community variation characterized by shifts in abundances

of specific taxa have also been observed in near-atmosphere air samples collected from forested, suburban, and agricultural areas (Bowers et al., 2011). Interestingly, the dominant groups identified in air samples – *Burkholderiales* and *Actinobacteria* – were also the most abundant groups identified in this study and were not correlated with any other abiotic parameter with the exception of a negative correlation between *Burkholderiales* abundance and cumulative (three-day) antecedent rainfall. The lack of correlations of these orders may result from their general ubiquity and high relative abundance, although it is important to note that the majority of variation is seen amongst taxa present in only moderate or minor abundances.

Although bacterial community structure could not be significantly associated with major land coverage by local analyses, abundances of many orders were correlated with land coverage when analyzed in the context of the entire dataset. Results of DFA suggested that only a few orders were significant in explaining variation in community structure based on patterns of land coverage, and several of these may be more useful targets to identify major non-point sources of contamination to the river or to identify biotic interactions influencing variation in community membership and/or structure. For example, increases in abundance of the orders *Aeromonadales* and *Nitrosomonadales* may serve as indicators for specifically urbanized contamination, while an increase in a specific order of *Gammaproteobacteria* may indicate the relative absence of anthropogenic impact. Lack of significant differences in abundance of many of these orders between developed and agricultural land could also imply that several groups are indicators of more generalized anthropogenic impacts (*i.e.* not specifically developed or agricultural, such as failure of septic systems), as has been previously suggested (Staley et al., 2013). Frequent matching of OTU sequences from orders associated with developed land cover to isolates from the wastewater treatment process suggest that these orders may be contributed as a result of effluent outfall. Similarly, matching of forest-associated OTUs with isolates from only freshwater bodies lends credit to the conclusion that these orders are reflective of more pristine, unimpacted conditions. It should be noted, however, that BLAST searching was far from exhaustive due to the size of the GenBank database, and isolates from many other sources may share identity with the OTU sequences queried. Use of computational algorithms such as SourceTracker (Knights et al., 2011) which employs an OTU-based approach characterizing source and sink communities will provide more objective determination of sources of OTUs, as demonstrated recently (Newton et al., 2013). However, use of these methods will require knowledge of specific sources and extensive characterization of their microbial communities, and this data was not collected in the current study.

Importantly, the concentration of *E. coli*, which is commonly used as an indicator of surface water quality and human health risk (United States Environmental Protection Agency, 2012, 2002), was not correlated with other measures of water quality (*e.g.* phosphorus concentration) or abundances of several orders that include human pathogens (*e.g.* *Enterobacteriales*) among the entire dataset. Furthermore, *E. coli* concentration was not significantly associated with the abundance of bacterial orders by local or Bayesian analyses, suggesting a poor relationship between this species and the overall bacterial community structure. *E. coli* concentrations measured here only twice exceeded the Environmental Protection Agency's one-time sampling threshold of 235 CFU 100 ml^{-1} (United States

Environmental Protection Agency, 2002), suggesting that risk of human pathogens during the sampling period may have been limited, although the presence of pathogens or of fecal contamination during this study was not determined. However, previous studies have suggested that factors including nutrient concentrations, land coverage, and the surrounding bacterial community are all potential factors associated with pathogen presence and activity (Viau et al., 2011; Williams et al., 2012), therefore *E. coli* may be an even poorer indicator of water quality than previously thought.

This study highlights the complexity of factors that influence bacterial community structure locally and regionally in a complex riverine ecosystem. Among all analyses, organic carbon and TDS were observed to be among the primary environmental factors influencing both diversity and the abundance of specific bacterial orders. These parameters were also regionally associated with specific land cover types suggesting that specific anthropogenic impacts alter the chemistry of this riverine ecosystem and contribute non-indigenous bacteria resulting in shifts in the overall bacterial community. Furthermore, this study is among the first to suggest that specific bacterial orders in the water column may be indicative of specific types of non-point source contamination, and may serve as more informative indicators of ecosystem impairment than traditional indicator bacteria. Further interrogation of the associations and networks proposed here will better allow regulatory agencies and resource managers to determine if contamination is a result of relatively local, potentially point sources, or may be due to an accumulation of chemicals from more diffuse, point and non-point sources upstream.

Acknowledgements

Funding for this study was provided by the Minnesota Environment and Natural Resources Trust Fund, at the recommendation of the Legislative-Citizen Commission on Minnesota Resources (LCCMR). This work was carried out using computing resources at the University of Minnesota Supercomputing Institute.

Conflict of interest

No conflict of interest declared.

References

Aronesty, E. (2013). Comparison of sequencing utility programs. *Open Bioinforma. J.* 7, 1–8. doi:10.2174/1875036201307010001.

Balogh, S. J., Engstrom, D. R., Almendinger, J. E., McDermott, C., Hu, J., Nollet, Y. H., Meyer, M. L., and Johnson, D. K. (2009). A sediment record of trace metal loadings in the Upper Mississippi River. *J. Paleolimnol.* 41, 623–639.

Balogh, S. J., Engstrom, D. R., Almendinger, J. E., Meyer, M. L., and Johnson, D. K. (1999). History of mercury loading in the Upper Mississippi River reconstructed from the sediments of Lake Pepin. *Environ. Sci. Technol.* 33, 3297–3302.

Bowers, R. M., McLetchie, S., Knight, R., and Fierer, N. (2011). Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. *ISME J.* 5, 601–12. doi:10.1038/ismej.2010.167.

Brandsma, J., Martinez, J. M., Slagter, H. A., Evans, C., and Brussaard, C. P. D. (2013). Microbial biogeography of the North Sea during summer. *Biogeochemistry* 113, 119–136.

Bray, J. R., and Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27, 325–349.

Byappanahalli, M., and Fujioka, R. (2004). Indigenous soil bacteria and low moisture may limit but allow faecal bacteria to multiply and become a minor population in tropical soils. *Water Sci. Technol.* 50, 27–32.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–4. doi:10.1038/ismej.2012.8.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117–143.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Marsh, T., Garrity, G. M., et al. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37, D141–D145.

Comte, J., and del Giorgio, P. A. (2009). Links between resources, C metabolism and the major components of bacterioplankton community structure across a range of freshwater ecosystems. *Environ. Microbiol.* 11, 1704–16. doi:10.1111/j.1462-2920.2009.01897.x.

Drury, B., Rosi-Marshall, E., and Kelly, J. J. (2013). Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. *Appl. Environ. Microbiol.* 79, 1897–1905.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200.

Ellis, J. B. (2006). Pharmaceutical and personal care products (PPCPs) in urban receiving waters. *Environ. Pollut.* 144, 184–189.

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.

Fisher, D., Steiner, J., Endale, D., Stuedemann, J., Schomberg, H., Franzluebbers, A., and Wilkinson, S. (2000). The relationship of land use practices to surface water quality in the Upper Oconee Watershed of Georgia. *For. Ecol. Manage.* 128, 39–48. doi:10.1016/S0378-1127(99)00270-4.

Fortunato, C. S., and Crump, B. C. (2011). Bacterioplankton community variation across river to ocean environmental gradients. *Microb. Ecol.* 62, 374–382.

Fortunato, C. S., Herfort, L., Zuber, P., Baptista, A. M., and Crump, B. C. (2012). Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. *ISME J.* 6, 554–563.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J. (2011). Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogramm. Eng. Remote Sens.* 77, 858–864.

Gibbons, S. M., Jones, E., Bearquiver, A., Blackwolf, F., Roundstone, W., Scott, N., Hooker, J., Madsen, R., Coleman, M. L., and Gilbert, J. a (2014). Human and environmental impacts on river sediment microbial communities. *PLoS One* 9, e97435. doi:10.1371/journal.pone.0097435.

Gilbert, J. A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., Somerfield, P. J., Huse, S., and Joint, I. (2009). The seasonal structure of microbial communities in the Western English Channel. *Environ. Microbiol.* 11, 3132–3139.

Gilbert, J. A., Steele, J. A., Caporaso, J. G., Steinbrueck, L., Reeder, J., Temperton, B., Huse, S., McHardy, A. C., Knight, R., Joint, I., et al. (2012). Defining seasonal marine microbial community dynamics. *ISME J.* 6, 298–308.

Grossart, H. P. (2010). Ecological consequences of bacterioplankton lifestyles: changes in concepts are needed. *Environ. Microbiol. Rep.* 2, 706–714.

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., Bushman, F. D., Knight, R., and Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* 8, 761–U107.

Larsen, P. E., Field, D., and Gilbert, J. A. (2012). Predicting bacterial community assemblages using an artificial neural network approach. *Nat. Methods* 9, 621–5. doi:10.1038/nmeth.1975.

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235.

Miller, J. D., Schoonover, J. E., Williard, K. W. J., and Hwang, C. R. (2011). Whole catchment land cover effects on water quality in the lower Kaskaskia River watershed. *Water, Air, Soil Pollut.* 221, 337–350. doi:10.1007/s11270-011-0794-9.

Mizrahi-Man, O., Davenport, E. R., and Gilad, Y. (2013). Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS One* 8, e53608. doi:10.1371/journal.pone.0053608.

Newton, R. J., Bootsma, M. J., Morrison, H. G., Sogin, M. L., and McLellan, S. L. (2013). A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. *Microb Ecol* 65, 1011–1023. doi:10.1007/s00248-013-0200-9.

Pereira, W. E., and Hostettler, F. D. (1993). Nonpoint-source contamination of the Mississippi River and its tributaries by herbicides. *Environ. Sci. Technol.* 27, 1542–1552.

Portillo, M. C., Anderson, S. P., and Fierer, N. (2012). Temporal variability in the diversity and composition of stream bacterioplankton communities. *Environ. Microbiol.* 14, 2417–28. doi:10.1111/j.1462-2920.2012.02785.x.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W. G., Peplies, J., and Glockner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196.

Ruan, Q., Dutta, D., Schwalbach, M. S., Steele, J. A., Fuhrman, J. A., and Sun, F. (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* 22, 2532–8. doi:10.1093/bioinformatics/btl417.

Russell, T. A., and Weller, L. (2012). State of the River Report: Water Quality and River Health in the Metro Mississippi River. Saint Paul, MN Available at: http://stateoftheriver.com/state-of-the-river-report/.

Schilling, K. E., Chan, K. S., Liu, H., and Zhang, Y. K. (2010). Quantifying the effect of land use land cover change on increasing discharge in the Upper Mississippi River. *J. Hydrol.* 387, 343–345.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi:AEM.01541-09 [pii] 10.1128/AEM.01541-09.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–504. doi:10.1101/gr.1239303.

Smith, V. A., Yu, J., Smulders, T. V., Hartemink, A. J., and Jarvis, E. D. (2006). Computational Inference of Neural Information Flow Networks. *PLoS Comput. Biol.* 2, e161. doi:10.1371/journal.pcbi.0020161.

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., Arrieta, J. M., and Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proc. Natl. Acad. Sci. U. S. A.* 103, 12115–12120.

Staley, C., Unno, T., Gould, T. J., Jarvis, B., Phillips, J., Cotner, J. B., and Sadowsky, M. J. (2013). Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. *J. Appl. Microbiol.* 115, 1147–1158. doi:doi:10.1111/jam.12323.

Steffen, M. M., Li, Z., Effler, T. C., Hauser, L. J., Boyer, G. L., and Wilhelm, S. W. (2012). Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. *PLoS One* 7, e44002. doi:10.1371/journal.pone.0044002.

Tu, J. (2011). Spatial and temporal relationships between water quality and land use in northern Georgia, USA. *J. Integr. Environ. Sci.* 8, 151–170.

United States Environmental Protection Agency (2002). EPA-823-B-02-003. Implementation Guidance for Ambient Water Quality Criteria for Bacteria.

United States Environmental Protection Agency (2012). Recreational water quality criteria. 820-F-12-058. *United States Environ. Prot. Agency Off. Water*.

United States Environmental Protection Agency (2007). Science Advisory Board (SAB) Hypoxia Panel Draft Advisory Report. 2013. Available at: http://www.epa.gov/sab/pdf/11-19-07_hap_draft.pdf.

Unno, T., Di, D. Y., Jang, J., Suh, Y. S., Sadowsky, M. J., and Hur, H. G. (2012). Integrated online system for a pyrosequencing-based microbial source tracking method that targets *Bacteroidetes* 16S rDNA. *Environ. Sci. Technol.* 46, 93–98. doi:10.1021/es201380c.

Unno, T., Jang, J., Han, D., Kim, J. H., Sadowsky, M. J., Kim, O. S., Chun, J., and Hur, H. G. (2010). Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. *Environ. Sci. Technol.* 44, 7777–7782. doi:10.1021/es101500z.

Verreydt, D., De Meester, L., Decaestecker, E., Villena, M.-J., Van Der Gucht, K., Vannormelingen, P., Vyverman, W., and Declerck, S. A. J. (2012). Dispersal-mediated trophic interactions can generate apparent patterns of dispersal limitation in aquatic metacommunities. *Ecol. Lett.* 15, 218–26. doi:10.1111/j.1461-0248.2011.01728.x.

Viau, E. J., Goodwin, K. D., Yamahara, K. M., Layton, B. A., Sassoubre, L. M., Burns, S. L., Tong, H.-I. I., Wong, S. H. C., Lu, Y. A., and Boehm, A. B. (2011). Bacterial pathogens in Hawaiian coastal streams - associations with fecal indicators, land cover, and water quality. *Water Res.* 45, 3279–3290. doi:10.1016/j.watres.2011.03.033.

Wang, S.-Y., Sudduth, E. B., Wallenstein, M. D., Wright, J. P., and Bernhardt, E. S. (2011). Watershed urbanization alters the composition and function of stream bacterial communities. *PLoS One* 6, e22972. doi:10.1371/journal.pone.0022972.

Wiener, J. G., and Sandheinrich, M. B. (2010). Contaminants in the Upper Mississippi River: historic trends, responses to regulatory controls, and emerging concerns. *Hydrobiologia* 640, 49–70.

Williams, A. P., Quilliam, R. S., Thorn, C. E., Cooper, D., Reynolds, B., and Jones, D. L. (2012). Influence of land use and nutrient flux on metabolic activity of *E. coli* O157 in river water. *Water, Air, Soil Pollut.* 223, 3077–3083. doi:10.1007/s11270-012-1090-z.

Xia, L. C., Steele, J. A., Cram, J. A., Cardon, Z. G., Simmons, S. L., Vallino, J. J., Fuhrman, J. A., and Sun, F. (2011). Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. *BMC Syst. Biol.* 5 Suppl 2, S15. doi:10.1186/1752-0509-5-S2-S15.

Zhang, S. Y., Zhang, Q. A., Darisaw, S., Ehie, O., and Wang, G. D. (2007). Simultaneous quantification of polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), and pharmaceuticals and personal care products (PPCPs) in Mississippi River water, in New Orleans, Louisiana, USA. *Chemosphere* 66, 1057–1069.

Table 1. Percentages of land cover of hydrological units sampled.

Site*	Distance (km)†	Developed, Open	Developed, Low	Developed, Med	Developed, High	Sum Developed‡	Dedciduous Forest	Evergreen Forest	Mixed Forest	Sum Forest§	Pasture	Cultivated	Sum Agriculture¶
Itasca[a]	0	2.51	0.65	0.17	0.08	3.42	34.50	9.99	4.23	48.72	5.65	1.01	6.66
St. Cloud[b]	263	5.95	1.95	1.26	0.40	9.56	16.45	1.11	0.05	17.61	16.96	38.93	55.89
Clearwater[b]	271	5.95	1.95	1.26	0.40	9.56	16.45	1.11	0.05	17.61	16.96	38.93	55.89
Twin Cities[c]	311	8.57	14.22	6.28	2.16	31.22	14.13	1.14	0.05	15.32	16.58	11.40	27.99
MN River[d]	NA	4.97	4.95	2.70	1.23	13.86	7.50	0.20	0.02	7.73	10.45	58.49	68.94
Confluence[c]	313	8.57	14.22	6.28	2.16	31.22	14.13	1.14	0.05	15.32	16.58	11.40	27.99
Hastings[c]	330	8.57	14.22	6.28	2.16	31.22	14.13	1.14	0.05	15.32	16.58	11.40	27.99
St. Croix River[e]	NA	4.52	1.62	0.61	0.20	6.95	27.01	2.07	0.54	29.62	23.73	22.05	45.78
Red Wing[f]	362	5.32	3.47	1.64	0.44	10.87	19.06	0.21	0.05	19.32	12.21	43.94	56.14
LaCrescent[g]	401	3.76	4.24	1.73	0.42	10.14	40.46	1.92	0.15	42.53	10.95	22.05	33.00
Zumbro River[h]	NA	5.72	2.37	0.70	0.25	9.04	9.59	0.12	0.01	9.71	11.42	55.66	67.07

*Samples with the same superscript (a-h) are within the same HUC boundaries.

†Distance from the headwaters. Samples marked NA are major confluent rivers.

‡Sum of all developed (open – high) area.

§Sum of deciduous, evergreen, and mixed forest area.

¶Sum of pasture and cultivated land.

Table 2. Summary of analytes measured among all sampling sites during both years of study.

Class	Analyte	2011				2012			
		Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Bacteria (CFU 100 ml^{-1})	*E. coli*	76.82	105.28	5.00	300.00	15.05	27.16	ND*	93.50
Physical Parameters	Temperate (°C)	18.18	2.32	12.00	21.00	21.52	1.76	18.20	23.80
	pH	7.72	0.28	7.20	8.10	7.59	0.30	6.89	7.89
	Cumulative Rainfall (mm)	7.39	7.54	0.00	21.85	13.44	14.84	0.00	43.69
	72 h Rainfall (mm)	1.78	2.73	0.00	7.37	4.66	8.29	0.00	21.84
	48 h Rainfall (mm)	3.05	4.85	0.00	11.18	3.53	11.26	0.00	38.61
	24 h Rainfall (mm)	2.56	5.72	0.00	17.53	5.24	6.67	0.00	17.78
Nutrients (mg L^{-1})	Organic carbon	6.20	1.92	1.59	8.99	8.72	3.58	2.06	14.47
	Ammonium	0.07	0.02	0.05	0.11	0.05	0.02	0.04	0.09
	Nitrate/nitrite	2.50	2.36	0.05	7.57	1.97	1.97	0.16	6.19
	Total phosphorus	0.19	0.06	0.08	0.28	0.11	0.01	0.09	0.13
	Orthophosphate	0.05	0.02	0.02	0.09	0.06	0.03	0.02	0.13
	TDS†	79.88	32.39	21.87	139.42	44.19	18.51	13.35	72.13
Ions or Metals (mg L^{-1})	Al	0.13	0.11	ND	0.41	0.42	0.30	0.08	1.18
	B	0.05	0.03	0.02	0.10	ND	ND	ND	ND
	Ca	65.43	25.93	17.79	112.79	36.72	14.17	11.50	57.75
	Cd	ND	ND	ND	ND	0.02	ND	0.02	0.02
	Cr	ND	0.01	ND	0.02	ND	ND	ND	ND
	Cu	0.06	0.16	ND	0.57	ND	ND	ND	0.01
	Fe	0.30	0.15	0.09	0.49	0.86	0.44	0.04	1.59
	K	1.97	0.89	0.90	3.48	2.17	0.83	0.79	3.52
	Mg	26.19	14.10	6.39	55.54	13.54	6.37	3.63	25.00
	Mn	0.05	0.02	0.02	0.08	0.06	0.02	0.02	0.09
	Na	11.75	5.48	3.62	21.39	5.44	2.72	1.63	9.91
	Ni	ND	ND	ND	ND	ND	ND	ND	ND
	P	ND	ND	ND	ND	0.97	2.20	0.08	7.79
	Pb	ND	ND	ND	ND	ND	ND	ND	ND
	Zn	0.02	0.03	ND	0.10	0.10	0.06	0.03	0.21

Antibiotics (ng ml^{-1})	Erythromycin	ND	ND	ND	ND	ND	ND	ND	ND
	Monensin	ND	ND	ND	ND	ND	ND	ND	ND
	Sulfumathoxazole	ND	ND	ND	ND	ND	ND	ND	ND
Pesticides (ng ml^{-1})	Acetochlor	9.59	15.10	ND	43.80	67.01	72.50	ND	231.00
	Atrazine	2.71	4.38	ND	14.30	4.51	1.42	3.07	7.24
	Carbaryl	ND	ND	ND	ND	0.61	1.05	0.26	3.89
	D-atrazine	NA[‡]	NA	NA	NA	ND	ND	ND	ND
	Iprodione	ND	ND	ND	ND	ND	ND	ND	ND
	Metolachlor	ND	ND	ND	ND	141.64	132.06	ND	374.00
Pharmaceuticals (ng ml^{-1})	Acetaminophen	ND	ND	ND	ND	ND	ND	ND	ND
	Caffeine	4.49	3.65	ND	11.60	10.28	7.93	3.50	32.70
	Ibuprofen	ND	ND	ND	ND	ND	ND	ND	ND
Endocrine disrupters (ng ml^{-1})	4-Nonylphenol	ND	ND	ND	ND	ND	ND	ND	ND
	Daidzein	ND	ND	ND	ND	ND	ND	ND	ND
	Carbamazepine	ND	ND	ND	ND	0.56	1.81	ND	6.20
	Fomonentin	ND	ND	ND	ND	ND	ND	ND	ND
	Genistein	ND	ND	ND	ND	ND	ND	ND	ND
	meta-Chlorophenylpiperazine	28.77	52.97	ND	162.00	ND	ND	ND	ND
	Zeranol	ND	ND	ND	ND	ND	ND	ND	ND
Personal care products (ng ml^{-1})	Cotinine	1.35	0.71	ND	2.56	ND	ND	ND	ND
	DEET	ND	ND	ND	ND	5.77	7.74	ND	21.40

[*]ND: Analyte was not detected.

[†]Total dissolved solids.

[‡]NA: Analyte was not measured.

Table 3. Correlation coefficients relating analyte concentrations with major land coverage patterns observed.

Class	Analyte	Developed	Forested	Pasture
Bacterium	*E. coli*	-0.090 (0.472)	0.108 (0.389)	**-0.303 (0.013)***
Physical Parameters	pH	**0.379 (0.002)**	**-0.577 (< 0.001)**	**-0.443 (< 0.001)**
Nutrients	Carbon	0.104 (0.406)	**0.246 (0.046)**	**0.695 (< 0.001)**
	Ammonium	0.154 (0.218)	0.030 (0.812)	0.161 (0.197)
	Nitrate/nitrite	**0.412 (0.001)**	**-0.657 (< 0.001)**	**-0.384 (0.001)**
	Total phosphorus	0.023 (0.853)	-0.089 (0.479)	0.071 (0.570)
	Orthophosphate	**0.497 (< 0.001)**	**-0.532 (< 0.001)**	**-0.412 (0.001)**
	TDS	**0.312 (0.011)**	**-0.518 (< 0.001)**	**-0.465 (< 0.001)**
Ions	Al	**0.297 (0.015)**	**-0.397 (0.001)**	-0.137 (0.274)
	B	0.214 (0.085)	-0.099 (0.427)	-0.047 (0.711)
	Ca	**0.282 (0.022)**	**-0.457 (< 0.001)**	**-0.482 (< 0.001)**
	Cd	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)
	Cr	0.070 (0.575)	0.140 (0.264)	-0.073 (0.563)
	Cu	**-0.399 (0.001)**	**0.421 (< 0.001)**	0.116 (0.354)
	Fe	0.118 (0.346)	-0.152 (0.222)	0.225 (0.070)
	K	**0.721 (< 0.001)**	**-0.637 (< 0.001)**	**-0.253 (0.041)**
	Mg	**0.381 (0.002)**	**-0.486 (< 0.001)**	**-0.476 (< 0.001)**
	Mn	**0.292 (0.017)**	-0.71 (0.569)	**0.261 (0.035)**
	Na	**0.469 (< 0.001)**	**-0.497 (< 0.001)**	-0.232 (0.060)
	P	0.001 (0.992)	0.023 (0.853)	-0.090 (0.470)
	Zn	-0.173 (0.165)	**0.391 (0.001)**	0.046 (0.715)
Pesticide	Acetochlor	**0.421 (< 0.001)**	-0.146 (0.243)	-0.220 (0.076)
	Atrazine	0.234 (0.058)	**-0.375 (0.002)**	-0.093 (0.456)
	Carbaryl	-0.041 (0.745)	-0.132 (0.290)	0.001 (0.992)
	Metolachlor	0.238 (0.054)	**-0.283 (0.021)**	0.038 (0.764)
Pharmaceutical	Caffeine	0.075 (0.549)	0.017 (0.895)	0.008 (0.947)
Endocrine disrupter	Carbamazepine	-0.211 (0.089)	**-0.279 (0.023)**	-0.181 (0.145)
Personal care products	Cotinine	0.190 (0.126)	0.028 (0.826)	0.137 (0.272)
	DEET	0.157 (0.207)	**-0.263 (0.033)**	0.244 (0.049)

*P values are shown in parentheses. Bold values indicate statistically significant correlations.

Table 4. Standardized canonical discriminant function coefficients for taxonomic orders best associated with land coverage type by DFA.

Taxonomic order	**DFA1 (72.2%)**	**DFA2 (27.8%)**
Methylophilales	1.797	1.242
Cyanobacteria (unclassified)	0.652	0.705
Rhodobacterales[*]	2.314	0.097
Gammaproteobacteria incertae sedis[†]	-1.357	0.603
Rhizobiales	0.058	0.688
Synergistales[†]	-3.667	-1.091
Chromatiales	0.976	0.924
Bacillales	1.266	1.218
Verrucomicrobia Subdivision3 *genera incertae sedis*[*]	0.502	-1.461
Prolixibacter[†]	0.877	-0.137
Aeromonadales[*]	0.697	-0.815
Legionellales	-2.333	-1.290
Bacteroidales	-0.005	1.680
Nitrosomonadales[*]	0.308	-1.486

[*]Order was significantly more abundant at sites with primarily developed land coverage ($P \leq 0.036$).

[†]Order was significantly more abundant at sites with primarily forested land coverage.

Table 5. Identification of representative OTUs via BLAST search. Results of the top 5 BLAST matches are shown and numbers in parenthesis indicate the number of hits to multiple isolates from the same source.

Order	Species ID	Isolation Sources
Rhodobacterales	uncultured *Rhodobacter* sp.	oilfield-produced water
	uncultured bacterium	wastewater treatment plant biofilter
	uncultured bacterium	oil-contaminated groundwater
	Roseinatronobacter sp.	alkaline hypersaline lake
	uncultured bacterium	water; Hungary
Gammaproteobacteria incertae sedis	uncultured bacterium (2)	Big Lake; Mljet, Dubrovnik, Croatia
	uncultured Methylotenera sp.	Holocene marine sediment
	uncultured betaproteobacterium	freshwater biofilm
	uncultured bacterium	municipal drinking water system, raw water influent
Synergistales	uncultured bacterium (2)	Arctic thaw pond water
	uncultured bacterium (2)	Lake Mizugaki; Yamanashi, Japan
	uncultured bacterium	deep-water sponge (*Baikalospongia intermedia*)
Verrucomicrobia Subdivision 3 *incertae sedis*	uncultured bacterium	municipal drinking water system, tap water
	uncultured bacterium	water, 17th Street Canal, New Orleans, LA, USA
	uncultured bacterium	Lake Poyang; China
	uncultured bacterium	soil; Harvard Forest, MA, USA
	uncultured bacterium	soil; Adulam, Israel
Prolixibacter	uncultured bacterium	Lake Mizugaki; Yamanashi, Japan
	uncultured bacterium	municipal drinking water system, raw water influent
	uncultured bacterium	Amazon River, Brazil
	uncultured bacterium	Las Cumbres Lake; Panama
	uncultured bacterium	diseased leaf; Lake Taihu; China
Aeromonadales	*Aeromonas veronii*	pond loach (*Misgurnus anguillicaudatus*)
	Aeromonas salmonicida (2)	channel catfish (*Ictalurus punctatus*)
	Aeromonas dhakensis	membrane bioreactor activated sludge
	Aeromonas hydrophila	membrane bioreactor activated sludge
Nitrosomonadales	uncultured bacterium	Yong Ding River; Beijing China
	uncultured bacterium	typha rhizosphere; Bai River; Beijing China
	uncultured bacterium	municipal drinking water distribution system
	uncultured bacterium	activated sludge

uncultured betaproteobacterium	activated sludge

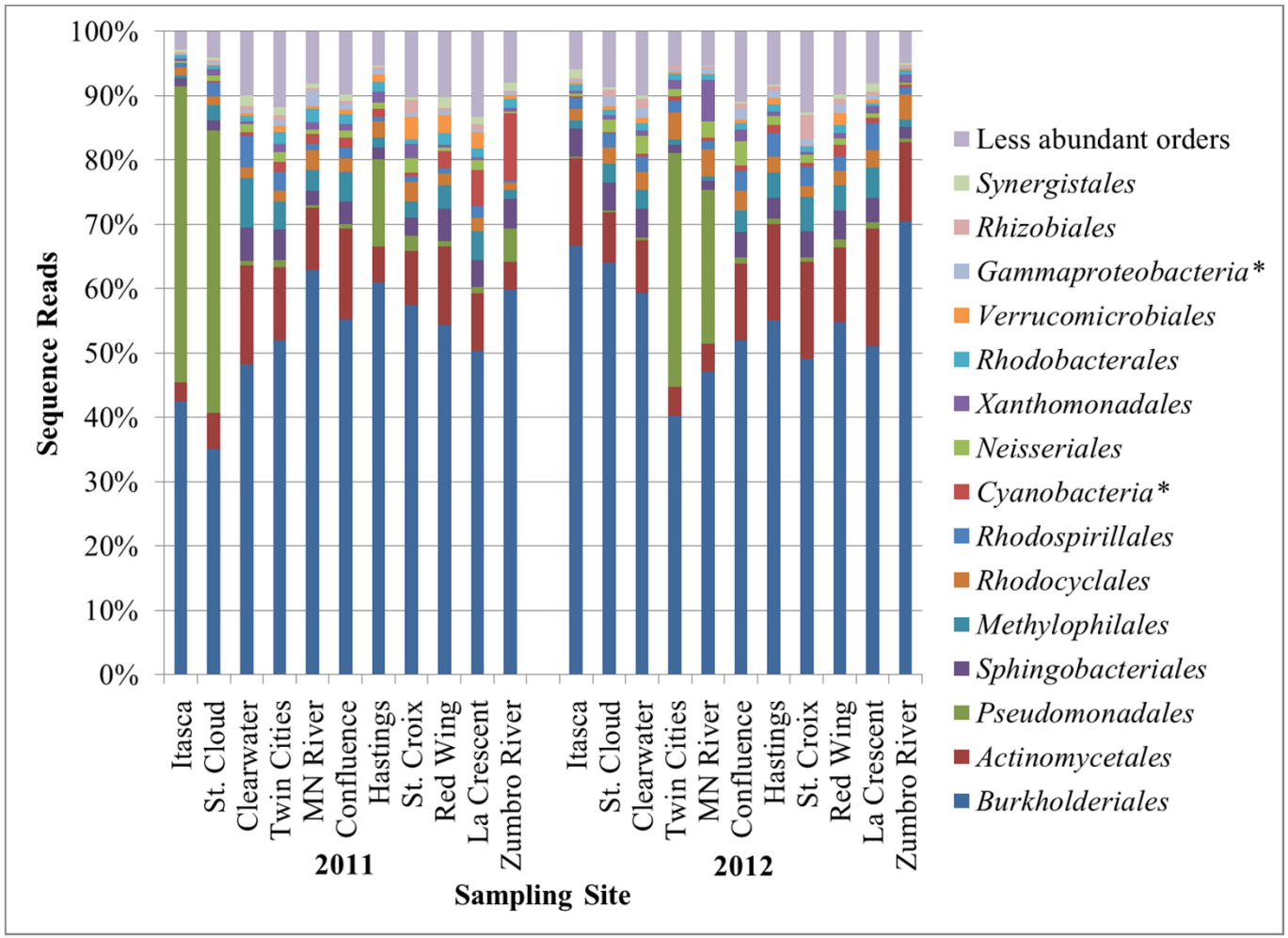
Figure legends

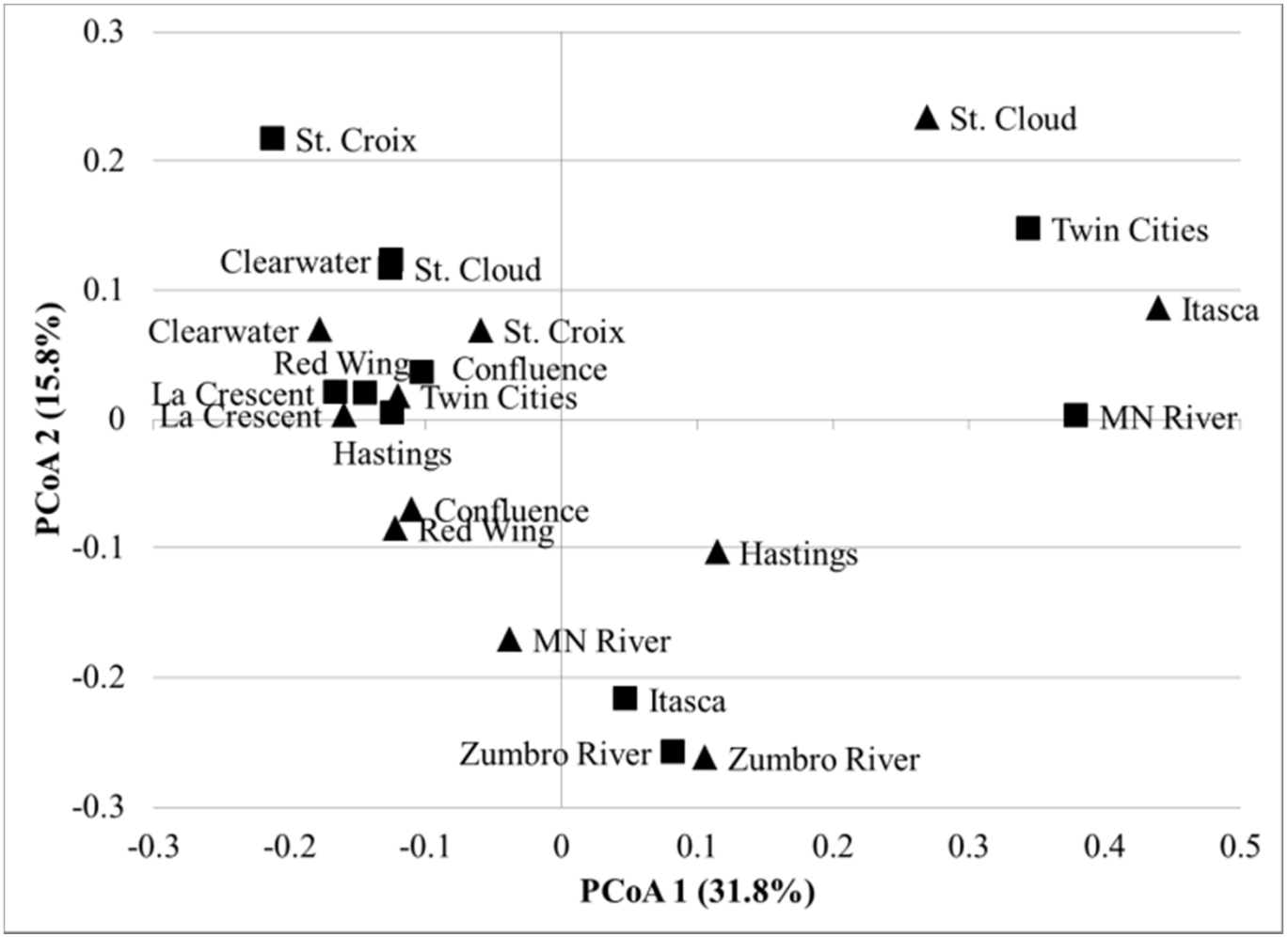
Figure 1. Distribution of the most abundant orders identified at sampling sites among triplicate samples at each site. Asterisks (*) denote orders designated *incertae sedis*.

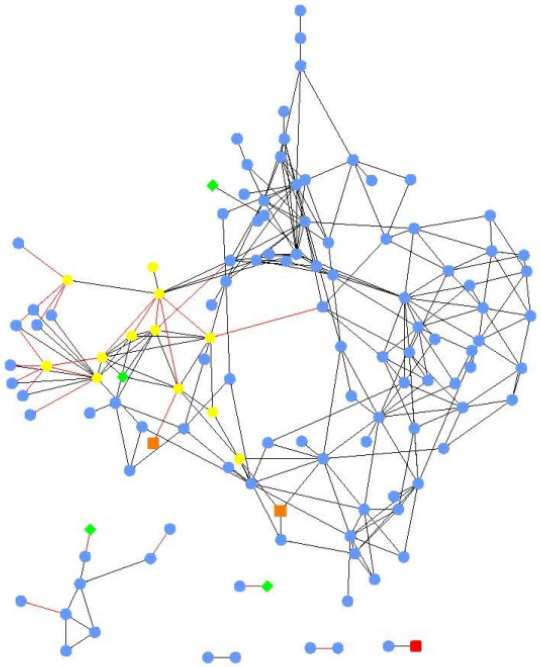
Figure 2. Principal coordinate analysis of bacterial communities from 2011 (▲) and 2012 (■) samples ($r^2 = 0.848$). All replicates were merged for ordination only; statistics were calculated with replicates separated. A total of 20 axes were necessary to explain all variation.

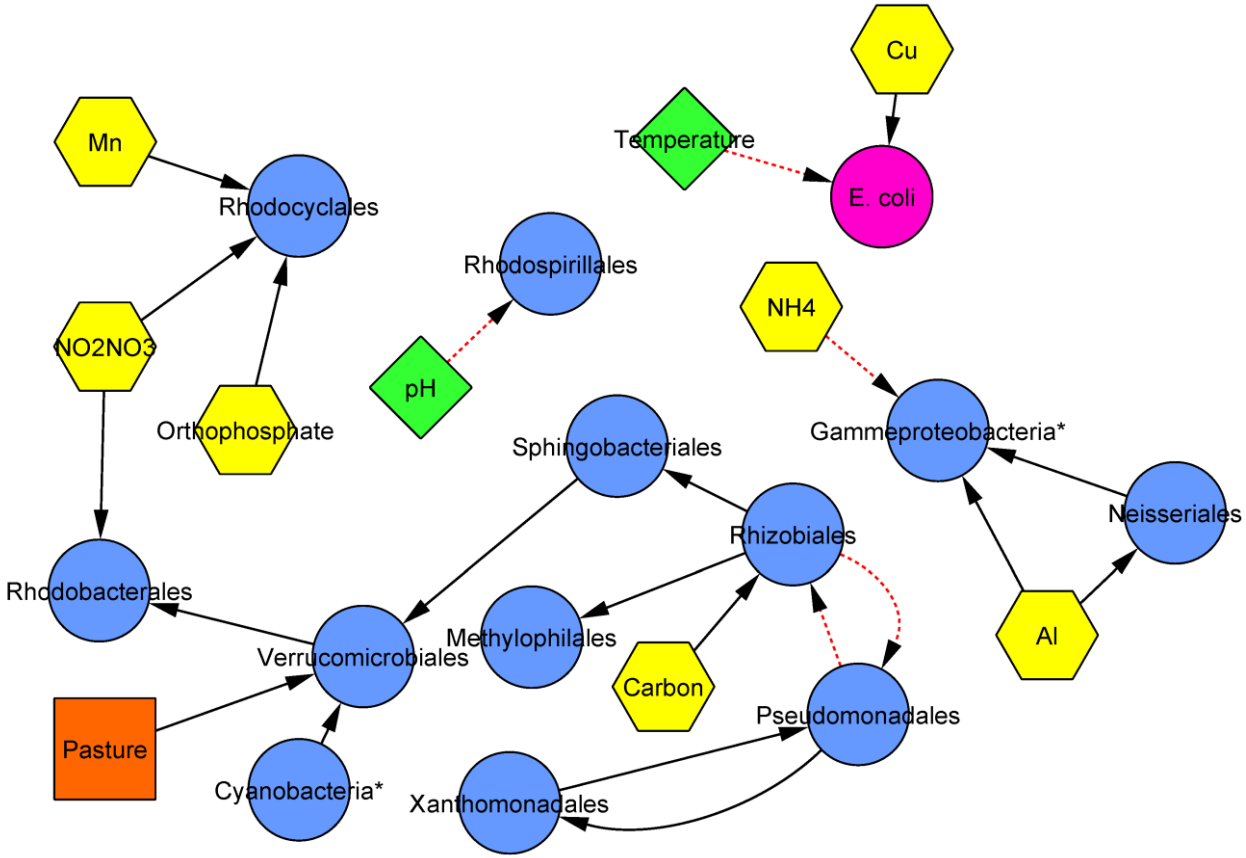
Figure 3. Local similarity analysis network of bacterial orders (blue circles), physicochemical parameters (green diamonds), chemicals (nutrients and ions; yellow hexagons), xenobiotic compounds (red squares), and land use (orange squares). All relationships were significant ($P < 0.05$, $q < 0.003$). Black edges indicate positive local similarity scores and red edges are negative, length is arbitrary.

Figure 4. Consensus inferred Bayesian network relating bacterial orders (blue circles), physicochemical parameters (green diamonds), chemicals (nutrients and ions; yellow hexagons), *E. coli* (pink circle), and land use (orange squares). Solid black lines indicate positive associations while dashed red lines indicate negative associations and arrows are directed from parent to child. Edge length is arbitrary. Asterisks (*) designate orders that are unclassified or designated *incertae sedis*.


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Sequence Reads
Itasca
St. Cloud
Clearwater
Twin Cities
MN River
Confluence
Hastings
St. Croix
Red Wing
La Crescent
Zumbro River
2011
2012
Sampling Site
Less abundant orders
Synergistales
Rhizobiales
Gammaproteobacteria*
Verrucomicrobiales
Rhodobacterales
Xanthomonadales
Neisseriales
Cyanobacteria*
Rhodospirillales
Rhodocyclales
Methylophilales
Sphingobacteriales
Pseudomonadales
Actinomycetales
Burkholderiales


0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
PCoA 2 (15.8%)
PCoA 1 (31.8%)
St. Croix
St. Cloud
Twin Cities
Clearwater
St. Cloud
Itasca
Clearwater
St. Croix
Red Wing
Confluence
La Crescent
Twin Cities
MN River
La Crescent
Hastings
Confluence
Red Wing
Hastings
MN River
Itasca
Zumbro River
Zumbro River




Cu
Mn
Temperature
Rhodocyclales
E. coli
Rhodospirillales
NH4
NO2NO3
pH
Orthophosphate
Sphingobacteriales
Gammeproteobacteria*
Neisseriales
Rhizobiales
Rhodobacterales
Methylophilales
Al
Verrucomicrobiales
Carbon
Pseudomonadales
Pasture
Cyanobacteria*
Xanthomonadales

Supplementary Material

Bacterial Community Structure is Indicative of Specific Chemical Inputs in the Upper Mississippi River

Christopher Staley[1], Trevor J. Gould[1,2], Ping Wang[1], Jane Phillips[2], James B. Cotner[3], and Michael J. Sadowsky[1,4,*]

[1]BioTechnology Institute,

[2]Biology Program,

[3]Department of Ecology, Evolution and Behavior,

[4]Department of Soil, Water and Climate, University of Minnesota, St. Paul, MN

[*]Corresponding Author: Michael J. Sadowsky, BioTechnology Institute, University of Minnesota, 140 Gortner Lab, 1479 Gortner Ave, Saint Paul, MN 55108; Phone: (612)-624-2706, Email: sadowsky@umn.edu

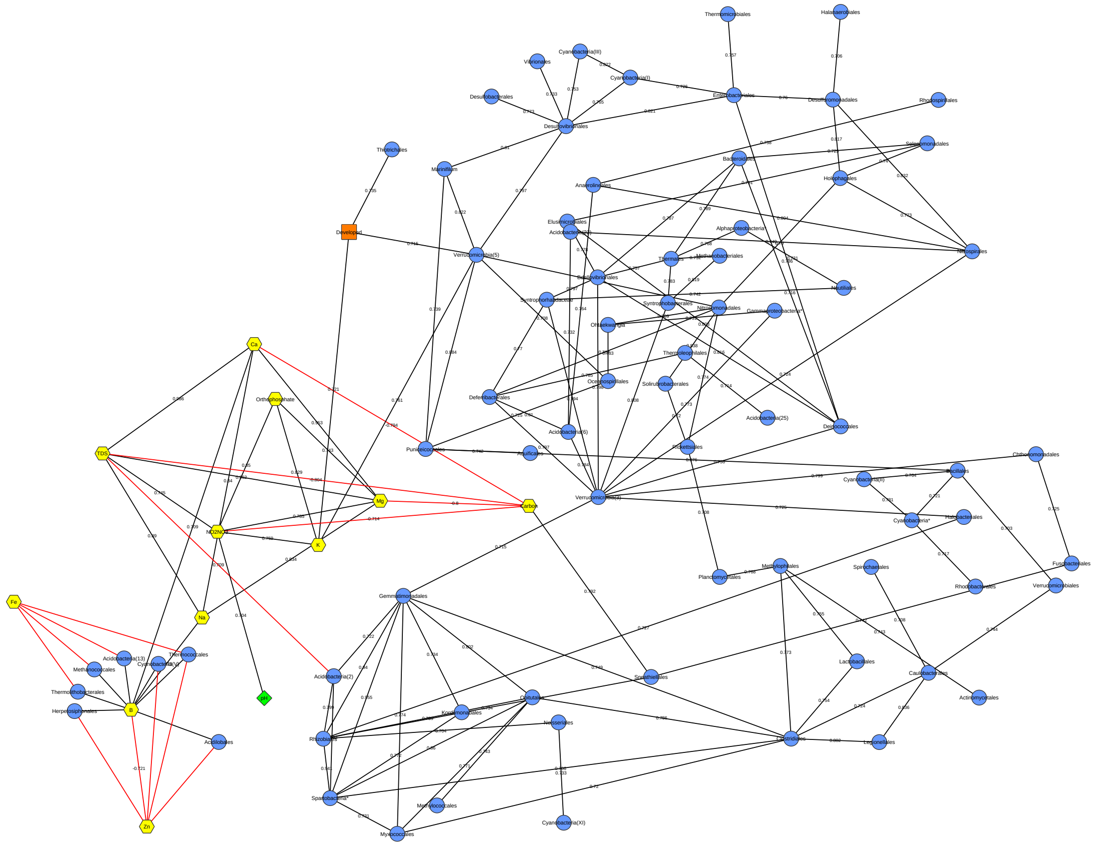
Supplementary Figures

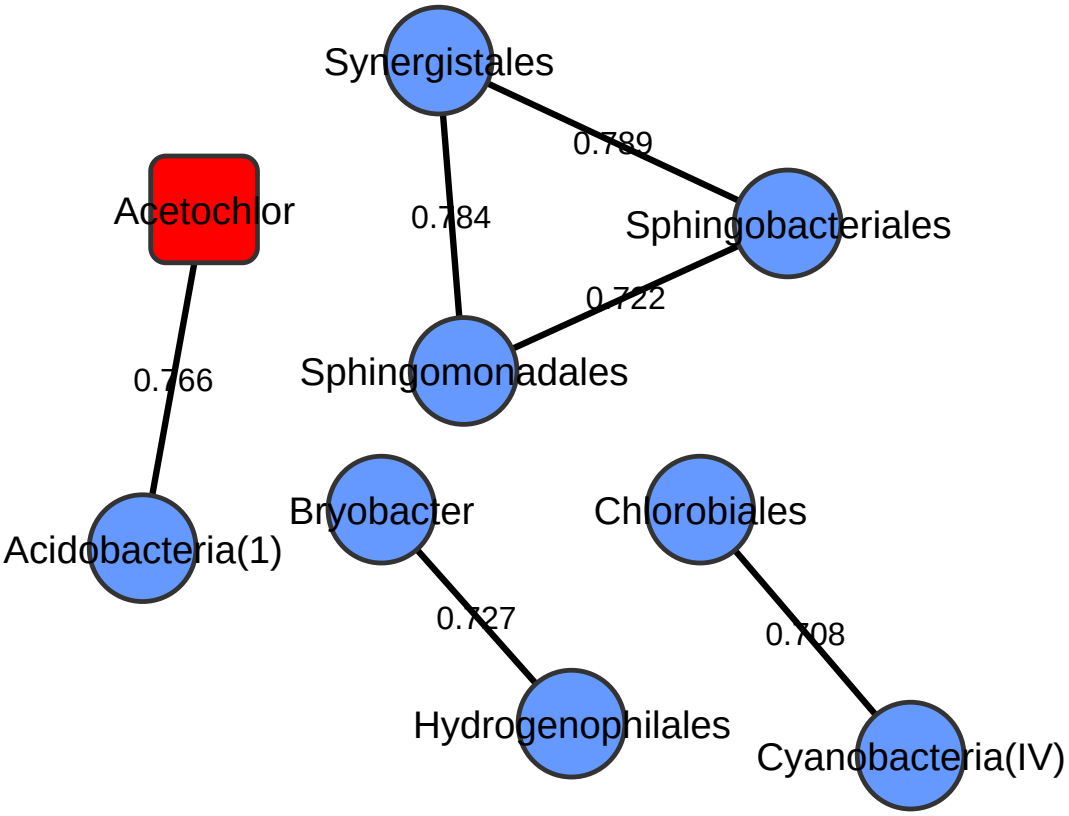
Figure S1. Primary network analysis of well-correlated, significant, local similarity relationships ($P < 0.05$, $Q < 0.003$, $-0.7 <$ Spearman's $r < 0.7$) among bacterial orders (blue circles), physicochemical parameters (green diamonds), chemicals (nutrients and ions; yellow hexagons), and land coverage (orange squares). Black edges indicate positive local similarity scores and dashed, red edges are negative, edge length is arbitrary. Spearman's r values are shown on the edges, except when orders were only detected in one sample and could not be analyzed for correlation. Numbers in parentheses indicate group (*Acidobacteria*), family (*Cyanobacteria*), or subdivision (*Verrucomicrobia*) and asterisks (*) indicate orders unclassified or designated *incertae sedis*.

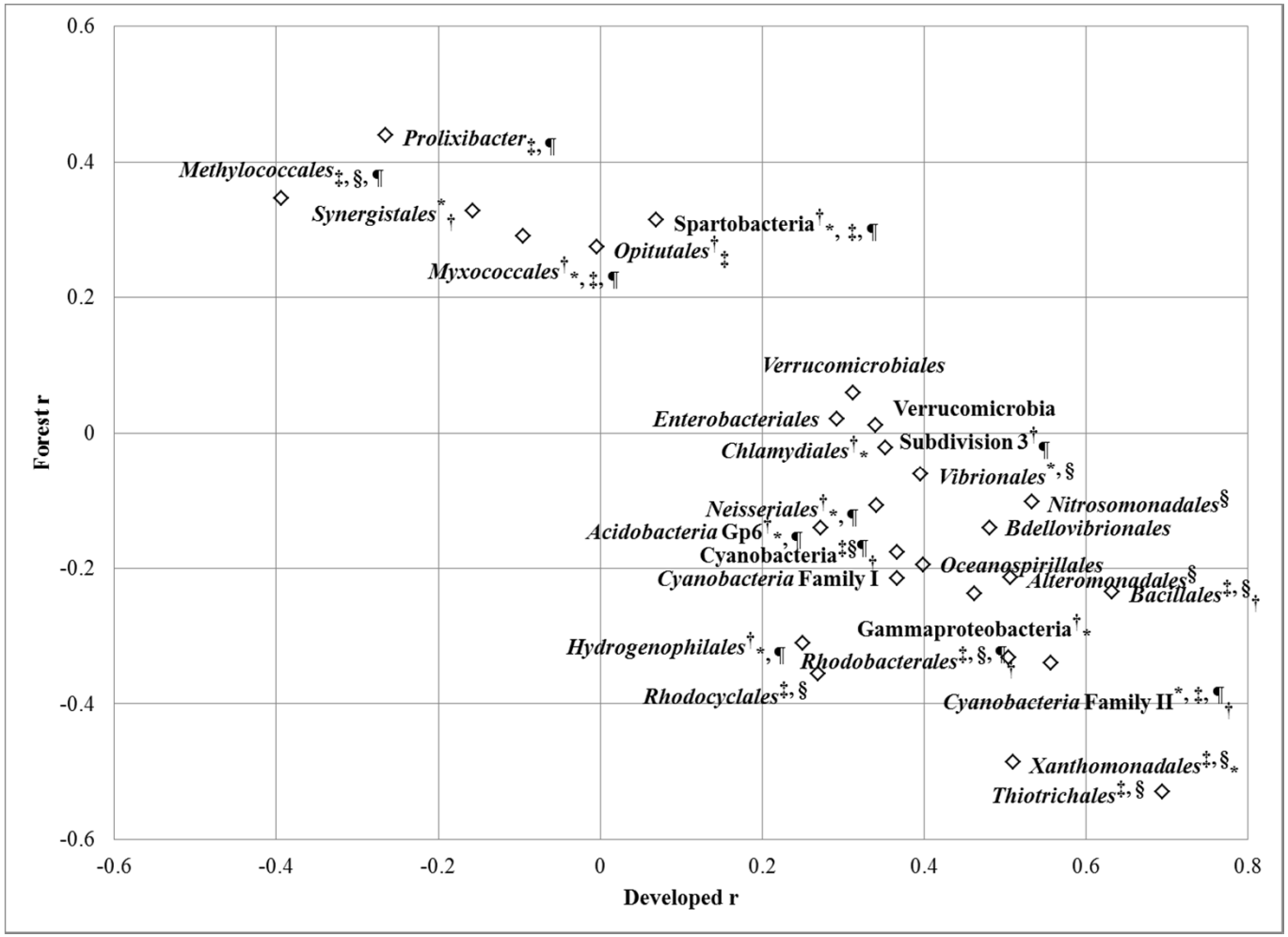
FigureS2. Smaller networks of well-correlated, significant local similarity relationships ($P < 0.05$, $Q < 0.003$, $-0.7 <$ Spearman's $r < 0.7$) among bacterial orders (blue circles) and xenobiotic compounds (red squares). Black edges indicate positive local similarity scores and edge length is arbitrary. Spearman's r values are shown on the edges. Numbers in parentheses indicate group (*Acidobacteria*) or family (*Cyanobacteria*).

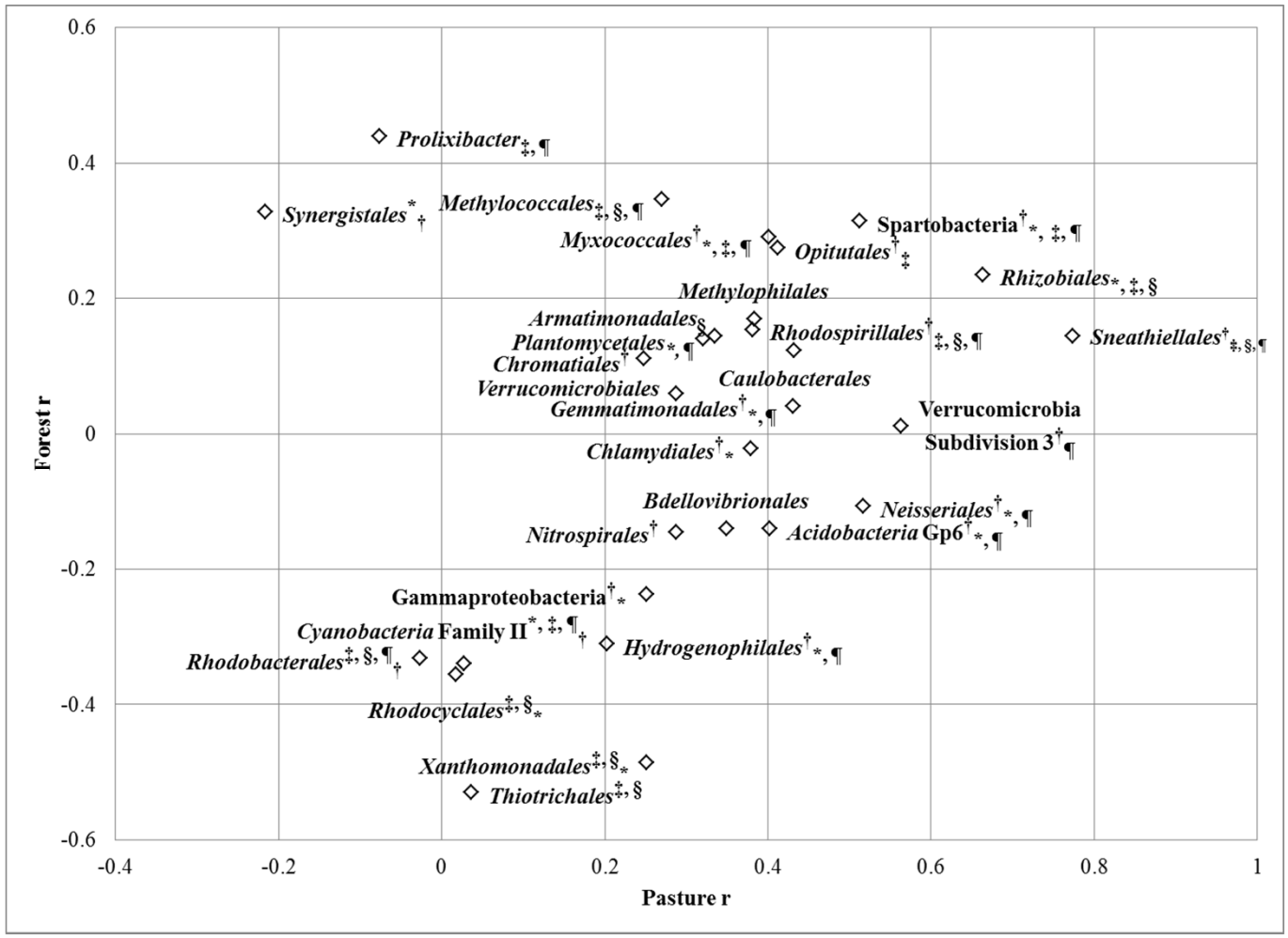
Figure S3. Correlation coefficients (r) relating relative abundance of orders to total developed or total forested area within HUC boundaries. Total developed and forested land area were negatively correlated ($r = -0.581$, $P < 0.001$). Superscripts indicate significant positive correlations while subscripts indicate negative correlations ($\alpha = 0.05$). Correlations are shown for *E. coli* (*), carbon (†), nitrate/nitrite (‡), orthophosphate (§), and TDS (¶). Non-italic names indicate orders *incertae sedis.*

Figure S4. Correlation coefficients (r) relating relative abundance of orders to pasture or total forested area within HUC boundaries. Superscripts indicate significant positive correlations while subscripts indicate negative correlations ($\alpha = 0.05$). Correlations are shown for *E. coli* (*), carbon (†), nitrate/nitrite (‡), orthophosphate (§), and TDS (¶). Non-italic names indicate orders *incertae sedis.*




Synergistales
0.789
Acetochlor
0.784
Sphingobacteriales
0.722
Sphingomonadales
0.766
Bryobacter
Chlorobiales
Acidobacteria(1)
0.727
0.708
Hydrogenophilales
Cyanobacteria(IV)


0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
Forest r
Developed r
Prolixibacter‡, ¶
Methylococcales‡, §, ¶
Synergistales*, †
Spartobacteria†*, ‡, ¶
Opitutales†‡
Myxococcales†*, ‡, ¶
Verrucomicrobiales
Verrucomicrobia
Enterobacteriales
Chlamydiales†*
Subdivision 3†¶
Vibrionales*, §
Nitrosomonadales§
Neisseriales†*, ¶
Bdellovibrionales
Acidobacteria Gp6†*, ¶
Cyanobacteria‡§¶†
Oceanospirillales
Cyanobacteria Family I
Alteromonadales§
Bacillales‡, §†
Gammaproteobacteria†*
Hydrogenophilales†*, ¶
Rhodobacterales‡, §, ¶
Rhodocyclales‡, §
Cyanobacteria Family II*, ‡, ¶†
Xanthomonadales‡, §*
Thiotrichales‡, §


0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Forest r
Pasture r
Prolixibacter‡, ¶
Synergistales*, †
Methylococcales‡, §, ¶
Myxococcales†*, ‡, ¶
Spartobacteria†*, ‡, ¶
Opitutales†‡
Rhizobiales*, ‡, §
Methylophilales
Armatimonadales§
Plantomycetales*, ¶
Rhodospirillales†‡, §, ¶
Sneathiellales†‡, §, ¶
Chromatiales†
Caulobacterales
Verrucomicrobiales
Gemmatimonadales†*, ¶
Verrucomicrobia Subdivision 3†¶
Chlamydiales†*
Bdellovibrionales
Neisseriales†*, ¶
Nitrospirales†
Acidobacteria Gp6†*, ¶
Gammaproteobacteria†*
Cyanobacteria Family II*, ‡, ¶†
Hydrogenophilales†*, ¶
Rhodobacterales‡, §, ¶†
Rhodocyclales‡, §*
Xanthomonadales‡, §*
Thiotrichales‡, §

356

Relationship Between Land Use and Anthropogenic Factors Influencing Bacterial Community Structure in the Upper Mississippi River

Christopher Staley[1], Tatsuya Unno[1], Trevor J. Gould[1,2], Bruce Jarvis[1,2], Jane Phillips[2], James B. Cotner[3], and Michael J. Sadowsky[4]

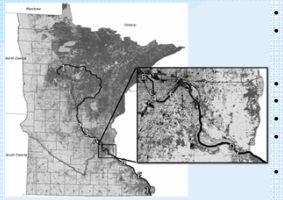
[1]BioTechnology Institute, University of Minnesota, St. Paul, MN; [2]Biology Program, University of Minnesota, Minneapolis, MN, [3]Dept of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN [4]Dept of Soil, Water, and Climate, University of Minnesota, St. Paul, MN

Contact:
1479 Gortner Ave, Suite 140
St. Paul, MN 55108
(p) 612-625-1722
cmstaley@umn.edu

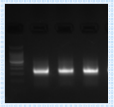
Introduction

- The Mississippi River is the largest watershed in the United States serving major roles in transportation, recreation, and drinking water supply.
- Metagenomic characterization of riverine and marine MCS have revealed stable, seasonally-driven fluctuations in the relative abundance of dominant taxa[3,4].
- Spatial and seasonal variation have been previously demonstrated to have greater impact on MCS than did sampling depth in a freshwater lake[5].
- The River is also known to be impacted by point and non-point sources of pollution including agricultural runoff as well as wastewater and industrial plant discharge[6].
- Previous studies in our laboratory have indicated that metagenomic community profiling may be useful in determining sources of fecal pollution to riverine systems[8].
- We hypothesize that anthropogenic impacts on upstream microbial communities will influence the MCS on sites downstream.
- The effects of sampling volume (cell pellets representing 0.33, 1, 2, and 6 L) and filter size (0.22 and 0.45 µm) on estimated MCS remain relatively under-explored and was also investigated.

Methods



- Filter water
- Elutriate cells
  cell pellets reflect 6 L or volume specified
- DNA extraction
- PCR (16S rDNA, V6 region)
- Gel extraction
- Amplicon poolings
- Illumina sequencing (MiSeq)



mothur

Sequence processing using mothur[7]: quality trimming, alignment, chimera removal (UCHIME[2]), sample normalization (sub-sampling), classification against RDP database[1], alpha/beta diversity analysis

References

1. Cole, JR, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res*, 37, D141-D145.
2. Edgar RC, et al. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27:2194-2200.
3. Fortunato CS, et al. (2012) Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. *ISME J* 2012, 6:554-563.
4. Gilbert JA, et al. (2012) Defining seasonal marine microbial community dynamics. *ISME J* 2012, 6:298-308.
5. Mueller-Spitz SR, et al. (2009) Temporal and spatial variability in nearshore bacterioplankton communities of Lake Michigan. *FEMS Microbiol Ecol*, 67:511-522.
6. Pereira, WE and Hostettler, FD (1993) Nonpoint-source contamination of the Mississippi River and its tributaries by herbicides. *Environ Sci Tech*, 27, 1542-1552.
7. Schloss PD, et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*, 75:7537-7541.
8. Unno, T, et al. (2010) Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. *Environ Sci Tech*, 44, 7777-7782.

Acknowledgments

Funding for this project was provided by the American Recovery and Reinvestment Act of 2009 and the Minnesota Environment and Natural Resources Trust Fund as recommended by the Legislative-Citizen Commission on Minnesota Resources. This work was carried out in part using computing resources at the University of Minnesota Supercomputing Institute.

Results

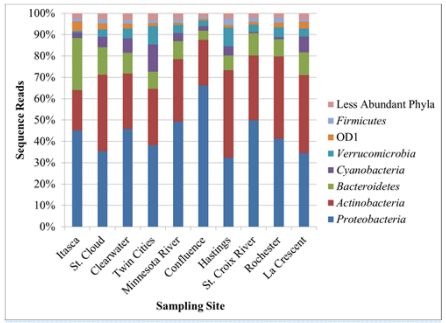


Figure 1. Abundant phyla identified at sampling sites. A total of 32 phyla were identified.

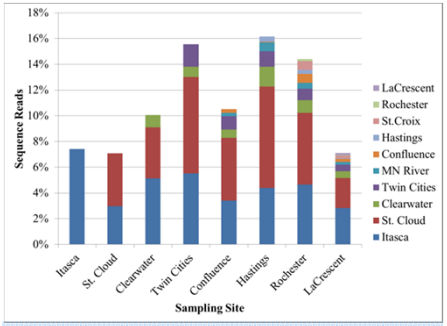


Figure 2. Distribution of sequence reads not shared among all sampling sites along the main branch of the Mississippi River. OTUs were assigned to sites at which they were first detected (farthest upstream). A mean of 90.5 ± 2.5% of sequence reads were shared among all sites.

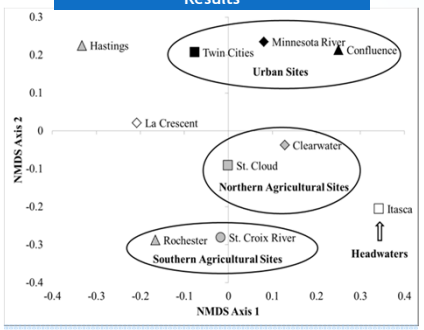


Figure 3. NMDS plot of total MCS calculated using the Bray-Curtis similarity coefficient. AMOVA analysis revealed nearly significant clustering by surrounding land use (primarily forest, agriculture, or urban) (P = 0.07)

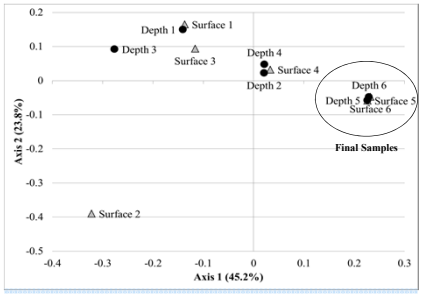


Figure 4. Principal coordinate analysis of biweekly surface and depth samples. Independent clustering of the final two surface and depth samples was supported by AMOVA analysis (P < 0.001).

Table 1. Summary of diversity indices among different sample volumes. S_{obs} is the number of OTUs and NP_Shannon is the non-parametric Shannon index.

Volume Filtered (L)	S_{obs}	Shannon	NP_Shannon	Simpson
0.33	9218	4.9	4.9	0.03
1.00	8990	4.8	4.8	0.04
2.00	8490	4.7	4.7	0.05
6.00	8432	4.6	4.6	0.05
Mean	8783	4.7	4.8	0.04
Standard Deviation	383	0.1	0.1	0.01

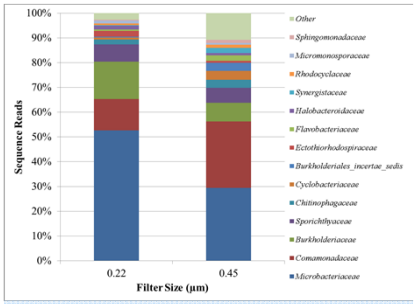


Figure 5. Distribution of the most abundant families in a sample filtered through a 0.45 µm pore-size filter followed by filtration through a 0.22 µm pore-size filter.

RIVERINE COMMUNITY STRUCTURE

- A small number of abundant OTUs were present at all sampling sites (12.3 ± 2.5%); these represented 90.5 ± 2.5% of sequence reads at each site.
- Community diversity was negatively correlated with distance from the headwaters at Itasca (r = -0.786, P = 0.021).
- Abundance of candidate division TM7, associated with pristine forest soils, was also negatively correlated with distance from the headwaters (r = -0.838, P = 0.009).
- Urban areas and those associated with river confluences (*e.g.* Confluence and Rochester sites) had significantly different MCS than other sites by weighted UniFrac ($P \leq$ 0.039).
- No significant differences among sites were revealed by unweighted UniFrac (P = 1) suggesting OTU presence/absence does not vary significantly among sites.
- The majority of non-ubiquitous OTUs were introduced at St. Cloud and were detected at relatively high proportions through downstream sites.
- A shift in MCS was observed among samples collected in late summer (P < 0.001).
- MCS at surface and depth in later summer, but not in earlier samples, differed significantly by weighted UniFrac (P < 0.001).
- 98.6 ± 0.7% of sequence reads were shared among paired surface and depth samples.
- 92.7 ± 1.4% of sequence reads were shared among all samples collected for biweekly and depth comparison.

EFFECT OF METHODOLOGICAL VARIATION

- Among OTUs from all sample volumes (n = 13,433), only 4.7 ± 0.6% of OTUs were unique to a specific volume.
- 31.0% of OTUs were common to all sample volumes representing 99.0 ± 0.1% of sequence reads.
- All diversity indices were within 10% of the mean with the exception of the Simpson index for 6 L, which deviated by 12.4%.
- Among OTUs collected using both filter sizes (n = 4,968), 13.6% were unique to the 0.22 µm filter and 39.7% of OTUs were unique to the 0.45 µm filter.
- 99.7% and 98.7% of sequence reads were shared among both filters for the 0.22 and 0.45 µm filters, respectively.

Discussion

- The MCS of the Upper Mississippi River in Minnesota is dominated by a small number of highly abundant OTUs.
- Shifts in MCS result primarily from variation in relative abundance of OTUs rather than presence/absence, as was reported for the English Channel[4].
- Variation in MCS appears to be influence by anthropogenic impacts and a small number of OTUs may indicate specific sources of contamination, as previously suggested[8].
- MCS in the Upper Mississippi River shows variation that may be related to seasonal dynamics as was previously described in a riverine ecosystem[3].
- Sampling volume, depth, and filter size minimally influence the estimated community structure of river water samples collected in this study area.

Species Sorting Dynamics in the Bacterial Community of the Upper Mississippi River are Influenced by Land Use and Sediment Resuspension

Christopher Staley[1], Trevor J. Gould[1,2], Ping Wang[1], Jane Phillips[2], James B. Cotner[3], and Michael J. Sadowsky[1,4]

[1]BioTechnology Institute, University of Minnesota, St. Paul, MN; [2]Biology Program, University of Minnesota, Minneapolis, MN, [3]Dept of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN [4]Dept of Soil, Water, and Climate, University of Minnesota, St. Paul, MN

Contact:
1479 Gortner Ave, Suite 140
St. Paul, MN 55108
612-625-1722
cmstaley@umn.edu

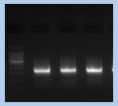
Introduction

- Variation in aquatic bacterial communities is associated with changes in physicochemical parameters (*e.g.* temperature, day length, and nutrient concentrations)[1,2].
- Shifts in community structure are seasonally reproducible[2] and these dynamics are more influential than sampling depth[3].
- Persistence of microbial constituents in marine, and potentially other aquatic reservoirs, suggests that taxonomic variation is due to environmental selection[4].
- Metacommunity theory has been proposed to explain interconnections of local communities[5], which are most likely effected by species sorting and mass effects.
- Environmental reservoirs and source-specific contamination (*e.g.* sediment resuspension and runoff/discharge) may also influence microbial communities[6].
- We hypothesize that
  - The bacterial community in the Upper Mississippi River is primarily shaped by species sorting dynamics;
  - Reproducible seasonal patterns in community structure occur as a result;
  - Environmental reservoirs and land use patterns are associated with changes in nutrient concentration and community structure.

Methods

- Sample in summer 2010-2012
- Filter water
- Elutriate cells (cell pellets reflect ~6 L)
- DNA extraction
- 16S PCR (967F/1046R, V6 region)
- Gel extraction
- Amplicon pooling
- Chemical analysis (Research Analytical Labs)
- Illumina sequencing (MiSeq/HiSeq)
- Triplicate (pseudo)replicates





mothur

Sequence processing using mothur[7]: quality trimming, alignment (SILVA[8]), chimera removal (UCHIME[9]), sample normalization (sub-sampling), classification against RDP database[10], alpha/beta diversity analysis; Sediment contribution was determined using SourceTracker[11]

References

1. Gilbert, J.A., Field, D., Swift, P., *et al.* (2009) *Environ Microbiol* 11, 3132-39.
2. Gilbert, J.A., Steele, J.A., Caporaso, J.G., *et al.* (2012) *ISME J* 6, 298-308.
3. Mueller-Spitz, S.R., Goetz, G.W. & McLellan, S.L. (2009) *FEMS Microbiol Ecol* 67, 511-22.
4. Caporaso, J.G., Paszkiewicz, K., Field, D., *et al.* (2012) *ISME Journal* 6, 1089-1093.
5. Leibold, M.A., Holyoak, M., Mouquet, N., *et al.* (2004) *Ecol Lett* 7, 601-13.
6. Newton, R.J., Bootsma, M.J., Morrison, *et al.* (2013) *Microb Ecol* 65, 1011-23.
7. Schloss, P.D., Westcott, S.L., Ryabin, T., *et al.* (2009) *Appl Environ Microbiol* 75, 7537-41.
8. Pruesse, E., Quast, C., Knittel, K., *et al.* (2007) *Nucleic Acids Res* 35, 7188-96.
9. Edgar, R.C., Haas, B.J., Clemente, J.C., *et al.* (2011) *Bioinformatics* 27, 2194-2200.
10. Cole, J.R., Wang, Q., Cardenas, E., *et al.* (2009) *Nucleic Acids Res.* 37, D141-45.
11. Knights, D., Kuczynski, J., Charlson, E.S., *et al.* (2011) *Nat Methods* 8, 761-U107.

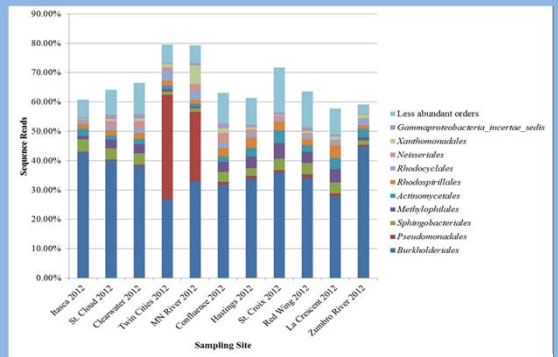


Figure 1. Distribution of orders of OTUs found to differ significantly by site in 2012 via Kruskal-Wallis test.

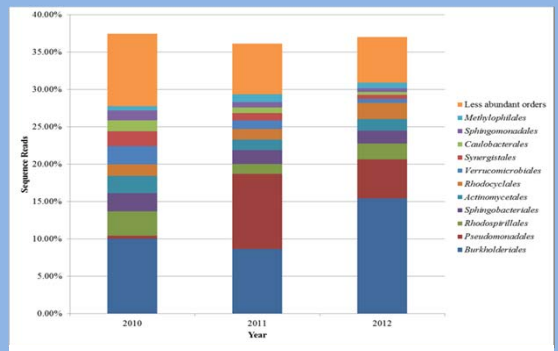


Figure 2. Distribution of orders of OTUs found to differ significantly by year via Kruskal-Wallis test.

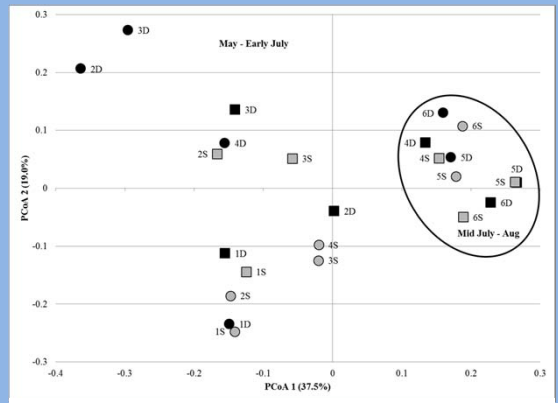


Figure 3. Principal coordinate analysis of surface and depth samples. The r-squared value relating ordination to the distance matrix for both axes is 0.87. Symbols represent samples collected in 2011 (circles), 2012 (squares), from the surface (S, shaded) and at depth (D, filled). Numbers refer to the order in which samples were collected.

RESULTS

COMMUNITY STRUCTURE

- A mean of ~1500 OTUs were identified in water vs. ~4700 in sediment.
- Within-year community variation was not significantly different among sampling sites (ANOSIM $P \geq 0.080$).
- Differences in communities among years showed significant variation in structure ($P \leq 0.008$).
- OTU abundances varied more by site than by year (Figs 1 & 2).
- In 2011 and 2012 communities at the Twin Cities site showed reproducible community structure in late summer that differed significantly from earlier samples (AMOVA $P < 0.001$; Fig 3).
- Impacts from sediments showed moderate influences on community structure (Table 1)

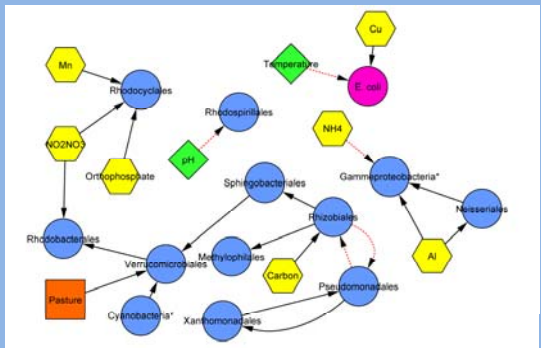


Figure 4. Consensus inferred Bayesian network relating bacterial orders, physicochemical parameters, chemicals, *E. coli*, and land use. Solid black lines indicate positive associations while dashed red lines indicate negative associations and arrows are directed from parent to child. Edge length is arbitrary. Asterisks (*) designate orders that are unclassified or designated *incertae sedis*.

COMMUNITY DYNAMICS

- Physicochemical parameters, land cover types, and relative abundances of bacterial orders were all significantly intercorrelated.
- Multiple linear regression (MLR) relating temperature, pH, rainfall, and distance from the headwaters revealed that only temperature and rainfall were significantly related to community diversity ($P \leq 0.048$).
- Among the most abundant phyla, abundances of *Actinobacteria* and *Bacteroidetes* were not related to any of the above parameters by MLR, and distance was only significantly associated with abundance of *Verrucomicrobia* ($\beta = 0.209$, $P = 0.045$).
- Bayesian modeling revealed specific bacterial orders were associated with nutrient and chemical concentrations, land use, and associations amongst themselves (Fig 4).
- Discriminant function analysis found that fourteen orders, most notably *Rhodobacterales*, *Gammaproteobacteria incertae sedis*, *Synergistales*, *Verrucomicrobia* Subdivision 3, *Prolixibacter*, *Aeromonadales*, and *Nitrosomonadales*, were significantly associated with major land cover types (*i.e.* developed, agricultural, or forested).

Table 1. Contribution of sediment to OTUs in water for 2012 samples.

Site	% Contribution
Itasca	9.0 ± 3.0
St. Cloud	13.3 ± 0.6
Clearwater	18.7 ± 0.6
Twin Cities	23.0 ± 1.7
MN River	57.3 ± 9.5
Confluence	20.0 ± 1.0
Hastings	18.7 ± 1.5
St. Croix River	14.3 ± 3.1
Red Wing	21.0 ± 1.7
La Crescent	11.0 ± 2.6
Zumbro River	13.7 ± 3.8

CONCLUSIONS

- While some dispersal dynamics were observed, variation in bacterial community structure is primarily associated with shifts in nutrient and physicochemical parameters.
- There is evidence that reproducible community dynamics result in similar community structure in the river in late summer each year.
- Sediments, and potentially other environmental reservoirs, can have considerable influences on the bacterial community composition in the water.
- Assessment of bacterial community structure may lend evidence to determine sources of non-point source pollution.

ACKNOWLEDGEMENTS

Funding for this project was provided, in part, by the American Recovery and Reinvestment Act of 2009 and the Minnesota Environment and Natural Resources Trust Fund, as recommended by the Legislative-Citizen Commission on Minnesota Resources. This work was carried out using computing resources at the University of Minnesota Supercomputing Institute.