ABSTRACT

Microbial Community Gene and Environmental Relationships in Phosphorus Biogeochemical Cycling within Streams

Erick S. LeBrun, Ph.D.

Mentor: Sanghoon Kang, Ph.D.

Microbial communities of Bacteria, Archaea, and Fungi are known to play integral roles in phosphorus (P) biogeochemical cycles. P is a vital ecosystem nutrient due to its utilization in many of the biological molecules and processes necessary for life. P is generally considered a limiting nutrient and a sudden influx or overabundance of it can cause drastic ecosystem effects. Stream systems are particularly sensitive to P inputs, primarily anthropogenic inputs, and suffer sever effects such as eutrophication as a result. The enclosed chapters take an in-depth look at microbial communities from all three taxonomic groups within stream and unique wetland environments representing P gradients or potentially unique P environments. We first characterize and investigate the relationships of microbes existing in the open water column to their local environment as well as environmental P. We then build on this foundation using functional metagenomics to explore microbial P cycle gene relationships to each other and then directly to environmental P. Throughout this process, we utilize frontier technologies, methods, and statistics to help elucidate these complex relationships. We further introduce new methods and analyses such as the System Relation Overview of Gene

Grouping (SROGG) and compare and contrast some existing methods for investigation like computational functional predictions and functional microarray analyses. Compiled, these studies offer a clearer picture of the quasi-mechanics of environmental microbial P cycling. They also highlight the magnitude of gap that exists relating to this type of study and offers a doorway along with tools for further research into the field.

Microbial Community Gene and Environmental Relationships in Phosphorus Biogeochemical Cycling within Streams

by

Erick S. LeBrun, B.S.

A Dissertation

Approved by the Department of Biology

_____

Dwayne D. Simmons, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Doctor of Philosophy

Approved by Dissertation Committee

_____

Sanghoon Kang, Ph.D., Chairperson

_____

Robert Doyle, Ph.D.

_____

William C. Hockaday, Ph.D.

_____

Ryan S. King, Ph.D.

_____

J. Thad Scott, Ph.D.

Accepted by the Graduate School

May 2018

_____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I acknowledge the support of the Baylor University Office of Research, the Department of Biology, and the Center for Reservoir and Aquatic Systems Research (CRASR). I thank the unyielding support of my advisor and mentor, Sanghoon Kang. I would also like to thank my chapter coauthors, Jeffery A. Back, Sanghoon Kang, Ryan S. King, and D. Lee Taylor.

DEDICATION

To Mary and Nyssa for their patience, understanding, and support.

CHAPTER ONE

Introduction

Background

Phosphorus (P) is one of three primary nutrients of importance in living systems alongside Carbon (C) and Nitrogen (N) (Miltner and others 1998). The nutrient P is important in numerous processes necessary for life including the production of ATP, nucleic acids, and phospholipids. There is a large body of evidence to suggest that P is the limiting nutrient in a number of systems (Schindler 1977; Smith 1984; Krom et al. 1991; Elser et al. 2007; Vitousek et al. 2010). Much of this work has been focused on freshwater ecosystems. An influx of anthropogenic P to a system can have significant impact on an environment and ecosystem, particularly a freshwater ecosystem resulting in adverse responses such as eutrophication (Sharpley et al. 1994; Correll 1998; Anderson et al. 2002). Anthropogenic sources are primarily wastewater along with agricultural fertilization and ranching (Correll 1998; Bennett et al. 2001) and these inputs are predicted only to increase in the future (Tilman et al. 2001). In consideration of this, it is important that we understand the interactions and response of ecosystems to the related increase in P.

It is well accepted that microbes are integral to the P cycle with environmental bacteria, archaea, and fungi being responsible for solubilization and mobilization of P. However, these mechanisms are little understood (Van Der Heijden et al. 2008; Richardson and Simpson 2011). Because of their unique P biogeochemical abilities and

foundational role in organic P cycling, microbes make an excellent target for studying P relationships in ecosystems.

Study of microbial communities and phosphorus to date has been largely focused on the field of agriculture and plant interactions in soil microbial communities (Van Der Heijden et al. 2008; Sharma et al. 2013; Fox et al. 2014). Microbial inoculations have shown potential for improving efficiency in P uptake by crops and improving the amount of P waste and fertilization cost involved in agriculture as well as to assist with growth in biochar soils used for carbon sequestration (Sharma et al. 2013; Fox et al. 2014).

Microbial relationships to P in environmental settings are also a topic of interest and are potentially more important than the agricultural questions as these studies both inform us on the environment and can be extrapolated for practical applications. The bulk of work on microbial communities in natural freshwaters has focused on the interactions of nutrients and biomass (Cotner and Biddanda 2002; Scott et al. 2012; Gorniak et al. 2013; Xu et al. 2013; Godwin et al. 2016). The mechanisms of P utilization and cycling in individual microbial taxa has occurred on specific genes such as *ppx* and *ppk* (Brewis et al. 1993; Qi et al. 1997; Prágai and Harwood 2002; Dyhrman et al. 2006; Kamat et al. 2011) and specific taxa (e.g. *Acinetobacter*) (Jansson 1993; Saralov et al. 2000; Prágai and Harwood 2002; Ghaffar et al. 2017). These studies cover only a very small proportion of the microbial taxonomic diversity involved in P cycling in nature. The majority of this work has also been conducted in simple and controlled environments than what is often seen in nature. Applying the findings from lab based studies to complex microbial communities consisting of large numbers of microbial taxa and populations in an environmental setting can prove difficult. This is made all the more

difficult because these microbial communities can be considered to consist of bacteria, archaea, and fungi with fungi requiring a differentiated approach due to variation in their genomics from bacteria and archaea.

There exists then several gaps in our understanding of microbial P biogeochemical processes. There is a gap in our understanding of microbial community structure and function relationships to P inputs in stream systems for both bacteria/archaea and fungi. A gap also exists in understanding the complex interactions and mechanisms involved in P biogeochemical cycling at the community and system levels. At the root of these issues exists a gap in the current reference material and lack of a generally accepted methodology for answering questions related to environmental microbial community structure and function for P specific nutrient cycling. Because of this, the microbial input to P biogeochemical systems is often written off as too complex and microbial interactions are summarized as general mineralization and solubilization (Richardson and Simpson 2011).

Project and Core Chapter Objectives

The chapters enclosed here are designed to address some of the afore mentioned gaps through experimental design, analysis, and interpretation. The approach taken is to first build the foundation of knowledge necessary to explore microbial community relationships to a range of environmental variables in a P sensitive system using genomics. Next is to build on this foundation looking at microbial relationships specifically to P cycle functions through a mechanistic lens. Then finally, to pave the way for future research in this field through the comparison of tools for researching nutrient specific microbial populations and relationships.

Chapter two takes a look at microbial structure, function, and relationships in a large stream system representing a gradient of P levels. This chapter is focused towards the goal of building the necessary foundation of knowledge for further study and looks specifically at bacteria and archaea. Structure is explored using next generation sequencing (NGS) of 16S rRNA genes to ascertain taxonomic composition and structure. Community function is constructed using computationally predicted functional metagenomes (PFMs). The chapter outlines the relationship of structure and function to multiple environmental variables including total phosphorus (TP) along with land usage metrics. Strong relationships exist between community structure and function in the chapter study. However, the relationships for structure and function differ significantly acting in a "decoupled" manner as outlined in more detail in the chapter.

Chapter three is also focused on the goal of building the necessary foundation for further exploring microbial community relationships to P using genomics. However, this chapter focuses on fungal community structure and functional relationships to environmental variables including TP in the same gradient system as chapter one. Structure is explored using NGS of ITS2 regions to ascertain taxonomic composition and structure. Community function is constructed using computationally predicted functional metagenomes (PFMs) using different software than was utilized for bacteria/archaea study. This chapter describes how fungi detected in the system are related to TP exclusively among collected environmental variables. A large number of terrestrial fungi were detected. The relationship found is not likely to be causative in either direction but rather representative of responses to non-collected catchment factors impacting each. The fungi found in the study are likely transient and allochthonous being present in the

context of dispersal. This suggests a potentially overlooked means of dispersal for terrestrial fungi.

Chapter four begins the next step of the project plan to look at microbial community relationships to P biogeochemical cycling through a mechanistic lens. In chapter four, whole community metagenomics sequencing is utilized to construct empirical functional profiles for communities from the same P gradient system and looking specifically at genes related to P cycling. The chapter uses a new, purpose created analysis called a System Relational Overview of Gene Groupings (SROGG) to illustrate relationships between functional gene groupings, communities, and P. The SROGG provides a quasi-mechanistic view of P utilization in the system. The chapter study finds that relationship involving orthophosphates (PO_4) and PO_4 levels themselves are overwhelmingly prevalent in the system. This is true even of more oligotrophic sites. Relationships to organophosphates show a much more site specific distribution that is not as strongly related to the levels of TP, PO_4, or organophosphate as it is to other catchment properties such as wastewater inputs.

Performing a study similar to the one in chapter four can be extremely costly and work intensive. To that end, chapter five compares to methods of obtaining microbial community function other than whole metagenome sequencing. The chapter looks at the ability of PFMs and another method, microarray detection, to detect diversity and community homogeneity in a unique basalt-soil wetland forest system. The project goal addressed by this chapter is testing and comparing potential tools for utilization in future nutrient specific genomic study. Chapter five illustrates how PFMs are likely the better

choice between the two methods for utilization is studies like the one presented in chapter three.

Author Contributions

Chapter one was written by Erick S. LeBrun (ESL) with input from Sanghoon Kang (SK). For chapter two, ESL and SK designed the experiment. ESL performed the experiment. Ryan S. King (RSK) oversaw sampling and field design of the study. Jeffery A. Back (JAB) conducted the nutrient analysis and participated in study sampling. ESL wrote the chapter with significant input and feedback from RSK, JAB and SK. For chapter three, ESL and SK designed the experiment. ESL performed the experiment. RSK oversaw sampling and field design of the study. JAB conducted the nutrient analysis and participated in study sampling. ESL wrote the chapter with significant input and feedback from D. Lee Taylor (DLT), RSK, JAB, and SK. For chapter four, ESL and SK designed the experiment. ESL performed the experiment. RSK oversaw sampling and field design of the study. JAB conducted the nutrient analysis and participated in study sampling. ESL wrote the chapter with significant input and feedback from RSK, JAB, and SK. For chapter five, ESL and SK designed the experiment. ESL performed the experiment. ESL wrote the chapter with significant input and feedback from SK. Chapter six was written by ESL.

Chapter References

Anderson DM, Glibert PM, Burkholder JM. 2002. Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. Estuaries 25:704–726. doi:10.1007/BF02804901.

Bennett EM, Carpenter SR, Caraco NF. 2001. Human Impact on Erodable Phosphorus and Eutrophication: A Global Perspective Increasing accumulation of phosphorus in soil threatens rivers, lakes, and coastal oceans with eutrophication. BioScience 51:227–234. doi:10.1641/0006-3568(2001)051[0227:HIOEPA]2.0.CO;2.

Brewis ND, Street AJ, Prescott AR, Cohen PT. 1993. PPX, a novel protein serine/threonine phosphatase localized to centrosomes. EMBO J. 12:987–996.

Correll DL. 1998. The Role of Phosphorus in the Eutrophication of Receiving Waters: A Review. J. Environ. Qual. 27:261. doi:10.2134/jeq1998.00472425002700020004x.

Cotner JB, Biddanda BA. 2002. Small Players, Large Role: Microbial Influence on Biogeochemical Processes in Pelagic Aquatic Ecosystems. Ecosystems 5:105–121. doi:10.1007/s10021-001-0059-3.

Dyhrman ST, Chappell PD, Haley ST, Moffett JW, Orchard ED, Waterbury JB, Webb EA. 2006. Phosphonate utilization by the globally important marine diazotroph Trichodesmium. Nature 439:68–71. doi:10.1038/nature04203.

Elser JJ, Bracken ME, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, Ngai JT, Seabloom EW, Shurin JB, Smith JE. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecol. Lett. 10:1135–1142.

Fox A, Kwapinski W, Griffiths BS, Schmalenberger A. 2014. The role of sulfur- and phosphorus-mobilizing bacteria in biochar-induced growth promotion of Lolium perenne. FEMS Microbiol. Ecol. 90:78–91. doi:10.1111/1574-6941.12374.

Ghaffar S, Stevenson RJ, Khan Z. 2017. Effect of phosphorus stress on Microcystis aeruginosa growth and phosphorus uptake. PLoS One San Franc. 12. doi:http://dx.doi.org/10.1371/journal.pone.0174349.

Godwin CM, Whitaker EA, Cotner JB. 2016. Growth rate and resource imbalance interactively control biomass stoichiometry and elemental quotas of aquatic bacteria. Ecology. [accessed 2017 Feb 13]. http://onlinelibrary.wiley.com/doi/10.1002/ecy.1705/full.

Gorniak A, Więcko A, Cudowski A. 2013. Fungi biomass in lowland rivers of north-eastern Poland: Effects of habitat conditions and nutrient concentrations. Pol J Ecol 61:749–758.

Jansson M. 1993. Uptake, Exchange, and Excretion of Orthophosphate in Phosphate-Starved Scenedesmus quadricauda and Pseudomonas K7. Limnol. Oceanogr. 38:1162–1178.

Kamat SS, Williams HJ, Raushel FM. 2011. Intermediates in the transformation of phosphonates to phosphate by bacteria. Nature 480:570–573. doi:10.1038/nature10622.

Krom MD, Kress N, Brenner S, Gordon LI. 1991. Phosphorus limitation of primary productivity in the eastern Mediterranean Sea. Limnol. Oceanogr. 36:424–432.

Miltner RJ, others. 1998. Primary nutrients and the biotic integrity of rivers and streams. Freshw. Biol. 40:145–158.

Prágai Z, Harwood CR. 2002. Regulatory interactions between the Pho and σB-dependent general stress regulons of Bacillus subtilis. Microbiology 148:1593–1602. doi:10.1099/00221287-148-5-1593.

Qi Y, Kobayashi Y, Hulett FM. 1997. The pst operon of Bacillus subtilis has a phosphate-regulated promoter and is involved in phosphate transport but not in regulation of the pho regulon. J. Bacteriol. 179:2534–2539. doi:10.1128/jb.179.8.2534-2539.1997.

Richardson AE, Simpson RJ. 2011. Soil microorganisms mediating phosphorus availability update on microbial phosphorus. Plant Physiol. 156:989–996.

Saralov A, Mol'kov D, Bannikova O, Solomennyĭ A, Chikin S. 2000. [Intracellular accumulation of monomer precursors of polyphosphates and polyhydroxyalkanoates from Acinetobacter calcoaceticus and Escherichia coli]. Mikrobiologiia 70:737–744.

Schindler DW. 1977. Evolution of phosphorus limitation in lakes. Science 195:260–262.

Scott T, Cotner J, LaPara T. 2012. Variable Stoichiometry and Homeostatic Regulation of Bacterial Biomass Elemental Composition. Front. Microbiol. 3. doi:10.3389/fmicb.2012.00042. [accessed 2017 Feb 13]. http://journal.frontiersin.org/article/10.3389/fmicb.2012.00042/abstract.

Sharma SB, Sayyed RZ, Trivedi MH, Gobi TA. 2013. Phosphate solubilizing microbes: sustainable approach for managing phosphorus deficiency in agricultural soils. SpringerPlus 2:1.

Sharpley AN, Chapra SC, Wedepohl R, Sims JT, Daniel TC, Reddy KR. 1994. Managing agricultural phosphorus for protection of surface waters: Issues and options. J. Environ. Qual. 23:437–451.

Smith SV. 1984. Phosphorus versus nitrogen limitation in the marine environment. Limnol. Oceanogr. 29:1149–1160.

Tilman D, Fargione J, Wolff B, D'Antonio C, Dobson A, Howarth R, Schindler D, Schlesinger WH, Simberloff D, Swackhamer D. 2001. Forecasting Agriculturally Driven Global Environmental Change. Science 292:281–284. doi:10.1126/science.1057544.

Van Der Heijden MG, Bardgett RD, Van Straalen NM. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecol. Lett. 11:296–310.

Vitousek PM, Porder S, Houlton BZ, Chadwick OA. 2010. Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen–phosphorus interactions. Ecol. Appl. 20:5–15.

Xu X, Thornton PE, Post WM. 2013. A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. Glob. Ecol. Biogeogr. 22:737–749. doi:10.1111/geb.12029.

CHAPTER TWO

Microbial Community Structure and Function Decoupling Across a Phosphorus Gradient in Streams

This chapter published as: LeBrun ES, King RS, Back JA, Kang S. 2018. Microbial Community Structure and Function Decoupling Across a Phosphorus Gradient in Streams. Microb Ecol 75:64–73. doi:10.1007/s00248-017-1039-2.

Abstract

Phosphorus (P) is a key biological element with important and unique biogeochemical cycling in natural ecosystems. Anthropogenic phosphorus inputs have been shown to greatly affect natural ecosystems, and this has been shown to be especially true of freshwater systems. While the importance of microbial communities in the P cycle is widely accepted, the role, composition, and relationship to P of these communities in freshwater systems still hold many secrets. Here, we investigated combined bacterial and archaeal communities utilizing 16S ribosomal RNA (rRNA) gene sequencing and computationally predicted functional metagenomes (PFMs) in 25 streams representing a strong P gradient. We discovered that 16S rRNA community structure and PFMs demonstrate a degree of decoupling between structure and function in the system. While we found that total phosphorus (TP) was correlated to the structure and functional capability of bacterial and archaeal communities in the system, turbidity had a stronger, but largely independent, correlation. At TP levels of approximately 55 μg/L, we see sharp differences in the abundance of numerous ecologically important taxa related to vegetation, agriculture, sediment, and other ecosystem inhabitants.

Introduction

Microbial communities are extremely important in environmental and biogeochemical processes in nature with diversity being a key component [1–3]. It is well established that in addition to being P consumers, microbial communities affect the amount of available P for other ecosystem inhabitants through accessing P that is unavailable to other organisms and making it available [2–5]. Recent culture-based work looking at microbial communities and stoichiometric relationships between nutrients in freshwaters has shown that community assemblies shift dependent on phosphorus availability as well [6, 7]. This demonstrates a complex relationship where microbial community assemblages affect environmental P and environmental P affects microbial community assemblages in return [8, 9].

Shifts in environmental P may have impacts on community structure, diversity, and functional capability as well [10]. The results of previous studies exploring the relationship between species composition and function using a wide range of methods have been inconsistent. If we consider a strong correlation between community structure and function to be "coupled," and a disconnect between the two to be decoupled, several studies have found community structure and function coupling [11, 12], others show decoupling [10, 13, 14], and many show a mixture of results [15, 16]. A mixture of coupling and decoupling or varying degrees of coupling seems the most likely response as some studies finding coupling such as in Garcia-Palacios et al. [11] hint at decoupled aspects and studies finding decoupling do not present the decouple as a complete disconnect. A large degree of uncertainty exists in lotic systems about the resiliency of microbial communities to environmental stress [17]. A decoupling of structure and

function could indicate potential differences in community resiliency to environmental conditions. Functional redundancy of aquatic microbial communities may also be a good indicator of community resiliency to environmental stress [18, 19].

Environment-driven microbial community structuring in aquatic systems has been studied extensively since the 1970s. Until recently, this study has largely been conducted using microscopy and low-resolution community fingerprinting methods such as DGGE and TRFLP while relatively little work has been performed using high-throughput approaches such as next generation sequencing (NGS) technologies to better capture complex microbial communities [20]. Studies of microbial diversity have been a keystone of aquatic ecosystem research, but lotic systems have been underrepresented when compared to marine and lentic systems [21]. River and stream ecosystems are sensitive to anthropogenic P inputs often resulting in eutrophication under increased P levels [22–24]. Decreasing the gap in understanding between the P in lotic systems and microbial community structure has the potential to act as the foundation for bringing the mechanisms of this relationship to light.

The purpose of this study was to examine microbial community structure, function, and diversity across a total phosphorus (TP) gradient in freshwater streams. We hypothesized that (1) in an environment displaying a strong TP gradient, TP would be a strong correlate of community structure and (2) the overall functional capability of the sampled communities would be strongly coupled to their structural assembly in their relationship to TP. We tested these hypotheses by using next generation sequencing to look at microbial community structure and predicted functional capability in a system of streams representing a strong TP gradient.

Methods

Sampling

The study area was a collection of streams and rivers along the Oklahoma-Arkansas border, an area with known P enrichment problems [25–27]. We selected sampling sites with TP levels ranging from 18 to 163 µg/L. Sampling was performed in October of 2014. At each of the 25 sampling sites (Appendix A: Fig. A.S1), 50 mL of water was collected approximately 10 cm below the water surface and vacuum filtered with approximately 10 kPa of pressure through two filter sizes. A stacked filtration through a 1-µm PALL-type A/E glass fiber filter was conducted and then a PALL SUPOR 200 0.2-µm filter for bacterial and archaeal collection. Only the 0.2-µm filters were used in this study. Filters were stored at 4 °C less than 7 days for transport and then transferred to −80 °C until processed for DNA extraction.

Environmental Data

Total phosphorus (TP), total N (TN), sestonic chlorophyll-a, total suspended solids (TSS), turbidity, pH, specific conductance, dissolved oxygen, and temperature were measured at each site at the same time when bacterial and archaeal samples were collected. Molar concentrations of carbon (C), P, and N were calculated from dissolved organic carbon (DOC), TP, and TN and used to calculate ratios of C/P and C/N.

Unfiltered water samples were used for TP and TN. TP was digested with persulfate in an autoclave at 121 °C for an hour then analyzed on a Lachat 8500 series 2 using the ascorbic acid-molybdate method [28]. TN was digested with persulfate and

sodium hydroxide (NaOH) in an autoclave at 121 °C for an hour. Samples were cooled and boric acid solution was added, then the sample was analyzed on a Lachat 8500 series 2 using the sulfanilamide cadmium reduction method [28]. For sestonic chlorophyll-a, 1 L of water was filtered onto a 0.45-µm filter then frozen. A 90% ethanol extraction was performed in a heated (78 °C) water bath for 10 min [29]. Total suspended solids (TSS), turbidity, pH, specific conductance, dissolved oxygen, and temperature were measured at each site at the same time when bacterial and archaeal samples were collected using a YSI EXO2 multiparameter data sonde (Yellow Springs, OH).

TP was measured at each site in August 2014 and October 2014 because of the likelihood that antecedent P conditions would drive bacterial and archaeal community composition as estimated by the October 2014 sampling. The October 2014 sampling event was a low water event resulting in abnormally high point data TP levels at some sites; thus, an average TP spanning August and October sampling events was used for subsequent analyses. This average was more representative of typical stream TP levels while still capturing the increase in TP leading into sampling.

Catchments were delineated in ArcGIS version 10 by utilizing the geographic coordinates of each stream reach to define a catchment outlet. We estimated the contributing area of each outlet on the basis of flow accumulation values derived from a 30-m digital elevation model (DEM) from the USGS National Elevation Dataset. Land use in each catchment was estimated from the National Land Cover Data (NLCD) raster, which was based on the 2011 Landsat Thematic Mapper.

Library Preparation

DNA was extracted from the 0.2-µm filters using a Mo Bio PowerWater® DNA extraction kit with the manufacturer's protocol. Quantification of extractions was performed using a Qubit 3.0 fluorometric system, and samples showing no DNA extraction were discarded from the study at this time. DNA yields were less than 1 ng/µL causing us some initial concern about coverage but resulted in sequence depth of ~70,000 up to ~250,000 sequences per site and good sampling coverage per the rarefaction curves generated downstream (Appendix A: Fig. A.S2.A).

An initial PCR amplification of the 16S ribosomal RNA (rRNA) gene V4–V5 region was conducted using 515F forward and 926R reverse primers [30] modified to include adapters for future indexing. PCR was done using 2X Platinum™ Green Master Mix from Invitrogen. PCR specifications were 1 cycle for HotStart step at 94 °C for 2:00 min, then 30 cycles of 94 °C denaturation step for 0:45 min, 50 °C annealing step for 1:00 min, and 72 °C elongation step for 1:30 min. Successful PCR was identified through electrophoresis gel. PCR cleanup was conducted using an Agencourt AMPure XP kit and with the standard protocol from Beckman Coulter Life Sciences. Final PCR product quantification was conducted using Qubit 3.0.

A second round of PCR amplification was run in order to add unique indices to each sample as well as Illumina sequencing adapters. PCR was again done using 2X Platinum™Green Master Mix kit from Invitrogen. PCR specifications for the second round of PCR were 1 cycle for HotStart step at 94 °C for 2:00 min, then 8 cycles of 94 °C denaturation step for 0:45 min, 59 °C annealing step for 1:00 min, and 72 °C elongation step for 1:30 min. Cleanup and quantification were performed in the same manner as the

first round of PCR. Samples were then pooled so that 10 ng of DNA from each sample was present in the final library.

Sequencing

Sequencing was performed on an Illumina MiSeq system using a MiSeq Reagent Kit v3 2X300 with paired-end reads. Libraries were spiked with 20% PhiX Control.

Sequence Processing

Initial sequence processing including filtering and demultiplexing was conducted through the 16S metagenomics pipeline in Illumina BaseSpace. Paired-end read fastq files for each sample were extracted for downstream processing. Additional sequence processing was done using Quantitative Insights Into Microbial Ecology (QIIME) [31]. Paired-end reads were aligned first using the join-fastq algorithm from ea-utils [32]. Resulting sequences were then filtered at a Phred score of 20. Chimeric sequences were identified and removed using the USEARCH algorithm [33]. OTU picking was performed open reference using uclust against the Greengenes 13_8 database with a 0.97 similarity cutoff followed by de novo OTU picking [34]. Alpha diversity metrics were computed by QIIME as part of this pipeline using a rarefaction depth of 70,500 allowing use of all sites included in this study.

Construction of Predicted Functional Metagenomes

Predicted functional metagenomes (PFMs) were constructed from 16S rRNA gene sequence data using Phylogenetic Investigation of Communities by Reconstruction of

Unobserved States (PICRUSt) [35]. OTU tables were normalized by count, and the PFMs were generated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [36, 37]. PFMs were then grouped at KEGG level 3 representing the most diverse and specific functional groupings available.

Statistical Analysis Software

All analyses were performed in the R software package v.3.2.3 [38] with appropriate packages and scripts. OTU table BIOM files from QIIME were either exported to tab delimited format directly from QIIME or were imported for use in R using the phyloseq package version 1.14.0 in R [39].

Ordinations, Models, and Comparisons

Non-metric multidimensional scaling (NMDS) ordinations were constructed using the vegan package version 2.4–0 [40] to describe community dissimilarity in unconstrained space. OTU tables were log10(x + 1) transformed, and distances were computed as both Bray-Curtis and UniFrac. NMDS for Bray-Curtis distance 16S rRNA data and PFMs were performed on three axes, and 16S rRNA UniFrac distance was also performed on three axes. Redundancy analysis (RDA) models were also built in the vegan package in order to describe the community structure ordinations in environmentally constrained space. NMDS and RDA plots were created using the ggplot2 package in R [41]. Environmental gradients were built on NMDS ordinations using ordisurf from the vegan package. ordisurf uses general additive model (GAM) model building to overlay environmental variables in the ordination space [42]. RDA model selection was performed by starting with an initial model including all collected

variables and manually removing collinear variables in an effort to maximize adjusted R^2 and minimize the magnitude of difference between R^2 and adjusted R^2.

16S rRNA and PFMdata was compared with a Mantel test using the mantel function in the vegan package with the Pearson method and 1000 permutations in order to test for similarity between the data sets [43]. PROcrustean randomization TEST of community environment concordance (PROTEST), a potentially more sensitive detection method than a Mantel test, was also used to compare NMDS ordinations as well as RDAs using the protest function in the vegan package [44]. PROTEST uses scaling and rotations to maximize alignment in ordinations as a multivariate measure of concordance in species abundance and environmental datasets along with permutation-based significance testing [44]. All PROTEST analyses were performed with 1000 permutations.

Environmental Groupings and Analysis

Because preliminary environmental fitting against NMDS ordinations as well as RDA models suggested that turbidity and TP were both important in the system; sites were divided into turbidity and TP groups of Low, Med, and High for each variable independently. Group cutoffs were determined at apparent breakpoints in the distribution of our collected turbidity and TP data. Low for TP was set for sites below 40 µg/L and High as above 70 µg/L. Low for turbidity was set for sites below 1 Nephelometric Turbidity Units (NTU) and High as above 2.0 NTU. The designation of groupings for each site can be viewed in Appendix B: Table B.S1. Multivariate ANOVA-type analyses including PERMANOVA and ANOSIM were then performed with 1000 permutations

using the vegan package to test for significant relationships between and among groupings.

Network Construction and Analysis

In order to visualize site relatedness and clustering, networks were constructed from phyloseq imported data using the network package version 1.13.0 and igraph package version 1.0.1 in R and Bray-Curtis distance [45–47]. The Bray-Curtis cutoff used for each network was the smallest value that allowed for the inclusion of all 25 sites. For 16S rRNA data, this value was 0.72, and for PFM data it was 0.028. Network clustering for each network was performed using a spin-glass model and simulated annealing via the cluster_spinglass function in the igraph package. Clustering from networks was compared using the normalized mutual information method (NMI) via the compare function in igraph package. Heatmaps were built using the heatmap.2 function in the gplots package version 3.0.1 [48]. Heatmap dendrograms were built using Bray-Curtis distance and uclust WPGMA clustering. Taxa that do not represent more than 1.5% of relative abundance in any site were removed post clustering for visual clarity.

Total Phosphorus Range of Greatest Impact

Threshold analysis for TP and turbidity was performed using the TITAN 2.1 package in R [49]. Threshold Indicator Taxa Analysis (TITAN) identifies environmental variable values maximizing taxa frequency and abundance using bootstrapping to identify reliable indicator taxa and the sum of indicator taxa value Z-scores to identify the environmental values representing the peak of increase or decline of the taxa [50]. The number of bootstraps performed in our TITAN 2.1 analyses was 500. For genus-level

analysis, referenced genera observed more than three times were used, and for OTU-level analysis, OTUs observed more than 15 times were used. The same analysis was performed for turbidity as well.

Functional Redundancy Testing

Functional redundancy can be represented by the relationship between functional and species diversity with less correlation representing more redundancy [19, 51]. In order to test for functional redundancy in the system, Shannon diversity scores were computed for all sites in both the 16S rRNA and PFM data sets. Shannon diversity scores for each data set were plotted against each other, and potential correlations were explored using a GAM through the mgcv package version 1.8- 11 in R [52]. GAM models were also used to model Shannon diversity correlation with turbidity for both 16S rRNA and PFM data.

Results

Environmental Relationship to Community Structure and Function

Vector fitting of environmental variables onto the NMDS ordinations identified turbidity and TP as the strongest and most significant environmental factors. The NMDS ordinations overlaid with turbidity and TP help to show trends of placement in ordination space with environment gradients for both 16S rRNA data and PFM data (Fig. 2.1). Contours for both turbidity and TP fit well with the distribution of sites in ordination space albeit in somewhat orthogonal directions. The turbidity gradient represents an almost linear fit to the ordination space distribution for the PFM data whereas more

smoothing was involved in the 16S rRNA data. 16S rRNA UniFrac and Bray-Curtis ordinations were very similar with a PROTEST correlation statistic of 0.88 ($p = 0.001$).

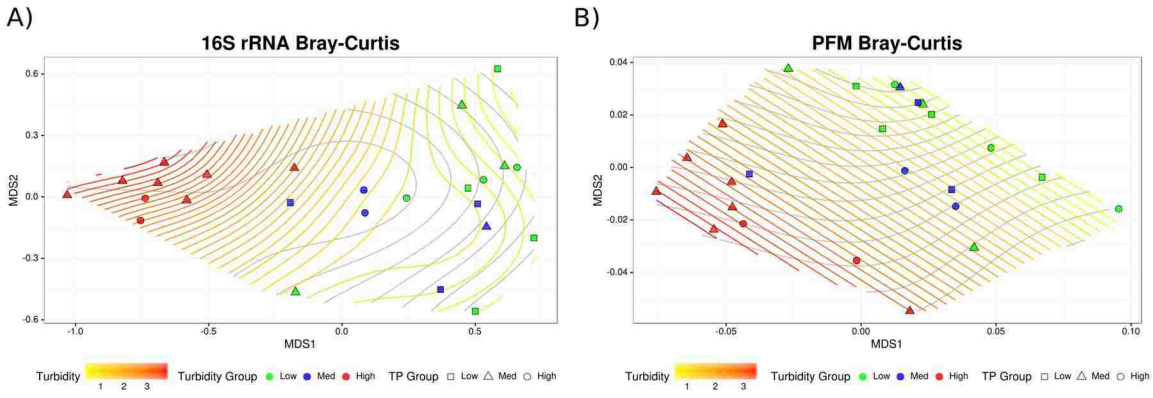


Figure 2.1: NMDS ordinations. A) 16S rRNA community with Bray-Curtis distance. Yellow to red gradient represents turbidity environmental fit, and gray lines represent TP (stress 0.067). B) Predicted functional metagenome with Bray-Curtis distance. Yellow to red lines represent turbidity environmental fit, and gray lines represent TP (stress 0.030)

The RDA model for 16S rRNA data included turbidity, TN, pasture land coverage, and C/P ratios (Fig. 2.2.a). TP was collinear with TN so it was removed from the model. The model had an adjusted R^2 value of 0.107, and an ANOVA-like permutation test for the constraining variables was significant with a p value of 0.001. The RDA model for the PFM data included turbidity, TP, developed land coverage, C/N ratios, and C/P ratios (Fig. 2.2.b). The model-adjusted R^2 was 0.394, and an ANOVA-like permutation test for the constraining variables was significant with a p value of 0.001.

PERMANOVA and ANOSIM on the 16S rRNA data for TP grouping were both significant. Only PERMANOVA was significant for TP grouping on the PFM data where ANOSIM was not (Table 2.1). PERMANOVA and ANOSIM on the 16S rRNA data for turbidity groupings were also both significant. The analyses were also significant for the PFMs and turbidity groupings (Table 2.1)

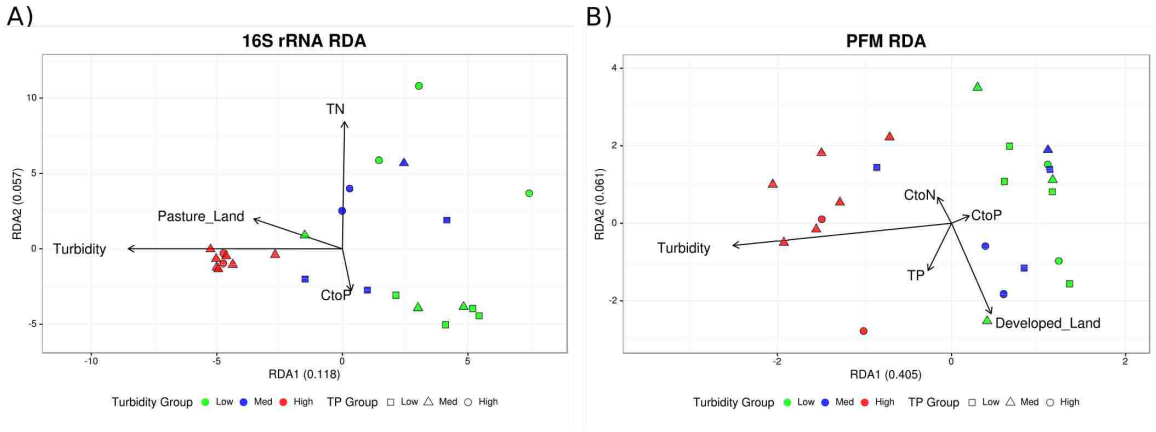


Figure 2.2: RDA model constrained ordinations. A) RDA model for the 16S rRNA community (adj. $R^2 = 0.106$, $p = 0.001$). B) RDA model for PFMs (adj. $R^2 = 0.380$, $p = 0.001$)

Table 2.1: Results from multivariate analyses.

Statistical approaches		Turbidity Grouping		TP Grouping	
		16S rRNA	PFM	16S rRNA	PFM
ANOSIM	*R*	0.636	0.432	0.213	0.102*
	p	**<0.001**	**0.002**	**0.013**	0.077*
PERMANOVA	*F*	3.401	7.92	1.735	2.710
	p	**0.002**	**0.002**	**0.013**	**0.037**

* Not significant ($\alpha = 0.05$).

Structure and Function Relationship

The OTU table and PFM, and dissimilarity matrices of each, were prepared for the structural and predicted functional capacity aspects of the communities. A Mantel test of 16S rRNA and PFM data returns no significant correlation between the dissimilarity matrices of the two data sets. The PROTEST correlation statistic for the RDA models was 0.57 ($p = 0.01$), and PROcrustes error plotting shows varying magnitudes and directions of change in ordination space (Appendix A: Fig. A.S3).

Linear and GAM modeling proved unproductive for diversity relationship to TP. A GAM model for 16S rRNA diversity shows a significant negative association between turbidity and Shannon diversity index values (R^2 adj. = 0.322, p = 0.008) (Appendix A: Fig. A.S4.A). A GAM model of PFM diversity against turbidity shows a negative association with turbidity as well (R^2 adj. = 0.593, $p < 0.001$) (Appendix A: Fig. A.S4.B).

16S rRNA and PFM networks by site relatedness were also visually quite different with very different clustering. In addition to being visually different, network clustering differed statistically with an NMI score of 0.430. Networks for 16s rRNA and PFMs appeared to both cluster well visually with turbidity groupings rather than TP groupings (Appendix A: Fig. A.S5). Many nodes differ in linked partners, and clusters differed in composition, number of connections to other clusters, and connection sites.

Total Phosphorus Range of Greatest Impact

TITAN 2.1 identified five separate indicator genera decreasing in abundance (decreasers) and eight indicator genera increasing in abundance (increasers) in response to TP levels. OTU-based analysis identified a large number of OTUs in both the increaser and decreaser categories. Increasers and decreasers intersect at just below 55 μg/L of TP (Fig. 2.3). The 95% confidence intervals for the analysis identify a range from 52 to 58 μg/L TP for decreasers and 74.5 to 79.6 μg/L for increasers. In terms of sheer number of taxa, decreasers were much more prevalent than increasers. A TITAN analysis performed on turbidity indicated a strong threshold value right around 2.0 NTU mostly of decreasers, but interestingly, *Sediminibacterium* saw a significant increase (Appendix A: Fig. A.S6).

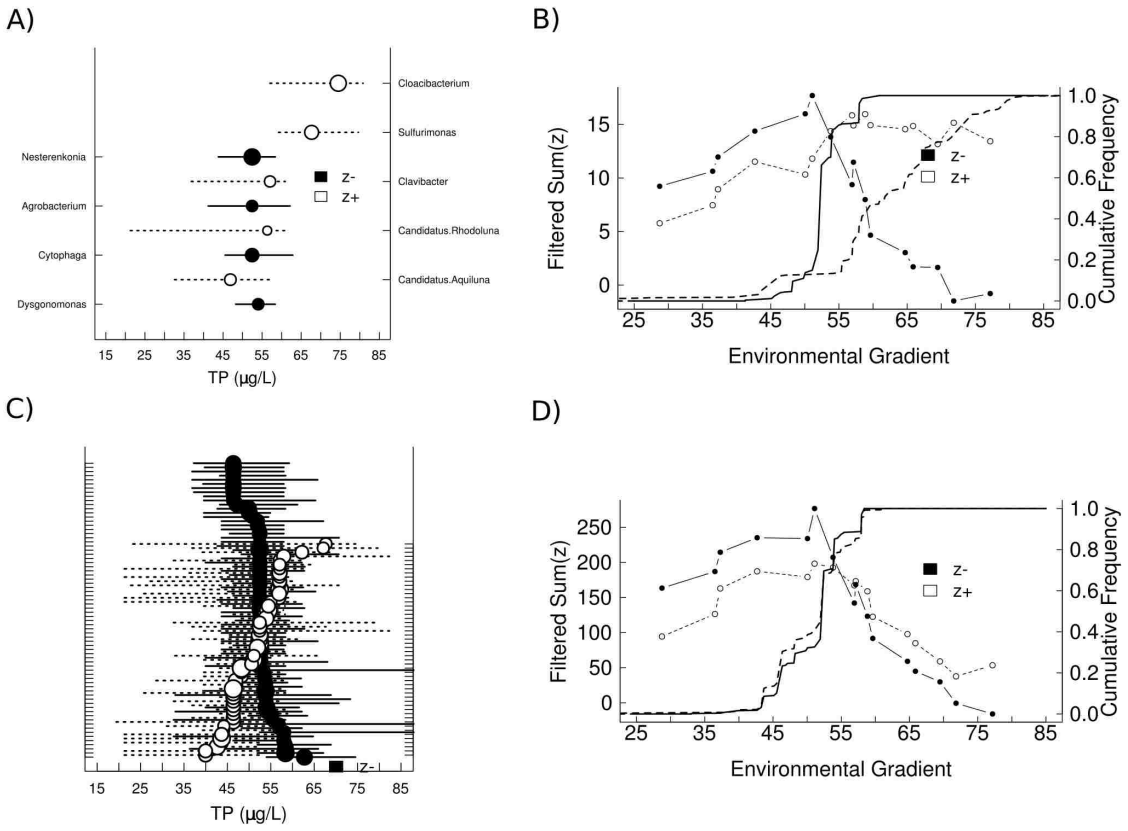


Figure 2.3: TITAN 2.1 analysis. A) Plot of pure and reliable indicator taxa along the TP gradient. Black symbols correspond to genera that declined with increasing TP (z−), whereas open symbols correspond to those that increased (z+). Symbols are sized in proportion to the magnitude of the response (z-score). Horizontal lines represent 5th and 95th quantiles of values of TP resulting in the largest change in genus zscores among 1000 bootstrap replicates. B) Plot of sum z-scores for genus level taxa. Steep slopes indicate major change points in abundance. C) Plot of pure and reliable OTUs changing abundance around the threshold region. D) Plot of sum z-scores for OTU-level taxa

Diversity and Community Taxonomy

The heatmap in Fig. 2.4 illustrates abundance trends across sites at the order level. Several unique genera displayed strong differences in relative abundance between groupings based on TP. Relative abundance of *Acinetobacter* from the order *Pseudomonales* varied with TP grouping increasing from very little (0.89%) in high TP sites to relatively much more in low TP sites (9.84%) (Appendix A: Fig. A.S2.B).

Bacteria of the family *Cytophagaceae* were also much more abundant in high TP sites (Appendix A: Fig. A.S2.B). Abundance of bacteria of the genus *Fluvicola* from the order *Flavobacteriales* also increased with TP showing greater abundance in higher groupings.

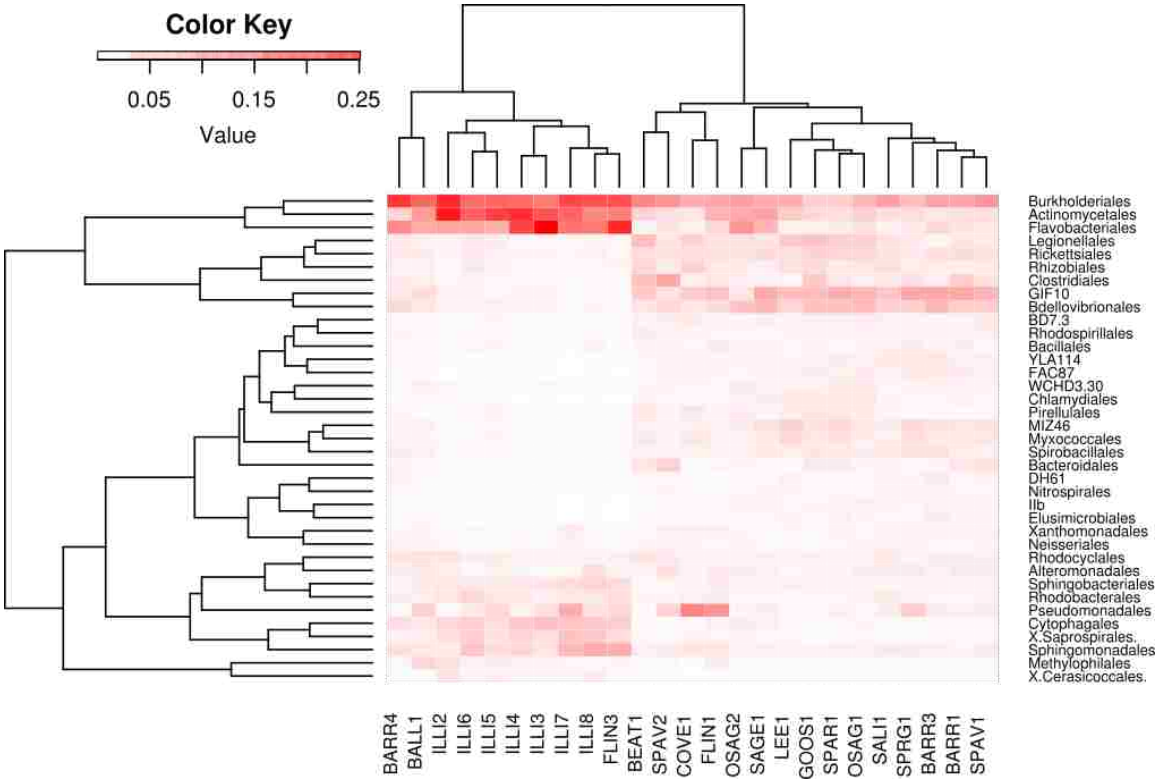


Figure 2.4: Heatmap of 16S community structure at the Order level by site (taxa observed >1.5% of total abundance only)

Discussion

Turbidity

Turbidity was an important factor in our analyses despite utilizing only the 0.2- to 1-µm fraction which likely removed many larger particle-associated bacteria. We would expect this to limit a large amount of particle associated bacteria, but the turbidity factor

still had a strong relationship to both the structure and the functional capacity of the studied communities.

The primary ecological effect of turbidity in stream systems is to limit light availability [53]. Interactions between light availability and phosphorus on algal growth have previously been demonstrated as important for algal growth and community composition [54]. Turbidity also introduces sediment to the system and, along with it, particle-associated organisms. We did see soil-associated taxa such as Pseudomonas, Cytophaga, Micrococcus, Bacillus, and Agrobacterium present in our communities despite the focus on the 0.2- to 1-μm fraction. It is important to note, however, that the highest turbidity Nephelometric Turbidity Unit (NTU) values from the study sites were still relatively low (0.5–3.5 NTU) when compared to large rivers, lakes, or estuaries that are considered turbid, where turbidity typically exceeds >10–20 NTU [55]. Sampling for this study also occurred under base flow conditions representing the likely lowest turbidity conditions.

These factors make it difficult to separate out precisely what factors turbidity may be representing. Turbidity showed no correlation to total chlorophyll, and although it did correlate with total suspended solids and was collinear with several terrestrial type factors such as catchment area and flow, that is not the whole picture. Turbidity can be a poor measurement of sedimentation [56], and preliminary data exploration shows that sediment was likely not as prevalent in this sampling event as it is during other times of the year with similar turbidity measures. Total suspended solids only represented a small proportion of the turbidity measure in our data and was lower ($x = 2.0$ mg/L, $s = 1.75$)

than 10 historical measurements on all sites over the course of 2 years ($x = 3.7$ mg/L, $s = 6.87$) within the same system as well.

The water column for the most turbid sites possessed an uncharacteristic and unidentified gray tint to the naked eye that was not readily apparent during other months of the year. This may be indicative of a biological factor such as the bacterial colonization of mineral particles or bacterial aggregates increasing turbidity scores. Therefore, turbidity here is probably a complex mixture of terrestrial and biological factors but further work is needed in order to explain this phenomenon.

Structure and Function Relationship

In addition to visual differences, the results of multiple analyses including the Mantel test, PROcrustes, and PROTEST analyses together indicate that structure and predicted functional capacity likely have features that are both coupled and decoupled in this system. Rather than seeing complete coupling or decoupling, we instead see indications of a degree of some decoupling. This appears to support the inconsistent results seen in previous studies and may mean a different result in terms of the degree of coupling when focusing on different specific ecosystem functional traits as opposed to the entire functional metagenome [15, 16].

Diversity relationship to turbidity appears to be one strongly coupled facet in this system. This indicates a potentially large environmental impact on taxa diversity with ongoing pollution and eutrophication of streams and rivers [57] which according to our results could lead to reduced functional diversity as well because the two appear to share a coupled diversity response. This is in contrast to the diversity relationship to TP which resulted in no productive linear or GAM models.

Functional Redundancy

We considered one possible explanation for any decoupling that we saw between structure and function to be functional redundancy. The GAM model for redundancy indicates a strong relationship between functional capability and species diversity in this system (Appendix A: Fig. A.S7). There does not appear to be much functional redundancy which could imply that environmental changes affecting diversity could have major implications for these communities.

The similarity of 16S rRNA UniFrac and Bray-Curtis ordinations seems to indicate that these communities have a strong evolutionary structural component and may maintain and develop unique communities over longer periods of time than would be expected in this type of fast moving lotic system. A recent study in catchment bacterioplankton found evidence that spatial factors are more important than temporal factors in determining community composition in lotic river systems supporting the development of unique community structures based on location and environmental parameters [58].

Nutrient Stoichiometric Relationships

The appearance of C/P and C/N ratios in our models is supported in previous research including research into changes in biomass incorporation based on C/P ratios [4]. The purpose of our study was to explore microbial communities across a gradient of TP, but the effects of C/P ratios should continue to be explored. Our results indicate only a relationship but not the direction of the relationship. Interestingly, the directions of TP and C/P ratios in the models were almost antithetical which is also supported in the

literature [59]. The same was true for TN and C/N ratios. DOC by itself was not a strong correlate in any of our analyses although possibly due to collinearity.

Total Phosphorus Range of Greatest Impact

The steep decline of taxa near the ~55-µg/L TP value identified by TITAN analysis fell in the middle of our "middle" grouping for TP. This value would suggest that for microbial communities, any lotic freshwater site with a TP level of greater than 55 µg/L might be considered as a high TP site. This number is higher than threshold values found for other taxa in other systems. Previous work on benthic periphyton has indicated a slightly lower threshold for benthic periphyton taxa using regression-tree analysis in a wide-scale study in Canada [60], and an application of TITAN on benthic periphyton in Texas revealed an assemblage level threshold of approximately 20 µg/L TP [61]. Previous study of algal biomass has also indicated a lower threshold of 25 µg/L depending on light availability; however, this refers only to soluble reactive phosphorus and not to total phosphorus [54].

Diversity and Community Taxonomy

QIIME diversity analysis and TITAN threshold analysis both identified taxa that could be of interest to P dynamics in this system. The QIIME diversity analysis identified *Acinetobacter* which has been known to accumulate polyphosphates and polyhydroxyalkanoates in lownutrient environments [62] allowing it to thrive in environments with relatively low concentrations of orthophosphates (e.g., [63]). The Lower TP sites may give *Acinetobacter* a competitive advantage over taxa that are more dependent on readily available P.

Within the threshold range of TP identified by TITAN analysis, we see a decrease in genus *Cytophaga*, a microbe associated with chitin consumption and with large organic matter decomposition [64] as well as *Agrobacterium*, *Nesterenkonia*, and *Dysgonomonas*. We see a significant increase in *Clavibacter*, a genus containing pythopathic pathogens affecting agriculture [65], and in *Cloacibacterium*, a genus originally isolated from wastewater in the central USA [66].

The heatmap in Fig. 2.4 illustrates how some sites are dominated by a few taxa where other sites are distributed much more evenly. The review performed by Zeglin [20] indicates a significant differentiation of many of these taxa's parental phyla across stream "compartments" with relatively consistent taxonomic abundance within compartments. However, our results would seem to indicate quite a bit of variation in those phyla within the same compartment (water column) among these different sites.

Conclusions

Although TP may not be the strongest factor involved in the structure of microbial communities in this system during this October 2014 sampling, it remains undoubtedly an important factor in this stream and river system representing a strong P gradient. Testing via three separate methods including MANOVA-like tests, ordination, and RDA modeling, all identify TP as a factor in community and functional capability assembly.

Turbidity displays a strong relationship to our system despite the potential limiting of large particle-associated microbes by using the 0.2- to 1-μm fraction. It is easy to see how a strong or complex factor such as turbidity, missing factors, or data resolution limitations may mask other important factors in these systems.

We found that TP is important to both community structure and the PFM of the combined bacterial and archaeal community in this lotic freshwater system but in a decoupled fashion which is contrary to what we expected. These results indicate that looking expressly at community structure without function may show an incomplete ecosystem picture.

Our results showing a lack of functional redundancy in this system and the potential impacts of diversity change illustrate why further study of lotic freshwater microbial communities in the context of P is important to our understanding of the microbial ecology and biogeochemical nutrient cycling in these systems.

It is important that we begin to look at the mechanisms for P biogeochemical cycling in lotic freshwater microbial communities beyond the characterization and modeling of structure, function, and diversity in order to better understand how they structure in relation to environmental P uptake and mobilization as members of a larger ecosystem. While these mechanisms are not addressed as part of this initial study, we have demonstrated that PFMs along with metagenome exploration may offer excellent tools to begin this investigation.

Acknowledgements

The authors thank Morgan Bettcher, Stephen Cook, Stephen Elser, Katherine Hooker, Lauren Housley, and Caleb Robbins for their help in collecting field samples. We also thank Owen Lind and J. Thad Scott for assistance with internal review. We acknowledge the research support by Baylor University Office of Research and Baylor University Center for Reservoir and Aquatic Systems Research (CRASR).

Conflict of Interest

The authors declare that they have no conflict of interest.

Statement of Data Availability

Sequence data that support the findings of this study have been deposited in GenBank with the BioProject accession code PRJNA350288. The environmental data that support the findings of this study are available from the corresponding author upon reasonable request.

Chapter References

1. Schlesinger WH, Bernhardt ES (2013) Front matter. In: Biogeochem, Third Ed. Academic Press, Boston, pp i–ii

2. Van Der Heijden MG, Bardgett RD, Van Straalen NM (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecol Lett 11:296–310

3. Cotner JB, Biddanda BA(2002) Small players, large role: microbial influence on biogeochemical processes in pelagic aquatic ecosystems. Ecosystems 5:105–121. doi:10.1007/s10021-001-0059-3

4. Scott T, Cotner J, LaPara T (2012) Variable stoichiometry and homeostatic regulation of bacterial biomass elemental composition. Front Microbiol 3. doi:10.3389/fmicb.2012.00042

5. Richardson AE, Simpson RJ (2011) Soil microorganisms mediating phosphorus availability update on microbial phosphorus. Plant Physiol 156:989–996

6. Godwin CM, Cotner JB (2014) Carbon: phosphorus homeostasis of aquatic bacterial assemblages is mediated by shifts in assemblage composition. Aquat Microb Ecol 73:245–258

7. Godwin CM, Whitaker EA, Cotner JB (2017) Growth rate and resource imbalance interactively control biomass stoichiometry and elemental quotas of aquatic bacteria. Ecology 98:820–829

8. Tapia-Torres Y, Rodríguez-Torres MD, Elser JJ et al (2016) How to live with phosphorus scarcity in soil and sediment: lessons from bacteria. Appl Environ Microbiol 82:4652–4662. doi:10.1128/AEM.00160-16

9. Su J-Q, Ding L-J, Xue K et al (2015) Long-term balanced fertilization increases the soil microbial functional diversity in a phosphorus-limited paddy soil. Mol Ecol 24:136–150. doi:10.1111/mec.13010

10. Louca S, Parfrey LW, Doebeli M (2016) Decoupling function and taxonomy in the global oceanmicrobiome. Science 353:1272–1277

11. García-Palacios P, Vandegehuchte ML, Shaw EA et al (2015) Are there links between responses of soil microbes and ecosystem functioning to elevated CO2, N deposition and warming? A global perspective. Glob Chang Biol 21:1590–1600

12. Logue JB, Stedmon CA, Kellerman AM et al (2016) Experimental insights into the importance of aquatic bacterial community composition to the degradation of dissolved organicmatter. ISME J 10:533–545

13. Pagaling E, Strathdee F, SpearsBMet al (2014) Community history affects the predictability of microbial ecosystem development ISME J 8:19–30. doi:10.1038/ismej.2013.150

14. Sunagawa S, Coelho LP, Chaffron S et al (2015) Structure and function of the global ocean microbiome. Science 348:1261359. doi:10.1126/science.1261359

15. Tilman D, Knops J, Wedin D et al (1997) The influence of functional diversity and composition on ecosystem processes. Science 277:1300–1302

16. Xu Z, Malmer D, Langille MG et al (2014) Which is more important for classifying microbial communities: who's there or what they can do. ISME J 8:2357–2359

17. Shade A, Peter H, Allison SD et al (2012) Fundamentals of microbial community resistance and resilience. Front Microbiol 3:417

18. Allison SD, Martiny JB (2008) Resistance, resilience, and redundancy in microbial communities. Proc Natl Acad Sci 105:11512–11519

19. Micheli F, Halpern BS (2005) Low functional redundancy in coastal marine assemblages. Ecol Lett 8:391–400

20. Zeglin LH (2015) Stream microbial diversity in response to environmental changes: review and synthesis of existing research. Front Microbiol 6:454

21. Zinger L, Gobet A, Pommier T (2012) Two decades of describing the unseen majority of aquatic microbial diversity. Mol Ecol 21: 1878–1896. doi:10.1111/j.1365-294X.2011.05362.x

22. Anderson DM, Glibert PM, Burkholder JM (2002) Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. Estuaries 25:704–726. doi:10.1007/BF02804901

23. Correll DL (1998) The role of phosphorus in the eutrophication of receiving waters: a review. J Environ Qual 27:261. doi:10.2134/jeq1998.00472425002700020004x

24. Sharpley AN, Chapra SC, Wedepohl R et al (1994) Managing agricultural phosphorus for protection of surface waters: issues and options. J Environ Qual 23:437–451

25. Green WR, Haggard BE (2001) Phosphorus and nitrogen concentrations and loads at Illinois River south of Siloam Springs, Arkansas, 1997–1999

26. Haggard BE, Soerens TS (2006) Sediment phosphorus release at a small impoundment on the Illinois River, Arkansas and Oklahoma, USA. Ecol Eng 28:280–287. doi:10.1016/j.ecoleng.2006.07.013

27. Haggard BE (2010) Phosphorus concentrations, loads, and sources within the Illinois River drainage area, northwest Arkansas, 1997–2008. J Environ Qual 39:2113–2120

28. American Public Health Association (1998) Standard methods for the examination of water and wastewater. APHA-AWWA-WEF, Washington, D.C
29. Biggs BJF, Kilroy C Stream periphyton monitoring manual. Prepared for the New Zealand Ministry for the Environment

30. Caporaso JG, Lauber CL, Walters WA et al (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621–1624

31. Caporaso JG, Kuczynski J, Stombaugh J et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336

32. Aronesty E (2013) Comparison of sequencing utility programs. Open Bioinform J 7

33. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461

34. DeSantis TZ, Hugenholtz P, Larsen N et al (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072

35. Langille MG, Zaneveld J, Caporaso JG et al (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31:814–821

36. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30

37. Kanehisa M, Sato Y, Kawashima M, et al (2015) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res gkv1070

38. R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

39. McMurdie PJ, Holmes S (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217

40. Oksanen J, Blanchet FG, Kindt R, et al (2016) Vegan: community ecology package. R package version 2.4–1

41. Wickham H (2006) ggplot: an implementation of the grammar of graphics. R Package Version 210

42. Marra G, Wood SN (2011) Practical variable selection for generalized additive models. Comput Stat Data Anal 55:2372–2387

43. Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst Zool 35:627–632

44. Jackson DA (1995) PROTEST: a PROcrustean randomization TEST of community environment concordance. Ecoscience 2:297–303

45. Butts CT (2008) Network: a package for managing relational data in R. J Stat Softw 24:1–36

46. Butts CT, Hunter D, Handcock M, et al (2015) Network: Classes for Relational Data

47. Csardi G, Nepusz T (2006) The igraph software package for complex network research. Int J Complex Syst 1695:1–9

48. Warnes GR, Bolker B, Bonebakker L, et al (2016) gplots:Various R programming tools for plotting data. R Package Version 301 2

49. Baker ME, King RS, Kahle D (2015) TITAN2: Threshold Indicator Taxa Analysis (Version 2.1). https://rdrr.io/cran/TITAN2/. Accessed 20 June 2016

50. Baker ME, King RS (2010) A new method for detecting and interpreting biodiversity and ecological community thresholds. Methods Ecol Evol 1:25–37

51. Miki T, Yokokawa T, Matsui K (2014) Biodiversity and multifunctionality in a microbial community: a novel theoretical approach to quantify functional redundancy. Proc R Soc B 281: 20132498. doi:10.1098/rspb.2013.2498

52. Wood SN (2001)mgcv:GAMs and generalized ridge regression for R. R News 1:20–25

53. Lloyd DS, Koenings JP, Laperriere JD (1987) Effects of turbidity in fresh waters of Alaska. N Am J Fish Manag 7:18–33

54. Hill WR, Fanta SE, Roberts BJ (2009) Quantifying phosphorus and light effects in stream algae. Limnol Oceanogr 54:368–380

55. Chapman DV, Unesco and World Health Organization and United Nations Environment Programme (eds) (1996) Water quality assessments: a guide to the use of biota, sediments, and water in environmental monitoring. E & FN Spon, London

56. Davies-Colley RJ, Smith DG (2001) Turbidity suspended sediment, and water clarity: a review

57. Torsvik V, Øvreås L, Thingstad TF (2002) Prokaryotic diversity— magnitude, dynamics, and controlling factors. Science 296:1064–1066. doi:10.1126/science.1071698

58. Read DS, Gweon HS, Bowes MJ et al (2015) Catchment-scale biogeography of riverine bacterioplankton. ISME J 9:516–526

59. Cardinale BJ, Hillebrand H, HarpoleWS et al (2009) Separating the influence of resource "availability" from resource "imbalance" on productivity–diversity relationships. Ecol Lett 12:475–487. doi:10.1111/j.1461-0248.2009.01317.x

60. Chambers PA,McGoldrick DJ, Brua RB et al (2012) Development of environmental thresholds for nitrogen and phosphorus in streams. J Environ Qual 41:7–20

61. Taylor JM, King RS, Pease AA, Winemiller KO (2014) Nonlinear response of stream ecosystem structure to low-level phosphorus enrichment. Freshw Biol 59:969–984

62. Singh V, Stapleton RD (2002) Biotransformations bioremediation technology for health and environmental protection. Elsevier Science Ltd., Amsterdam. http://0-www.sciencedirect.com.fama.us.es/science/publication?issn=00796352&volume=36

63. Saralov AI, Mol'kov DV, Bannikova OM, Solomennyĭ AP, Chikin SM (2001) Intracellular accumulation of monomer precursors of polyphosphates and polyhydroxyalkanoates from *Acinetobacter calcoaceticus* and *Escherichia coli*. Mikrobiologiia 70:737–744

64. Cottrell MT, Kirchman DL (2000) Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low-and high-molecular-weight dissolved organic matter. Appl Environ Microbiol 66:1692–1697

65. Gartemann K-H, Kirchner O, Engemann J et al (2003) *Clavibacter michiganensis* subsp. *michiganensis*: first steps in the understanding of virulence of a gram-positive phytopathogenic bacterium. J Biotechnol 106:179–191

66. Allen TD, Lawson PA, Collins MD et al (2006) *Cloacibacterium normanense* gen. nov., sp. nov., a novel bacterium in the family *Flavobacteriaceae* isolated from municipal wastewater. Int J Syst. Evol Microbiol 56:1311–1316. doi:10.1099/ijs.0.64218-0

CHAPTER THREE

Rivers May Constitute an Overlooked Avenue of Dispersal for Terrestrial Fungi

This chapter published as: LeBrun ES, Taylor DL, King RS, Back JA, Kang S. 2018. Rivers may constitute an overlooked avenue of dispersal for terrestrial fungi. Fungal Ecol. 32:72-79. doi.org/10.1016/j.funeco.2017.12.003.

Abstract

While fungi are intimately associated with substrates in freshwater systems, the role of fungi in the open water column is less well defined. Using next generation sequencing of 0.2 μm–1 μm filtered water columns samples, we detected abundant and diverse fungal sequences across 25 stream and river sites in the Ozark region of Oklahoma and Arkansas. Fungal communities were only weakly related to stream environmental metrics with the exception of total phosphorus (TP). We infer from our results that TP is acting as a proxy for unique catchment effects. We observed patterns of dominant community taxa at higher taxonomic groupings but lower taxonomic groupings were site specific. OTU functional assignment showed the majority of sequences to be related to plant and animal pathogens, and some saprotrophs. The likely allochthonous origin and strong site specificity of these fungi suggest overlooked dispersal via lotic waterways, which may have important biogeographic consequences for fungi.

Introduction

Fungi in aquatic ecosystems have been extensively studied, typically focusing on substrate surfaces such as allochthonous leaf litter in freshwater in the context of decomposition (Suberkropp and Klug, 1976; Nikolcheva and Bärlocher, 2005; Sridhar

et al., 2008). One of the topics that has been largely neglected is fungi in the water column not associated with substrates or sediment. One study that assessed fungal biomass in the upper 1 m of the water column in 32 temperate streams in Poland found a significant correlation between fungal biomass and total nitrogen (N) and phosphorus (P) in water using regression and Pearson correlation analysis (Gorniak et al., 2013). Community structure and role were not directly investigated but direct fungal participation in water column nutrient cycling was hypothesized (Gorniak et al., 2013).

Another possible explanation for fungal presence in the water column is simply the deposition of hyphal fragments or other potential propagules from air, upstream water, detritus deposition, and root to stream contact. Water column dispersal is well studied for aquatic hyphomycetes (Ingold, 1942; Thomas et al., 1991; Suberkropp and Wallace, 1992; Sridhar and Bärlocher, 1994) but such inputs could provide an overlooked means of dispersal for terrestrial fungi as well. In fact, fungus-like Oomycete plant pathogens of the genus *Phytophthora* are well known to disperse via river systems (Li, 2016). Recent studies clearly show that some true fungi are dispersal-limited (Peay et al., 2012; Cline and Zak, 2014; Peay and Bruns, 2014), leading to strong biogeographic distribution patterns (Taylor and Bruns, 1999; Peay et al., 2010).

Most work on dispersal in terrestrial fungi has focused on movement of aerial spores (Brown and Hovmøller, 2002; Pashley et al., 2012; Savage et al., 2012; Grinn-Gofroń and Bosiacka, 2015), including next generation sequencing studies of indoor air (Amend et al., 2010; Adams et al., 2013). While aquatic hyphomycetes have received attention with respect to diversity of decomposers and macroscopic life stages (Fabre, 1998a, 1998b, 1998c), there have been few studies of the roles of river systems in the

dispersal of true fungi in general, particularly by very small propagules. Given the fact that streams and rivers serve as ecological aggregators of processes throughout their watershed catchments (Frissell et al., 1986; Allan, 2004; Bormann and Likens, 2012), and that riverine dispersal is important in numerous other taxa, e.g. fish, insects, reptiles, and plants (Maguire, 1963; Bermingham and Avise, 1986; Bunn and Hughes, 1997; Bernatchez and Wilson, 1998; Miller et al., 2002; Santamaría, 2002; Petersen et al., 2004; Pellegrino et al., 2005; Vanschoenwinkel et al., 2008), the lack of fungal studies represents a major gap.

A potential hurdle to investigating dispersal via the water column is determining whether or not fungi detected are active aquatic community members or are transient and inactive. There is a large body of evidence linking fungi to the phosphorus (P) cycle in soils (Bolan, 1991; Schachtman et al., 1998; Van der Heijden et al., 2008). High total phosphorus (TP) is also an indicator of excessive nutrients from catchments feeding into streams and rivers (Schindler, 1977; Carpenter et al., 1998) and has a large impact on aquatic systems (Bennett et al., 2001; Anderson et al., 2002; Hart et al., 2004). Exploring fungi across a gradient of TP allows identification of relationships of free living fungal communities to P in the water column.

The bulk of water-column particulate matter consists of eroded soils and particulate organic matter (Schlesinger and Melack, 1981; Waters, 1995; Bilotta and Brazier, 2008). Fungi are both ubiquitous in soils and are directly involved in colonizing and decomposing organic matter in streams (Christensen, 1989; Gessner, 1997). To focus on dispersal via small fungal cells, fragments, and spores in the water column, it is desirable to exclude fungi associated with particles using a method like size filtering (APHA, 1998). Here, we analyze total fungal diversity in a microscopic fraction (0.2–1.0 μm) across a river system

spanning a range of TP. While many fungal cells are larger than 1 μm, we anticipated good detection of fungi through small cells, spores, and cell fragments. The ecological gradient is representative of differences in multiple catchment properties such as vegetation and nutrient cycling across the varied watersheds enhancing the exploration of relationships between fungal communities and the environment in the system.

Methods

Sampling

Extraction of genomic DNA (gDNA) from water-column filter samples was described in detail in LeBrun et al., (2017) (LeBrun et al., 2017). The study area was a collection of mid-order (3rd-5th) streams and rivers along the Oklahoma-Arkansas border, an area with known P enrichment problems (Fig. D.S1) (Green and Haggard, 2001; Haggard, 2010; Haggard and Soerens, 2006). Sampling was performed in October of 2014. Additional site characteristics are also available through a study by Cook et al., 2017 (*in press*) where data on these sites was collected at regular intervals over 2 y. The sampling sites represented a gradient of TP levels from 7 to 173 μg/L. Sample processing involved a stacked filtration of 50 mL of water from ∼10 cm below the surface in the water column through a 1 μm glass fiber filter and then a 0.2-μm filter. Only components collected on the 0.2-μm filters were used in extracting the gDNA for this study (i.e. the 0.2–1.0 μm size fraction).

Environmental data

Environmental data including dissolved organic carbon (DOC), total phosphorus (TP), total N (TN), C:N ratio, C:P ratio, sestonic chlorophyll-a, total suspended solids

(TSS), turbidity, pH, dissolved oxygen (DO), temperature, and specific conductance in stream/river water along with catchment size and catchment and land use factors including percentage pasture land, percentage impervious cover, percentage developed land were collected as reported in LeBrun et al., (2017) (LeBrun et al., 2017). In brief, water chemistry was measured using YSI EXO2 multiparameter data sonde (Yellow Springs, OH) and standardized water testing (APHA, 1998). Catchments were delineated using ArcGIS and land usage was estimated from the National Land Cover Data (NLCD) raster (ESRI, 2011. *ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.*).

Library preparation

Library preparation for this study started with the gDNA collected in LeBrun et al. (2017). An initial PCR amplification of the ITS2 region was conducted using 5.8S_Fun and ITS4_Fun primers (Taylor et al., 2016) modified to include adapters for future indexing. PCR was performed using 2× Platinum™ Green Master Mix from Invitrogen. PCR specifications were 1 cycle for HotStart step at 94 °C for 2 min followed by 30 cycles of 94 °C denaturation step for 45 s, 50 °C annealing step for 1 min, and 72 °C elongation step for 1.5 min. Successful PCR was identified through 1% agarose gel electrophoresis run at 70 V for 40 min. PCR cleanup was conducted using an Agencourt AMPure XP kit (Beckman Coulter Life Sciences) following the manufacturer's protocol. Final PCR product quantification was conducted using a Qubit 3.0 system.

A second round of PCR amplification was run to add unique indices to each sample as well as Illumina sequencing adapters. PCR was again performed using 2× Platinum™ Green Master Mix. PCR specifications for the second round of PCR were

1 cycle for HotStart step at 94 °C for 2 min followed by 8 cycles of 94 °C denaturation step for 45 s, 59 °C annealing step for 1 min, and 72 °C elongation step for 1.5 min. Cleanup and quantification were performed in the same manner as the first round of PCR. Samples were then pooled so that 10 ng of DNA from each sample was present in final library.

Sequencing was performed on an Illumina MiSeq system using a MiSeq Reagent Kit v3 2 × 300 with paired-end reads. Libraries were spiked with 20% phiX control.

Sequence processing

Demultiplexing was conducted through Illumina BaseSpace. Paired-end read fastq files for each sample were extracted for downstream processing. Additional sequence processing was carried out using Quantitative Insights Into Microbial Ecology (QIIME) version 1.9.1 (Caporaso et al., 2010). Paired-end reads were combined using the fastq-join algorithm from eautils (Aronesty, 2013). Un-paired reads were discarded at this time. The resulting sequences were then filtered with a maximum unacceptable Phred quality score of 20. Chimeric sequences were identified and removed using the UCHIME algorithm within USEARCH (Edgar, 2010). Operational taxonomic unit (OTU) picking was performed via open reference with a 0.50 pre-filter using uclust against the dynamic UNITE database version 7 with a 0.94 similarity cutoff. Singleton sequences were removed during OTU picking and taxonomy was assigned with the UNITE database as reference. Reads identified as Plantae or Protista were then manually removed from the resulting OTU table via filtering.

Functional information in the form of guild assignment to OTUs was performed using the online version of FUNGuild (Nguyen et al., 2016). FUNGuild parses OTUs into “guilds” or “functional groupings”based on their taxonomic assignments and

ecological data extracted from the literature (Nguyen et al., 2016). Guilds are representative of species, whether related or unrelated, that exploit the same class of environmental resources in a similar way.

Statistical analysis software

All analyses were performed in the R software package v.3.2.3 (R Core Team, 2015) using various packages and scripts as identified. OTU table BIOM files from QIIME were either exported to tab delimited format directly from QIIME or imported for use in R using the phyloseq package version 1.14.0 in R (McMurdie and Holmes, 2013).

Diversity, ordinations, models, and comparisons

Diversity metrics for Shannon (H'), Simpson (1-D), and Inverse Simpson (1/D) indices were calculated using the vegan package version 2.4–0 (Oksanen et al., 2016). Sample overlap was calculated using the Morisita-Horn index and bootstrapping (n = 200) in the vegetarian package version 1.2 (Charney and Record, 2012). Heatmaps were built using the heatmap.2 function in the gplots package version 3.0.1 (Warnes et al., 2016). Heatmap dendrograms were built using Bray-Curtis distance. Taxa that do not represent more than 1 percent of relative abundance in any site were removed post clustering for visual clarity.

Non-metric multidimensional scaling (NMDS) ordinations were constructed to describe community dissimilarity in unconstrained space using the vegan package. NMDS using Bray-Curtis distance was performed for ITS2 and Guild datasets. NMDS plots were created using the ggplot2 package in R (Wickham, 2006). Environmental gradients were built on NMDS ordinations using ordisurf from the vegan package. Ordisurf uses

Generalized Additive Modeling (GAM) model building to overlay environmental variables in the ordination space (Marra and Wood, 2011).

ITS2 and Guild data were compared with a Mantel test using the mantel function in the vegan package with the Pearson correlation method and 1000 permutations in order to test for similarity between the data sets (Smouse et al., 1986). PROcrustean randomization TEST of community environment concordance (PROTEST), a potentially more sensitive detection method than a Mantel test, was also used to compare NMDS ordinations in the vegan package (Jackson, 1995). PROTEST uses scaling and rotations to maximize alignment in ordinations as a multivariate measure of concordance in datasets along with permutation based significance testing (Jackson, 1995). All PROTEST analyses were performed with 1000 permutations. Mantel and PROTEST analyses were also used to investigate relationships between fungal community assemblages and the bacterial community assemblages from LeBrun et al. (2017) for the 23 sites shared between the two studies.

Environmental groupings and analysis

Due to the experimental design focusing on a TP gradient, sites were divided into groups of Low, Med, and High TP. Group cutoffs were determined at apparent breakpoints in the distribution of collected TP data. Low for TP was set for sites below 40 µg/L and High as above 70 µg/L. The designation of groupings for each site can be viewed in Table D.S1. Multivariate ANOVA (MANOVA) like non-parametric analyses including PERMANOVA and ANOSIM were then performed with 1000 permutations using the vegan package to test for significant TP group related dispersion and variation.

Network construction and analysis

To visualize taxa relatedness and clustering, networks were constructed from data imported through phyloseq using the network package version 1.13.0 and igraph package version 1.0.1 in R with Bray-Curtis distances (Butts et al., 2015; Butts et al., 2008; Csardi and Nepusz, 2006). Clustering for each network was performed using a spin-glass model and simulated annealing via the cluster_spinglass function in the igraph package.

Additional modeling and testing

Redundancy Analysis (RDA) models were built in the vegan package (Oksanen et al., 2016) in order to describe the community structure in environmentally constrained space for both OTU and FUNGuild data. RDA model selection was performed by starting with an initial model including all collected variables and manually removing collinear variables in an effort to maximize adjusted R^2 and minimize the magnitude of difference between R^2 and adjusted R^2.

Generalized Additive Models (GAMs) were built individually for each collected environmental variable against Shannon, Simpson, and inverse Simpson diversity scores calculated using relative abundance data. GAMs were built using the mgcv package (Wood, 2001) for multiple environmental metrics.

Indicator species analysis was performed using the IndVal function in R from the labdsv package (Roberts, 2013) with 2:6 clusters and 1000 iterations. Indicator species analysis identifies important taxa for typologies created from any classification procedure independently from the classification method (Dufrêne and Legendre, 1997).

Threshold analysis for TP and turbidity was performed using the TITAN 2.1 package in R (Baker et al., 2015). Threshold Indicator Taxa ANalysis (TITAN) identifies

environmental variable values maximizing taxa frequency and abundance using bootstrapping to identify reliable indicator taxa and the sum of indicator taxa value Z scores to identify the environmental values representing the peak of increase or decline of the taxa (Baker and King, 2010). The number of bootstraps performed in our TITAN analysis was 200. The genus taxonomic level was used and only taxa observed more than 3 times across all sites were used.

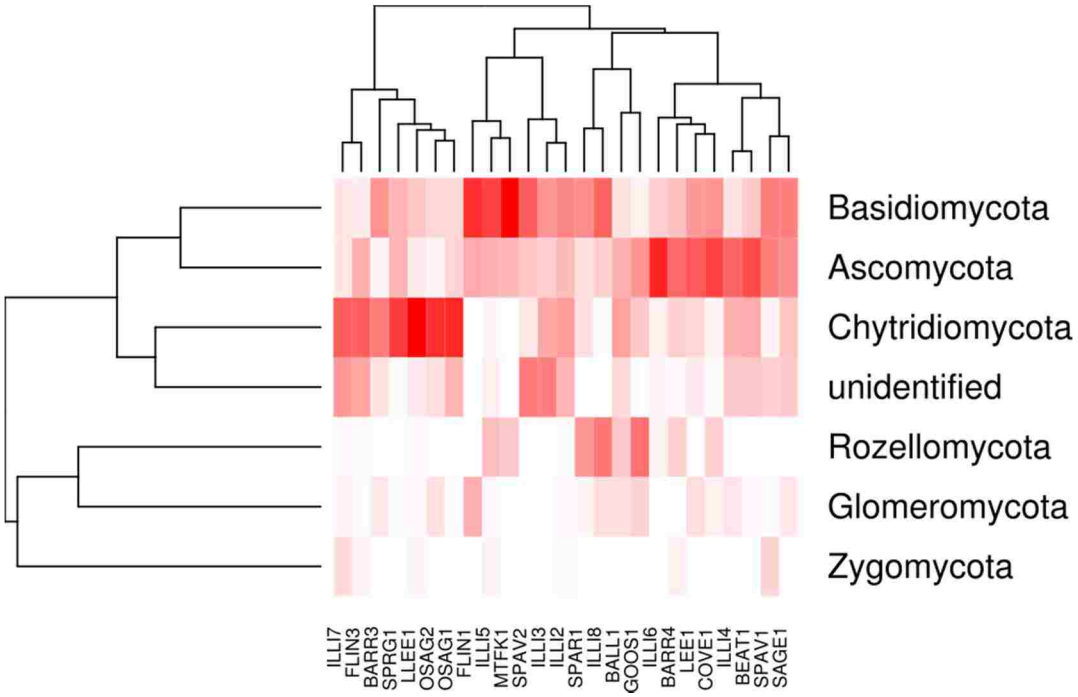


Figure 3.1: Heatmap of community structure at the Phylum level by site. Dendrograms are constructed with Bray-Curtis distance.

Results

OTU counts and site diversity

We identified from 48 to 168 fungal taxa at each site across the 25-site system (mean = 85, SD = 33.25) (Table D.S1). Shannon index values ranged from 1.59 to 3.36 (mean = 2.35, SD = 0.53), Simpson index values ranged from 0.66 to 0.94 (mean = 0.84,

SD = 0.08), and inverse Simpson index values ranged from 2.98 to 17.25 (mean = 7.99, SD = 3.90) (Table D.S1). The Morisita-Horn overlap index was $C_D = 0.179$ (SE = 0.0003). The heatmap of taxa at the Phylum level shows groupings of sites primarily dominated by one of the phyla *Basidiomycota*, *Ascomycota*, or *Chytridiomycota* (Fig. 3.1).

Network analysis

Network analysis showed small, distinct clusters of ecologically related taxa at the genus level (Fig. 3.2). The majority of taxa represented OTUs that made up less than 1% of total abundance and so will be referred to as "rare" taxa for this study. The relationships shown are between taxa throughout the sites.

Guild designations

42% of ITS2 OTUs were classified into guilds by FUNGuild. The majority of taxa fell into animal and plant pathogen guilds as well as unidentified saprotrophs (Fig.3.3).

NMDS ordinations and RDA models

For NMDS ordination with taxonomic data, sites were dispersed fairly evenly through ordination space with no apparent clusters of sites. However, a pattern of sites positioning in a related manner emerges when considering TP groupings (Fig. 3.4). The TP gradient fit to the ordination using GAM explained 32.6% of deviance with $p = 0.02$. NMDS ordination with the FUNGuild data showed slightly more separation between potential groups of sites (Fig. D.S2). The TP gradient fit to the ordination using GAM explained 48.2% of deviance with $p < 0.01$. Results from PERMANOVA and ANOSIM were both significant for TP groupings with $F = 1.48$ ($p = 0.02$) and $R = 0.131$ ($p = 0.03$).

The Mantel test between the fungal ITS2 data and the bacterial 16S NMDS data from LeBrun et al. (2017) for the 23 overlapping sites showed no significant correlation ($r_m = -0.033$, $p = 0.94$). Procrustean PROTEST comparison between fungal and bacterial NMDS ordinations had a correlation statistic of 0.48 with $p = 0.013$. The Procrustes error plot is shown in Fig. D.S3.

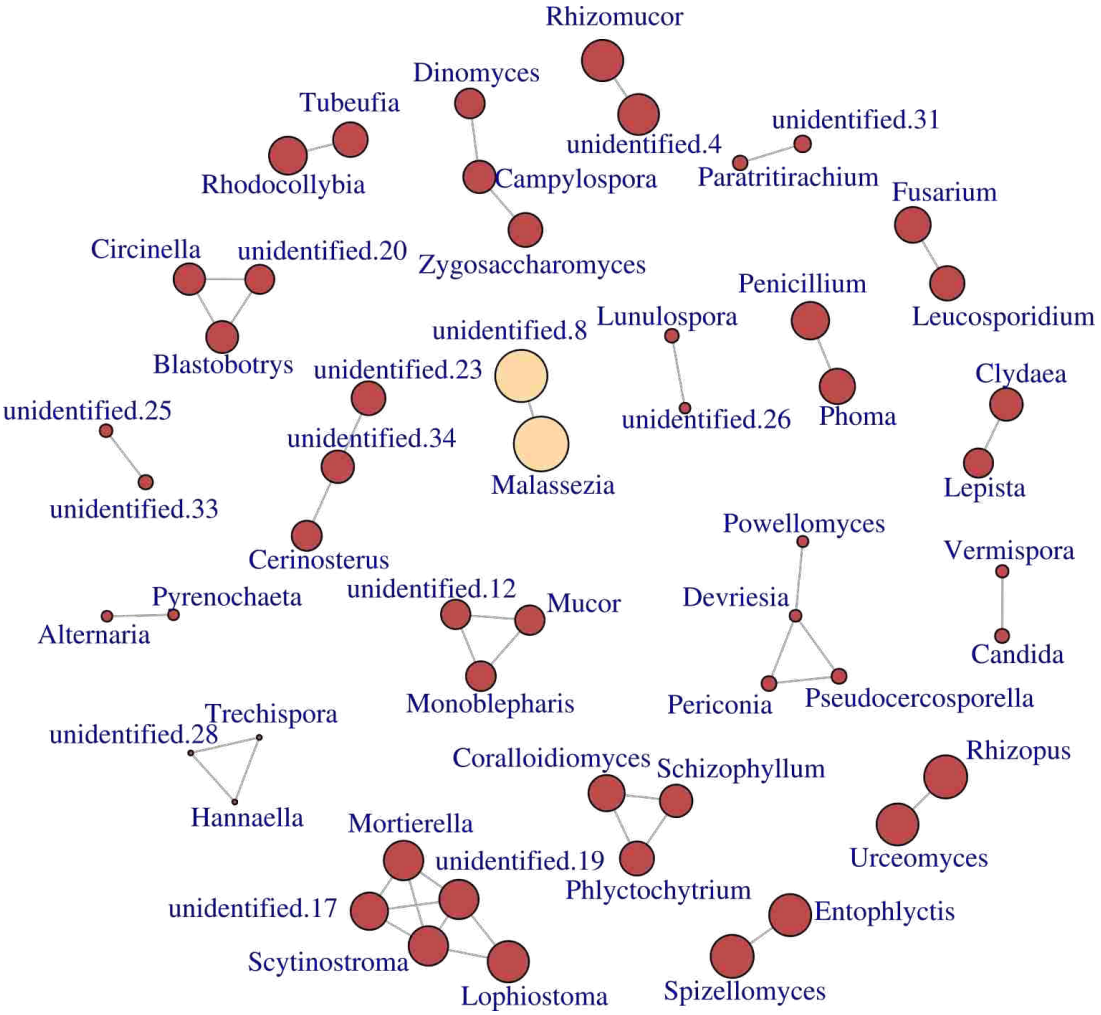


Figure 3.2: Taxonomic network generated using Bray-Curtis distances. Isolated taxa have been removed leaving only Taxa with at least one connecting edge. Node size represents total abundance of that taxon on a log scale. Taxa with red nodes represent less than 1% of total abundance while yellow nodes indicate taxa not considered "rare". Taxa labeled "unidentified.xy" were unable to be classified at the genus level.

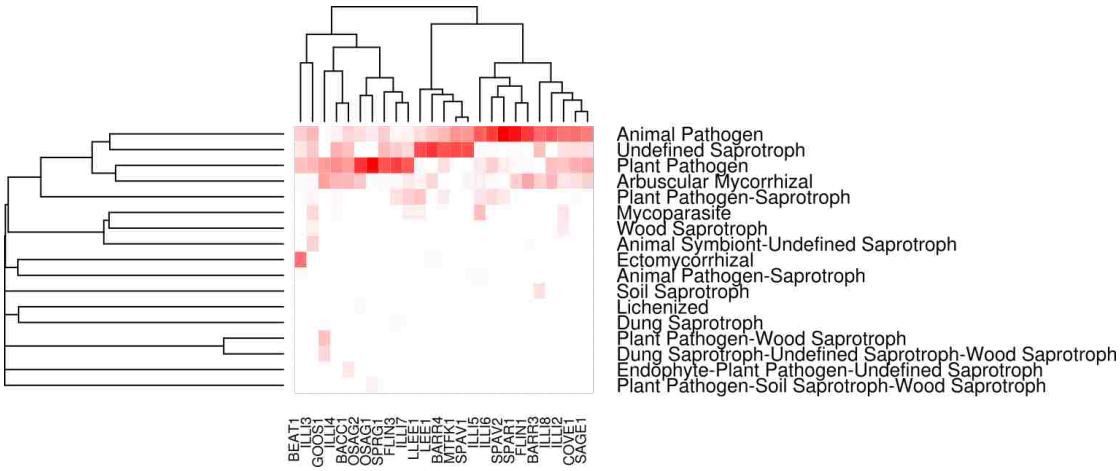


Figure 3.3: Heatmap of FUNGuild identified guild abundance by site. Dendrograms are constructed with Bray-Curtis distance.

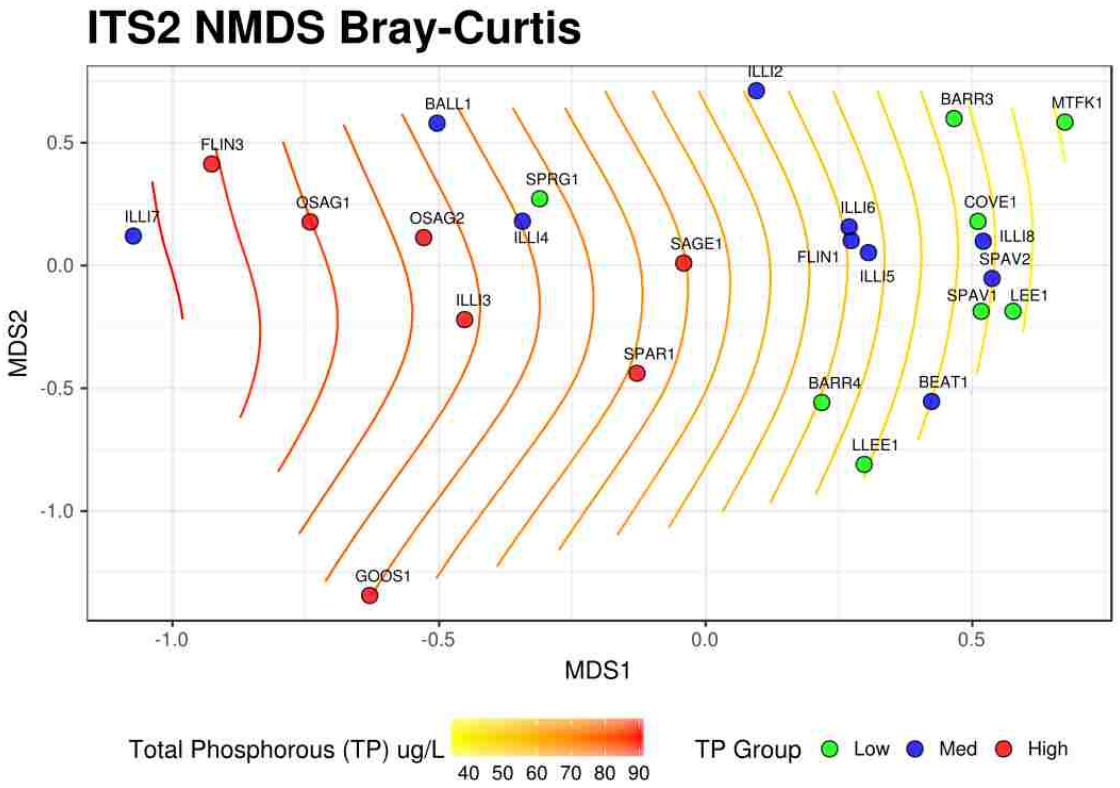


Figure 3.4: NMDS ordination of ITS2 community with Bray-Curtis distance (Stress 0.155). Gradient represents environmental TP fit to ordination using GAM.

Attempts to model taxonomic community structure using Redundancy Analysis (RDA) and GAMs with collected environmental and land use variables were unproductive as none of the environmental variables resulted in a significant model. The final RDA model for FUNGuild data included TP, carbon to P ratio (C:P), DOC, DO, temperature, and pH with an adjusted R^2 of 0.345 and $p = 0.014$ (Fig. 3.5). The direction of C:P was the antithesis of to DOC and TP, indicating a relationship differing from TP or DOC, and so it was kept in the model.

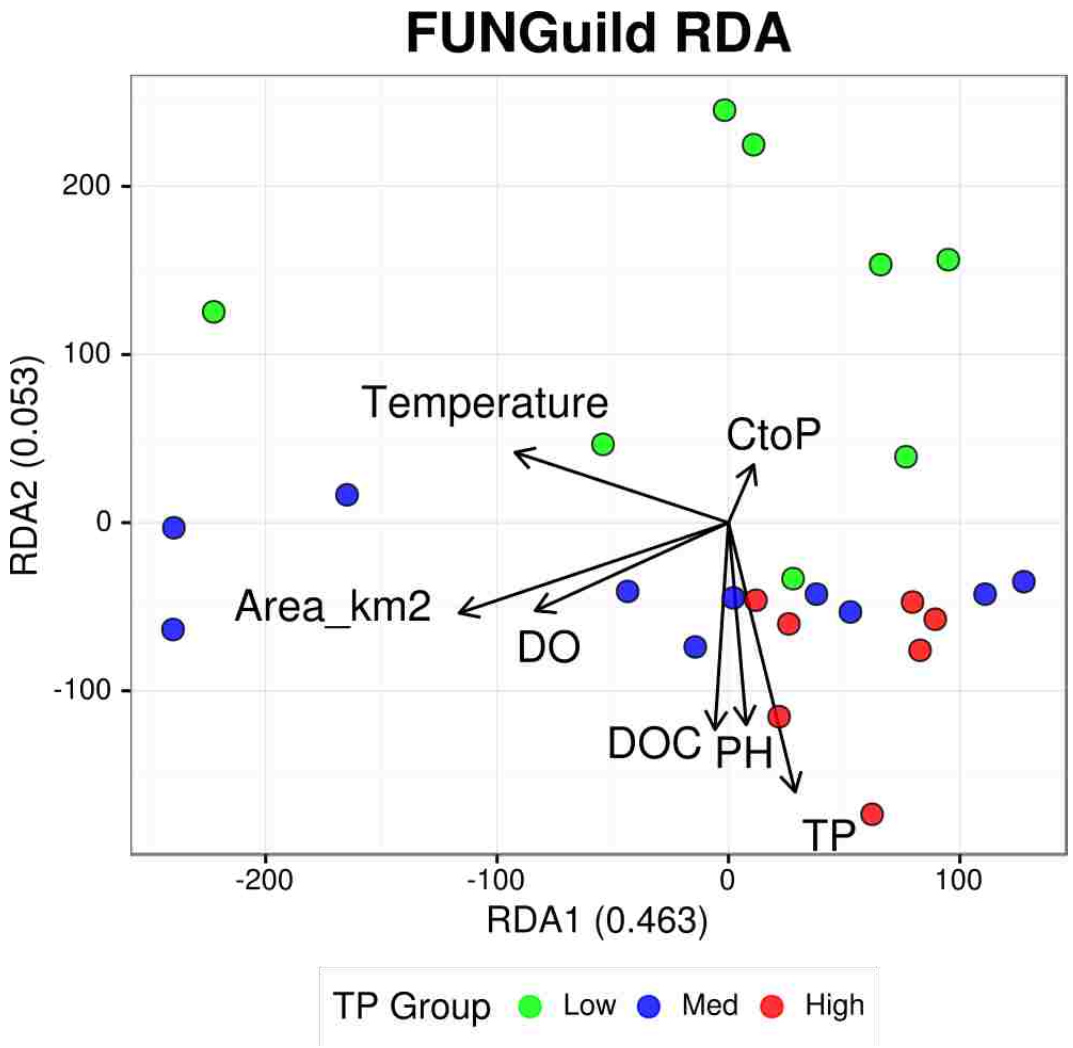


Figure 3.5 RDA model of environmental variables and FUNguild designations at each site (adj. R2 = 0.345, p = 0.014).

Indicator species and total phosphorous gradient relationship

Indicator species analysis was able to identify a few indicator taxa in our system; however, the results were not consistent across multiple runs and clustering levels, and identified taxa were few and most only weakly significant. Two taxa that were consistently identified were the genus *Hygrocybe* ($d = .9804$, $p = .014$) at lower clustering levels (2–6 k-means clusters) and the genus *Entophlyctis* ($d = .948$, $p = 0.039$) at higher clustering levels (6–10 k-means clusters). Unfortunately, the reason for these species as indicators for the relevant sites remains elusive, although the log abundance of *Entophlyctis* was weakly related to TP by GAM (Deviance explained = 9.33%, $p < 0.001$). TITAN was unable to identify any reliable indicator taxa or change points in system related to the TP gradient.

Discussion

Site specificity of fungal assembly

Taxa in the system showed a high site specificity. Visually, site distributions in ordination space were widely dispersed and taxonomic heatmaps (Fig. D.S4 – Fig. D.S8) illustrate the increasing site specificity at lower taxonomic levels. This increase is to be expected; however, even at the class level, sites are distinct. We also found quantitative support for this site specificity. For Morisita-Horn, $C_D = 0$ represents a system of samples with no overlap, while $C_D = 1$ represents a system of samples with complete overlap (Jost, 2007). Thus, the observed value of $C_D = 0.179$ indicates very little taxonomic overlap. Little community overlap in the system likely affected network analysis with the majority of taxa being removed from the final network due to nodes having no edges along with making it difficult for indicator species and TITAN analyses to identify indicator taxa. In

a previous study on bacteria and archaea within the same system, we were able to establish that there were no overwhelming effects of stream connectivity or downstream flow in the system (LeBrun et al., 2017). The site specificity in the current study also illustrates a lack of effects caused by any site flow connectivity (Fig. D.S8).

Taxa relationships identified by network analysis

The majority of identified relationships involve rare taxa (Fig. 3.2); taxa sharing a relationship tended to be found in the same abundances within the system. This finding could be indicative of taxa replacing each other in roles within the different catchments or may be an artifact of the network construction due to site specificity. Either way, these relationships warrant additional investigation. These rare taxa showed no relationship to environmental variables in the stream when separated from the overall community.

Nutrient cycling and environmental interactions

Analyses investigating fungal relationships to TP provided mixed results. The GAMs built with the ordinations for both taxonomic and FUNGuild data show a significant relationship between TP and placement in ordination space. It must be taken into account that both ordinations had stress values bordering on the "suspect" range (0.155 & 0.104). However, a relationship to TP is supported by the significant PREMANOVA and ANOSIM tests. The taxonomic composition of the detected fungi did not correlate with any variables in RDA. Diversity was also unrelated to environmental metrics as shown by the lack of significant GAM models.

Functional composition of the fungal assemblage via FUNGuild designations appears to be more closely tied to collected environmental variables than was taxonomic

composition. The TP gradient explains a higher percentage of deviance in the ordinations and the RDA model ties the functional designations to catchment and stream variables. The decoupling between taxa assemblage and function makes sense in light of the specificity of taxa at each site. However, the primary functional designations in the system, animal/plant pathogens and saprotrophs, do not intuitively apply to non-substrate (i.e. not on detritus, other organic material, or sediment) water column processes, meaning these fungi are likely transient.

Bacterial/archaeal communities in this system have previously been connected to wider ecological data (LeBrun et al., 2017) and the detected fungal communities are only weakly related to the bacterial communities. The small but surprisingly significant PROTEST correlation score is probably indicative of catchment effects, spatial autocorrelation, or a relationship to an allochthonous, terrestrial subset within the bacterial community. Ordinal TP GAMs explain significant site organization in ordination space for both fungi and bacteria. PROTEST is known to be more sensitive than a Mantel test (Jackson, 1995) and so is capable of capturing these types of minor effects; however, the Mantel test was insignificant and the Procrustes error plot shows no cohesive trends or patterns (Fig. D.S3).

We infer from the collective results that a correlation does exist between the fungi detected at each site and TP but that the fungi are likely transient and allochthonous rather than being active water column community members. TP likely represents unique catchment features such as vegetation, that are difficult to identify due to a large number of environmental factors that could not be included in this study such as vegetation. Stream TP is a product of numerous catchment factors, so covariation of these factors with TP is

very possible (King et al., 2005). This allochthonous origin is further supported by the identification of several taxa within the system such as the *Hygrocybe* identified in indicator species analysis. *Hygrocybe* is a terrestrial agaric with spore sizes larger than expected to pass through our 1-μM filter. Hence, it is unclear why it was so readily detected or whether we were capturing DNA from spores or cell fragments. While our findings corroborate prior findings of a TP correlation with fungi in the water column (Gorniak et al., 2013), our best guess based on our analyses of community structure and predicted function is that these fungi are not major participants in nutrient cycling within the water column itself.

Relationship of P to pathogens

In both constrained and unconstrained ordinations of FUNGuild data, there is a clear distinction between placement of Low and High TP grouped sites with Med grouped sites somewhat mixed in with one or the other. An increase in the abundance of some bacterial pathogens coincided with increased TP in a previous study of the same system (LeBrun et al., 2017). There appears to be a high potential for factors that result in high P (e.g. agricultural pollution or waste water inputs) to have a relationship to the presence and abundance of both bacterial and fungal pathogens. Pathogens in soil can cause negative density dependence in communities of vegetation (Laliberté et al., 2015). Distribution of vegetation has a strong relationship to both P levels and soil microbial communities (Langille et al., 2013). All of these inputs affect waterways (Bormann and Likens, 2012). While we have not established causation in the relationship between P and pathogens, there exists a potential feedback loop of pathogens affecting P and P affecting pathogens within

catchments and waterways. This is likely a complex relationship that would require further study to fully understand as our data only hint at such a possibility.

The water column as a dispersal medium

There are multiple lines of evidence suggesting an allochthonous origin for fungal taxa detected in this study. The identified taxa are not likely active in stream nutrient cycles or processes and they are only loosely related to bacterial communities in the system, if at all. Our focus on a smaller size fraction favors detection of small cells and fragments. In guild analysis, most taxa were identified as pathogens and saprotrophs. Organisms in these categories have potential benefits from aquatic dispersal as streams are rife with the detritus for saprotroph colonization and have access to vegetation and animals for pathogen colonization.

Phytophthora, although not a true fungus, offers an excellent example of pathogen delivery to terrestrial plants via streams. Although *Phytophthora* is known for infecting agricultural and wild plants on land, its primary method of dispersal is through the water column and monitoring and detection is carried out directly in streams (Sutton et al., 2009; Hulvey et al., 2010). A fungal pathogen dispersing through an aquatic system is less dependent on chance than it is through aerial dispersal. The water column might act as a vehicle of pathogen delivery directly to a host, be it plant or animal. In addition, aquatic dispersal offers the advantages of protection from drought and UV stress that apply strongly to aerial dispersal. Further investigation of fungal pathogen dispersal patterns through natural streams and rivers seems warranted, as anthropogenic inputs to streams continue to grow (Søndergaard and Jeppesen, 2007).

Conclusions

Here we have documented a surprisingly large number of fungal taxa occurring in the 0.2 μm - 1 μm fraction of the water column. These taxa mostly consist of pathogens and saprotrophs from the *Basidiomycota, Ascomycota*, and *Chytridiomycota* but their species composition varied greatly by site. We infer that these taxa are likely present due to deposition from allochthonous sources. The site specificity, diversity, and abundance of terrestrial fungi suggest an overlooked means of dispersal that could promote or reinforce biogeographic patterns in terrestrial fungal communities.

Conflicts of interest

The authors declare no conflict of interest.

Statement of data availability

Sequence data that support the findings of this study have been deposited in GenBank with the BioProject accession code PRJNA350288. The environmental data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

The authors thank Morgan Bettcher, Stephen Cook, Stephen Elser, Katherine Hooker, Lauren Housley and Caleb Robbins for their help in collecting field samples. We acknowledge the research support by Baylor University Office of Research and Baylor University Center for Reservoir and Aquatic Systems Research (CRASR).

Adams R.I., Miletto M., Taylor J.W. and Bruns T.D., Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances, *ISME J. Lond* **7**, 2013, 1262–1273, https://doi.org/10.1038/ismej.2013.28.

Allan J.D., Landscapes and Riverscapes: the influence of land use on stream ecosystems, *Annu. Rev. Ecol. Evol. Systemat.* **35**, 2004, 257–284, https://doi.org/10.1146/annurev.ecolsys.35.120202.110122.

Amend A.S., Seifert K.A., Samson R. and Bruns T.D., Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics, *Proc. Natl. Acad. Sci.* **107**, 2010, 13748–13753.

Anderson D.M., Glibert P.M. and Burkholder J.M., Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences, *Estuaries* **25**, 2002, 704–726, https://doi.org/10.1007/BF02804901.

APHA, Standard Methods for the Examination of Water and Wastewater, 1998, Am. Public Health Assoc. APHA Wash; DC USA.

Aronesty E., Comparison of sequencing utility programs, *Open Bioinforma* 2013, J. 7(1) 1-8.

Baker M.E. and King R.S., A new method for detecting and interpreting biodiversity and ecological community thresholds, *Methods Ecol. Evol.* **1**, 2010, 25–37.

Baker M.E., King R.S. and Kahle David, TITAN2: Threshold Indicator Taxa Analysis, 2015, Version 2.1.

Bennett E.M., Carpenter S.R. and Caraco N.F., Human impact on erodable phosphorus and eutrophication: a global perspective increasing accumulation of phosphorus in soil threatens rivers, lakes, and coastal oceans with eutrophication, *BioScience* **51**, 2001, 227–234, 10.1641/0006-3568(2001)051[0227:HIOEPA]2.0.CO;2.

Bermingham E. and Avise J.C., Molecular zoogeography of freshwater fishes in the southeastern United States, *Genetics* **113**, 1986, 939–965.

Bernatchez L. and Wilson C.C., Comparative phylogeography of nearctic and palearctic fishes, *Mol. Ecol.* **7**, 1998, 431–452.

Bilotta G.S. and Brazier R.E., Understanding the influence of suspended solids on water quality and aquatic biota, *Water Res.* **42**, 2008, 2849–2861, https://doi.org/10.1016/j.watres.2008.03.018.

Bolan N.S., A critical review on the role of mycorrhizal fungi in the uptake of phosphorus by plants, *Plant Soil* **134**, 1991, 189–207, https://doi.org/10.1007/BF00012037.

Bormann F.H. and Likens G., Pattern and Process in a Forested Ecosystem: Disturbance, Development and the Steady State Sased on the Hubbard Brook Ecosystem Study, 2012, Springer Science & Business Media.

Brown J.K. and Hovmøller M.S., Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease, *Science* **297**, 2002, 537–541.

Bunn S.E. and Hughes J.M., Dispersal and recruitment in streams: evidence from genetic studies, *J. North Am. Benthol. Soc.* **16**, 1997, 338–346, https://doi.org/10.2307/1468022.

Butts C.T., et al., network: a package for managing relational data in R, *J. Stat. Software* **24**, 2008, 1–36.

Butts C.T., Hunter D., Handcock M., Bender-deMoll Skye and Horner Jeffery, Network: Classes for Relational Data, 2015.

Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Pena A.G., Goodrich J.K., Gordon J.I., et al., QIIME allows analysis of high-throughput community sequencing data, *Nat. Methods* **7**, 2010, 335–336.

Carpenter S.R., Caraco N.F., Correll D.L., Howarth R.W., Sharpley A.N. and Smith V.H., Nonpoint pollution of surface waters with phosphorus and nitrogen, *Ecol. Appl.* **8**, 1998, 559–568.

Charney N. and Record S., Vegetarian: Jost Diversity Measures for Community Data, 2012.

Christensen M., A view of fungal Ecology, *Mycologia* **81**, 1989, 1–19, https://doi.org/10.2307/3759446.

Cline L.C. and Zak D.R., Dispersal limitation structures fungal community assembly in a long-term glacial chronosequence, *Environ. Microbiol.* **16**, 2014, 1538–1548.

Cook S.C., Housley L., Back J.A. and King R.S., Freshwater eutrophication drives sharp reductions in temporal beta diversity, in press. *Ecology* 2017, https://doi.org/10.1002/ecy.2069.

Csardi G. and Nepusz T., The igraph software package for complex network research, *InterJournal Complex Syst* **1695**, 2006, 1–9.

Dufrêne M. and Legendre P., Species assemblages and indicator species: the need for a flexible asymmetrical approach, *Ecol. Monogr.* **67**, 1997, 345–366, 10.1890/0012-9615(1997)067[0345:SAAIST]2.0.CO;2.

Edgar R.C., Search and clustering orders of magnitude faster than BLAST, *Bioinformatics* **26**, 2010, 2460–2461.

ESRI, ArcGIS Desktop: Release 10, 2011, Environmental Systems Research Institute; Redlands, CA.

Fabre E., Aquatic hyphomycetes in three rivers of southwestern France. III. Relationships between communities spatial and temporal dynamics, *Canadian journal of botany* **76**, 1998a, 115–121.

Fabre E., Aquatic hyphomycetes in three rivers of southwestern France. II. Spatial and temporal differences between species, *Can. J. Bot.* **76**, 1998b, 107–114, https://doi.org/10.1139/b97-156.

Fabre E., Aquatic hyphomycetes in three rivers of southwestern France. I. Spatial and temporal changes in conidial concentration, species richness, and community diversity, *Can. J. Bot.* **76**, 1998c, 99–106, https://doi.org/10.1139/b97-155.

Frissell C.A., Liss W.J., Warren C.E. and Hurley M.D., A hierarchical framework for stream habitat classification: viewing streams in a watershed context, *Environ. Manage* **10**, 1986, 199–214, https://doi.org/10.1007/BF01867358.

Gessner M.O., Fungal biomass, production and sporulation associated with particulate organic matter in streams, *Limnética* **13**, 1997, 33–44.

Gorniak A., Więcko A. and Cudowski A., Fungi biomass in lowland rivers of north-eastern Poland: effects of habitat conditions and nutrient concentrations, *Pol. J. Ecol.* **61**, 2013, 749–758.

Green W.R. and Haggard B.E., Phosphorus and Nitrogen Concentrations and Loads at Illinois River South of Siloam Springs, 2001, 1997–1999, Arkansas.

Grinn-Gofroń A. and Bosiacka B., Effects of meteorological factors on the composition of selected fungal spores in the air, *Aerobiologia* **31**, 2015, 63–72, https://doi.org/10.1007/s10453-014-9347-1.

Haggard B.E., Phosphorus concentrations, loads, and sources within the Illinois River drainage area, northwest Arkansas, 1997–2008, *J. Environ. Qual.* **39**, 2010, 2113–2120.

Haggard B.E. and Soerens T.S., Sediment phosphorus release at a small impoundment on the Illinois River, Arkansas and Oklahoma, *USA. Ecol. Eng* **28**, 2006, 280–287.

Hart M.R., Quin B.F. and Nguyen M., Phosphorus runoff from agricultural land and direct fertilizer effects, *J. Environ. Qual.* **33**, 2004, 1954–1972.

van Der Heijden M.G., Bardgett R.D. and Van Straalen N.M., The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems, *Ecol. Lett.* **11**, 2008, 296–310.

Hulvey J., Gobena D., Finley L. and Lamour K., Co-occurrence and genotypic distribution of Phytophthora species recovered from watersheds and plant nurseries of eastern Tennessee, *Mycologia* **102**, 2010, 1127–1133.

Ingold C.T., Aquatic hyphomycetes of decaying alder leaves, *Trans. Br. Mycol. Soc.* **25**, 1942, 339IN1–417IN6.

Jackson D.A., PROTEST: a PROcrustean randomization TEST of community environment concordance, *Ecoscience* **2**, 1995, 297–303.

Jost L., Partitioning diversity into independent alpha and beta components, *Ecology* **88**, 2007, 2427–2439, https://doi.org/10.1890/06-1736.1.

King R.S., Baker M.E., Whigham D.F., Weller D.E., Jordan T.E., Kazyak P.F. and Hurd M.K., Spatial considerations for linking watershed land cover to ecological indicators in streams, *Ecol. Appl.* **15**, 2005, 137–153, https://doi.org/10.1890/04-0481.

Laliberté E., Lambers H., Burgess T.I. and Wright S.J., Phosphorus limitation, soil-borne pathogens and the coexistence of plant species in hyperdiverse forests and shrublands, *New Phytol.* **206**, 2015, 507–521, https://doi.org/10.1111/nph.13203.

Langille M.G., Zaneveld J., Caporaso J.G., McDonald D., Knights D., Reyes J.A., Clemente J.C., Burkepile D.E., Thurber R.L.V., Knight R., et al., Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences, *Nat. Biotechnol.* **31**, 2013, 814–821.

LeBrun E.S., King R.S., Back J.A. and Kang S., Microbial community structure and function decoupling across a phosphorus gradient in streams, *Microb. Ecol.* 2017, 1–10, https://doi.org/10.1007/s00248-017-1039-2.

Li D.-W., Biology of Microfungi, 2016, Springer.

Maguire B., The passive dispersal of small aquatic organisms and their colonization of isolated bodies of water, *Ecol. Monogr.* **33**, 1963, 161–185, https://doi.org/10.2307/1948560.

Marra G. and Wood S.N., Practical variable selection for generalized additive models, *Comput. Stat. Data Anal.* **55**, 2011, 2372–2387.

McMurdie P.J. and Holmes S., phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data, *PLos One* **8**, 2013, e61217.

Miller M.P., Blinn D.W. and Keim P., Correlations between observed dispersal capabilities and patterns of genetic differentiation in populations of four aquatic insect species from the Arizona White Mountains, *U.S.A. Freshw. Biol* **47**, 2002, 1660–1673, https://doi.org/10.1046/j.1365-2427.2002.00911.x.

Nguyen N.H., Song Z., Bates S.T., Branco S., Tedersoo L., Menke J., Schilling J.S. and Kennedy P.G., FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild, *Fungal Ecol* **20**, 2016, 241–248, https://doi.org/10.1016/j.funeco.2015.06.006.

Nikolcheva L.G. and Bärlocher F., Seasonal and substrate preferences of fungi colonizing leaves in streams: traditional versus molecular evidence, *Environ. Microbiol.* **7**, 2005, 270–280.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H. and Wagner H., Vegan: Community Ecology Package, 2016, R package version 2.4-1.

Pashley C.H., Fairs A., Free R.C. and Wardlaw A.J., DNA analysis of outdoor air reveals a high degree of fungal diversity, temporal variability, and genera not seen by spore morphology, *Fungal Biol.* **116**, 2012, 214–224.

Peay K.G. and Bruns T.D., Spore dispersal of basidiomycete fungi at the landscape scale is driven by stochastic and deterministic processes and generates variability in plant–fungal interactions, *New Phytol.* **204**, 2014, 180–191, https://doi.org/10.1111/nph.12906.

Peay K.G., Bidartondo M.I. and Arnold A.E., Not every fungus is everywhere: scaling to the biogeography of fungal–plant interactions across roots, shoots and ecosystems, *New Phytol.* **185**, 2010, 878–882.

Peay K.G., Schubert M.G., Nguyen N.H. and Bruns T.D., Measuring ectomycorrhizal fungal dispersal: macroecological patterns driven by microscopic propagules, *Mol. Ecol.* **21**, 2012, 4122–4136.

Pellegrino K.C.M., Rodrigues M.T., Waite A.N., Morando M., Yassuda Y.Y. and Sites J.W., Phylogeography and species limits in the Gymnodactylus darwinii complex (Gekkonidae, Squamata): genetic structure coincides with river systems in the Brazilian Atlantic Forest, *Biol. J. Linn. Soc.* **85**, 2005, 13–26, https://doi.org/10.1111/j.1095-8312.2005.00472.x.

Petersen I., Masters Z., Hildrew A.G. and Ormerod S.J., Dispersal of adult aquatic insects in catchments of differing land use, *J. Appl. Ecol.* **41**, 2004, 934–950, https://doi.org/10.1111/j.0021-8901.2004.00942.x.

R Core Team, R: a Language and Environment for Statistical Computing, 2015, R Foundation for Statistical Computing; Vienna, Austria.

Roberts D.W., Labdsv: Ordination and Multivariate Analysis for Ecology, 2013, R package version 1.6-1. h ttp. CRAN R-Proj. Orgpackage Labdsv.

Santamaría L., Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment, *Acta Oecol.* **23**, 2002, 137–154, https://doi.org/10.1016/S1146-609X(02)01146-3.

Savage D., Barbetti M.J., MacLeod W.J., Salam M.U. and Renton M., Seasonal and diurnal patterns of spore release can significantly affect the proportion of spores expected to undergo long-distance dispersal, *Microb. Ecol.* **63**, 2012, 578–585.

Schachtman D.P., Reid R.J. and Ayling S.M., Phosphorus uptake by plants: from soil to cell, *Plant Physiol.* **116**, 1998, 447–453.

Schindler D.W., Evolution of phosphorus limitation in lakes, *Science* **195**, 1977, 260–262.

Schlesinger W.H. and Melack J.M., Transport of organic carbon in the world's rivers, *Tellus* **33**, 1981, 172–187, https://doi.org/10.3402/tellusa.v33i2.10706.

Smouse P.E., Long J.C. and Sokal R.R., Multiple regression and correlation extensions of the Mantel test of matrix correspondence, *Syst. Zool.* **35**, 1986, 627–632.

Søndergaard M. and Jeppesen E., Anthropogenic impacts on lake and stream ecosystems, and approaches to restoration, *J. Appl. Ecol.* **44**, 2007, 1089–1094.

Sridhar K.R. and Bärlocher F., Viability of aquatic hyphomycete conidia in foam, *Can. J. Bot.* **72**, 1994, 106–110.

Sridhar K.R., Bärlocher F., Wennrich R., Krauss G.-J. and Krauss G., Fungal biomass and diversity in sediments and on leaf litter in heavy metal contaminated waters of Central Germany, *Fundam. Appl. Limnol. Für Hydrobiol.* **171**, 2008, 63–74.

Suberkropp K. and Klug M.J., Fungi and bacteria associated with leaves during processing in a woodland stream, *Ecology* **57**, 1976, 707–719.

Suberkropp K. and Wallace J.B., Aquatic hyphomycetes in insecticide-treated and untreated streams, *J. North Am. Benthol. Soc.* **11**, 1992, 165–171, https://doi.org/10.2307/1467382.

Sutton W., Hansen E.M., Reeser P.W. and Kanaskie A., Stream monitoring for detection of Phytophthora ramorum in Oregon tanoak forests, *Plant Dis.* **93**, 2009, 1182–

Taylor D.L. and Bruns T.D., Community structure of ectomycorrhizal fungi in a Pinus muricata forest: minimal overlap between the mature forest and resistant propagule communities, *Mol. Ecol.* **8**, 1999, 1837–1850.

Taylor D.L., Walters W.A., Lennon N.J., Bochicchio J., Krohn A., Caporaso J.G. and Pennanen T., Accurate estimation of fungal diversity and abundance through improved lineage-specific primers optimized for Illumina amplicon sequencing, *Appl. Environ. Microbiol.* **82**, 2016, 7217–7226, https://doi.org/10.1128/AEM.02576-16.

Thomas K., Chilvers G.A. and Norris R.H., A dynamic model of fungal spora in a freshwater stream, *Mycol. Res.* **95**, 1991, 184–188.

Van Der Heijden, M.G., Bardgett, R.D., Van Straalen, N.M., 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecol. Lett. 11, 296–310.

Vanschoenwinkel B., Gielen S., Vandewaerde H., Seaman M. and Brendonck L., Relative importance of different dispersal vectors for small aquatic invertebrates in a rock pool metacommunity, *Ecography* **31**, 2008, 567–577, https://doi.org/10.1111/j.0906-7590.2008.05442.x.

Warnes G.R., Bolker B., Bonebakker L., Gentleman R., Huber W., Liaw A., Lumley T., Maechler M., Magnusson A., Moeller S., et al., Gplots: Various R Programming Tools for Plotting Data, 2016, R Package Version 301 2.

Waters T.F., Sediment in Streams: Sources, Biological Effects, and Control, American Fisheries Society Monograph, 1995, American Fisheries Society; Bethesda, Md, 0362–1715 ;7.

Wickham H., ggplot: an Implementation of the Grammar of Graphics, 2006, R Package Version 210.

Wood S.N., mgcv: GAMs and generalized ridge regression for R, *R. News* **1**, 2001, 20–25.

CHAPTER FOUR

A Metagenome-based Investigation of Gene Relationships for Non-substrate Associated Microbial Phosphorus Cycling in the Water Column of Streams and Rivers

This chapter is a preprint of a paper currently in review for publication: LeBrun ES, King RS, Back JA, Kang S. 2018. A metagenome-based investigation of gene relationships for non-substrate associated microbial phosphorus cycling in the water column of streams and rivers. In review.

Abstract

Phosphorus (P) is a nutrient of primary importance in all living systems and it is especially important in streams and rivers which are sensitive to anthropogenic P inputs and eutrophication. Microbes are accepted as the primary mineralizers and solubilizers of P improving bioavailability for organisms at all trophic levels. Here we use a genomics approach with metagenome sequencing of 24 temperate streams and rivers representing a total P (TP) gradient to identify relationships between functional genes, functional gene groupings, P, and organisms within the P biogeochemical cycle. Combining information from network analyses, functional groupings, and system P levels, we have constructed a System Relational Overview of Gene Groupings (SROGG) which is a cohesive system level representation of P cycle gene and nutrient relationships. Using SROGG analysis in concert with other statistical approaches, we found that the compositional makeup of P cycle genes is strongly correlated to environmental P whereas absolute abundance of P genes shows no significant correlation to environmental P. We also found orthophosphate (PO_4^{3-}) to be the dominant factor correlating with system P cycle gene composition with

little evidence of a strong organic phosphorous correlation present even in more oligotrophic streams.

Introduction

Phosphorus (P) is one of the three primary nutrients of importance in living systems alongside Carbon (C) and Nitrogen (N) [1]. The nutrient P is important in numerous processes necessary for life including the production of ATP, nucleic acids, and phospholipids. There is a large body of evidence to suggest that P is the limiting nutrient in a number of systems [2–6]. An influx of anthropogenic P to a system can have significant impact on an environment and ecosystem, particularly a freshwater ecosystem resulting in eutrophication [7–9]. The bulk of organisms can only utilize orthophosphates (PO_4^{3-}) for biological processes [10–13] and there are limited natural inputs of PO_4^{3-} to freshwater systems; the primary source of PO_4^{3-} being eroded mineral rocks. Due to the limitation of natural PO_4^{3-} inputs into systems, microbial populations are believed to play an important role in P biogeochemical cycling due to their ability to mineralize organic phosphorus making it available to other organisms and their ability to solubilize P by changing the properties of systems to favor the release of bioavailable PO_4^{3-} [14].

Anthropogenic sources are primarily wastewater along with agricultural fertilization and ranching [8, 15] and these inputs are predicted only to increase in the future [16]. In consideration of this, it is important that we understand the interactions and response of ecosystems to the related increase in P. Because of their unique P biogeochemical abilities and foundational role in organic P cycling, microbes make an excellent target for studying P relationships in ecosystems. There has already been much work towards understanding microbes and P. The bulk of work on microbial

communities in natural freshwaters has focused on the interactions of P, N, C, and biomass [17–21] or on chemical properties using methods like isotope detection of nutrients [22, 23]. More detailed mechanistic study has occurred on specific genes [24–28] and specific organisms (e.g. *Acinetobacter*) [24, 29–31] although the bulk of this work has been conducted in more simple and controlled environments than what is often seen in nature. Unfortunately, these findings are difficult to translate to complex microbial communities consisting of large numbers of microbial taxa and populations in an environmental setting. Microbial organisms exist in nature as members of complex community amalgamations rather than monocultures suggesting biogeochemical pathways should be considered in the whole community metagenome rather than the genome of any specific isolated organism in order to understand microbial community ecosystem interactions. There exists then a gap in our understanding of microbial P biogeochemical processes at the community and system levels in terms of mechanisms. In fact, the microbial input to P biogeochemical systems is often written off as too complex and microbial interactions are summarized as general mineralization and solubilization [12].

Metagenome sequencing offers an opportunity to address the gap in community and system microbial P biogeochemical with a quasi-mechanistic approach utilizing genomics and bioinformatics. Here, we use the microbial metagenomes of 24 temperate stream sites to investigate the abundance and composition of microbial P cycle related genes and their relationships to P and to each other within the system. The 24 streams and rivers represent a gradient of total P (TP) and PO_4^{3-}. We focus on the non-substrate associated bacterial/archaeal community in the water column by filtering out particle-

associated communities [32, 33]. Looking at functional gene groupings representing P biogeochemical cycle pathways within communities as a whole rather than isolated to specific organisms we target an approach between the general "mineralization and solubilization" and the "specific genes within specific organisms" in order to build a meaningful representation of microbial P cycling at the community and system level. We hypothesized that gene composition and abundance would be strongly associated with TP and that functions related to organic or inorganic P would exhibit strong relationships to levels of those nutrients accordingly.

Methods

Sampling and Sequencing

The study area was a collection of streams and rivers along the Oklahoma-Arkansas border, an area with known P enrichment problems (Figure E.S1)[34–36]. The sampling sites represented a gradient of TP levels from 7 µg/L to 173 µg/L. The genomic DNA (gDNA) used in this study came from LeBrun et al. (2017)[32]. In brief, samples were collected ~10 cm below the surface in the water column and 50 mL of water was run through a stacked filtration of a 1 µm glass fiber filter and then a 0.2 µm membrane filter for each site. Only components collected on the 0.2 µm filters were used in extracting the gDNA. gDNA was sent to Molecular Research LP (MR DNA) (Shallowater, TX) for metagenome sequencing. gDNA first underwent linear amplification performed by MR DNA and then was sequenced using an Illumina HiSeq 2500 2x150bp paired end reads to a minimum depth of 10 million reads per sample.

Sequence Processing and Data Extraction

Resulting raw read data provided by MR DNA was then processed in house. Paired-end reads were aligned and filtered at a Phred score of 20 using the join-fastq algorithm from eautils [37]. Unpaired reads were discarded at this time. Remaining reads were then aligned using DIAMOND version 0.9.10 [38] against the NCBI NR database as downloaded on June, 15 2017 [39]. Mapping using MEGAN GI to taxonomy and GI to SEED was then performed on the alignments using MEGAN 6 community edition [40]. Taxonomic and SEED results were extracted from MEGAN 6 at multiple hierarchical levels in tab-delimited format for downstream analysis.

SEED Classified Groupings

SEED analysis resulted in three tiers of classifications for P cycling with the most general being simply "Phosphorus Utilization" and the most specific being known genes and enzymes. The six nodes of the mid-tier SEED classifications for P were selected as the functional "gene groupings" for this study. They include "P Uptake" for the uptake of PO_4^{3-}, "Phosphoenolpyruvate Phosphomutase" for the synthesis of phosphonates, "Alkylphosphate Utilization" for the processing of alkyl phosphates, "Phosphonate Metabolism" for phosphonate utilization, "High Affinity Phosphate Transporter and Control of PHO Regulon" (HAPHO) for intracellular transport and polyphosphate storage, and "Phosphate Metabolism" for general PO_4^{3-} related metabolic processes. A detailed breakdown of gene inclusions for each group is available in Supplementary Table 1. A seventh grouping, "Total Membrane Transport" was also included in order to quantify a relationship between water column bacteria/archaea and particle associated P. Note that some genes are found in multiple functional groups potentially increasing group

relatedness. We found capturing these relationships to be desirable as functions utilizing the same genes should be more related and our analyses downstream largely do not show an overpowering effect.

Phosphorus Data

Filtered water samples (0.45 µm) were analyzed for dissolved PO_4-P on a Lachat QC 8500 series 2 using the ascorbic acid-molybdate method [41]. Unfiltered water samples were acidified with H_2SO_4 (2 µl H_2SO_4 per ml sample) and digested with persulfate in an autoclave at 121 °C for an hour then analyzed on a Lachat QC 8500 series 2 using the ascorbic acid-molybdate method [41] for TP determination.

Statistical Analysis Software

All analyses were performed in the R software package v.3.2.3 [42] using the identified packages and scripts.

Network Analyses

Several relational networks were generated. All networks were constructed using network package version 1.13.0 and igraph package version 1.0.1 in R [43–45].

The first network constructed we refer to as the "intuitive" network. This network was constructed by hand assigning adjacency to P functional gene groupings as vertices based on the presence of shared genes or input and output molecules of the functional groups. Direction of relationships in this network was included using the same logic of output to input with an arrow representing the product input from the source and the tail representing the product output. Two way relational arrows were using in the case of

shared genes between functional groups or outputs and input being shared by both groupings.

The second network constructed we refer to as the "data-based" network and it was constructed using Bray-Curtis dissimilarity on the observed distribution of genes in the functional groupings within the system. This is a non-directional network to explore the dissimilarity relationship between functional groupings. We also constructed a third network using the same parameters with an inclusion of the "Total Membrane Transport" SEED group in the data based network to assess its relationships to TP and other P functional gene groupings.

The relationship between the data-based network and a non-directional version of the intuitive network was tested using a Mantel test with the Pearson correlation method and 1000 permutations on the representative adjacency matrices through the vegan package in R [46] in order to ascertain similarity between the data and the hypothesized relationships.

System Relational Overview of Gene Groupings (SROGG)

Using combined information from the system gene networks, we constructed a whole system representation we refer to as the System Relational Overview of Gene Groupings (SROGG). The purpose of the SROGG is to show a complex series of relationships including genes, pathways, and nutrients in one understandable community and system level view. It incorporates already known information on genes and pathways along with experimental system data to better explain microbial community P dynamics.

Information from all three networks was combined to create the SROGG. The majority of relationships in the SROGG came from the data-based network and the

magnitude of the edges. The relationships connecting Total Membrane Transport came from the inclusive network. Added to this are relationships identified in both the intuitive and data-based networks that we believed to be important but did not meet threshold values in the data-based network. We inferred directions of relationships from the intuitive network as the data-based network is non-directional. Lastly, we added hypothesized relationships that cannot be tested in the system due to data or experimental limitations.

For a visual representation of relative gene abundance in the context of TP and PO_4^{3-}, we included radar charts in our SROGG showing gene count numbers in each functional grouping relative to total number of P related genes detect at the site as well as relative TP and PO_4^{3-} levels. Radar charts were constructed using the fmsb package version 0.6.1in R [47].

Ordinations and Generalized Additive Models (GAMs)

Non-metric multidimensional scaling (NMDS) ordinations were constructed using Bray-Curtis dissimilarity through the vegan package for each functional gene grouping except for Phosphoenolpyruvate Phosphomutase as it did not possess enough individual genes to ordinate. Gradient GAM models for four variables, TP, PO_4^{3-}, PO_4:TP ratio, and non-PO_4 were built on the NMDS ordinations using ordisurf from the vegan package [48]. NMDS plots for ordinations with significant GAM models were created using the ggplot2 package in R [49]. Non-ordinal GAMs were also built using both absolute and relative gene counts in each functional grouping as well as absolute total P gene counts with TP, PO_4^{3-}, PO_4:TP ratio, and non-PO_4 using the mgcv package version 1.8-19 [50].

Taxonomic Bubble Plots

Bubble Plots for identifiable taxa associated with each functional gene group were constructed using the ggplot2 package. Taxonomic identities were assigned through MEGAN using the May, 2017 NCBI protein accession to taxonomy database [39]. Prior to plotting, taxonomic assignments were assessed manually and corrections were made to account for missing or incorrectly identified hierarchical taxonomic levels. The data contained a diverse mix of taxonomic calls terminating at multiple hierarchical levels so all information was included in bubble plots sorted in an orthological manner so that levels are more easily discernable.

Results

Networks

The Mantel test between the adjacency matrices of the intuitive and data-based networks resulted in a Mantel statistic of 0.67 with $p = 0.003$. The primary differences between the two networks (Figure 4.1) include a connection directly between TP and the Phosphoenolpyruvate Phosphomutase in the data-based network, a connection between Phosphoenolpyruvate Phosphomutase and HAPHO in the intuitive network, a connection between Phosphoenolpyruvate Phosphomutase and P Uptake in the intuitive network, and a connection between Phosphonate metabolism and HAPHO in the intuitive network. The Total Membrane Transport inclusive network was very similar to the data-based network with the addition of edges between Total Membrane Transport and P Uptake, P Metabolism, and HAPHO (Figure E.S2).

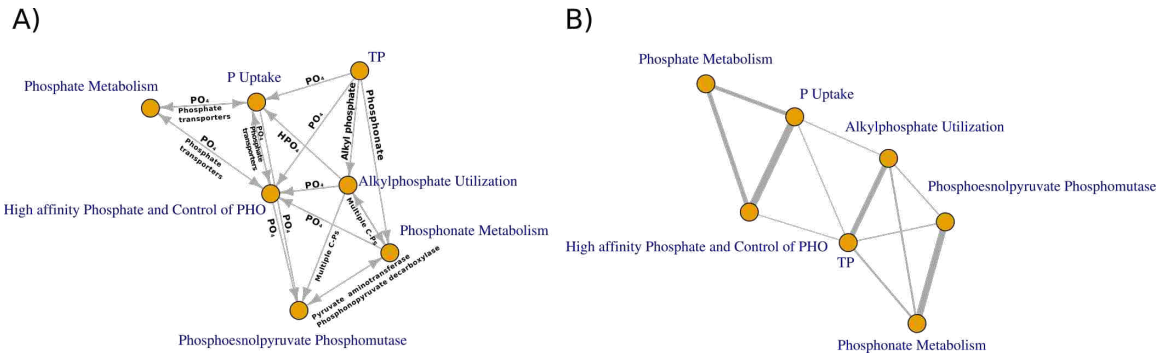


Figure 4.1: Networks: A) "Intuitive" network. Directions were chosen based on factors as labelled in the figure. B) “Data-based” network. Edges are weighted according to strength of the relationship (1 – Bray-Curtis dissimilarity). Bray-Curtis cutoff for visualization = 0.935.

System Relational Overview of Gene Groupings (SROGG)

We summarized the combination of information from the 3 generated networks as well as how we hypothesize P reserves and solubilized/insolubilized P relates to functional groupings (Figure 4.2). The relative presence radar charts show Phosphate Metabolism prevalence to be high across all sites regardless of P levels. Relative presence of genes in other functional groups tended to be more sporadically distributed across the range of P although general trends can be observed. For example, P Uptake appears relatively higher at lower P sites along with HAPHO and Total Membrane Transport. Alkylphosphate Utilization appears to be relatively higher at high P sites. Relative Phosphoenolpyruvate Phosphomutase shows no real trends in terms of total relative abundance nor does Phosphonate Metabolism.

NMDS Ordinations and GAMs

The ordinated composition of each functional gene group was fitted with TP, PO_4^{3-}, non-PO_4 and PO_4:TP ratio GAMs. All NMDS ordination and fit GAM results are shown in Table 4.1. TP and PO_4^{3-} resulted in several significant models while Non-PO_4

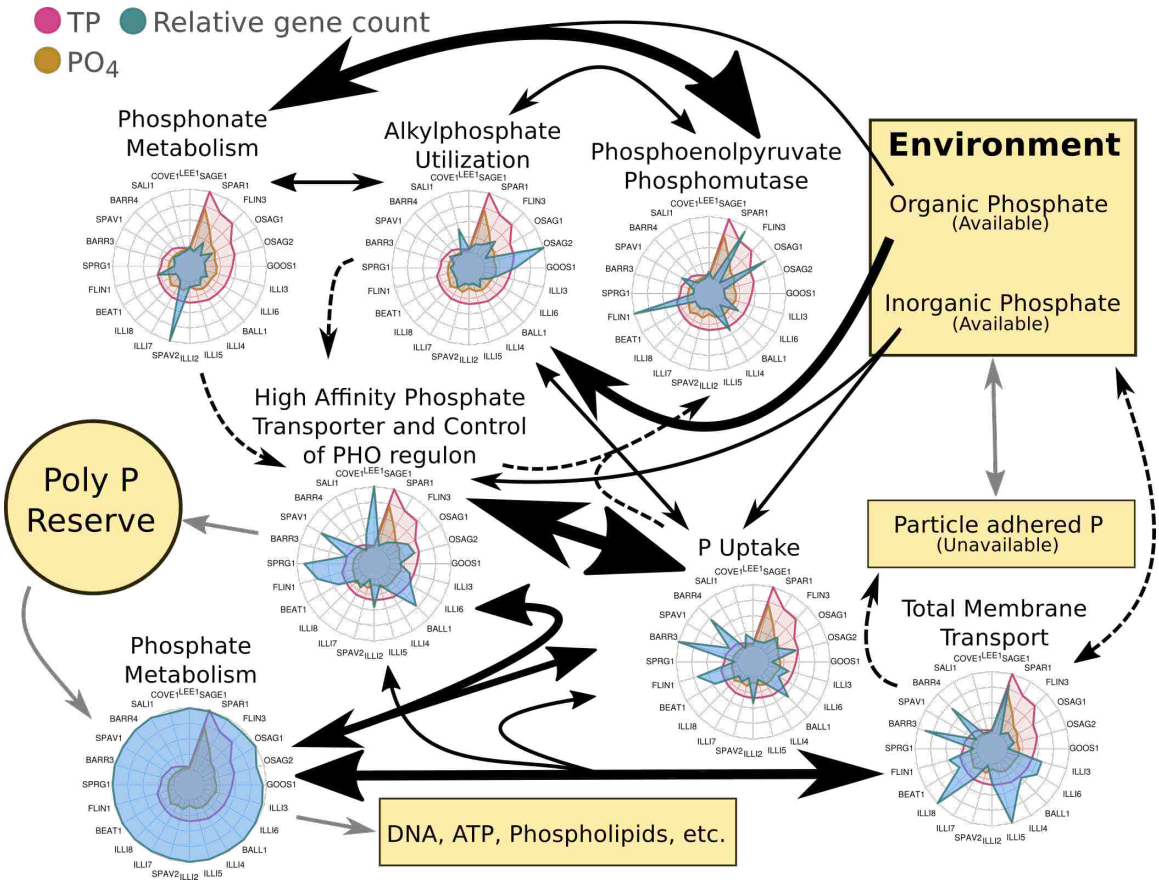


Figure 4.2: System Relational Overview of Gene Groupings (SROGG). Solid black lines are supported in and weighted by the data based network from Figure 4.1. Dotted lines are supported in the intuitive network and are believed important but fell below the threshold value in the data based network. Gray lines are hypothesized relationships as they involve objects or distinctions not included in the study. The outer ring of the radar charts represents the maximum value of TP or relative gene abundance independent of each other. PO43- is proportional to TP.

Table 4.1: Results from NMDS ordinations and fitted GAM models.

		Total P genes	Phosphonate Metabolism	Alkylphosphate Utilization	HAPHO	Phosphate metabolism	P uptake
K (Axes)		3	2	2	3	3	3
Stress value		0.107	0.06	0.104	0.06	0.106	0.06
TP							
	Deviance Explained	**54.7%**	3.75%	13.9%	**47.1%**	**48.6%**	0.38%
	p	**0.001**	0.368	0.114	**0.006**	**0.002**	0.383
	Model Axes	**1&3**	1&2	1&2	**1&2**	**2&3**	1&2
PO4							
	Deviance Explained	**61.3%**	37.2%	32.1%	**66%**	**61.3%**	**35.8%**
	p	**0.001**	0.092	0.07	**0.001**	**0.001**	**0.045**
	Model Axes	**2&3**	1&2	1&2	**1&2**	**2&3**	**2&3**
PO4:TP							
	Deviance Explained	~0%	~0%	~0%	3.84%	~0%	~0%
	p	~1	~1	~1	0.318	~1	~1
	Model Axes	NA	NA	NA	1&3	NA	NA
Non-PO4							
	Deviance Explained	**48%**	~0%	4.26%	8.78%	**33%**	1.67%
	p	**0.002**	~1	0.267	0.222	**0.023**	0.37
	Model Axes	**1&3**	NA	1&2	2&3	**2&3**	1&2

*Bold values indicate a significant result ($\alpha = 0.05$)

resulted in 2 significant models and PO_4:TP ratio resulted in none. The ordination and GAM for all P genes in the system are shown in Figure 4.3. Ordination and GAMs for all remaining significant results are provide in the Supplementary Figures (Figure E.S3 – E.S10). There were no significant models for the Phosphonate Metabolism or Alkylphosphate Utilization gene groupings.

Taxonomic Contributions to Functional Gene Groups

A significant portion 45.3% to 100% (mean = 90.1% SD = 11.9%) of detected genes at all sites for all functional groups (45.3% to 100% (mean = 90.1% SD = 11.9%) were identified to a taxonomic hierarchy although the hierarchical level of assignment varies greatly. The bubble plot showing taxa associated with Alkylphosphate Utilization

can be viewed in Figure 4.4. Bubble plots for all remaining functional gene groups are much larger and more complex and can be viewed in the Supplementary Figures (Figure E.S11-E.S15). Overall, the distribution of taxa across the TP gradient for the functional groups was diverse and showed no real trends or clusters with the exception of possibly *Acinetobacter*, which was seen much more frequently and low TP sites.

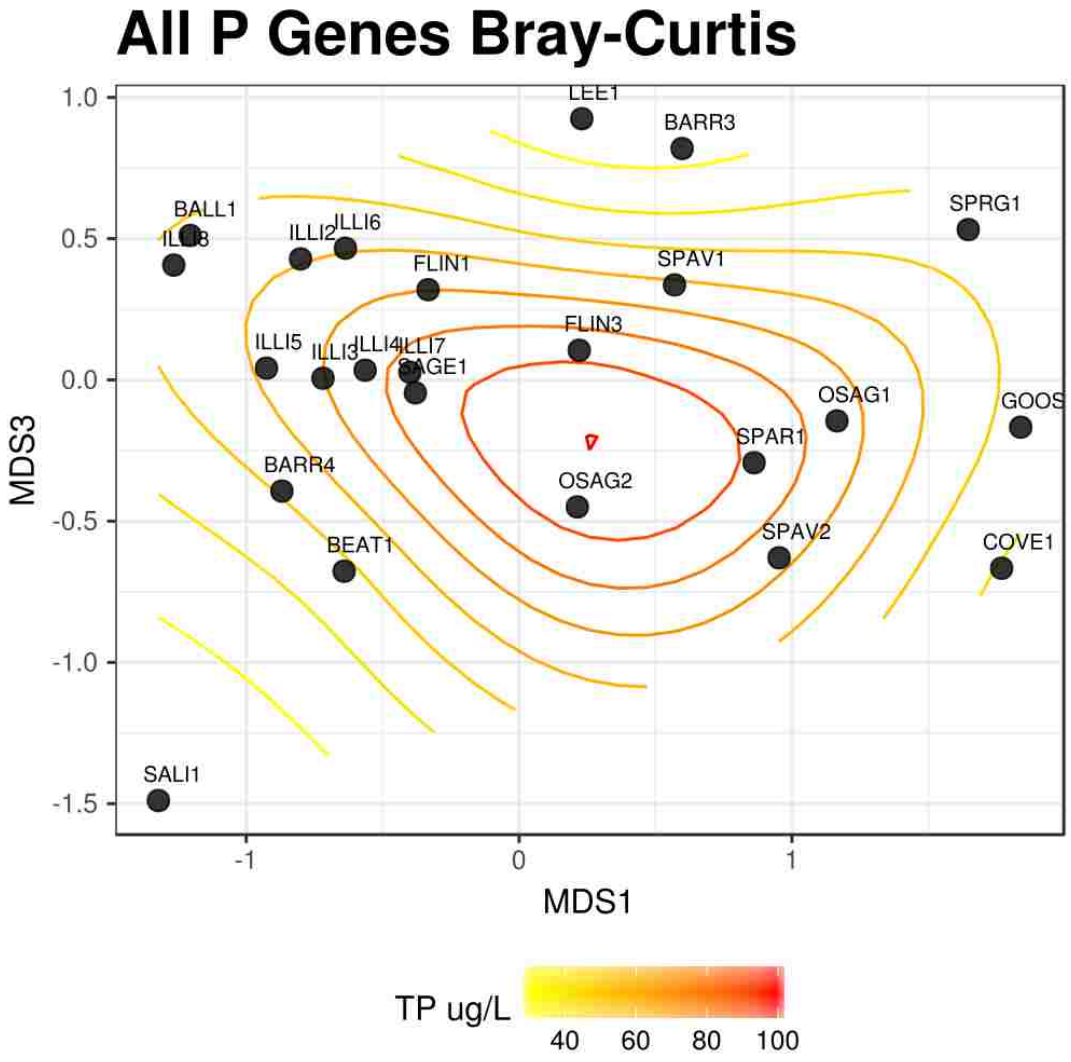


Figure 4.3: NMDS ordination of all P cycle related genes at each site (stress = 0.107). The TP gradient is fit using a GAM (Dev exp = 54.7%, p = 0.001).

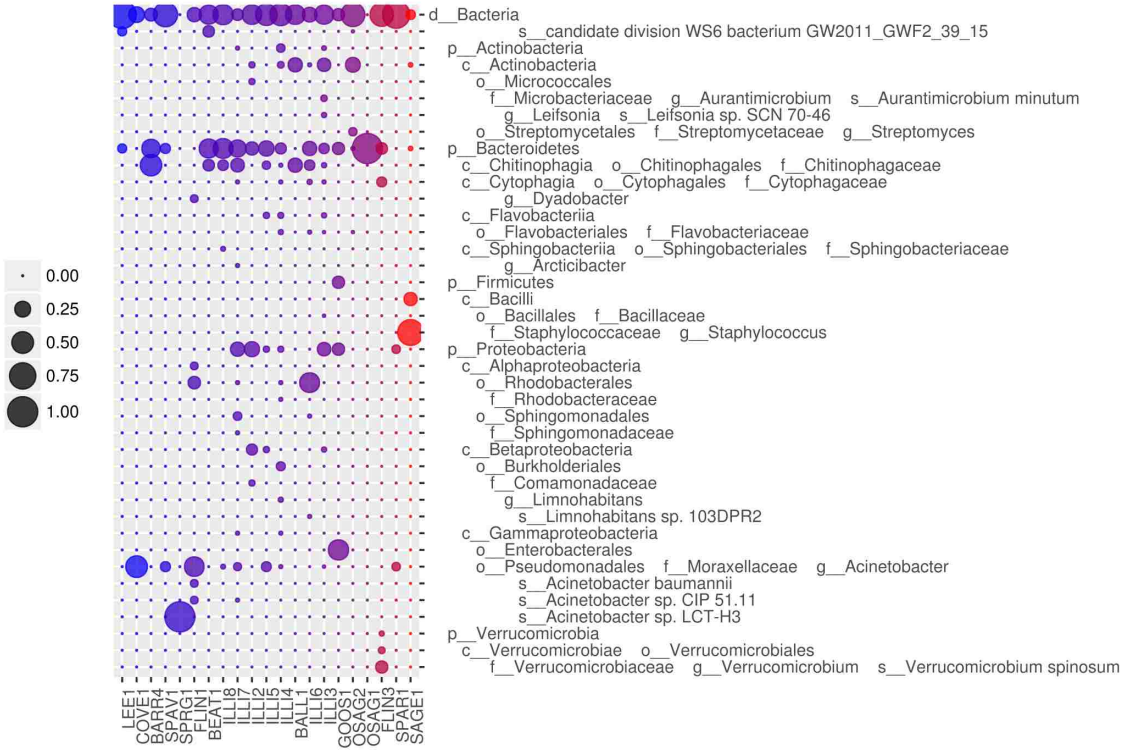


Figure 4.4: Taxa identified within the Alkylphosphate Utilization gene grouping. Only sites with detected genes in that gene grouping are shown. Sites are arranged from low TP (Blue) to high TP (Red).

Discussion

Intuitive and Data-Based Network Differences

A connection between TP and Phosphoenolpyruvate Phosphomutase was not included in the intuitive network because it was believed that although the Phosphoenolpyruvate Phosphomutase functional group produces organic phosphates, the relationship was not expected to be meaningful enough. Inclusions of the Phosphoenolpyruvate Phosphomutase edges in the intuitive network were mostly due to the idea that the PO_4^{3-} used in the pathway has to originate from somewhere, however, no significant explanatory relationships exist in the data-based network. Similarly, the edge between Phosphonate Metabolism and HAPHO was included in the intuitive network

because the resulting PO_4^{3-} from the functional group then has to be used by the system. Again, the relationship was not significant in the data-based network.

System Relational Overview of Gene Groupings

Strong relationships shown to HAPHO are all intuitive as the groupings provide or receive PO_4^{3-} in phosphorus metabolic processes. The strong relationships between some other groupings such as that between Phosphonate Metabolism and Phosphoenolpyruvate Phosphomutase may be explained by similar taxa possessing multiple organophosphate pathways; however, the strength of the Alkylphosphate Utilization gene grouping to TP levels is more difficult to explain. It is especially puzzling due to the fact the SROGG shows more relative Alkylphosphate Utilization genes, for organophosphate pathways, at higher TP/PO_4 levels. One potential explanation is that alkyl phosphate sources include pesticides used in agriculture [51, 52]. Presence of alkyl phosphate can be used to measure exposure to pesticides [53, 54] and other processed material such as flame retardants [55]. High Alkylphosphate Utilization sites in the study are all associated with significant wastewater discharge. This wastewater discharge includes human waste, pasture, and poultry processing waste. The strong presence of organisms capable of processing alkyl phosphates at these sites suggests that a large amount of alkyl phosphate is making it into these streams through wastewater discharge or that the organisms themselves originate from the wastewater treatment process. Direct alkyl phosphate measurements would confirm this along with identification of the source of alkyl phosphate.

Another interesting set of relationships are those of Total Membrane transport to PO_4^{3-} related gene groupings. Phosphate Metabolism, HAPHO, and P Uptake all possess

some form of transport genes and so make up some portion of the Total Membrane Transport gene grouping. What is surprising is how strong these relationships are. The Total Membrane Transport gene grouping is compiled using all known membrane transport genes from the metagenome data. P related genes are only a small portion of the total metagenome dataset. It is apparent from our findings though that Total Membrane Transport is strongly related to PO_4^{3-} gene groupings and in particular to Phosphate Metabolism.

A gene expression study with the cyanobacteria *Microcystis* und P stress showed strong up-regulation of PHO genes in the HAPHO group at low P levels [56]. The gradient design of the study captures enough ecological and evolutionary response to detect this in the metagenome as well. This can be seen in the distribution of relative HAPHO genes at low P sites as shown in the HAPHO radar plot of the SROGG (Figure 4.2).

Community Gene Composition and P

Our results clearly show that the compositional makeup of genes within communities is correlated to system P levels demonstrating that community genes are sorting differently based on P levels. While GAMs on ordinations of gene composition produced several significant models with high deviance explained, the GAMs performed on absolute abundance, relative abundance, and diversity of genes and groupings resulted in no significant models at all. This finding is enhanced visually in the taxonomic bubble plots. No consistent shifts of relative abundance are immediately apparent in any of the specific taxa associated with P cycle genes. *Acinetobacter* may be an exception to this as it looks to feature more heavily in the low P sites and in the gene compositions of

organophosphate related functional groupings. Previously, we identified an increased presence of *Acinetobacter* at the low P site in the system as well with 16S rRNA gene-based taxonomic data with computational functional predictions and hypothesized the taxa to be important in the P cycle at these sites due to its ability to accumulate polyphosphates and polyhydroxyalkanoates [32, 58].

From the same study, we also found several taxa that increase or decrease within the system around a threshold TP level of ~55 μg/L [32]. These findings may seem contradictory; however, the current study tells only that we do not see clear trends of taxonomic shift in specific taxa originating genes within the functional groups. The results suggest that there is importance to the compositional makeup of genes in the community independent of taxa which seems to corroborate our previously proposed decoupling of taxonomic structure and function within the system [32].

Importance of Organic Phosphorus

Our results show that functional group gene composition in this system is dominantly correlated with PO_4^{3-}. While we did not have direct measurements of different organophosphates, the non-PO_4 fraction of TP resulted in significant models for only Total P Genes and Phosphate Metabolism and deviance explained was much lower than models including PO_4^{3-}. Additionally, no P metrics from the system resulted in significant models for the functional groupings specific to organic phosphorous. However, when considering the grouping as a whole in the system, relationships with environmental P do exist as shown in the SROGG analysis.

There is undoubtedly an evolutionary impetuous to organisms possessing P mineralizing genes and related organisms in this system are likely to be using available

organophosphate sources. In NMDS GAM models, the fraction containing organophosphate (non-PO_4) fails to explain distributions in ordination space but the SROGG shows some strong relationships between organophosphate gene groupings and TP such as the relationship between TP and Alkylphosphate Utilization. Due to the identified relationship but inability to model, it is difficult to quantify the importance of organophosphate in community gene sorting within this system. Genes for processes associated with organophosphate processing are not constantly abundant at every site despite relative non-PO_4 being fairly invariable across all. However, there are definitely examples of sporadically distributed bacterial communities in the system with an abundance of these mechanisms available. These findings seem to suggest that the importance of mineralization to the system is context specific to other site properties and not strongly related to measured P levels. Instead it is likely related to factors not captured in this study such as the availability of specific organophosphates as suggested by the Alkylphosphate Utilization relationship or more complex ecological involvement like the need for production of phosphonolipids [59] or phosphonate dependent antibiotics [60].

Solubilization of P

Our results investigating solubilization using Total Membrane Transport should be viewed cautiously. Exudation of protons, mineral ions, organic acids, and other molecules can affect the solubility of particle-associated P [61, 62]. It is for this reason we targeted Total membrane transport as an indirect measure of solubilization. Because this measure is indirect at best, it is still difficult to realistically quantify solubilization here although it is safe to say some relationship exists based on our data. The solubility of

P in the system likely results from a complex mixture of factors including pH, particle and substrate-associated microbes, and turbidity or sedimentation levels just to name a few. Still, using total membrane transport gives us a straightforward and semi-quantitative way of indirectly gauging solubilization in some capacity within the system.

Ecosystem P Cycling and Genomics

Perhaps one of the most important takeaways from this study is that despite the extensive study of environmental P cycling, there is still a lot of work to be done. Certain genes inherently lend themselves better to taxonomic assignment than others [63]. However, the varied accuracy to which taxa were able to be assigned to our detected genes and the highly variable hierarchical level assignment highlight a need for continued work improving references for genomic P study. There is still much to be learned about how P biogeochemical cycling works at the whole community and system levels and how environmental relationships may be impacted in the future.

Conclusions

We conclude that although P biogeochemical cycling in microbial communities is complex, it can be meaningfully represented in a cohesive manner using the SROGG method. We infer from our results that the composition and distribution of P related genes is stronger in its relationship to natural P levels than total gene abundance. PO_4^{3-} is the dominant form of P studied in relation to the P gene composition of non-substrate associate microbial communities in this system. The fraction of P containing organic P does not show significant relation to gene sorting for any functional gene grouping including gene groupings specifically utilizing organic P. Although we present a cohesive

picture of genomic P relationships and gene sorting in a natural system here, there is still much to be learned about the mechanics at the system and community levels and further study including improvement of P related reference databases is still needed.

Acknowledgements

The authors thank Morgan Bettcher, Stephen Cook, Stephen Elser, Katherine Hooker, Lauren Housley and Caleb Robbins for their help in collecting field samples. We also thank Owen Lind and J. Thad Scott for assistance with internal review. We acknowledge the research support by Baylor University Office of Research and Baylor University Center for Reservoir and Aquatic Systems Research (CRASR).

Conflict of Interest

The authors declare no conflict of interest.

Statement of Data Availability

Sequence data that support the findings of this study have been deposited in GenBank with the BioProject accession code PRJNA350288. The environmental data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplementary information is available at the *Microbial Ecology Journal* website.

Chapter References

1. Miltner RJ, others (1998) Primary nutrients and the biotic integrity of rivers and streams. Freshw Biol 40:145–158

2. Schindler DW (1977) Evolution of phosphorus limitation in lakes. Science 195:260–262

3. Elser JJ, Bracken ME, Cleland EE, et al (2007) Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecol Lett 10:1135–1142

4. Vitousek PM, Porder S, Houlton BZ, Chadwick OA (2010) Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen–phosphorus interactions. Ecol Appl 20:5–15

5. Smith SV (1984) Phosphorus versus nitrogen limitation in the marine environment. Limnol Oceanogr 29:1149–1160

6. Krom MD, Kress N, Brenner S, Gordon LI (1991) Phosphorus limitation of primary productivity in the eastern Mediterranean Sea. Limnol Oceanogr 36:424–432

7. Anderson DM, Glibert PM, Burkholder JM (2002) Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. Estuaries 25:704–726 . doi: 10.1007/BF02804901

8. Correll DL (1998) The Role of Phosphorus in the Eutrophication of Receiving Waters: A Review. J Environ Qual 27:261 . doi: 10.2134/jeq1998.00472425002700020004x

9. Sharpley AN, Chapra SC, Wedepohl R, et al (1994) Managing agricultural phosphorus for protection of surface waters: Issues and options. J Environ Qual 23:437–451

10. Frossard E, Condron LM, Oberson A, et al (2000) Processes governing phosphorus availability in temperate soils. J Environ Qual Madison 29:15

11. Richardson AE (2001) Prospects for using soil microorganisms to improve the acquisition of phosphorus by plants. Funct Plant Biol 28:897–906

12. Richardson AE, Simpson RJ (2011) Soil microorganisms mediating phosphorus availability update on microbial phosphorus. Plant Physiol 156:989–996

13. Rier ST, Kinek KC, Hay SE, Francoeur SN (2016) Polyphosphate plays a vital role in the phosphorus dynamics of stream periphyton. Freshw Sci 35:490–502

14. Cosgrove DJ (1977) Microbial Transformations in the Phosphorus Cycle. In: Advances in Microbial Ecology. Springer, Boston, MA, pp 95–134

15. Bennett EM, Carpenter SR, Caraco NF (2001) Human Impact on Erodable Phosphorus and Eutrophication: A Global Perspective Increasing accumulation of phosphorus in soil threatens rivers, lakes, and coastal oceans with eutrophication. BioScience 51:227–234 . doi: 10.1641/0006-3568(2001)051[0227:HIOEPA]2.0.CO;2

16. Tilman D, Fargione J, Wolff B, et al (2001) Forecasting Agriculturally Driven Global Environmental Change. Science 292:281–284 . doi: 10.1126/science.1057544

17. Scott T, Cotner J, LaPara T (2012) Variable Stoichiometry and Homeostatic Regulation of Bacterial Biomass Elemental Composition. Front Microbiol 3: . doi: 10.3389/fmicb.2012.00042

18. Gorniak A, Więcko A, Cudowski A (2013) Fungi biomass in lowland rivers of north-eastern Poland: Effects of habitat conditions and nutrient concentrations. Pol J Ecol 61:749–758

19. Godwin CM, Whitaker EA, Cotner JB (2016) Growth rate and resource imbalance interactively control biomass stoichiometry and elemental quotas of aquatic bacteria. Ecology

20. Xu X, Thornton PE, Post WM (2013) A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. Glob Ecol Biogeogr 22:737–749 . doi: 10.1111/geb.12029

21. Cotner JB, Biddanda BA (2002) Small Players, Large Role: Microbial Influence on Biogeochemical Processes in Pelagic Aquatic Ecosystems. Ecosystems 5:105–121 . doi: 10.1007/s10021-001-0059-3

22. Boulion VV (2012) Assimilation and turnover time of phosphorus by size fractions of microplankton in lakes of different types. Inland Water Biol 5:304–309 . doi: 10.1134/S1995082912040049

23. Vadstein O, Olsen Y, Reinertsen H, Jensen A (1993) The Role of Planktonic Bacteria in Phosphorus Cycling in Lakes-Sink and Link. Limnol Oceanogr 38:1539–1544

24. Prágai Z, Harwood CR (2002) Regulatory interactions between the Pho and σB-dependent general stress regulons of Bacillus subtilis. Microbiology 148:1593–1602 . doi: 10.1099/00221287-148-5-1593

25. Qi Y, Kobayashi Y, Hulett FM (1997) The pst operon of Bacillus subtilis has a phosphate-regulated promoter and is involved in phosphate transport but not in regulation of the pho regulon. J Bacteriol 179:2534–2539 . doi: 10.1128/jb.179.8.2534-2539.1997

26. Brewis ND, Street AJ, Prescott AR, Cohen PT (1993) PPX, a novel protein serine/threonine phosphatase localized to centrosomes. EMBO J 12:987–996

27. Dyhrman ST, Chappell PD, Haley ST, et al (2006) Phosphonate utilization by the globally important marine diazotroph Trichodesmium. Nature 439:68–71 . doi: 10.1038/nature04203

28. Kamat SS, Williams HJ, Raushel FM (2011) Intermediates in the transformation of phosphonates to phosphate by bacteria. Nature 480:570–573 . doi: 10.1038/nature10622

29. Jansson M (1993) Uptake, Exchange, and Excretion of Orthophosphate in Phosphate-Starved Scenedesmus quadricauda and Pseudomonas K7. Limnol Oceanogr 38:1162–1178

30. Ghaffar S, Stevenson RJ, Khan Z (2017) Effect of phosphorus stress on Microcystis aeruginosa growth and phosphorus uptake. PLoS One San Franc 12: . doi: http://dx.doi.org/10.1371/journal.pone.0174349

31. Saralov A, Mol'kov D, Bannikova O, et al (2000) Intracellular accumulation of monomer precursors of polyphosphates and polyhydroxyalkanoates from Acinetobacter calcoaceticus and Escherichia coli. Mikrobiologiia 70:737–744

32. LeBrun ES, King RS, Back JA, Kang S (2018) Microbial Community Structure and Function Decoupling Across a Phosphorus Gradient in Streams. Microb Ecol 75:64–73 . doi: 10.1007/s00248-017-1039-2

33. LeBrun ES, Taylor DL, King RS, et al (2018) Rivers may constitute an overlooked avenue of dispersal for terrestrial fungi. Fungal Ecol In press:

34. Green WR, Haggard BE (2001) Phosphorus and nitrogen concentrations and loads at Illinois River south of Siloam Springs, Arkansas, 1997-1999

35. Haggard BE, Soerens TS (2006) Sediment phosphorus release at a small impoundment on the Illinois River, Arkansas and Oklahoma, USA. Ecol Eng 28:280–287 . doi: 10.1016/j.ecoleng.2006.07.013

36. Haggard BE (2010) Phosphorus concentrations, loads, and sources within the Illinois River drainage area, northwest Arkansas, 1997–2008. J Environ Qual 39:2113–2120

37. Aronesty E (2013) Comparison of sequencing utility programs. Open Bioinforma J 7:

38. Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60 . doi: 10.1038/nmeth.3176

39. NCBI Resource Coordinators (2017) Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res 45:D12–D17 . doi: 10.1093/nar/gkw1071

40. Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 17:377–386 . doi: 10.1101/gr.5969107

41. APHA (1998) Standard methods for the examination of water and wastewater. Am Public Health Assoc APHA Wash DC USA

42. R Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

43. Butts CT, others (2008) network: a Package for Managing Relational Data in R. J Stat Softw 24:1–36

44. Butts CT, Hunter D, Handcock M, et al (2015) Network: Classes for Relational Data

45. Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJournal Complex Syst 1695:1–9

46. Oksanen J, Blanchet FG, Kindt R, et al (2016) vegan: Community Ecology Package. R package version 2.4-1.

47. Nakazawa M (2017) fmsb: Functions for medical statistics book with some demographic data. R package version 0.6.1. CRAN R-Proj Orgpackage Fmsb

48. Marra G, Wood SN (2011) Practical variable selection for generalized additive models. Comput Stat Data Anal 55:2372–2387

49. Wickham H (2006) ggplot: An Implementation of the Grammar of Graphics. R Package Version 210

50. Wood S (2017) mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation

51. Cook AM, Daughton CG, Alexander M (1978) Phosphorus-containing pesticide breakdown products: quantitative utilization as phosphorus sources by bacteria. Appl Environ Microbiol 36:668–672

52. Pérez GL, Torremorell A, Mugni H, et al (2007) Effects of the Herbicide Roundup on Freshwater Microbial Communities: A Mesocosm Study. Ecol Appl 17:2310–2322 . doi: 10.1890/07-0499.1

53. Hernández F, Sancho JV, Pozo OJ (2002) Direct determination of alkyl phosphates in human urine by liquid chromatography/electrospray tandem mass spectrometry. Rapid Commun Mass Spectrom 16:1766–1773 . doi: 10.1002/rcm.790

54. Duggan A, Charnley G, Chen W, et al (2003) Di-alkyl phosphate biomonitoring data: assessing cumulative exposure to organophosphate pesticides. Regul Toxicol Pharmacol 37:382–395 . doi: 10.1016/S0273-2300(03)00031-X

55. Pedersen JA, Soliman M, Suffet IH (Mel) (2005) Human Pharmaceuticals, Hormones, and Personal Care Product Ingredients in Runoff from Agricultural Fields Irrigated with Treated Wastewater. J Agric Food Chem 53:1625–1632 . doi: 10.1021/jf049228m

56. Harke MJ, Gobler CJ (2013) Global Transcriptional Responses of the Toxic Cyanobacterium, Microcystis aeruginosa, to Nitrogen Stress, Phosphorus Stress, and Growth on Organic Matter. PLOS ONE 8:e69834 . doi: 10.1371/journal.pone.0069834

57. Barsdate RJ, Prentki RT, Fenchel T (1974) Phosphorus cycle of model ecosystems: significance for decomposer food chains and effect of bacterial grazers. Oikos 239–251

58. Stapleton RD, Singh VP (2002) Biotransformations: Bioremediation Technology for Health and Environmental Protection. Elsevier

59. Mukhamedova KS, Glushenkova AI (2000) Natural Phosphonolipids. Chem Nat Compd 36:329–341 . doi: 10.1023/A:1002804409503

60. Ternan NG, Grath JWM, Mullan GM, Quinn JP (1998) Review: Organophosphonates: occurrence, synthesis and biodegradation by microorganisms. World J Microbiol Biotechnol 14:635–647 . doi: 10.1023/A:1008848401799

61. Vazquez P, Holguin G, Puente ME, et al (2000) Phosphate-solubilizing microorganisms associated with the rhizosphere of mangroves in a semiarid coastal lagoon. Biol Fertil Soils 30:460–468 . doi: 10.1007/s003740050024

62. Rodríguez H, Fraga R (1999) Phosphate solubilizing bacteria and their role in plant growth promotion. Biotechnol Adv 17:319–339 . doi: 10.1016/S0734-9750(99)00014-2

63. Philippe H, Brinkmann H, Lavrov DV, et al (2011) Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. PLOS Biol 9:e1000602 . doi: 10.1371/journal.pbio.1000602

CHAPTER FIVE

A Comparison of Computationally Predicted Functional Metagenomes and Microarray Analysis for Microbial P Cycle Genes in a Unique Basalt-soil Forest

This chapter published as: LeBrun ES and Kang S. 2018. A comparison of computationally predicted functional metagenomes and microarray analysis for microbial P cycle genes in a unique basalt-soil forest. F1000Research. 7:179 (Version 1)(doi: 10.12688/f1000research.13841.1)

Abstract

Here we compared microbial results for the same Phosphorus (P) biogeochemical cycle genes from a GeoChip microarray and PICRUSt functional predictions from 16S rRNA data for 20 samples in the four spatially separated Gotjawal forests on Jeju Island in South Korea. The high homogeneity of microbial communities detected at each site allows sites to act as environmental replicates for comparing the two different functional analysis methods. We found that while both methods capture the homogeneity of the system, both differed greatly in the total abundance of genes detected as well as the diversity of taxa detected. Additionally, we introduce a more comprehensive functional assay that again captures the homogeneity of the system but also captures more extensive community gene and taxonomic information and depth. While both methods have their advantages and limitations, PICRUSt appears better suited to asking questions specifically related to microbial community P as we did here. This comparison of methods makes important distinctions between both the results and the capabilities of each method and can help select the best tool for answering different scientific questions.

Introduction

Relating the functionality of microbes to environmental factors is one of the primary goals in microbial ecology. With the advent of modern genomic technologies such as next generation sequencing and microarray hybridization there are more options than ever to test environmental community's genomics and functional capabilities. Metagenome sequencing is one of the most thorough and comprehensive methods currently available for looking at microbial community gene compositions (1–5) but can be costly and generate enormous data sets that require a large amount of work in processing, analysis, and storage. Two technologies currently in use for looking at community functional profiles that can be less expensive and more accessible than metagenome sequencing include computationally predicted functional metagenomes (PFMs) (6) and microarray analyses (7). These technologies both have known advantages and disadvantages (8) but investigation into how they compare in the same system is still needed.

Here we compare PFMs from Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) (6) to GeoChip (9) microarray data. While both methods are distinct, they can each be applied to an environmental community gene pool to estimate the presence and abundance of genes within the community genomic landscape related to function. Resulting datasets from each technique are tables showing counts of genes or functions as determined by either probes (microarray) or reference data (PFMs) and is therefore are directly comparable in the context of functional gene landscapes within the system. We utilize 20 sites in a unique basalt-soil Gotjawal forest on Jeju Island in Korea. Despite being both rocky, lava-

formed basalt and having dense vegetation (10), this forest is considered a wetland environment due to the homogenous, rocky soil and its capacity for absorbing water (11). All 20 sites, though spatially separated by distance of 5 km to 65 km (Figure G.S1), showed strong homogeneity in bacterial/archaeal community assemblies in 16S rRNA gene taxonomic analysis (Figure G.S2) and so act as replicates in this system for the current study. This makes it ideal for comparing the technologies. We specifically look at how these technologies perform related to the same phosphorus (P) cycle genes as the unique basalt-soil environment has the potential to be a unique P environment (12–14).

Methods

Data origination and processing

GeoChip 4.0 data for P cycle genes came from Kim et al. (15) For sequencing data, we started with raw sequencing files also from the study by Kim et al. (16). Paired-end reads were combined using the join-fastq algorithm from eautils (17). Un-paired reads were discarded at this time. Additional sequence processing was performed using Quantitative Insights Into Microbial Ecology (QIIME) version 1.9.1 (18). Sequences were then filtered with a maximum unacceptable Phred quality score of 20. Chimeric sequences were identified and removed using the UCHIME algorithm within USEARCH (19). Operational taxonomic unit (OTU) picking was performed via open reference using uclust against the Greengenes 13_8 database with a 0.97 similarity cutoff (20). Singleton sequences were removed during OTU picking and taxonomy was assigned with Greengenes 13_8 database as reference.

Only reads identified in closed reference picking were used for the PICRUSt analysis. Using PICRUSt (6), predicted functional metagenomes (PFMs) were constructed from the resulting 16S rRNA sequences. PFMs were generated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (21,22) as a functional reference.

Genes studied

The GeoChip 4.0 data provided probe data for genes identified as "*phytase*", "*ppk*", and "*ppx*". We identified these genes in the KEGG database to have the KEGG orthology (KO) numbers K01083 and K01093 for phytase, K00937 for *ppk*, and K01514 for *ppx* These KO numbers were the only PICRUSt results extracted for direct comparison. Additionally, we built another P assay in PICRUSt utilizing 417 KO numbers associated with P (Table S1).

Statistical analyses

All analyses were performed in the R software package v.3.2.3(23). The relationship between the PICRUSt and GeoChip data was tested using a Mantel test with the Pearson correlation method and 1,000 permutations through the vegan package (24). Non-metric multidimensional scaling (NMDS) ordinations were constructed using Bray-Curtis dissimilarity through the vegan package. A PROcrustean randomization TEST of community environment concordance (PROTEST), a potentially more sensitive detection method than a Mantel test, was also used to compare the NMDS ordinations to each other (25). Figures and plots were created using the ggplot2 package (26).

Results and discussion

Both PICRUSt and GeoChip appear to have captured the homogeneity of the system (Figure 5.1). PICRUSt captured much more diversity and depth in terms of taxa identified (Figure 5.1) and total counts (Figure 5.2) than GeoChip. PICRUSt identified organisms from 40 different phyla where GeoChip identified organisms from 15. Total counts at each site for the two methods were on a very different scale. When placed on a scale that shows the variation in each set of counts, it becomes apparent that the trends of total counts across sites do not match between methods (Figure G.S3). The Mantel test resulted in no significant statistic between the two data sets and Procrustes analysis confirmed this, showing no significant correlation either (Figure G.S4). The same analyses were performed with the data for each gene isolated and each of the three genes independently provided similar results of inconsistency between methods to the comparison of total gene datasets. There was no correlation between the datasets in Mantel or Procrustes analysis and gene counts and trends were markedly different.

The new PICRUSt assay with 417 P related genes captured the system homogeneity but with additional depth (Figure G.S5). The new assay identified organisms from 41 phyla similar to the smaller, comparative assay's 40 but also provided data counts per site ranging from ~70,000 to ~110,000. The PICRUSt dataset from the new assay not only represents what is likely a better dataset for answering community functional questions within the P cycle than the previous, comparative PICRUSt or GeoChip datasets but also illustrates an important difference between the two methods. While both methods could be considered "closed-format" technologies in that they are reliant on the available known references (8), the process of adapting or updating the two

methods contrasts. The method of using computational predictions is highly adaptable and allows for the easy inclusion or exclusion of additional genes (6). Improving or expanding the reference database that computational prediction can be achieved through simply updating the curated reference database. The microarray method is more involved including the identification, creation, and inclusion of specific target probes into the manufacturing of a microarray (7).

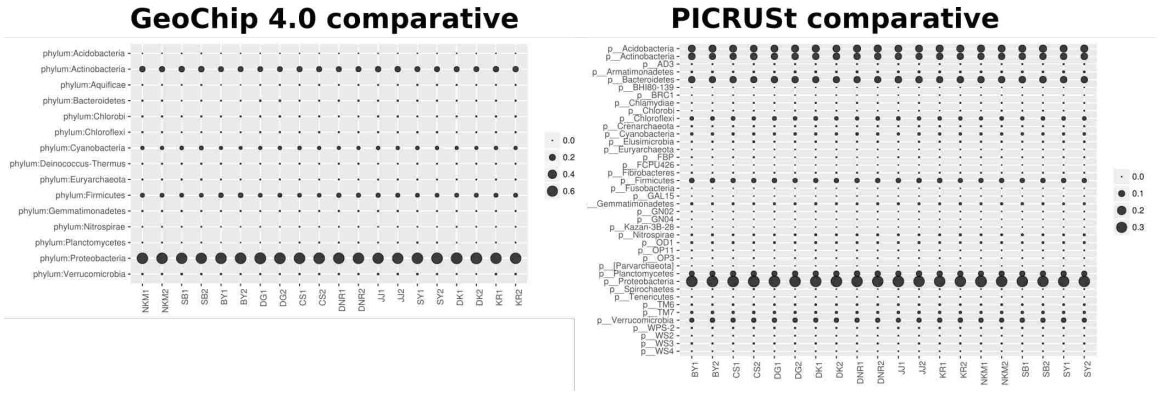


Figure 5.1: Bubble plots of taxa relative abundance detected by the GeoChip 4.0 array PICRUSt from 16S rRNA data for P cycle genes found on GeoChip array.

It is important to note that for our comparison we are specifically looking at functional genes within the P biogeochemical cycle. Both methods explored are designed for, and capable of looking a more comprehensive whole functional profile for communities. Computational functional prediction seems to be better suited to the task of viewing independent functional groupings as we did here. While microarrays have shown linear relationships to RNA and DNA levels in environmental systems(16,27), they are limited in coverage and small sequence divergence can affect quantitative capability (7). These quantitative limitations should be carefully considered in light of recent findings showing that the composition of P cycle genes in some microbial communities are more

closely related to environmental P levels than absolute abundance (1). Computational functional prediction again seems better equipped to handle questions related to functional gene composition due to the high specificity of probes to taxa and limited genes included in microarrays. It is also important to note that the data from both methods is representative of DNA present in microbial communities and not true expression levels or enzyme abundance.

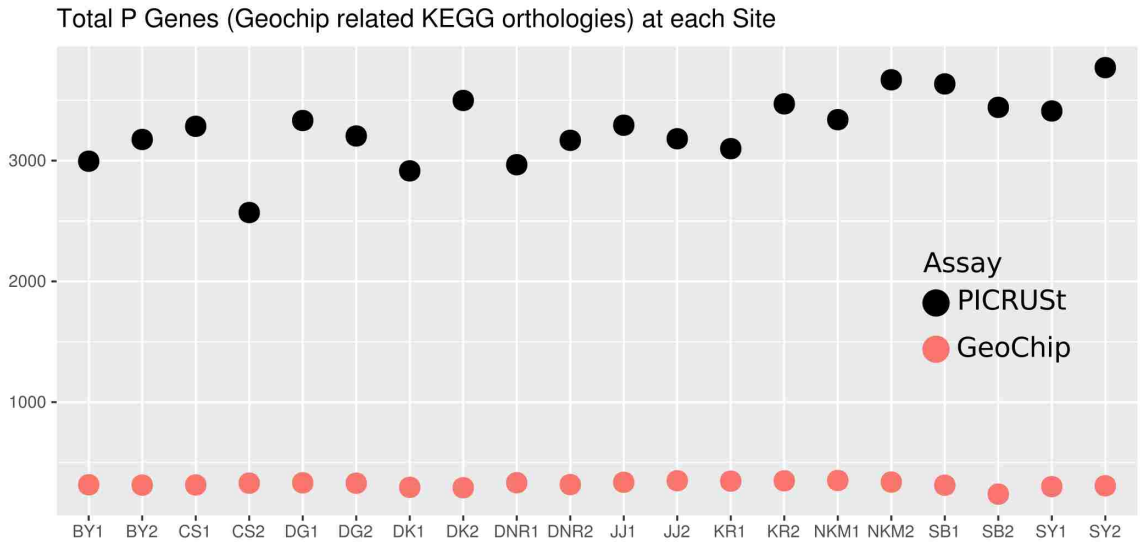


Figure 5.2: Plot of total P cycle gene counts as detected by PICRUSt and GeoChip at each site.

Conclusions

Computational functional prediction and microarray analysis of P cycle genes both captured system homogeneity. However, they did not agree in terms of capturing absolute abundance or taxonomic composition in P cycle genes. Computational functional prediction provided more count depth and taxonomic diversity than microarray analysis did. The ease with which computational functional prediction is adapted additionally allowed for the capture of additional genes and taxonomic diversity in P

function along with increased depth by expanding the PICRUSt assay to include 417 KO numbers related to P function instead of the original 4 used in the microarray comparison. While we compared two methods for the exploration of functional P cycle genes within microbial communities to each other, an additional comparison to whole metagenome data in a system would further validate either method.

Acknowledgements

The authors acknowledge Dr. Jong-Shik Kim (Geyongbuk Institute for Marine Bioindustry, Korea) for the data and the site map made available for this study.

Data availability

The sequence data used in this study was deposited in the NCBI Sequence Read Archive (SRA) under the BioSample accession numbers SAMN06049757 to SAMN06049776. The GeoChip microarray data used in this study is available in OSF: http://doi.org/10.17605/OSF.IO/AT93H28.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Chapter references

1. LeBrun ES, King RS, Back JA, Kang S. A metagenome-based investigation of gene relationships for non-substrate associated microbial phosphorus cycling in the water column of streams and rivers. Rev. 2018;

2. Daniel R. The soil metagenome – a rich resource for the discovery of novel natural products. Curr Opin Biotechnol. 2004 Jun 1;15(3):199–204.

3. Uchiyama T, Abe T, Ikemura T, Watanabe K. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat Biotechnol. 2004 Dec 19;23(1):nbt1048.

4. Inskeep WP, Jay ZJ, Tringe SG, Herrg\a ard MJ, Rusch DB, Committee YMPS, et al. The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. Front Microbiol. 2013;4.

5. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012 Sep 26;490(7418):nature11450.

6. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol. 2013;31(9):814–821.

7. He Z, Deng Y, Zhou J. Development of functional gene microarrays for microbial community analysis. Curr Opin Biotechnol. 2012 Feb 1;23(1):49–55.

8. Zhou J, He Z, Yang Y, Deng Y, Tringe SG, Alvarez-Cohen L. High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats. mBio. 2015 Feb 27;6(1):e02288-14.

9. Tu Q, Yu H, He Z, Deng Y, Wu L, Van Nostrand JD, et al. GeoChip 4: a functional gene-array-based high-throughput environmental technology for microbial community analysis. Mol Ecol Resour. 2014 Sep 1;14(5):914–28.

10. Kim J-S, Jung M-Y, Lee KC, Kim D-S, Ko S-H, Lee J-S, et al. The Archaea Community Associated with Lava-Formed Gotjawal Forest Soil in Jeju, Korea. J Agric Chem Environ. 2014;03(03):96.

11. Jang Y-C, Lee C-W. Gotjawal Forest in Jeju Island as an Internationally Important Wetland. J Wetl Res. 2009;11.

12. Toplis MJ, Libourel G, Carroll MR. The role of phosphorus in crystallisation processes of basalt: An experimental study. Geochim Cosmochim Acta. 1994 Jan 1;58(2):797–810.

13. Scheu S. Analysis of the microbial nutrient status in soil microcompartments: earthworm faeces from a basalt–limestone gradient. Geoderma. 1993 Mar 15;56(1):575–86.

14. Wells N, Saunders WMH. Soil studies using sweet vernal to assess element availability IV. Phosphorus. N Z J Agric Res. 1960 Apr 1;3(2):279–99.

15. Kim J-S, Kim D-S, Lee KC, Ko S-H, Lee J-S, King GM, et al. Microbial Community Structure and Functional Potential of Lava-Formed Gotjawal Soils in Jeju, Korea. Rev. 2017;

16. Wu L, Thompson DK, Li G, Hurt RA, Tiedje JM, Zhou J. Development and Evaluation of Functional Gene Arrays for Detection of Selected Genes in the Environment. Appl Environ Microbiol. 2001 Dec 1;67(12):5780–90.

17. Aronesty E. Comparison of sequencing utility programs. Open Bioinforma J [Internet]. 2013 [cited 2016 Oct 27];7(1). Available from: http://benthamopen.com/ABSTRACT/TOBIOIJ-7-1

18. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335–336.

19. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460–2461.

20. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72(7):5069–5072.

21. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.

22. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2015;gkv1070.

23. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. [Internet]. 2015. Available from: https://www.R-Project.org/

24. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. vegan: Community Ecology Package. R package version 2.4-1. 2016;

25. Jackson DA. PROTEST: a PROcrustean randomization TEST of community environment concordance. Ecoscience. 1995;2(3):297–303.

26. Wickham H. ggplot: An Implementation of the Grammar of Graphics. R Package Version 210 [Internet]. 2006 [cited 2016 Oct 27]; Available from: http://ftp.auckland.ac.nz/software/CRAN/src/contrib/Descriptions/ggplot.html

27. Rhee S-K, Liu X, Wu L, Chong SC, Wan X, Zhou J. Detection of Genes Involved in Biodegradation and Biotransformation in Microbial Communities by Using 50-Mer Oligonucleotide Microarrays. Appl Environ Microbiol. 2004 Jul 1;70(7):4303–17.

CHAPTER SIX

Conclusions

Summary of Findings and Final Thoughts

Chapter one and chapter two established some of the necessary groundwork for understanding stream system microbial communities, relationships, and dynamics in the open water column. The studies included in the two chapters also made several independent discoveries that were novel and interesting in their own right. Chapter one's findings of a decoupling between structure and function in communities helps support a current debate in the field on the connections between ecosystem inhabitants and what they are doing, or capable of doing functionally. This system and community view showing a disconnect between the taxonomic structure and functional capability of different microbial communities illustrates the importance of simultaneously exploring both structural and functional properties in any study. Study of independent taxa and functional relationships may miss these important relational nuisances. Chapter one also identifies a lack of functional redundancy in bacteria and archaea within the system. Functional diversity and redundancy is recognized as a key factor to maintain important functions and services of ecosystems (Laureto et al. 2015). The Ozark stream system being studied is not only a system sensitive to P inputs but is likely sensitive to many other potential disturbances as well.

The findings described in chapter two were surprising for a number of reasons. Not only were more fungi detectable than was expected, those fungi do not appear to

have been biologically active in terms of in stream nutrient cycling. That is not to say that they are without ecological importance. The fungi detected were largely from terrestrial sources and largely consisted of pathogens. Streams offer a potential avenue of dispersal for distribution to crops and livestock in addition to native natural ecosystem inhabitants. This means that while fungi are likely not heavily involved in movement of P within the open water column itself, they likely play a large role in the broader stream ecosystem including catchments. The broad ecological importance of fungi is not surprising considering their known important to numerous ecosystems (Harley 1971; Ingham et al. 1985; Heijden et al. 1998; Van Der Heijden and Horton 2009). In consideration of this, their apparent lack of involvement in more specific, suspended water column processes is somewhat more surprising.

Chapter three begins to look through a mechanistic lens at how microbial communities are equipped to handle P in streams systems for the P biogeochemical cycle specifically. Combining gene abundance data along with system network analyses, both empirical and hypothetical, using the SROGG begins to show us a meaningful picture of how the presence of theses genes relates to environmental P. We are also able to view how P cycle genes tend to sort relative to each other within communities. This comparison of the gene composition of entire microbial communities provides information on how the communities are related to a natural ecosystem and show relationships that would be difficult or impossible to identify utilizing other means due to their inherent complexity. Perhaps one of the most interesting findings in chapter three is the disconnect between the presence of organic P genes and the levels of organic and inorganic P present. Several sites with strong wastewater inputs showed a trend of high

organophosphate gene presence despite high TP and PO_4. This phenomena may very well be demonstrating the strong magnitude to which anthropogenic inputs impact natural microbial communities which would then influence greater ecosystems as a whole (Azam et al. 1983; Torsvik and Øvreås 2002; Schimel et al. 2007; Van Der Heijden et al. 2008).

Chapter four takes a very direct approach to suggesting alternative tools to whole community metagenomic sequencing for the type of work outlined in Chapter three. Metagenome sequencing can be expensive and generates massive amounts of data needing processing, storage, and analysis. Of the two tools explored in Chapter four, computationally predicted functional metagenomes and microarray analysis, the former appears to be the better starting point. PICRUSt created PFMs displayed better depth, diversity, and versatility for P cycle gene data than GeoChip 4.0. Both tools have viable applications to answering various scientific questions but for profiling and investigating nutrient specific genes within a system, PFM generation appears superior.

As a collection, the previous chapters make a major contribution to addressing the previously identified gaps of understanding microbial community structure and function relationships to P inputs in stream systems for bacteria, archaea, and fungi as well as understanding the complex interactions and mechanisms involved in P biogeochemical cycling at the community and system levels. However, they are just beginning to scratch the surface of potential discovery in the field. Perhaps the most meaningful contribution of the enclosed chapters is in demonstrating exactly how large these knowledge gaps are. The included chapter studies establish a solid foundation of knowledge and framework of tools for similar investigation in numerous other systems and for other nutrient cycles and functions.

Chapter References

Azam F, Fenchel T, Field JG, Gray JS, Meyer-Reil LA, Thingstad F. 1983. The Ecological Role of Water-Column Microbes in the Sea. Mar. Ecol. Prog. Ser. 10:257–263.

Harley JL. 1971. Fungi in Ecosystems. J. Appl. Ecol. 8:627–642. doi:10.2307/2402673.

Heijden MGA van der, Klironomos JN, Ursic M, Moutoglis P, Streitwolf-Engel R, Boller T, Wiemken A, Sanders IR. 1998. Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. Nature 396:69. doi:10.1038/23932.

Ingham RE, Trofymow JA, Ingham ER, Coleman DC. 1985. Interactions of Bacteria, Fungi, and their Nematode Grazers: Effects on Nutrient Cycling and Plant Growth. Ecol. Monogr. 55:119–140. doi:10.2307/1942528.

Laureto LMO, Cianciaruso MV, Samia DSM. 2015. Functional diversity: an overview of its history and applicability. Nat. Conserv. 13:112–116. doi:10.1016/j.ncon.2015.11.001.

Schimel J, Balser TC, Wallenstein M. 2007. Microbial Stress-Response Physiology and Its Implications for Ecosystem Function. Ecology 88:1386–1394. doi:10.1890/06-0219.

Torsvik V, Øvreås L. 2002. Microbial diversity and function in soil: from genes to ecosystems. Curr. Opin. Microbiol. 5:240–245. doi:10.1016/S1369-5274(02)00324-7.

Van Der Heijden MG, Bardgett RD, Van Straalen NM. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecol. Lett. 11:296–310.

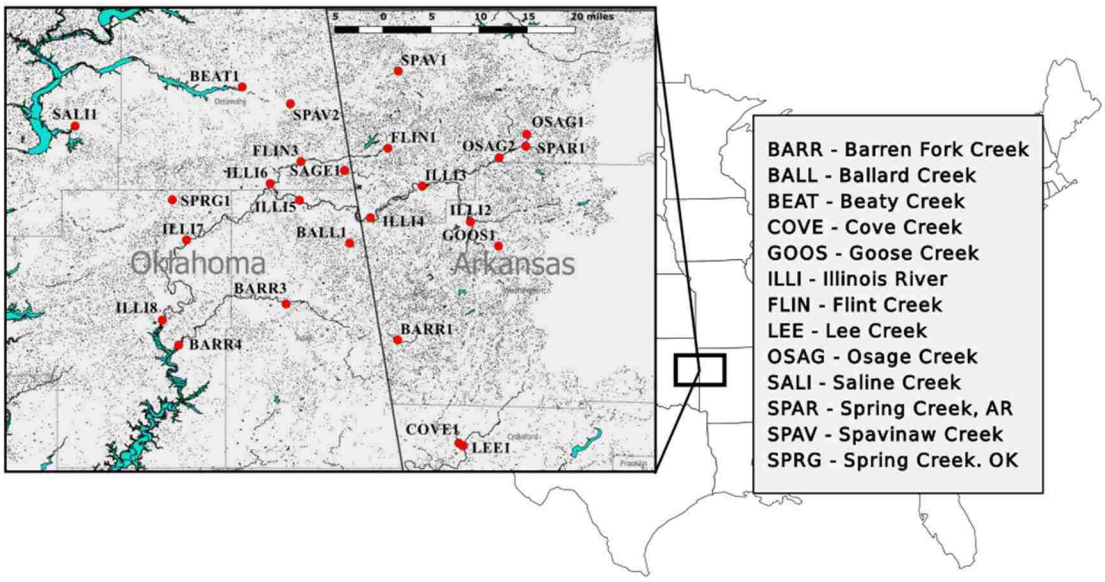
Van Der Heijden MGA, Horton TR. 2009. Socialism in soil? The importance of mycorrhizal fungal networks for facilitation in natural ecosystems. J. Ecol. 97:1139–1150. doi:10.1111/j.1365-2745.2009.01570.x.

APPENDICES

APPENDIX A

Supplementary Figures (Chapter two)

Figure A.S1: Map of sample sites in Oklahoma and Arkansas. Site mapping for was constructed in Quantum GIS version 2.14.7 [2] using the Watershed Boundary Dataset (WBD).[1]



1. Coordinated Effort between the United States Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS), the United States Geological Survey (USGS), and the Environmental Protection Agency (EPA). The Watershed Boundary Dataset (WBD) Was Created from a Variety of Sources from Each State and Aggregated into a Standard National Layer for Use in Strategic Planning and Accountability.

2. Quantum GIS Development Team (2009) Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project.

Figure A.S2: Alpha diversity plots. Only taxa referenced in main text are identified. A) Rarefaction curve for rarefaction at 70,500 depth. B) Taxonomic diversity bar chart with sites grouped by TP (Genus level). C) Taxonomic diversity bar chart with sites grouped by Turbidity (Genus level).

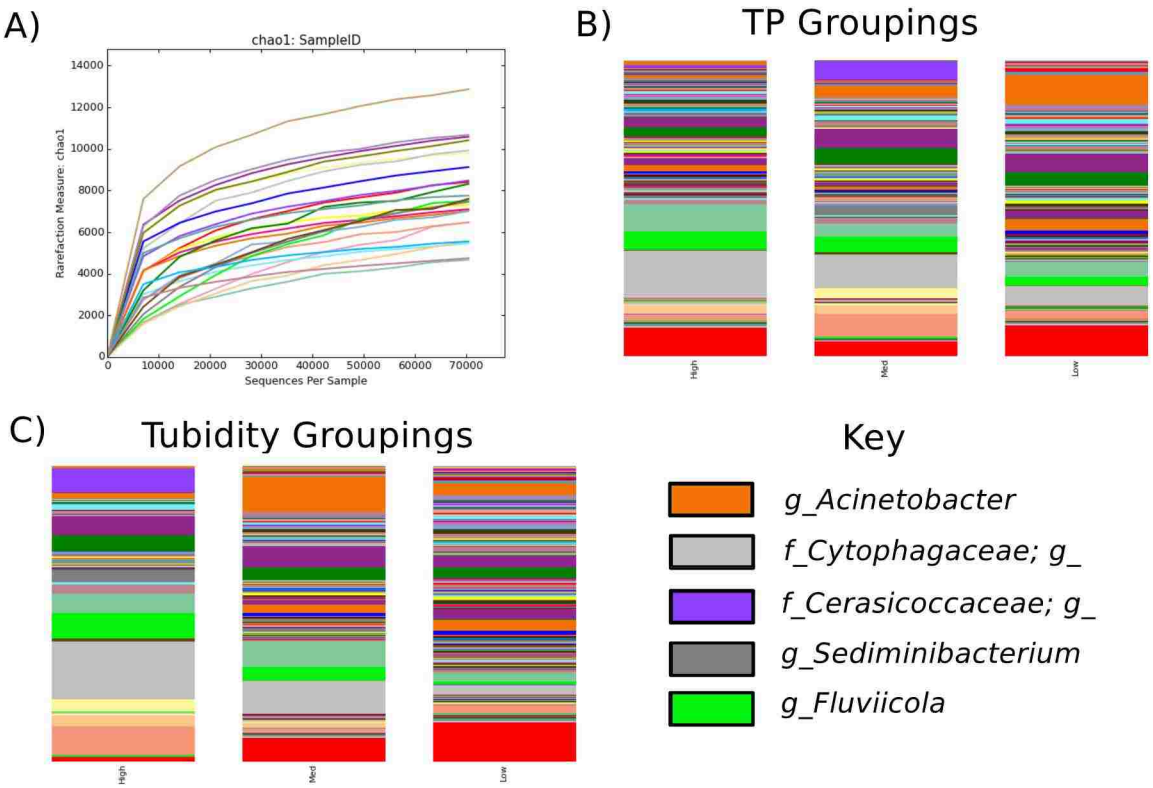


Figure A.S3: Procrustes error plot for RDA models for 16S rRNA and PFMs.

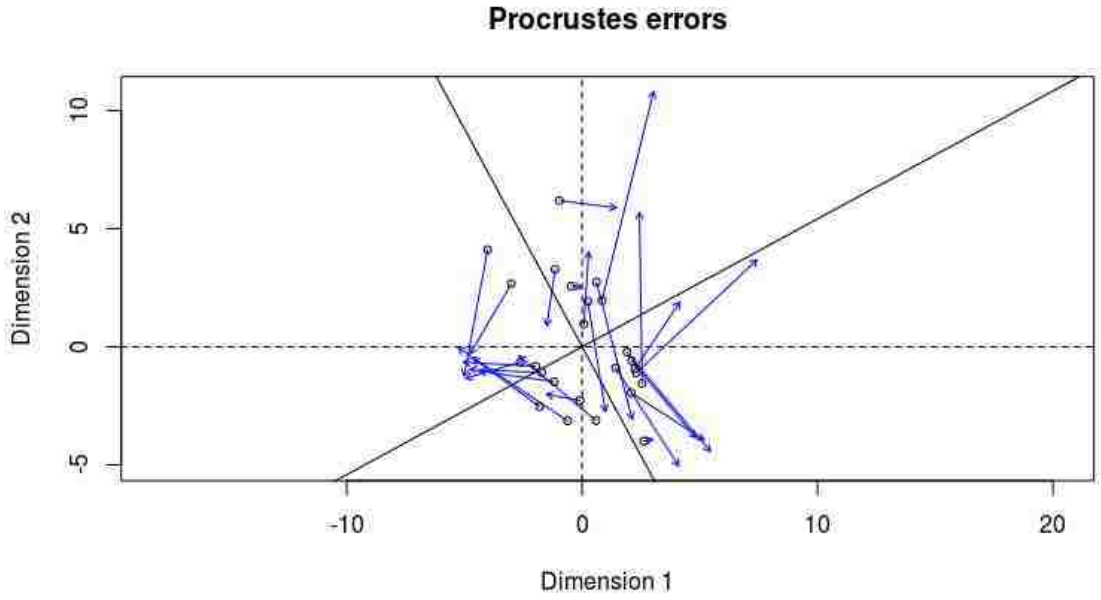


Figure A.S4: A) GAM model of OTU diversity (Shannon) against turbidity ($R^2_{adj.}$= 0.322, p = 0.008). B) GAM model of multifunctional diversity (DUNNO) against turbidity ($R^2_{adj.}$= 0.593, p < 0.001)

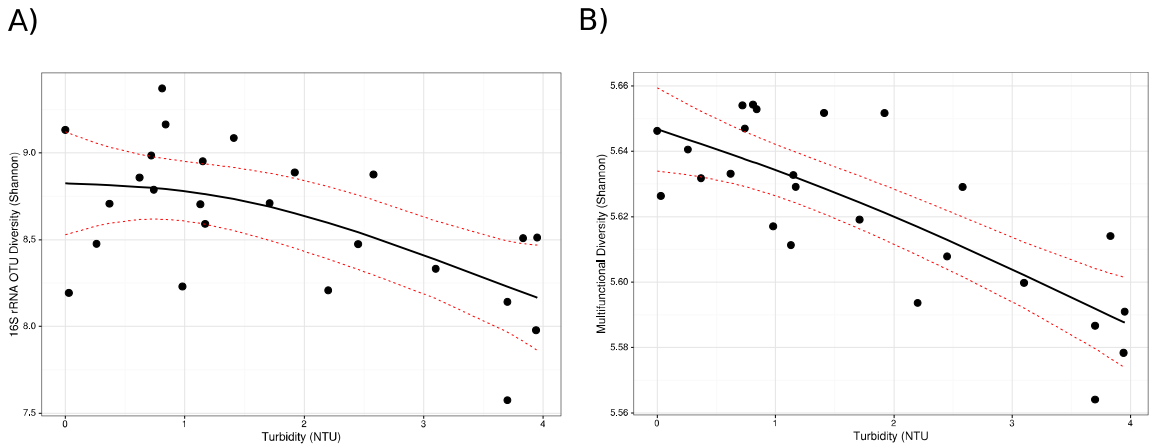
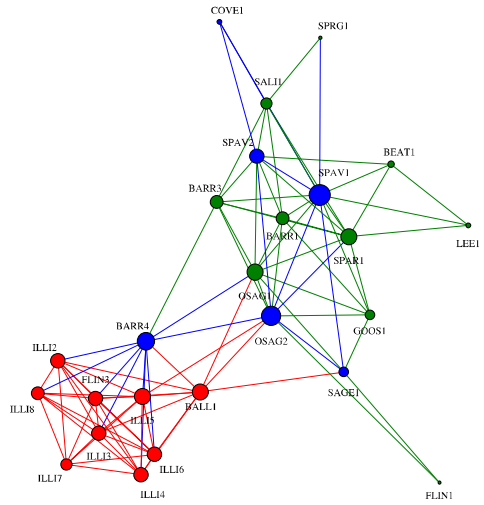


Figure A.S5: Network plots. A) Network of site relatedness from 16s rRNA data colored by turbidity group. Node size represents degree or number of connections for the node. B) Network of site relatedness from PFM data colored by turbidity group. Node size represents degree or number of connections for the node.

A)



B)

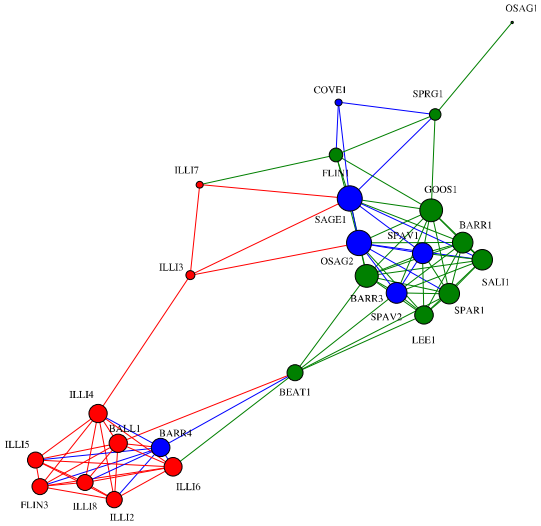


Figure A.S6: Turbidity Analysis. A) Plot of pure and reliable indicator taxa along the turbidity gradient. Black symbols correspond to genera that declined with increasing TP (z-), whereas open symbols correspond to those that increased (z+). Symbols are sized in proportion to the magnitude of the response (z-score). Horizontal lines represent 5th and 95th quantiles of values of turbidity resulting in the largest change in genera z-scores among 1000 bootstrap replicates. B) Plot of sumZ scores for genus level taxa. Steep slopes indicate major change points in abundance. C) Plot of pure and reliable OTUs. D) Plot of sumZ scores for OTU level taxa.

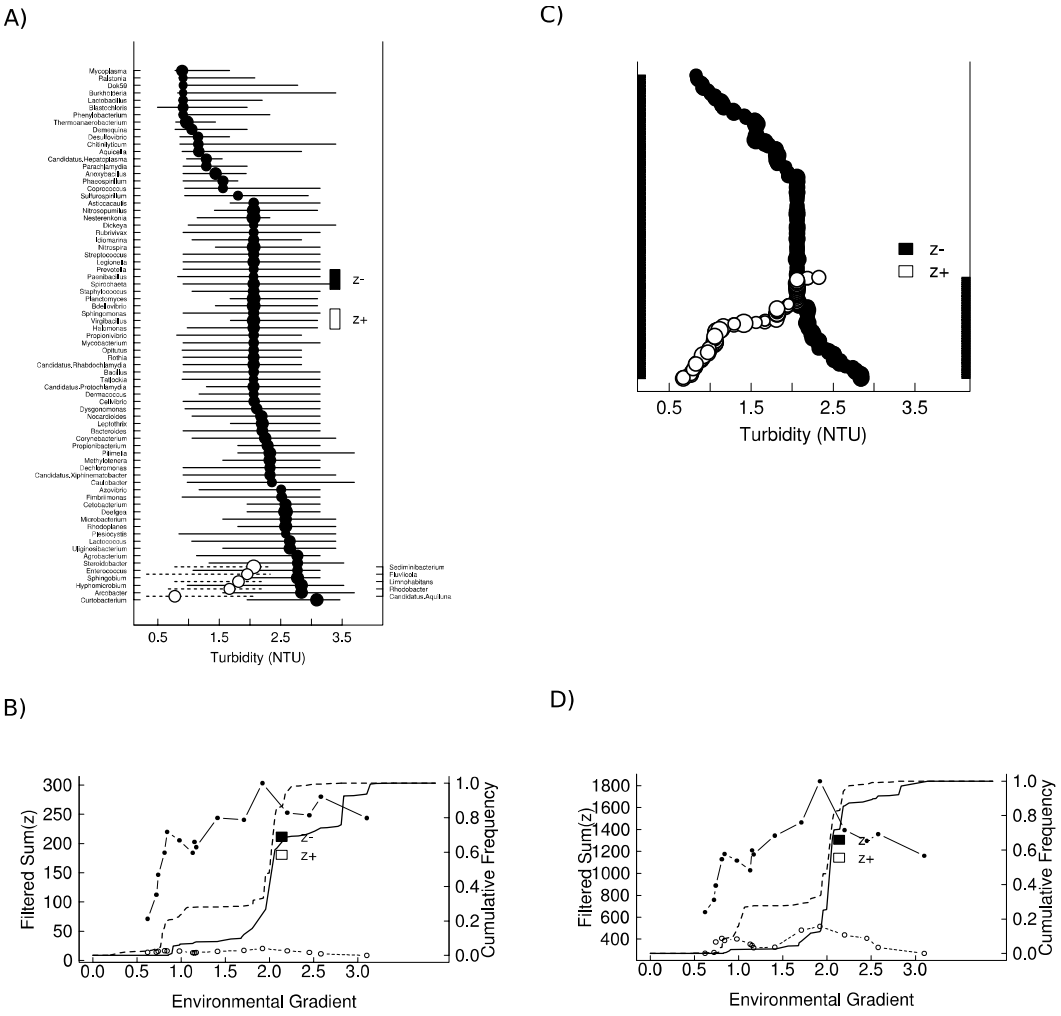


Figure A.S6: Plot of Shannon diversity index for site 16s rRNA OTU counts against Shannon diversity index for site predicted functional metagenome ($R^2 = 0.748$, $p < 0.001$).

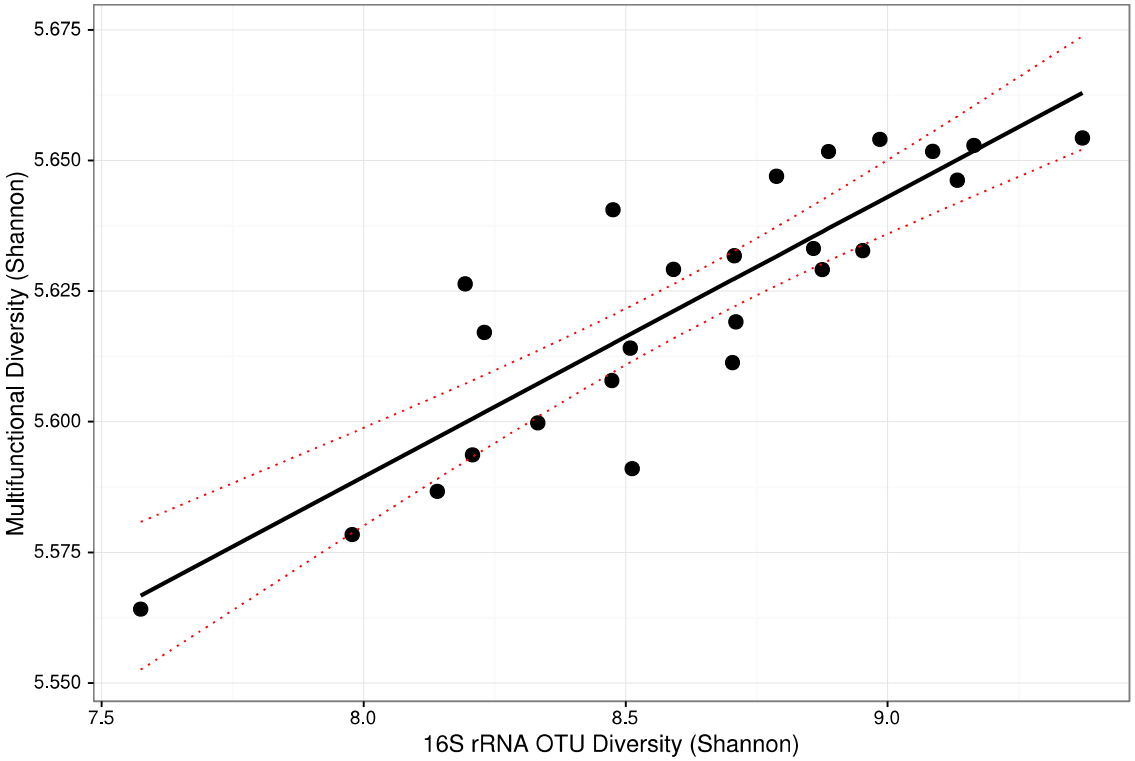


Figure A.S8: Trends for P values of statistical tests on other levels of PICRUSt data. A) ANOSIM results. B) PERMANOVA results. C) MRPP results.

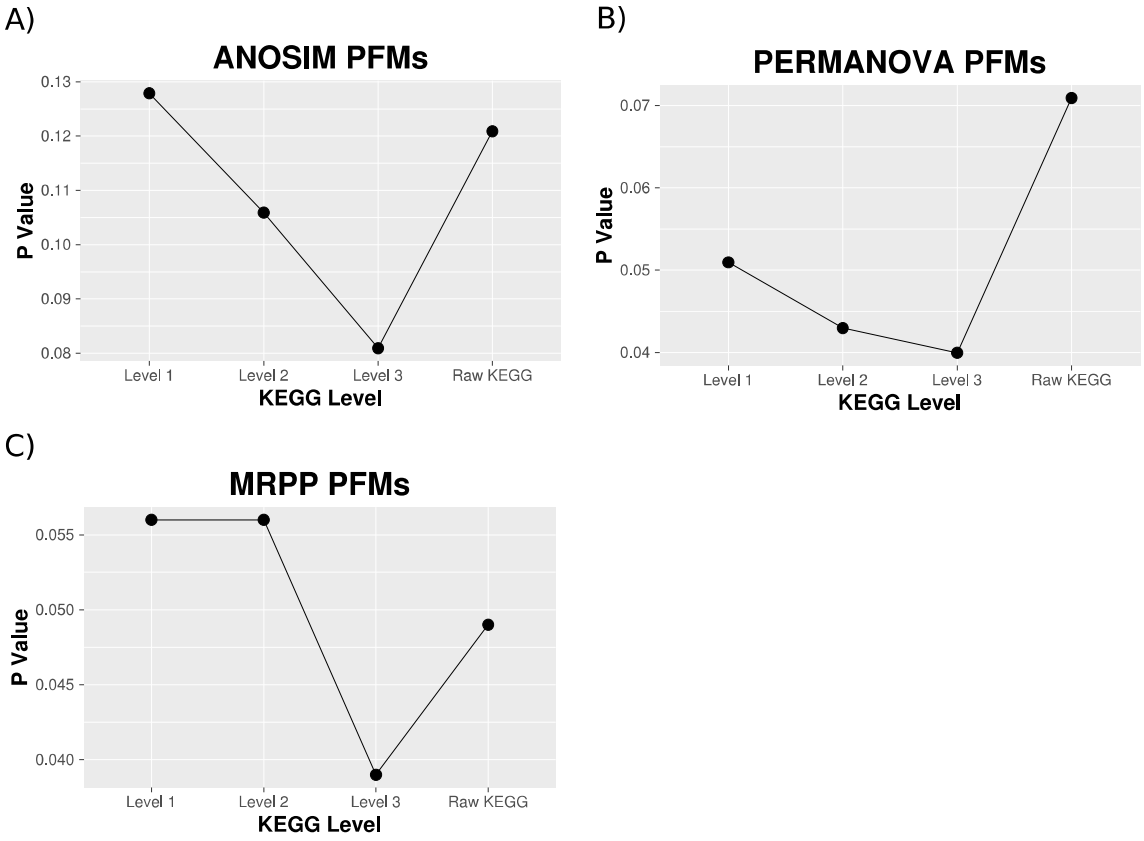
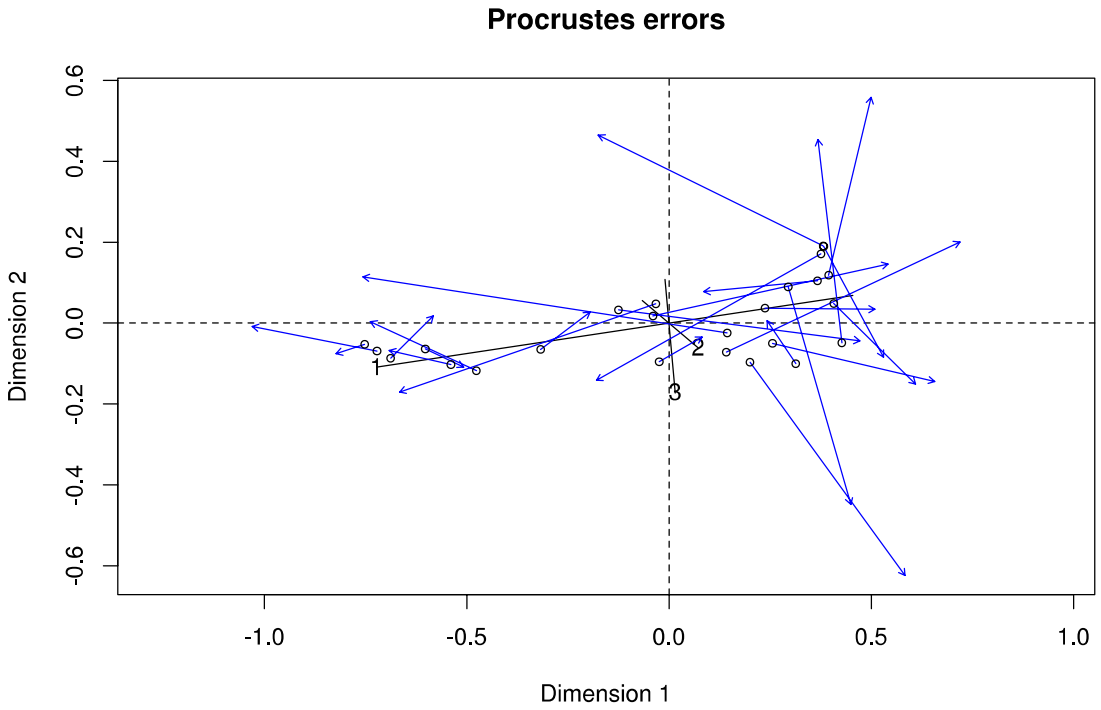


Figure A.S9: Procrustes error plot for 16S rRNA Bray-Curtis distance NMDS and Environmental Variable Bray Curtis NMDS ordinations.



APPENDIX B

Supplementary Tables (Chapter two)

Table B.S1: Listing of sampling sites along with designated Total Phosphorus (TP) and Turbidity grouping assignments.

Sample Id	TP Group	Turbidity Group
BALL1	Med	High
BARR1	Med	Low
BARR3	Low	Low
BARR4	Low	Med
BEAT1	Med	Low
COVE1	Low	Med
FLIN1	Med	Low
FLIN3	High	High
GOOS1	High	Low
ILLI2	Med	High
ILLI3	High	High
ILLI4	Med	High
ILLI5	Med	High
ILLI6	Med	High
ILLI7	Med	High
ILLI8	Med	High
LEE1	Low	Low
OSAG1	High	Low
OSAG2	High	Med
SAGE1	High	Med
SALI1	Low	Low
SPAR1	High	Low
SPAV1	Low	Med
SPAV2	Med	Med
SPRG1	Low	Low

APPENDIX C

Supplementary Results (Chapter two)

Importance of Functional Data Resolution

Testing and significance, particularly on the PFM data depended heavily on tier resolution of the data and availability and inclusion of different environmental metrics. By resolution of the data, we are referring to how aggregated the data is based on taxonomic level or function. This was largely not a problem for the OTU data as pure OTUs already offer good resolution for that data set and there were no discrepancies in the analyses where we did aggregate data by different taxonomic levels such as TITAN 2.1. Conceptually though, the OTU level would offer a better data resolution than higher taxonomic designations such as Phylum. OTUs were generated at 97 percent similarity but clustering at 99 percent may still offer even better data resolution and result in trends that were not seen at the 97 percent clustering level.

For our PFMs, we intentionally selected the KEGG level 3 data *a priori* from PICRUSt as it should offer the best resolution or aggregation to the lowest functionally meaningful designation available but in order to satisfy curiosity, we decided to explore the effects of using the other available data tiers as well. We found that had we used KEGG level 1 or KEGG level 2 groupings we would have seen diluted significance in testing (Figure S8). This helps confirm that our *a priori* selection of KEGG level 3 data was the best option available. The resolution issue may be why we had a close to, but nonsignificant result with regards to TP grouping in the ANOSIM analysis. Direct functional metagenome data may have resulted in even more definitive findings.

The inexpensive cost and ease of using PICRUSt make it an invaluable tool and it has been used effectively in many systems particularly in medical sciences [1–3]. The validity of PICRUSt data has been explored before and it has been shown to be

dependent on quality of the data and accessibility of available reference genomes [4]. In the case of our study, we were able to reference a large number of OTUs but many not to lower taxonomic levels like genus. Hartman et al. (2015) recently presented side by side measured and PICRUSt predicted metagenomes showing very similar data with a few notable exceptions such as genes involved in nitrate reduction. Credibility of PICRUSt PFMs is supported by this and similar results between this study and other recent studies [5, 6] though there is also evidence that linking taxa to function from 16S rRNA data should be approached cautiously [7].

Illinois River Effect

After seeing the Illinois River sites behaving similarly through several exploratory data analyses and due to the relatively large size of the Illinois River sites and their catchment areas, we were initially concerned about the potential for the Illinois River to impose a disproportionate influence on our results. In order to determine whether the Illinois River sites had an excessive influence on the results, NMDS ordinations were performed using Bray-Curtis distance on the full matrix of environmental data. Procrustes analysis including a PROTEST from the vegan package was then used to compare the 16S rRNA and environmental ordinations for similar trends and relationship in ordination space of Illinois River sites that would be indicative of an Illinois River effect.

In procrustes analysis, a PROTEST correlation statistic of 0.68 ($p<0.001$) did indicate a potential relationship between the full 16S rRNA and environmental metric ordinations. However, the Procrustes error plotting indicated random directions and magnitudes of travel for Illinois River sites in ordination space with no discernible trends or patterns indicating likely little overwhelming Illinois River effect on other analyses

(Figure A.S9). Additionally, on the heatmap of relative abundance showing general trends in taxa relative abundance among sites (Figure 2), BALL1, BARR4, and FLIN3 all cluster with Illinois River sites in WPGMA clustering further supporting the lack of a unique Illinois River effect.

Appendix C References

1. Cox LM, Yamanishi S, Sohn J, et al (2014) Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. Cell 158:705–721.

2. David LA, Maurice CF, Carmody RN, et al (2014) Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–563. doi: 10.1038/nature12820

3. Forsberg KJ, Patel S, Gibson MK, et al (2014) Bacterial phylogeny structures soil resistomes across habitats. Nature 509:612.

4. Xu Z, Malmer D, Langille MG, et al (2014) Which is more important for classifying microbial communities: who's there or what they can do. ISME J 8:2357–2359.

5. Louca S, Parfrey LW, Doebeli M (2016) Decoupling function and taxonomy in the global ocean microbiome. Science 353:1272–1277.

6. Sunagawa S, Coelho LP, Chaffron S, et al (2015) Structure and function of the global ocean microbiome. Science 348:1261359. doi: 10.1126/science.1261359

7. Logue JB, Stedmon CA, Kellerman AM, et al (2016) Experimental insights into the importance of aquatic bacterial community composition to the degradation of dissolved organic matter. ISME J 10:533–545.

APPENDIX D

Supplementary Material (Chapter three)

Supplementary Table

Table D.S1: Sample sites with total phosphorus grouping and diversity index values.

Sample ID	TP Grouping	OTUs	Shannon Diversity	Simpson Diversity	Inverse Simpson Diversity
BALL1	Med	102.00	3.04	0.94	15.69
BARR3	Low	56.00	1.84	0.79	4.83
BARR4	Low	113.00	3.00	0.93	13.40
BEAT1	Med	75.00	2.07	0.80	5.11
COVE1	Low	76.00	2.22	0.84	6.45
FLIN1	Med	56.00	1.59	0.74	3.90
FLIN3	High	140.00	3.00	0.89	9.37
GOOS1	High	63.00	2.17	0.84	6.08
ILLI2	Med	168.00	3.36	0.94	17.24
ILLI3	High	100.00	2.95	0.91	10.63
ILLI4	Med	143.00	2.67	0.87	7.58
ILLI5	Med	63.00	2.62	0.90	10.35
ILLI6	Med	122.00	1.84	0.66	2.98
ILLI7	Med	119.00	3.09	0.91	11.50
ILLI8	Med	61.00	1.77	0.78	4.56
LEE1	Low	64.00	2.52	0.90	10.32
LLEE1	Low	101.00	2.90	0.92	11.87
MTFK1	Low	65.00	1.72	0.79	4.85
OSAG1	High	61.00	1.72	0.69	3.22
OSAG2	High	62.00	2.35	0.86	6.91
SAGE1	High	55.00	2.39	0.88	8.64
SPAR1	High	48.00	2.09	0.85	6.68
SPAV1	Low	50.00	2.34	0.89	9.15
SPAV2	Med	87.00	1.80	0.78	4.56
SPRG1	Low	67.00	1.79	0.74	3.87
Mean		84.68	2.35	0.84	7.99
Standard Deviation		33.25	0.53	0.08	3.90

Supplementary Figures

Figure D.S1: Map of sample sites. Created in QGIS.

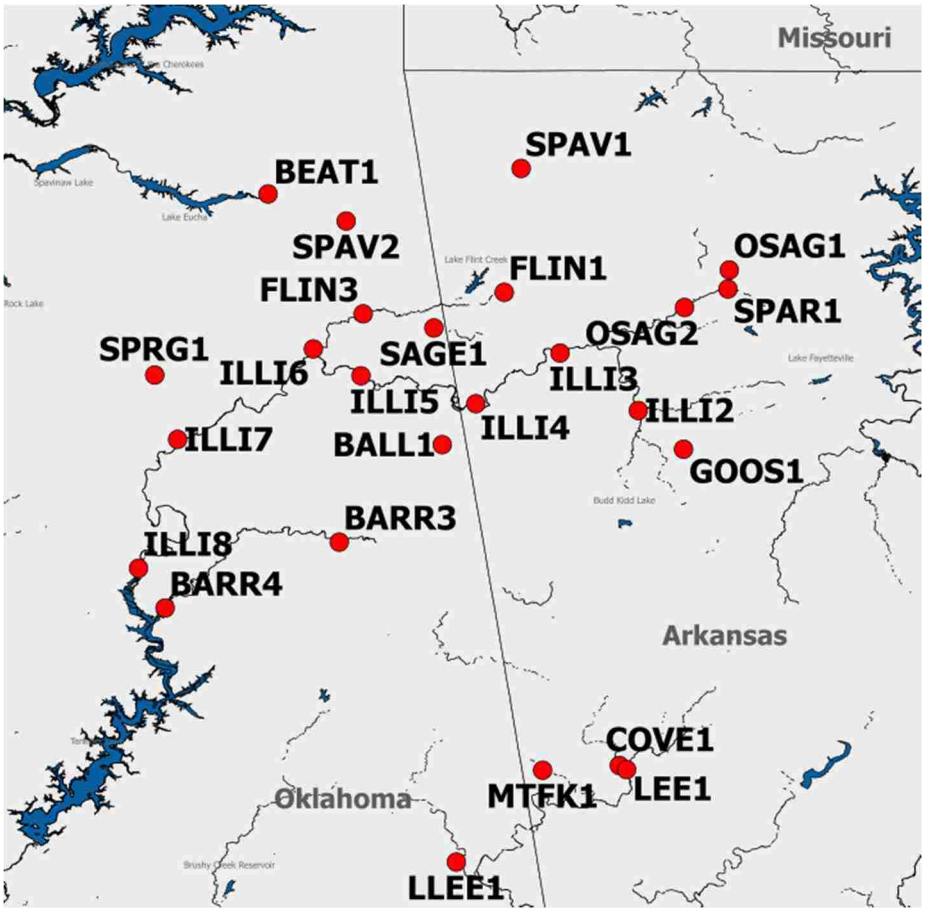


Figure D.S2: NMDS ordination of FUNGuild designations with Bray-Curtis distance (Stress = 0.104). Gradient represents environmental TP fit to ordination using GAM.

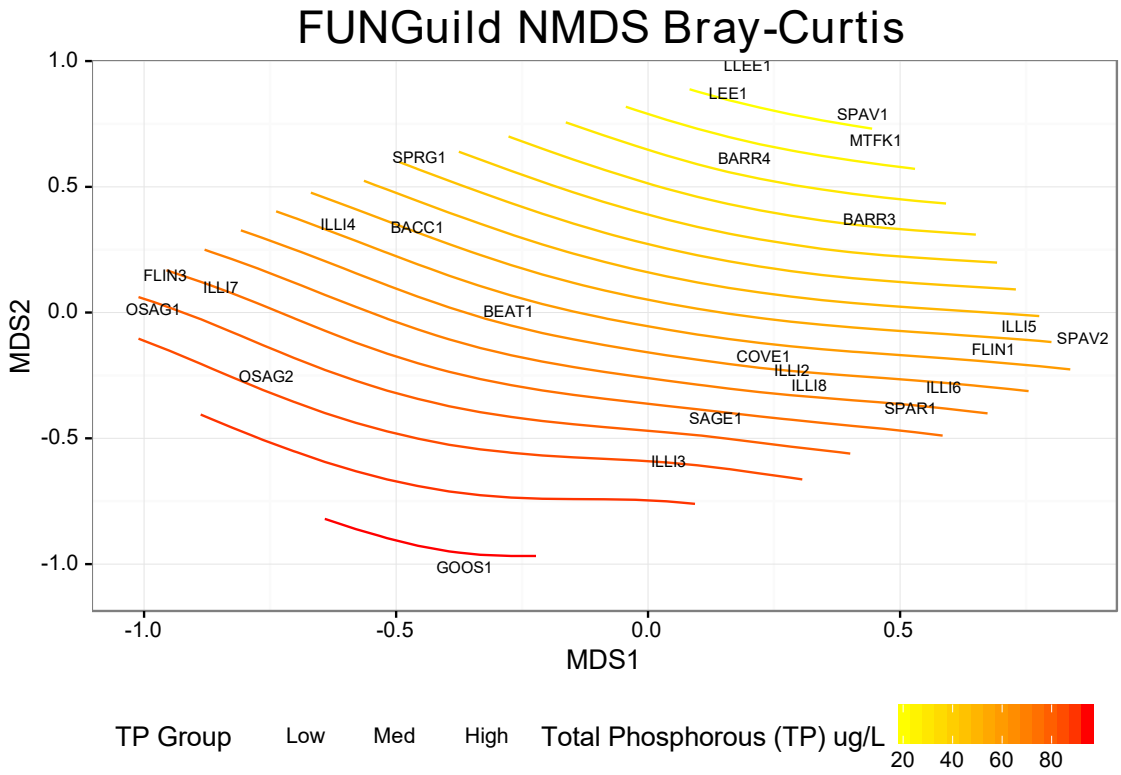


Figure D.S3: Procrustes error plot for comparison of fungal ITS2 and bacterial 16s ordinations for 23 overlapping sites.

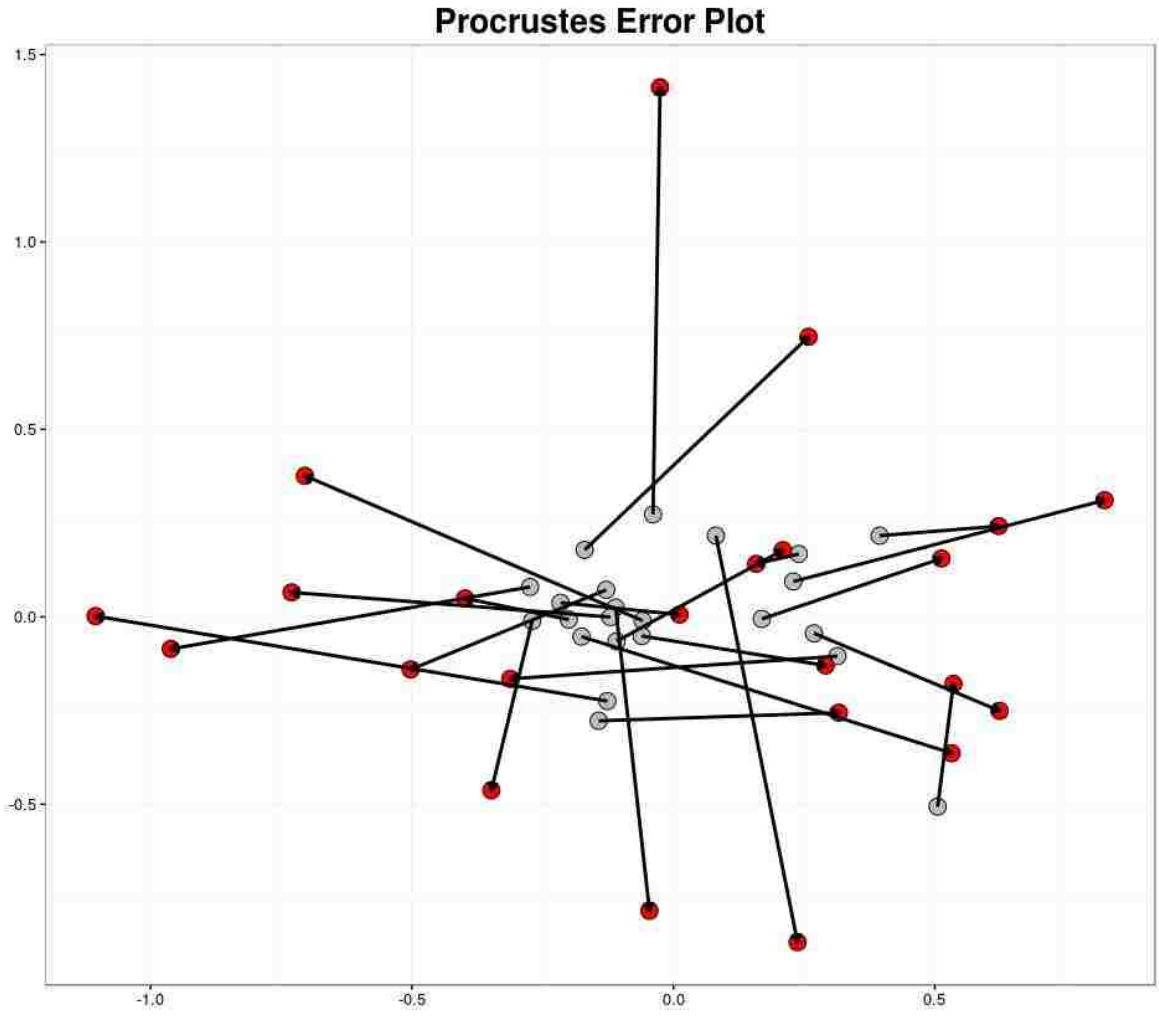


Figure D.S4: Heatmap generated at the Class level. Taxa observed <1% removed.

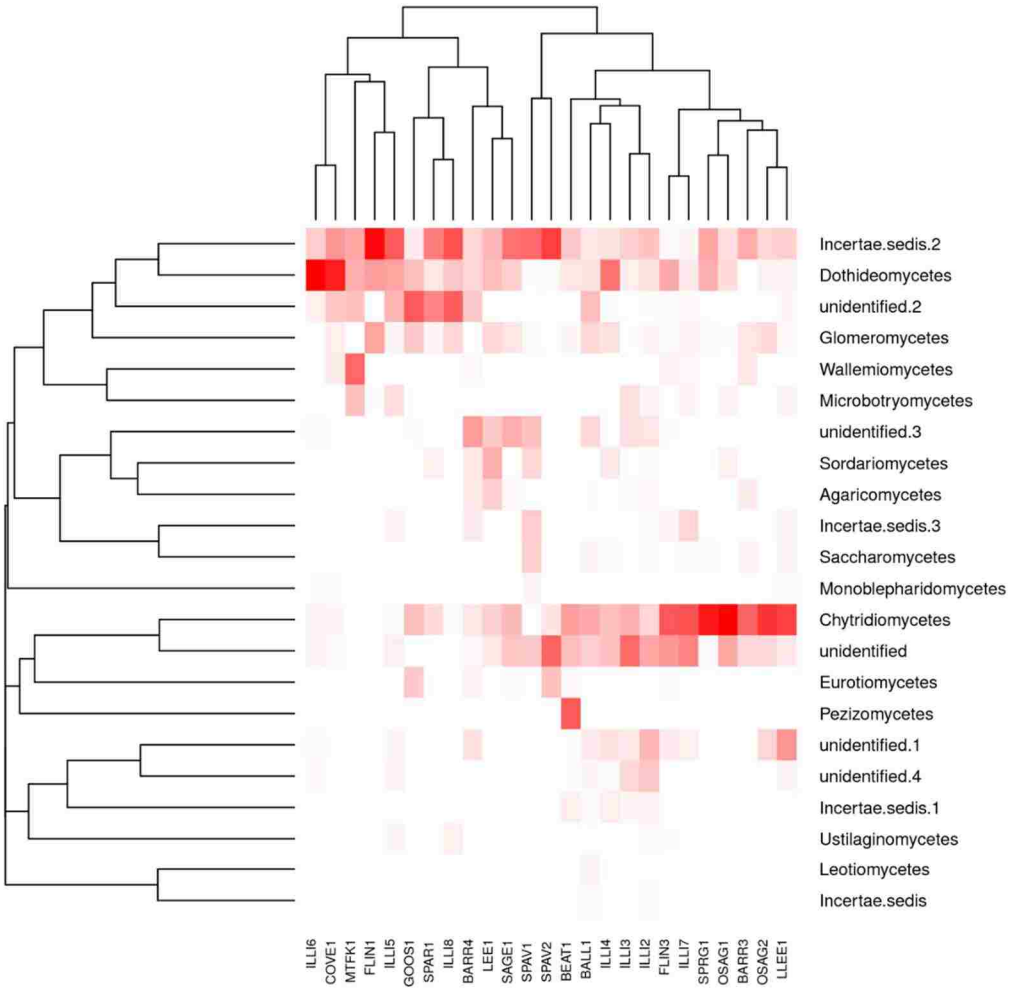


Figure D.S5: Heatmap generated at the Order level. Taxa observed <1% removed.

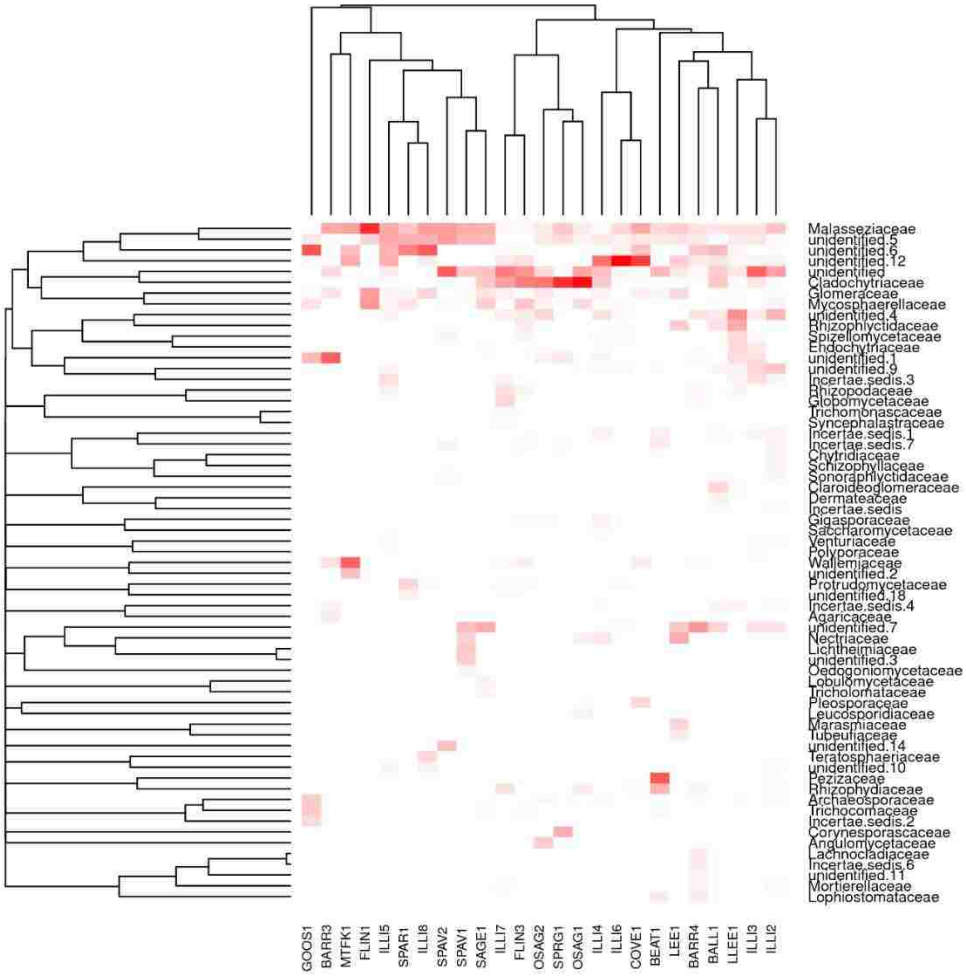


Figure D.S6: Heatmap generated at the Family level. Taxa observed <1% removed.

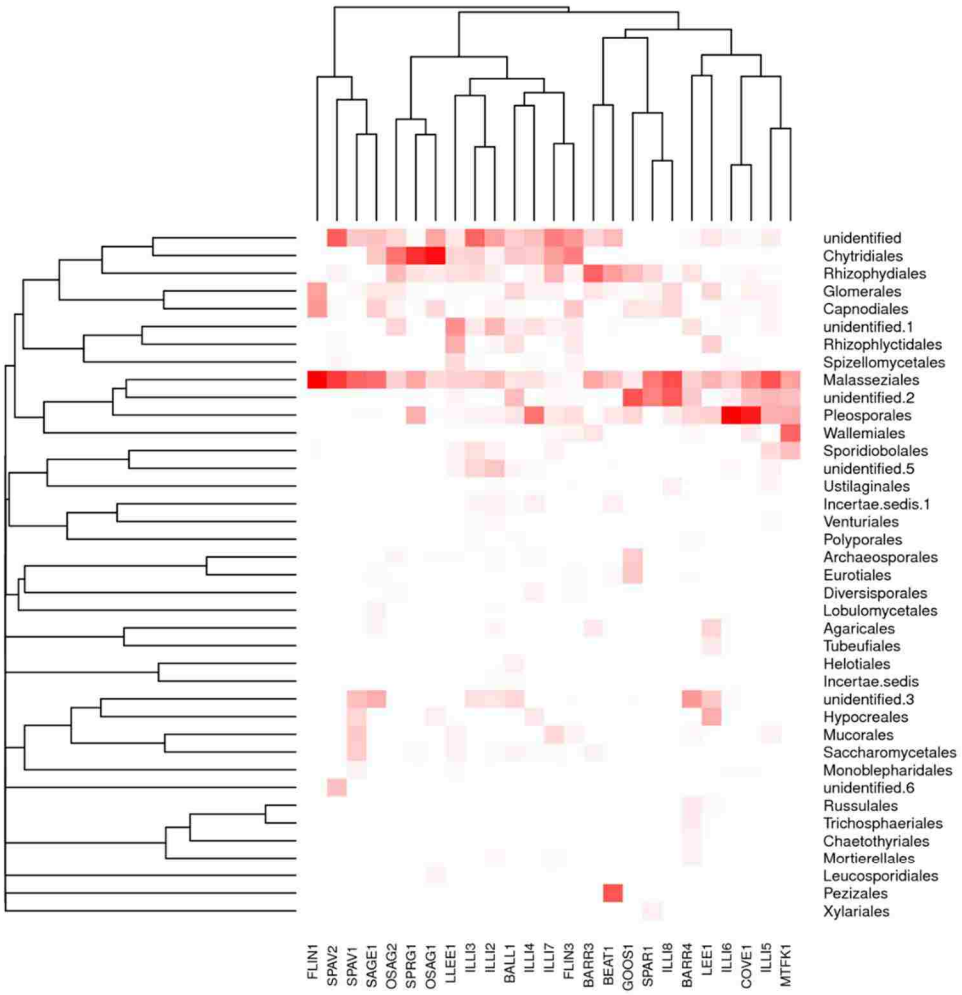


Figure D.S7: Heatmap generated at the Genus level. Taxa observed <1% removed.

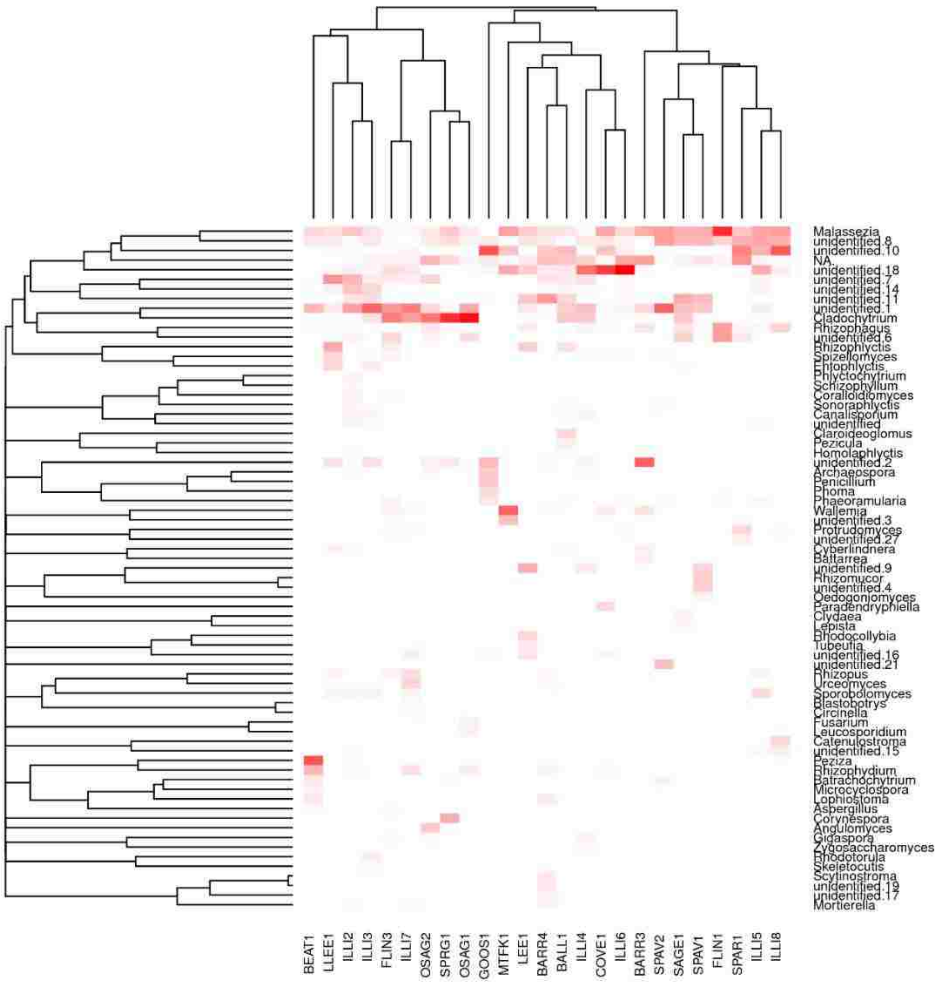
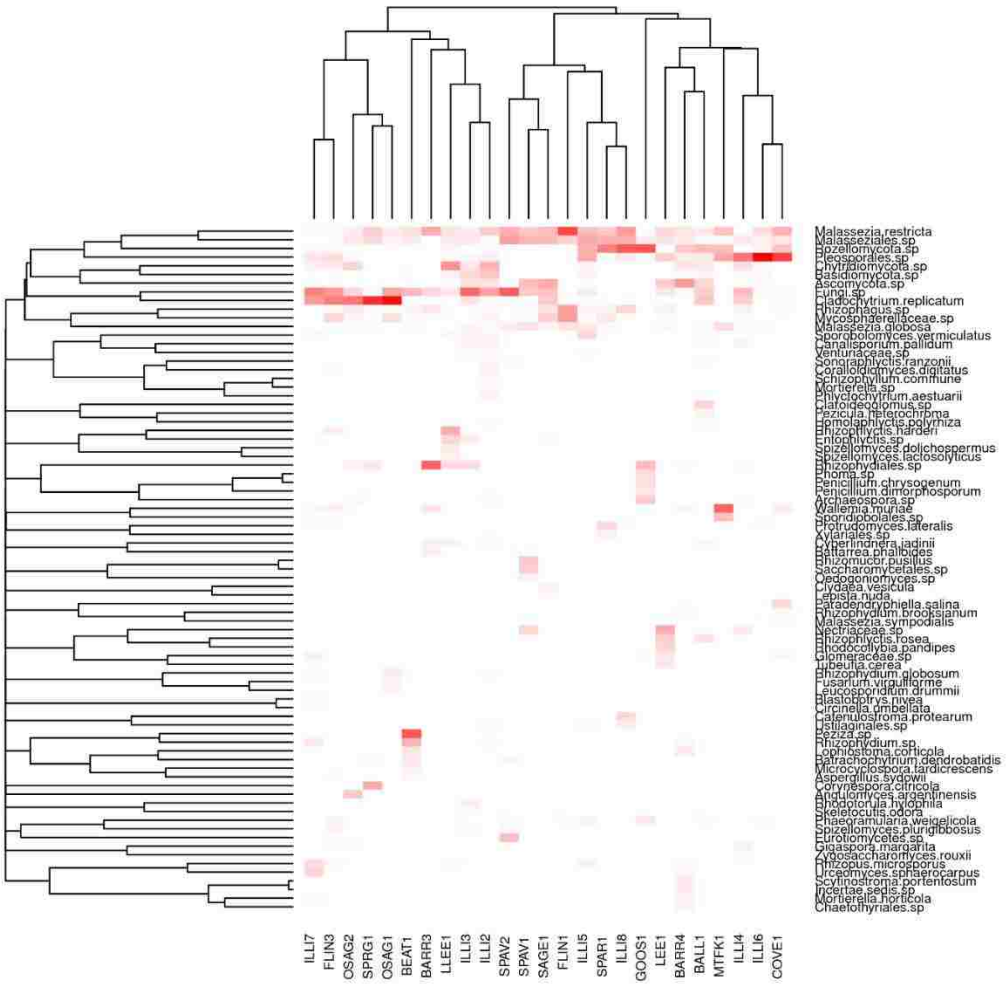


Figure D.S8: Heatmap generated at the OTU/Species level. Taxa observed <1% removed.



APPENDIX E

Supplementary Figures (Chapter four)

Figure E.S1: Map of sample sites. Built using QGIS.

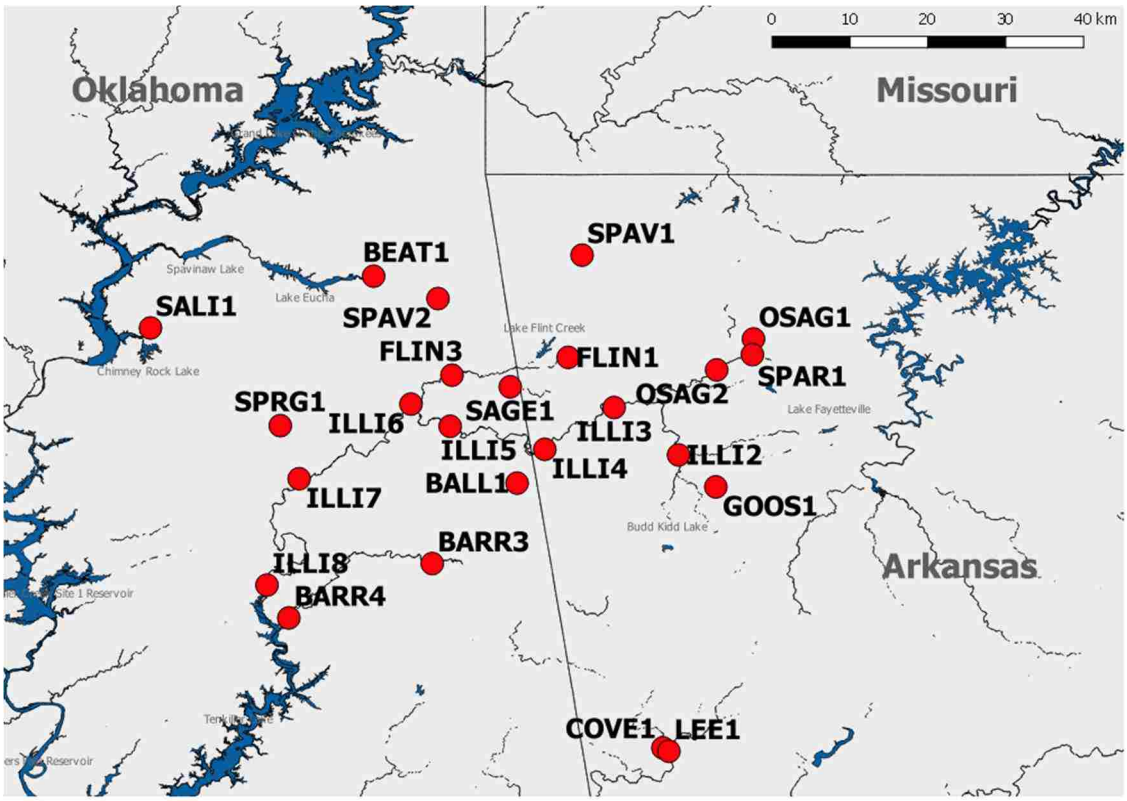


Figure E.S2: Network including Total Membrane Transport gene grouping in addition to the data based network from Figure 1. Edges are weighted by strength of relationship (1 – Bray-Curtis Dissimilarity).

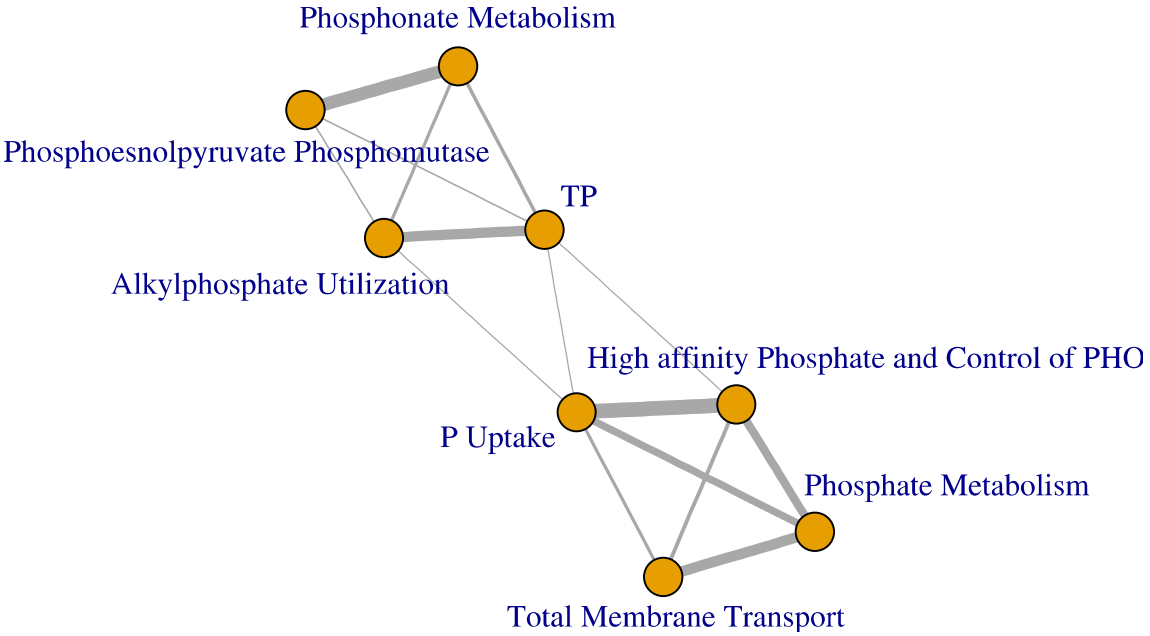


Figure E.S3: NMDS ordination of all P genes by site. The PO_4 gradient is fit using a GAM. See Table 1 for metrics.

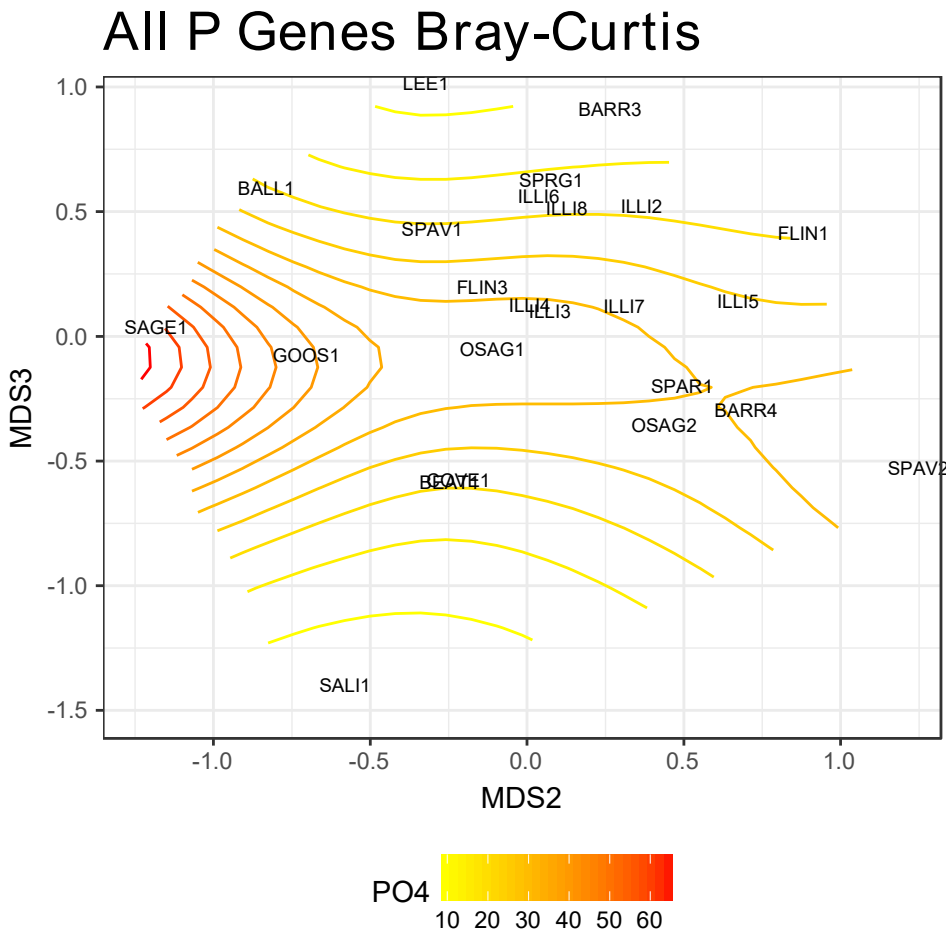


Figure E.S4: NMDS ordination of all P genes by site. The non-PO_4 gradient is fit using a GAM. See Table 1 for metrics.

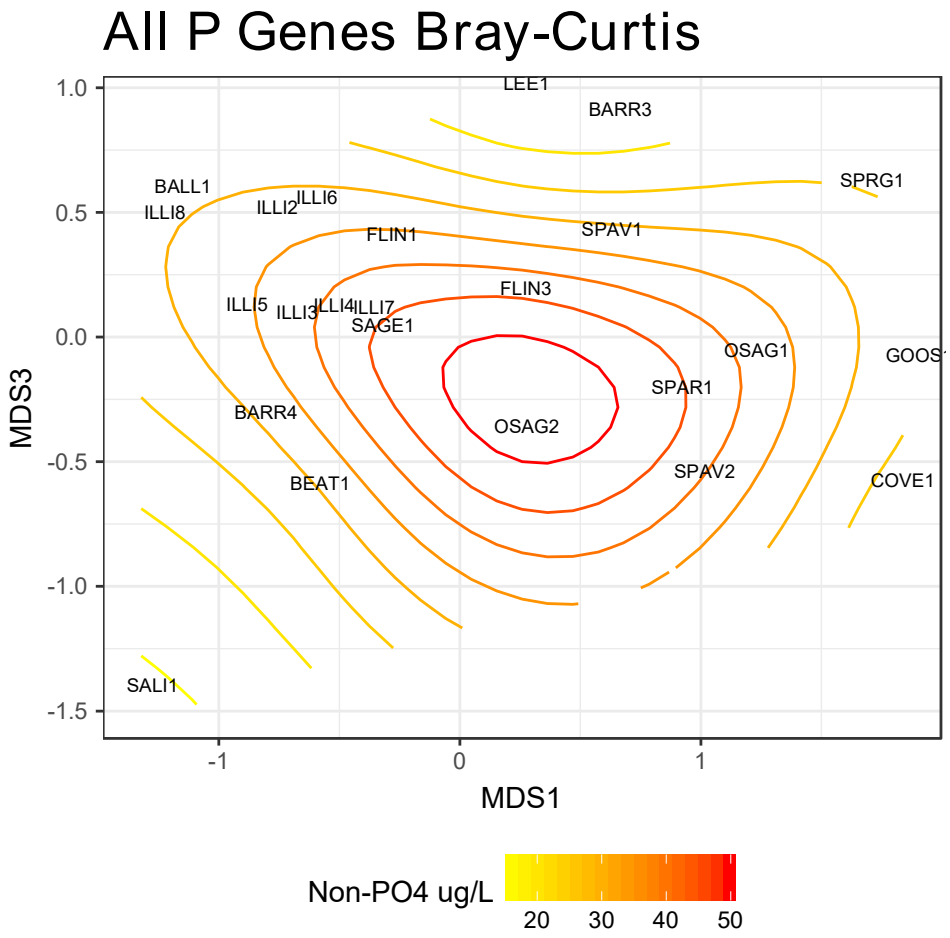


Figure E.S5: NMDS ordination of High Affinity Phosphate Transporter and Control of PHO Regulon genes by site. The TP gradient is fit using a GAM. See Table 1 for metrics.

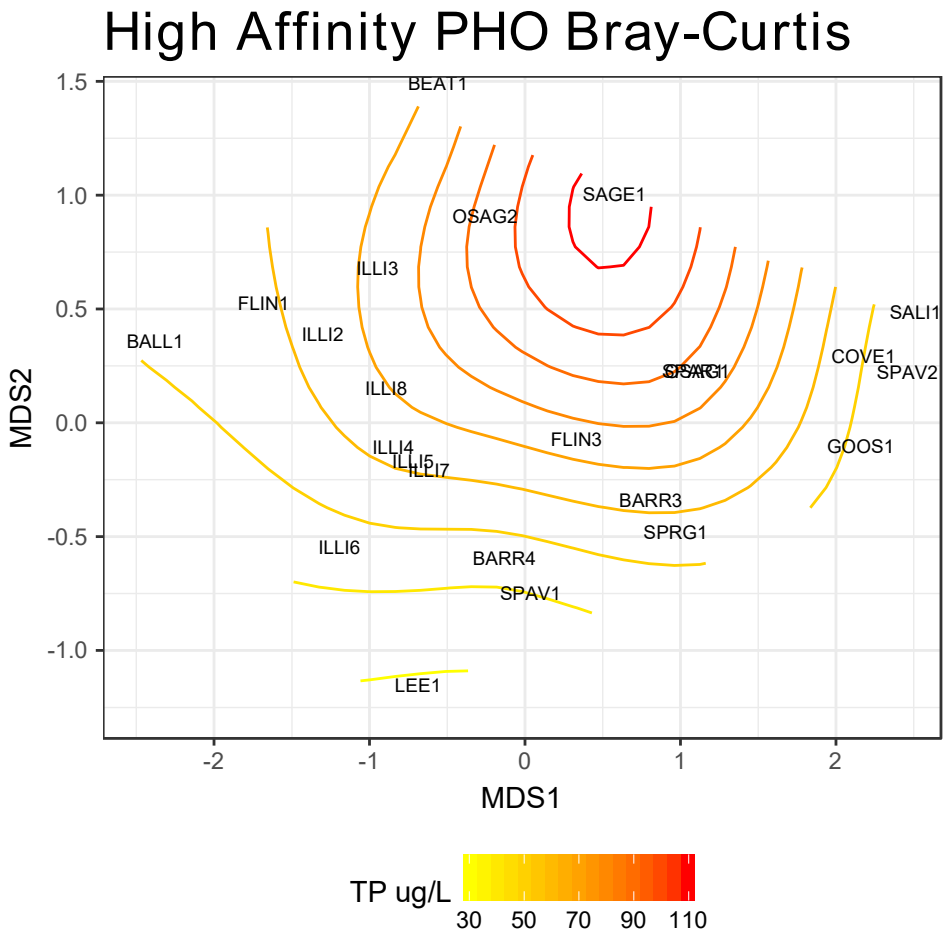


Figure E.S6: NMDS ordination of High Affinity Phosphate Transporter and Control of PHO Regulon genes by site. The PO_4 gradient is fit using a GAM. See Table 1 for metrics.

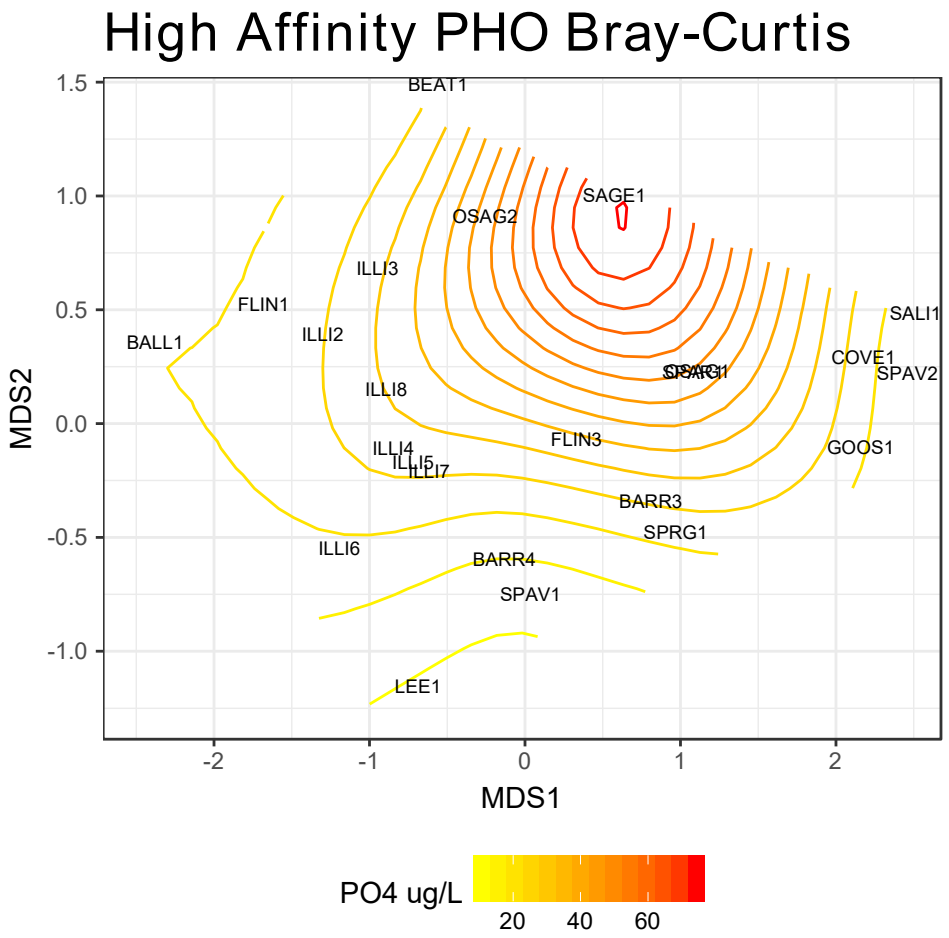


Figure E.S7: NMDS ordination of Phosphate Metabolism genes by site. The TP gradient is fit using a GAM. See Table 1 for metrics.

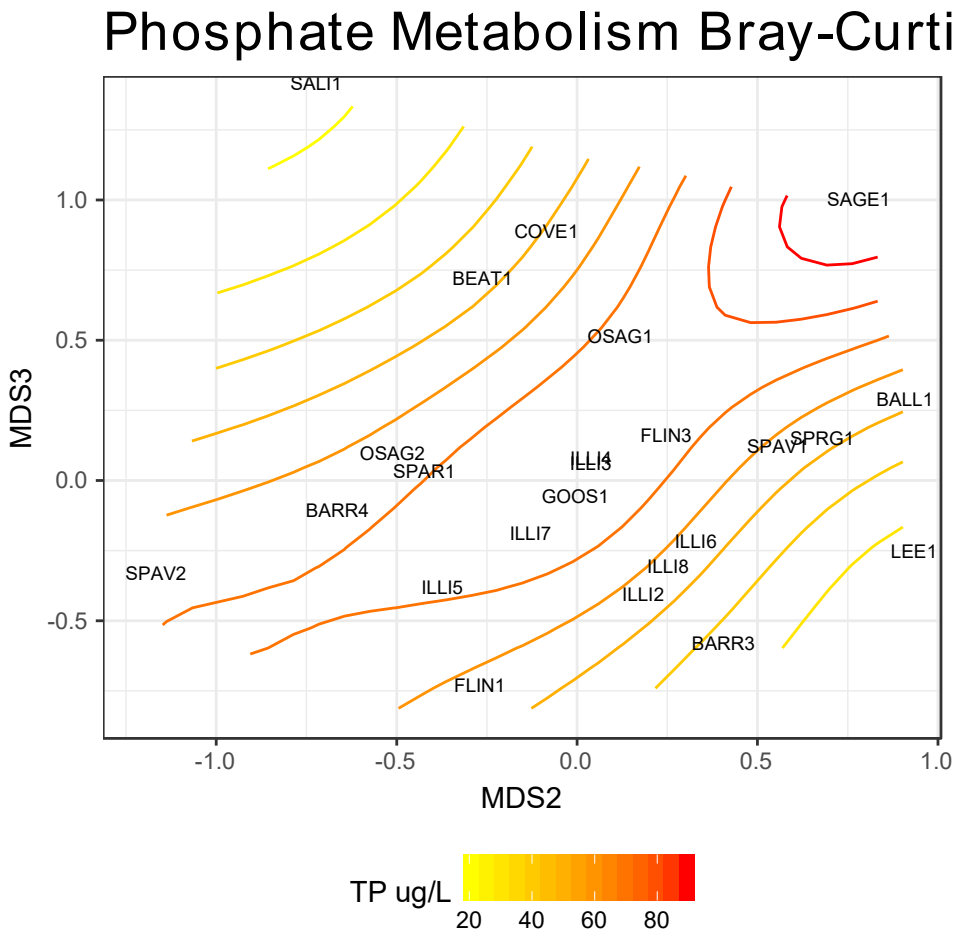


Figure E.S8 NMDS ordination of Phosphate Metabolism genes by site. The PO_4 gradient is fit using a GAM. See Table 1 for metrics.

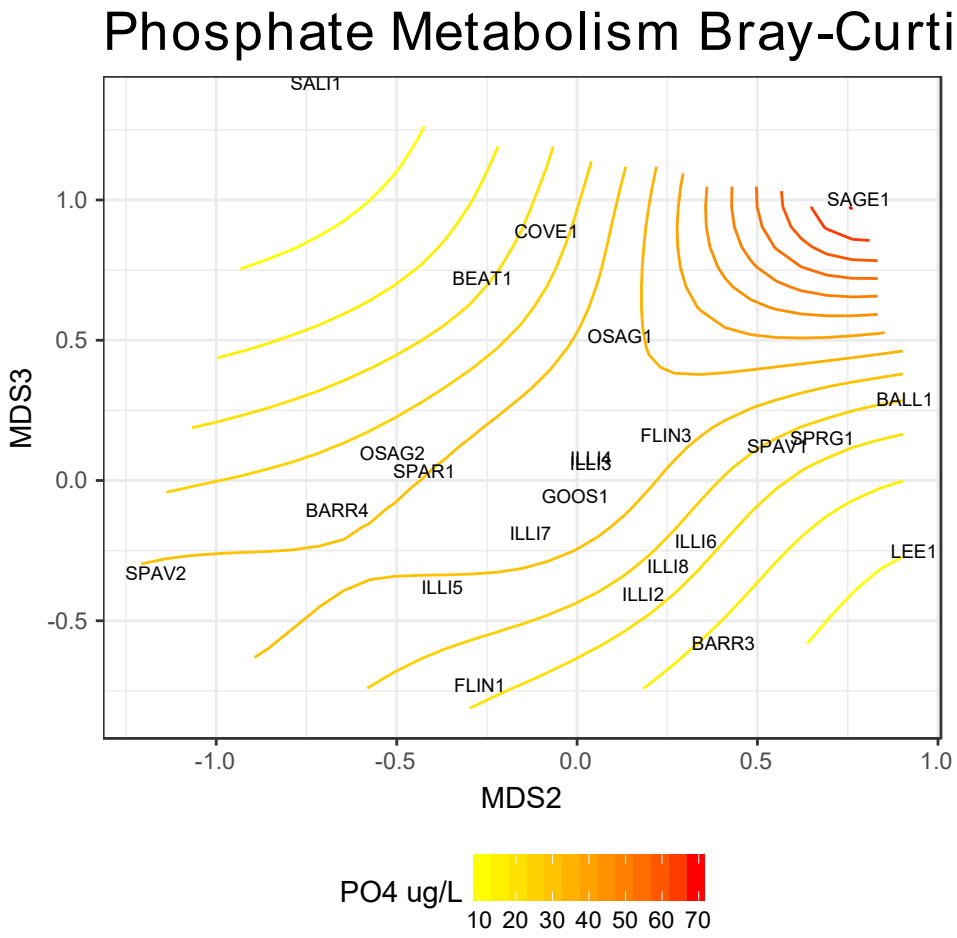


Figure E.S9: NMDS ordination of Phosphate Metabolism genes by site. The non-PO_4 gradient is fit using a GAM. See Table 1 for metrics.

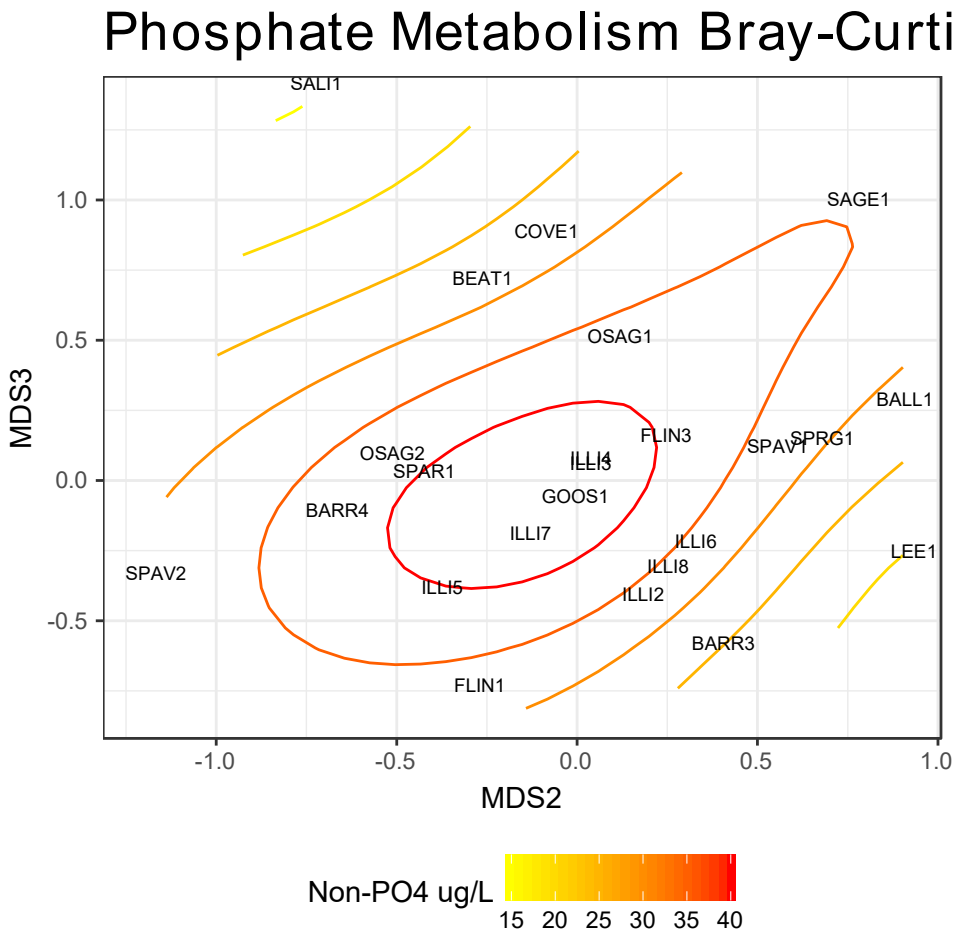


Figure E.S10: NMDS ordination of P Uptake genes by site. PO_4 gradient is fit using a GAM. See Table 1 for metrics.

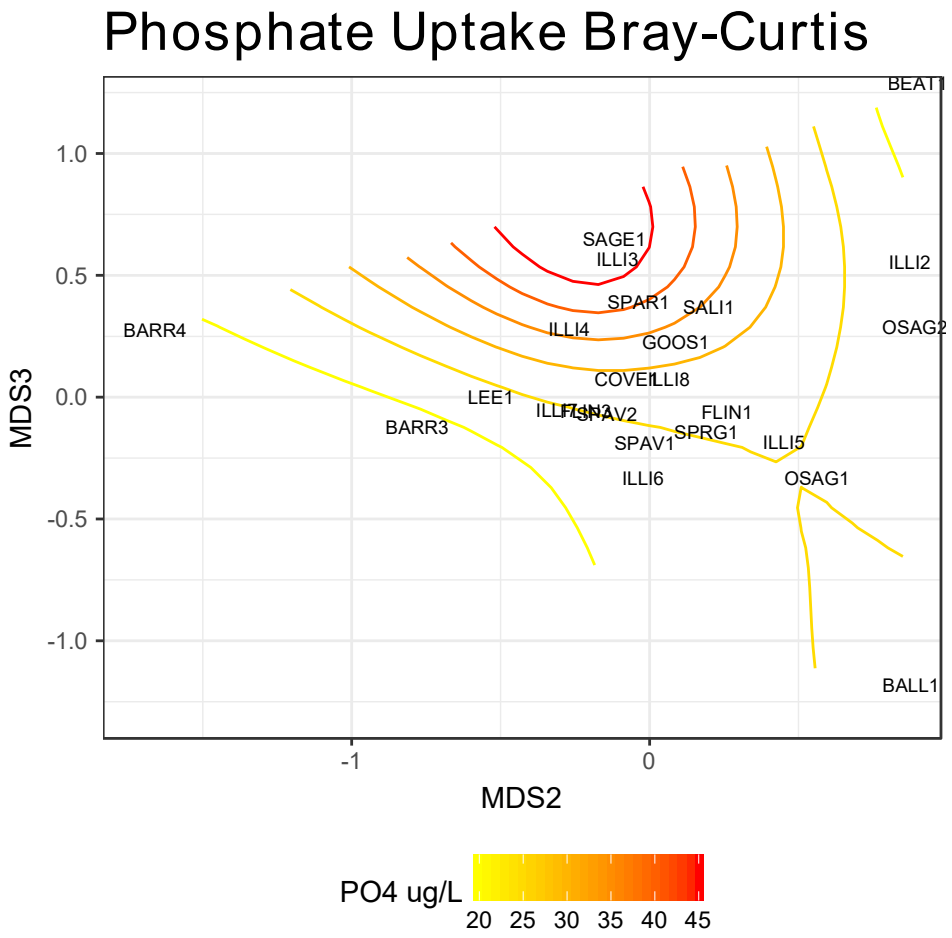


Figure E.S11: Bubble plot showing taxa identified in the Phosphonatate Metabolism gene group. Only sites with detected genes in the group are shown. Sites are arranged from low TP (Blue) to high TP (Red).

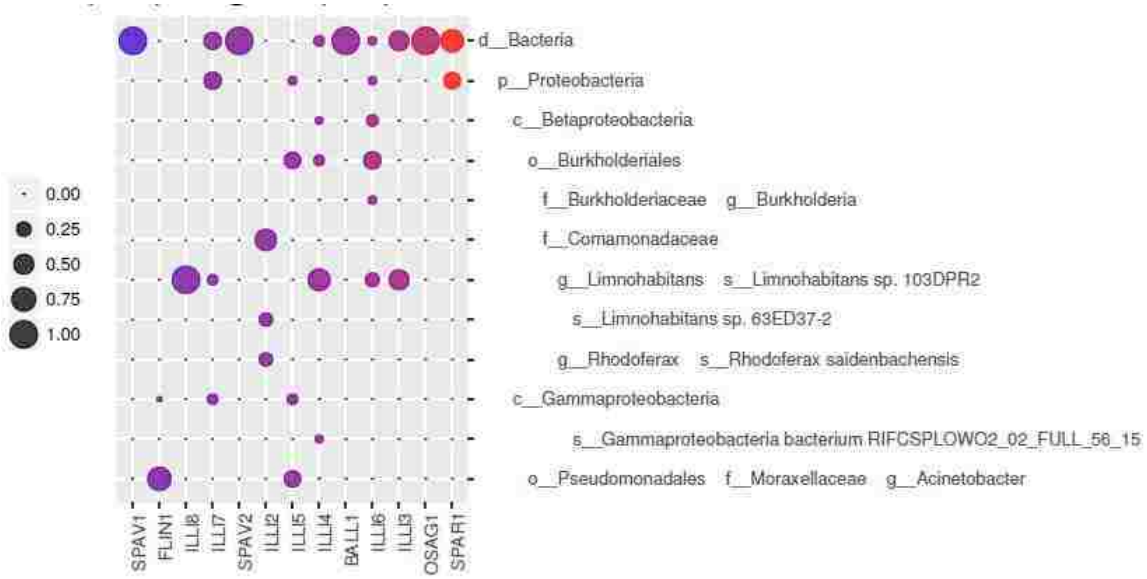


Figure E.S12: Bubble plot showing taxa identified in the Phosphoenolpyruvate Phosphomutase gene group. Only sites with detected genes in the group are shown. Sites are arranged from low TP (Blue) to high TP (Red).

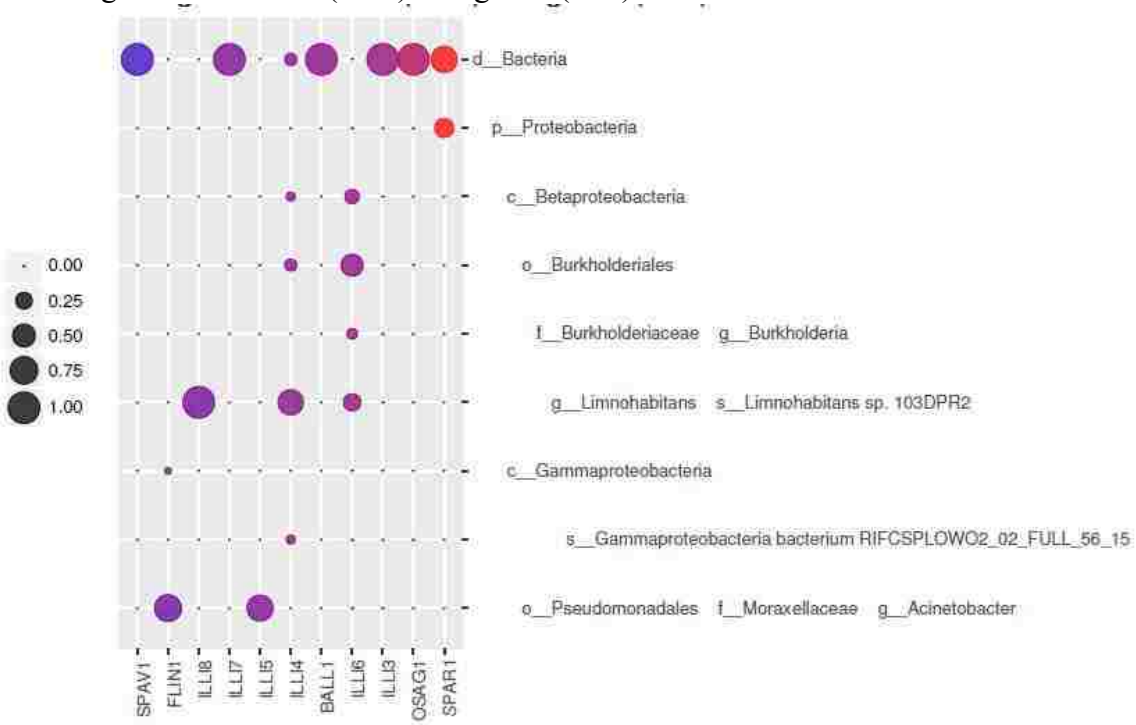


Figure E.S13: Bubble plot showing taxa identified in the High Affinity Phosphate Transporter and Control of PHO Regulon gene group. Only sites with detected genes in the group are shown. Sites are arranged from low TP (Blue) to high TP (Red). For full-size version, please contact the author with reasonable request.

Figure E.S14: Bubble plot showing taxa identified in the Phosphate Metabolism gene group. Only sites with detected genes in the group are shown. Sites are arranged from low TP (Blue) to high TP (Red). For full-size version, please contact the author with reasonable request.

Figure E.S15: Bubble plot showing taxa identified in the P Uptake gene group. Only sites with detected genes in the group are shown. Sites are arranged from low TP (Blue) to high TP (Red). For full-size version, please contact the author with reasonable request.

APPENDIX F

Supplementary Tables (Chapter four)

Supplementary Table

Table F.S1**:** Listing of Genes/Proteins assigned to each SEED functional grouping.

Seed Group	Gene/Protein
Alkylphosphate Utilization	
	• Alkylphosphonate utilization operon protein PhnA • Metal-dependent hydrolase involved in phosphonate metabolism • PhnB protein • PhnH protein • PhnI protein • PhnJ protein • PhnO protein • Phosphonates transport ATP-binding protein PhnK • Protein RcsF • Transcriptional regulator PhnF
High affinity phosphate transporter and control of PHO regulon	
	• Alkaline phosphatase synthesis transcriptional regulatory protein PhoP • Phosphate ABC transporter • Phosphate regulon sensor protein PhoR (SphS) (EC 2.7.13.3) • Phosphate regulon transcriptional regulatory protein PhoB (SphR) • Phosphate transport ATP-binding protein PstB (TC 3.A.1.7.1) • Phosphate transport system permease protein PstA (TC 3.A.1.7.1) • Phosphate transport system permease protein PstC (TC 3.A.1.7.1) • Phosphate transport system regulatory protein PhoU • Polyphosphate kinase (EC 2.7.4.1)
Phosphate Metabolism	
	• 1-acyl-sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51) • Alkaline phosphatase (EC 3.1.3.1) • Alkaline phosphatase like protein • Alkaline phosphatase synthesis transcriptional regulatory protein PhoP • Apolipoprotein N-acyltransferase (EC 2.3.1.-) • Exopolyphosphatase (EC 3.6.1.11) • Inorganic pyrophospatase PpaX (EC 3.1.3.18) • Inorganic pyrophosphatase (EC 3.6.1.1) • Integral membrane protein YggT • Low-affinity inorganic phosphate transporter • Magnesium and cobalt efflux protein CorC • Manganese-dependent inorganic pyrophosphatase (EC 3.6.1.1) • Metal-dependent hydrolase YbeY

- NAD(P) transhydrogenase alpha subunit (EC 1.6.1.2)
- NAD(P) transhydrogenase subunit beta (EC 1.6.1.2)
- Phosphatase
- Phosphate ABC transporter
- Phosphate regulon sensor protein PhoR (SphS) (EC 2.7.13.3)
- Phosphate regulon transcriptional regulatory protein PhoB (SphR)
- Phosphate starvation-inducible protein PhoH
- Phosphate transport ATP-binding protein PstB (TC 3.A.1.7.1)
- Phosphate transport regulator (distant homolog of PhoU)
- Phosphate transport system permease protein PstA (TC 3.A.1.7.1)
- Phosphate transport system permease protein PstC (TC 3.A.1.7.1)
- Phosphate transport system regulatory protein PhoU
- Phosphate-specific outer membrane porin OprP
- Polyphosphate kinase (EC 2.7.4.1)
- Predicted ATPase related to phosphate starvation-inducible protein PhoH
- Probable low-affinity inorganic phosphate transporter
- putative alkaline phosphatase-like protein
- Pyrophosphate-energized proton pump (EC 3.6.1.1)
- Pyrophosphate-specific outer membrane porin OprO
- response regulator in two-component regulatory system with PhoQ
- secreted alkaline phosphatase
- Sodium-dependent phosphate transporter
- Soluble pyridine nucleotide transhydrogenase (EC 1.6.1.1)

Phosphoenolpyruvate Phosphomutase

- 2-aminoethylphosphonate:pyruvate aminotransferase (EC 2.6.1.37)
- Phosphonopyruvate decarboxylase (EC 4.1.1.82)

Phosphonate Metabolism

- 2-aminoethylphosphonate:pyruvate aminotransferase (EC 2.6.1.37)
- Phosphonoacetaldehyde hydrolase (EC 3.11.1.1)
- Phosphonoacetate hydrolase (EC 3.11.1.2)
- Phosphonopyruvate decarboxylase (EC 4.1.1.82)

P Uptake

- Alkaline phosphatase (EC 3.1.3.1)
- Phosphate ABC transporter
- Phosphate transport ATP-binding protein PstB (TC 3.A.1.7.1)
- Phosphate transport system permease protein PstA (TC 3.A.1.7.1)
- Phosphate transport system permease protein PstC (TC 3.A.1.7.1)

APPENDIX G

Supplementary Material (Chapter five)

Supplementary Figures

Figure G.S1: Map of Gotjawal forest sample sites on Jeju Island, Korea. Originally published in Kim, J.-S. et al. Microbial Community Structure and Functional Potential of Lava-Formed Gotjawal Soils in Jeju, Korea. (In review)

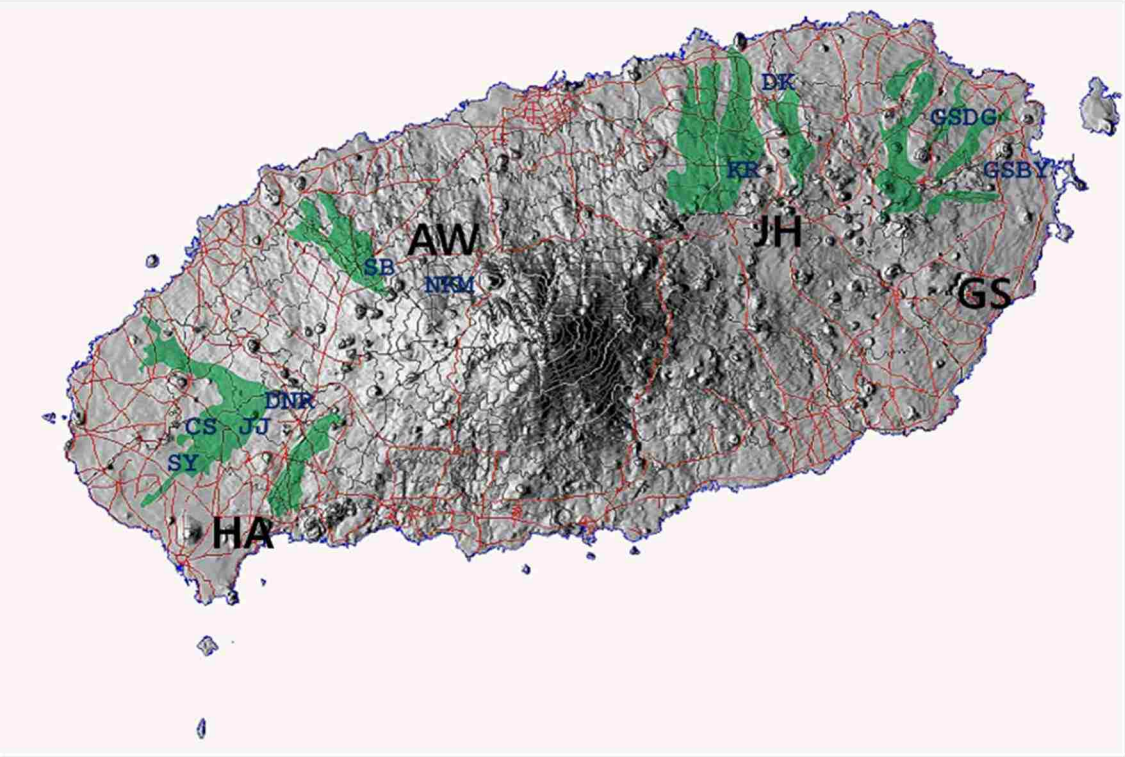
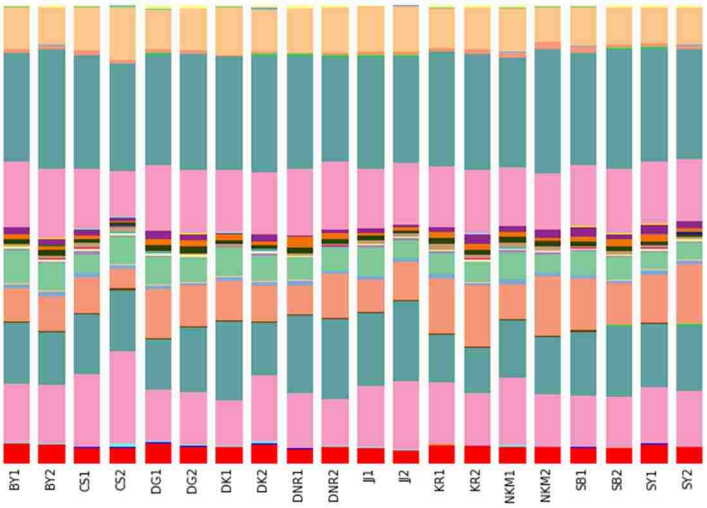


Figure G.S2: Bar plot of taxa at the phyla level in the system showing taxonomic homogeneity across sites. Rarefaction depth of 41,000 used.



Legend	Taxonomy	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
	Unassigned;Other	3.7%	4.4%	4.1%	3.5%	3.3%	4.6%	3.8%	3.7%	4.2%	3.3%	3.7%	3.5%	2.9%	4.0%	3.9%	3.6%	3.6%	3.5%	3.4%	4.3%	3.7%
	k__Archaea;p__Crenarchaeota	0.0%	0.0%	0.1%	0.1%	0.3%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Archaea;p__Euryarchaeota	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Archaea;p__[Parvarchaeota]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;Other	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__AD3	0.2%	0.3%	0.2%	0.1%	0.8%	0.3%	0.2%	0.1%	0.5%	0.1%	0.0%	0.1%	0.1%	0.2%	0.0%	0.4%	0.0%	0.1%	0.0%	0.0%	0.0%
	k__Bacteria;p__Acidobacteria	12.8%	12.8%	12.7%	15.9%	20.2%	11.2%	11.5%	9.9%	14.4%	11.9%	10.1%	13.4%	14.9%	13.6%	11.4%	14.6%	11.5%	11.1%	11.3%	12.3%	12.1%
	k__Bacteria;p__Actinobacteria	13.8%	13.2%	11.4%	13.0%	13.3%	10.9%	14.2%	17.3%	11.5%	17.1%	17.6%	15.8%	17.6%	10.4%	9.8%	12.6%	12.4%	13.9%	15.2%	13.9%	14.4%
	k__Bacteria;p__Armatimonadetes	0.2%	0.3%	0.3%	0.2%	0.3%	0.2%	0.1%	0.1%	0.2%	0.1%	0.1%	0.2%	0.2%	0.3%	0.3%	0.3%	0.3%	0.3%	0.2%	0.1%	0.1%
	k__Bacteria;p__BHI80-139	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__BRC1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Bacteroidetes	9.2%	7.1%	7.5%	7.8%	4.2%	10.7%	8.8%	8.7%	7.7%	6.4%	9.6%	7.1%	8.3%	12.0%	13.1%	7.4%	13.1%	11.3%	9.2%	10.4%	13.3%
	k__Bacteria;p__Chlamydiae	0.8%	1.0%	1.1%	1.0%	0.9%	0.8%	0.7%	0.9%	0.8%	1.1%	0.6%	0.8%	0.8%	0.9%	0.7%	0.9%	0.6%	0.6%	0.6%	0.9%	0.7%
	k__Bacteria;p__Chlorobi	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.3%	0.3%	0.2%	0.1%	0.1%	0.1%	0.2%	0.2%	0.1%	0.2%	0.1%	0.3%	0.2%
	k__Bacteria;p__Chloroflexi	5.1%	7.2%	6.1%	3.9%	6.2%	6.0%	5.4%	6.2%	5.4%	4.7%	5.1%	6.2%	3.9%	4.5%	4.3%	6.0%	3.9%	5.1%	5.0%	3.8%	3.9%
	k__Bacteria;p__Cyanobacteria	0.3%	0.4%	0.3%	0.6%	0.3%	0.2%	0.2%	0.2%	0.4%	0.2%	0.2%	0.5%	0.3%	0.3%	0.2%	0.4%	0.4%	0.3%	0.2%	0.3%	0.2%
	k__Bacteria;p__Elusimicrobia	0.4%	0.3%	0.5%	0.7%	0.5%	0.5%	0.4%	0.1%	0.5%	0.2%	0.2%	0.4%	0.3%	0.3%	0.5%	0.3%	0.5%	0.5%	0.3%	0.5%	0.4%
	k__Bacteria;p__FBP	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__FCPU426	0.1%	0.1%	0.1%	0.1%	0.1%	0.2%	0.1%	0.0%	0.2%	0.0%	0.0%	0.1%	0.0%	0.1%	0.1%	0.0%	0.1%	0.1%	0.0%	0.1%	0.1%
	k__Bacteria;p__Fibrobacteres	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%	0.1%	0.0%	0.1%	0.0%	0.1%
	k__Bacteria;p__Firmicutes	0.6%	0.5%	0.5%	1.1%	1.3%	0.3%	0.2%	0.5%	0.2%	0.5%	0.3%	0.7%	0.7%	1.2%	0.5%	1.0%	0.7%	0.3%	0.6%	0.5%	0.4%
	k__Bacteria;p__Fusobacteria	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__GAL15	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__GN02	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__GN04	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__GOUTA4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Gemmatimonadetes	1.0%	0.9%	1.2%	0.9%	0.9%	1.3%	1.4%	0.7%	0.8%	1.2%	1.0%	0.9%	0.7%	1.3%	1.1%	1.3%	1.1%	1.0%	1.2%	0.9%	0.9%
	k__Bacteria;p__Kazan-3B-28	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__MVP-21	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__NKB19	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Nitrospirae	1.1%	1.2%	1.0%	0.7%	0.2%	1.2%	1.4%	1.4%	0.8%	2.4%	1.4%	0.7%	0.7%	1.2%	1.4%	1.1%	0.5%	0.9%	1.4%	1.4%	0.8%
	k__Bacteria;p__OC31	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__OD1	1.3%	1.3%	1.4%	1.4%	0.9%	1.8%	1.2%	0.4%	1.4%	0.3%	0.7%	1.0%	0.5%	1.0%	2.1%	1.1%	1.8%	2.1%	1.1%	2.2%	1.5%
	k__Bacteria;p__OP11	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.1%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
	k__Bacteria;p__OP3	0.1%	0.1%	0.1%	0.1%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%
	k__Bacteria;p__Planctomycetes	13.5%	14.3%	15.3%	12.9%	10.1%	14.1%	14.0%	13.7%	13.4%	14.6%	14.6%	12.9%	13.6%	13.2%	13.9%	12.8%	12.1%	13.3%	14.2%	13.7%	13.5%
	k__Bacteria;p__Proteobacteria	24.7%	23.7%	25.9%	24.8%	23.2%	24.3%	25.1%	24.6%	25.6%	24.8%	23.1%	24.5%	23.2%	25.0%	25.1%	23.8%	27.1%	24.6%	26.0%	24.7%	23.9%
	k__Bacteria;p__SR1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Spirochaetes	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__TM6	0.1%	0.1%	0.1%	0.1%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
	k__Bacteria;p__TM7	0.8%	0.6%	0.8%	0.9%	0.7%	0.8%	0.7%	0.4%	0.7%	0.4%	0.7%	0.9%	0.8%	0.7%	0.9%	1.1%	1.5%	1.3%	0.9%	0.8%	0.9%
	k__Bacteria;p__Tenericutes	0.0%	0.1%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.2%	0.1%	0.0%	0.1%	0.0%	0.1%
	k__Bacteria;p__Verrucomicrobia	9.1%	9.2%	8.1%	9.5%	11.4%	8.7%	9.1%	10.4%	9.1%	9.4%	9.7%	9.8%	10.0%	8.6%	9.0%	9.5%	7.5%	8.5%	7.9%	7.8%	8.0%
	k__Bacteria;p__WPS-2	0.1%	0.1%	0.1%	0.3%	0.6%	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%	0.2%	0.1%	0.0%	0.0%	0.2%	0.3%	0.2%	0.0%	0.1%	0.2%
	k__Bacteria;p__WS2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%
	k__Bacteria;p__WS3	0.4%	0.4%	0.4%	0.2%	0.0%	0.8%	0.7%	0.4%	0.6%	0.8%	0.5%	0.1%	0.1%	0.6%	0.5%	0.5%	0.2%	0.2%	0.5%	0.6%	0.4%
	k__Bacteria;p__WS4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__WS5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__WS6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__[Caldithrix]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Figure G.S3: Counts of P genes by PICRUSt and GeoChip scaled independently.

Black = PICRUSt. Red = Geochip. Separated for scale.

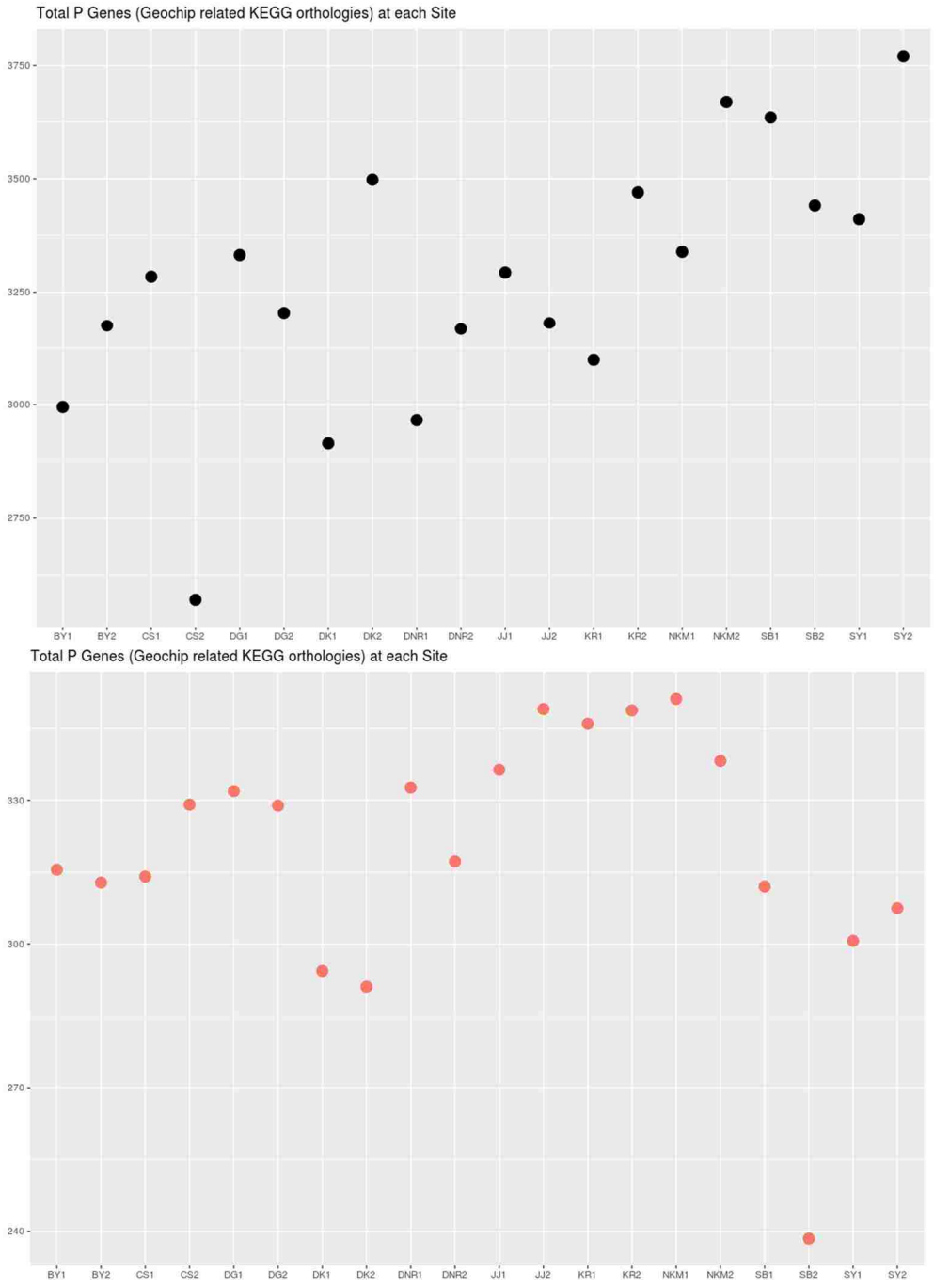


Figure G.S4: Procrustes Error Plot for Procrustes analysis between PICRUSt and GeoChip datasets.

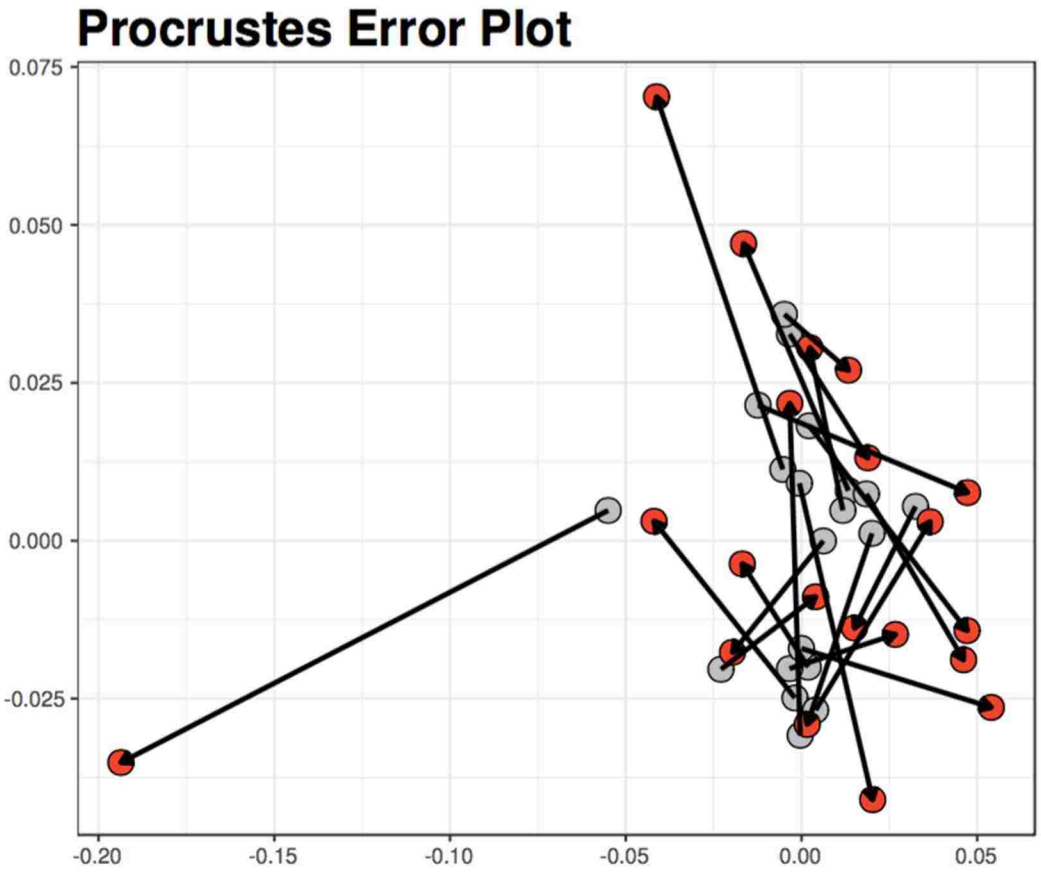
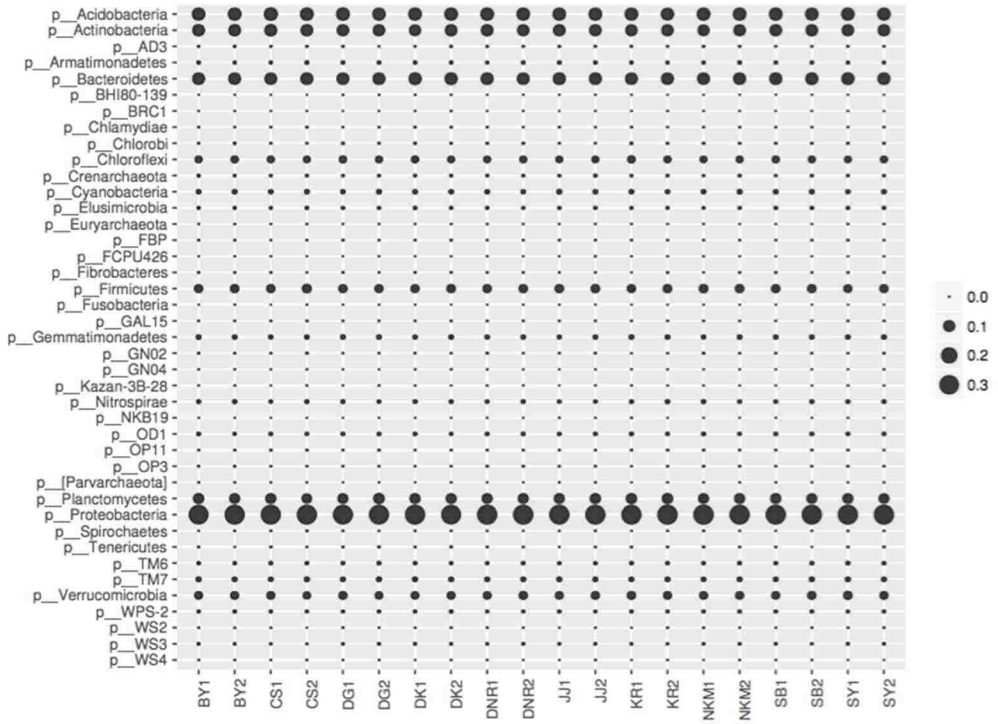
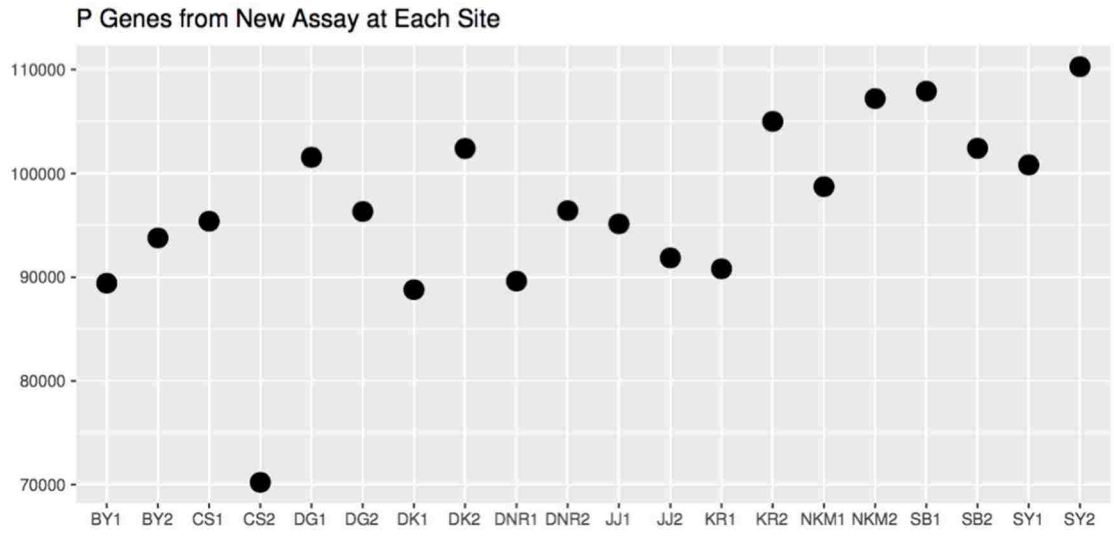


Figure G.S5: A) Bubble plot of taxa identified by new assay at each site. B) Total counts identified in new assay at each site by PICRUSt.

A)



B)



Supplementary Table

Table G.S1: KEGG Orthologies included in "new" PICRUSt P assay.

Gene/Enzyme	KEGG Orthology Numbers									
phnA	K06193	K19670								
phnM	K06162									
phnB	K04750									
phnH	K06165									
phnI	K06164									
phnO	K09994									
phnJ	K06163									
phnK	K05781									
RcsF	K06080									
phnF	K02043									
phoP	K07660	K07658								
pstC	K02037									
pstA	K02038									
phoR	K07636									
phoB	K07657	K07657								
pstB	K02036									
phoU	K05946									
ppk	K00937									
plsC	K00655									
phoA	K01077									
lnt	K03820									
ppX1	K01514									
ppaX	K06019									
ppa	K01507									
yggT	K02221									
pit	K03306									
corC	K06189									
ppaC	K15986									
ybeY	K07042									
pntA	K00324									
pntB	K00325									
nnt	K00323									
Phosphatase	K00906	K01514	K05307	K06366	K08743	K13988	K16054	K17571	K18046	K20220
	K01077	K01515	K05518	K06367	K08966	K14019	K16055	K17574	K18065	K20278
	K01078	K01517	K05602	K06368	K09474	K14165	K16340	K17575	K18078	K20389

	K01079	K01518	K05693	K06369	K09612	K14379	K16662	K17576	K18079	K20437
	K01080	K01519	K05694	K06478	K09880	K14394	K16667	K17579	K18203	K20439
	K01084	K01520	K05695	K06645	K10047	K14395	K16723	K17585	K18398	K20440
	K01085	K01521	K05696	K06776	K10147	K14410	K16833	K17594	K18424	K20551
	K01086	K01522	K05697	K06777	K10909	K14497	K16904	K17605	K18446	K20581
	K01087	K01524	K05698	K06778	K10916	K14501	K16910	K17607	K18447	K20827
	K01089	K01525	K05766	K06881	K11240	K14634	K17453	K17614	K18453	K20860
	K01090	K01526	K05866	K06896	K11532	K14803	K17457	K17615	K18498	K20861
	K01091	K01622	K05867	K06928	K11583	K14819	K17458	K17616	K18568	K20862
	K01092	K02226	K05978	K06949	K11584	K15422	K17459	K17617	K18649	K20866
	K01093	K02374	K05979	K07024	K11725	K15423	K17491	K17618	K18654	K20945
	K01094	K02446	K06018	K07026	K11751	K15424	K17499	K17619	K18693	K20979
	K01095	K02555	K06019	K07053	K11777	K15425	K17500	K17623	K18697	K21013
	K01096	K03084	K06116	K07189	K11915	K15426	K17501	K17816	K18998	K21055
	K01097	K03103	K06117	K07252	K11938	K15427	K17502	K17817	K18999	K21063
	K01098	K03270	K06124	K07293	K12152	K15494	K17503	K17879	K19028	K21064
	K01099	K03273	K06153	K07313	K12328	K15498	K17504	K18018	K19029	K21278
	K01100	K03426	K06162	K07314	K12329	K15499	K17505	K18019	K19030	K21302
	K01101	K03456	K06268	K07315	K12354	K15500	K17506	K18024	K19269	K21403
	K01102	K03574	K06269	K07658	K12584	K15501	K17507	K18025	K19270	K21503
	K01103	K03788	K06270	K07757	K12804	K15502	K17508	K18026	K19283	K21517
	K01104	K03841	K06352	K07758	K12944	K15503	K17509	K18027	K19284	K21797
	K01106	K04041	K06353	K07766	K12945	K15504	K17549	K18032	K19302	K21798
	K01107	K04342	K06354	K07817	K12977	K15529	K17550	K18033	K19581	K21814
	K01109	K04348	K06355	K08050	K12978	K15544	K17552	K18034	K19582	K21830
	K01110	K04354	K06356	K08067	K13084	K15637	K17553	K18035	K19704	K21946
	K01111	K04382	K06357	K08073	K13085	K15640	K17555	K18036	K19705	K22200
	K01112	K04457	K06358	K08075	K13086	K15728	K17556	K18037	K19708	K22223
	K01113	K04458	K06359	K08077	K13216	K15731	K17557	K18038	K19790	
	K01122	K04459	K06360	K08114	K13248	K15732	K17558	K18039	K19806	
	K01139	K04460	K06361	K08296	K13297	K15759	K17562	K18040	K19812	
	K01507	K04461	K06362	K08310	K13555	K15781	K17565	K18041	K20074	
	K01509	K04486	K06363	K08312	K13617	K15850	K17566	K18043	K20124	
	K01512	K04716	K06364	K08320	K13807	K15909	K17567	K18044	K20201	
	K01513	K04765	K06365	K08726	K13987	K15986	K17568	K18045	K20216	
phoH	K06217									
oprO_P	K07221									
hppA	K15987									
phoQ	K07637									
SLC20A	K14640									

sthA	K00322
phnW	K03430
EC4.1.1.82	K09459
phnX	K05306

COMPILED REFERENCES

Adams RI, Miletto M, Taylor JW, Bruns TD. 2013. Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. The ISME Journal; London 7:1262–1273. doi:http://dx.doi.org/10.1038/ismej.2013.28.

Allan JD. 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. Annual Review of Ecology, Evolution, and Systematics 35:257–284. doi:10.1146/annurev.ecolsys.35.120202.110122.

Allen TD, Lawson PA, Collins MD, Falsen E, Tanner RS. 2006. Cloacibacterium normanense gen. nov., sp. nov., a novel bacterium in the family Flavobacteriaceae isolated from municipal wastewater. International Journal of Systematic and Evolutionary Microbiology 56:1311–1316. doi:10.1099/ijs.0.64218-0.

Allison SD, Martiny JB. 2008. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences 105:11512–11519.

Amend AS, Seifert KA, Samson R, Bruns TD. 2010. Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. Proceedings of the National Academy of Sciences 107:13748–13753.

Anderson DM, Glibert PM, Burkholder JM. 2002. Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. Estuaries 25:704–726.

APHA. 1998. Standard methods for the examination of water and wastewater. American Public Health Association (APHA): Washington, DC, USA.

Aronesty E. 2013. Comparison of sequencing utility programs. The Open Bioinformatics Journal 7. [accessed 2016 Oct 27]. http://benthamopen.com/ABSTRACT/TOBIOIJ-7-1.

Azam F, Fenchel T, Field JG, Gray JS, Meyer-Reil LA, Thingstad F. 1983. The Ecological Role of Water-Column Microbes in the Sea. Marine Ecology Progress Series 10:257–263.

Baker ME, King RS. 2010. A new method for detecting and interpreting biodiversity and ecological community thresholds. Methods in Ecology and Evolution 1:25–37.

Baker ME, King RS, Kahle, David. 2015. TITAN2: Threshold Indicator Taxa Analysis (Version 2.1).

Barsdate RJ, Prentki RT, Fenchel T. 1974. Phosphorus cycle of model ecosystems: significance for decomposer food chains and effect of bacterial grazers. Oikos:239–251.

Bennett EM, Carpenter SR, Caraco NF. 2001. Human Impact on Erodable Phosphorus and Eutrophication: A Global PerspectiveIncreasing accumulation of phosphorus in soil threatens rivers, lakes, and coastal oceans with eutrophication. BioScience 51:227–234. doi:10.1641/0006-3568(2001)051[0227:HIOEPA]2.0.CO;2.

Bermingham E, Avise JC. 1986. Molecular zoogeography of freshwater fishes in the southeastern United States. Genetics 113:939–965.

Bernatchez L, Wilson CC. 1998. Comparative phylogeography of Nearctic and Palearctic fishes. Molecular Ecology 7:431–452.

Bilotta GS, Brazier RE. 2008. Understanding the influence of suspended solids on water quality and aquatic biota. Water Research 42:2849–2861. doi:10.1016/j.watres.2008.03.018.

Bolan NS. 1991. A critical review on the role of mycorrhizal fungi in the uptake of phosphorus by plants. Plant Soil 134:189–207. doi:10.1007/BF00012037.

Bormann FH, Likens G. 2012. Pattern and Process in a Forested Ecosystem: Disturbance, Development and the Steady State Based on the Hubbard Brook Ecosystem Study. Springer Science & Business Media.

Boulion VV. 2012. Assimilation and turnover time of phosphorus by size fractions of microplankton in lakes of different types. Inland Water Biol 5:304–309. doi:10.1134/S1995082912040049.

Brewis ND, Street AJ, Prescott AR, Cohen PT. 1993. PPX, a novel protein serine/threonine phosphatase localized to centrosomes. EMBO J 12:987–996.

Brown JK, Hovmøller MS. 2002. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. Science 297:537–541.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Meth 12:59–60. doi:10.1038/nmeth.3176.

Bunn SE, Hughes JM. 1997. Dispersal and Recruitment in Streams: Evidence from Genetic Studies. Journal of the North American Benthological Society 16:338–346. doi:10.2307/1468022.

Butts CT, Hunter D, Handcock M, Bender-deMoll, Skye, Horner, Jeffery. 2015. Network: Classes for Relational Data.

Butts CT, others. 2008. network: a Package for Managing Relational Data in R. Journal of Statistical Software 24:1–36.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature methods 7:335–336.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, et al. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal 6:1621–1624.

Cardinale BJ, Hillebrand H, Harpole WS, Gross K, Ptacnik R. 2009. Separating the influence of resource 'availability' from resource 'imbalance' on productivity–diversity relationships. Ecology Letters 12:475–487. doi:10.1111/j.1461-0248.2009.01317.x.

Carpenter SR, Caraco NF, Correll DL, Howarth RW, Sharpley AN, Smith VH. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecological applications 8:559–568.

Chambers PA, McGoldrick DJ, Brua RB, Vis C, Culp JM, Benoy GA. 2012. Development of environmental thresholds for nitrogen and phosphorus in streams. Journal of Environmental Quality 41:7–20.

Chapman DV, others. 1996. Water quality assessments: a guide to the use of biota, sediments, and water in environmental monitoring. [accessed 2017 Jan 24]. http://wwwlive.who.int/entity/water_sanitation_health/resourcesquality/watqualassess.pdf.

Charney N, Record S. 2012. vegetarian: Jost Diversity Measures for Community Data.

Christensen M. 1989. A View of Fungal Ecology. Mycologia 81:1–19. doi:10.2307/3759446.

Cline LC, Zak DR. 2014. Dispersal limitation structures fungal community assembly in a long-term glacial chronosequence. Environmental microbiology 16:1538–1548.

Cook AM, Daughton CG, Alexander M. 1978. Phosphorus-containing pesticide breakdown products: quantitative utilization as phosphorus sources by bacteria. Applied and Environmental Microbiology 36:668–672.

Cook SC, Housley L, Back JA, King RS. Freshwater eutrophication drives sharp reductions in temporal beta diversity. Ecology. doi:10.1002/ecy.2069.

Correll DL. 1998. The Role of Phosphorus in the Eutrophication of Receiving Waters: A Review. Journal of Environment Quality 27:261. doi:10.2134/jeq1998.00472425002700020004x.

Cosgrove DJ. 1977. Microbial Transformations in the Phosphorus Cycle. In: Advances in Microbial Ecology. Springer, Boston, MA. (Advances in Microbial Ecology). p. 95–134. [accessed 2017 Sep 15]. https://link.springer.com/chapter/10.1007/978-1-4615-8219-9_3.

Cotner JB, Biddanda BA. 2002. Small Players, Large Role: Microbial Influence on Biogeochemical Processes in Pelagic Aquatic Ecosystems. Ecosystems 5:105–121. doi:10.1007/s10021-001-0059-3.

Cottrell MT, Kirchman DL. 2000. Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low-and high-molecular-weight dissolved organic matter. Applied and Environmental Microbiology 66:1692–1697.

Csardi G, Nepusz T. 2006. The igraph software package for complex network research. InterJournal, Complex Systems 1695:1–9.

Daniel R. 2004. The soil metagenome – a rich resource for the discovery of novel natural products. Current Opinion in Biotechnology 15:199–204. doi:10.1016/j.copbio.2004.04.005.

Davies-Colley RJ, Smith DG. 2001. Turbidity suspended sediment, and water clarity: A review.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and environmental microbiology 72:5069–5072.

Dufrêne M, Legendre P. 1997. Species Assemblages and Indicator Species:the Need for a Flexible Asymmetrical Approach. Ecological Monographs 67:345–366. doi:10.1890/0012-9615(1997)067[0345:SAAIST]2.0.CO;2.

Duggan A, Charnley G, Chen W, Chukwudebe A, Hawk R, Krieger RI, Ross J, Yarborough C. 2003. Di-alkyl phosphate biomonitoring data: assessing cumulative exposure to organophosphate pesticides. Regulatory Toxicology and Pharmacology 37:382–395. doi:10.1016/S0273-2300(03)00031-X.

Dyhrman ST, Chappell PD, Haley ST, Moffett JW, Orchard ED, Waterbury JB, Webb EA. 2006. Phosphonate utilization by the globally important marine diazotroph Trichodesmium. Nature 439:68–71. doi:10.1038/nature04203.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.

Elser JJ, Bracken ME, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, Ngai JT, Seabloom EW, Shurin JB, Smith JE. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecology letters 10:1135–1142.

ESRI. ArcGIS Desktop: Release 10, 2011 Environmental Systems Research Institute; Redlands, CA.

Fabre E. 1998a. Aquatic hyphomycetes in three rivers of southwestern France. III. Relationships between communities spatial and temporal dynamics. Canadian journal of botany 76:115–121.

Fabre E. 1998b. Aquatic hyphomycetes in three rivers of southwestern France. II. Spatial and temporal differences between species. Canadian Journal of Botany 76:107–114. doi:10.1139/b97-156.

Fabre E. 1998c. Aquatic hyphomycetes in three rivers of southwestern France. I. Spatial and temporal changes in conidial concentration, species richness, and community diversity. Can. J. Bot. 76:99–106. doi:10.1139/b97-155.

Fox A, Kwapinski W, Griffiths BS, Schmalenberger A. 2014. The role of sulfur- and phosphorus-mobilizing bacteria in biochar-induced growth promotion of Lolium perenne. FEMS Microbiology Ecology 90:78–91. doi:10.1111/1574-6941.12374.

Frissell CA, Liss WJ, Warren CE, Hurley MD. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. Environmental Management 10:199–214. doi:10.1007/BF01867358.

Frossard E, Condron LM, Oberson A, Sinaj S, Fardeau JC. 2000. Processes governing phosphorus availability in temperate soils. Journal of Environmental Quality; Madison 29:15.

García-Palacios P, Vandegehuchte ML, Shaw EA, Dam M, Post KH, Ramirez KS, Sylvain ZA, Tomasel CM, Wall DH. 2015. Are there links between responses of soil microbes and ecosystem functioning to elevated CO2, N deposition and warming? A global perspective. Global change biology 21:1590–1600.

Gartemann K-H, Kirchner O, Engemann J, Gräfen I, Eichenlaub R, Burger A. 2003. Clavibacter michiganensis subsp. michiganensis: first steps in the understanding of virulence of a Gram-positive phytopathogenic bacterium. Journal of Biotechnology 106:179–191.

Gessner MO, Newell SY. 2002. Biomass, growth rate, and production of filamentous fungi in plant litter. Manual of environmental microbiology 2:390–408.

Ghaffar S, Stevenson RJ, Khan Z. 2017a. Effect of phosphorus stress on Microcystis aeruginosa growth and phosphorus uptake. PLoS One; San Francisco 12. doi:http://dx.doi.org/10.1371/journal.pone.0174349.

Ghaffar S, Stevenson RJ, Khan Z. 2017b. Effect of phosphorus stress on Microcystis aeruginosa growth and phosphorus uptake. PLoS One; San Francisco 12. doi:http://dx.doi.org/10.1371/journal.pone.0174349.

Godwin CM, Cotner JB. 2014. Carbon: phosphorus homeostasis of aquatic bacterial assemblages is mediated by shifts in assemblage composition. Aquatic Microbial Ecology 73:245–258.

Godwin CM, Whitaker EA, Cotner JB. 2016. Growth rate and resource imbalance interactively control biomass stoichiometry and elemental quotas of aquatic bacteria. Ecology. [accessed 2017 Feb 13]. http://onlinelibrary.wiley.com/doi/10.1002/ecy.1705/full.

Gorniak A, Więcko A, Cudowski A. 2013. Fungi biomass in lowland rivers of north-eastern Poland: Effects of habitat conditions and nutrient concentrations. Pol. J. Ecol 61:749–758.

Green WR, Haggard BE. 2001. Phosphorus and nitrogen concentrations and loads at Illinois River south of Siloam Springs, Arkansas, 1997-1999. [accessed 2016 Oct 27]. https://pubs.er.usgs.gov/usgspubs/wri/wri014217.

Grinn-Gofroń A, Bosiacka B. 2015. Effects of meteorological factors on the composition of selected fungal spores in the air. Aerobiologia 31:63–72. doi:10.1007/s10453-014-9347-1.

Haggard BE. 2010. Phosphorus concentrations, loads, and sources within the Illinois River drainage area, northwest Arkansas, 1997–2008. Journal of environmental quality 39:2113–2120.

Haggard BE, Soerens TS. 2006. Sediment phosphorus release at a small impoundment on the Illinois River, Arkansas and Oklahoma, USA. Ecological Engineering 28:280–287. doi:10.1016/j.ecoleng.2006.07.013.

Harke MJ, Gobler CJ. 2013. Global Transcriptional Responses of the Toxic Cyanobacterium, Microcystis aeruginosa, to Nitrogen Stress, Phosphorus Stress, and Growth on Organic Matter. PLOS ONE 8:e69834. doi:10.1371/journal.pone.0069834.

Harley JL. 1971. Fungi in Ecosystems. Journal of Applied Ecology 8:627–642. doi:10.2307/2402673.

Hart MR, Quin BF, Nguyen M. 2004. Phosphorus runoff from agricultural land and direct fertilizer effects. Journal of Environmental Quality 33:1954–1972.

He Z, Deng Y, Zhou J. 2012. Development of functional gene microarrays for microbial community analysis. Current Opinion in Biotechnology 23:49–55. doi:10.1016/j.copbio.2011.11.001.

Heijden MGA van der, Klironomos JN, Ursic M, Moutoglis P, Streitwolf-Engel R, Boller T, Wiemken A, Sanders IR. 1998. Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. Nature 396:69. doi:10.1038/23932.

Hernández F, Sancho JV, Pozo OJ. 2002. Direct determination of alkyl phosphates in human urine by liquid chromatography/electrospray tandem mass spectrometry. Rapid Commun. Mass Spectrom. 16:1766–1773. doi:10.1002/rcm.790.

Hill WR, Fanta SE, Roberts BJ. 2009. Quantifying phosphorus and light effects in stream algae. Limnology and Oceanography 54:368–380.

Hulvey J, Gobena D, Finley L, Lamour K. 2010. Co-occurrence and genotypic distribution of Phytophthora species recovered from watersheds and plant nurseries of eastern Tennessee. Mycologia 102:1127–1133.

Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. Genome Res. 17:377–386. doi:10.1101/gr.5969107.

Ingham RE, Trofymow JA, Ingham ER, Coleman DC. 1985. Interactions of Bacteria, Fungi, and their Nematode Grazers: Effects on Nutrient Cycling and Plant Growth. Ecological Monographs 55:119–140. doi:10.2307/1942528.

Ingold CT. 1942. Aquatic hyphomycetes of decaying alder leaves. Transactions of the British Mycological Society 25:339IN1–417IN6.

Inskeep WP, Jay ZJ, Tringe SG, Herrg\a ard MJ, Rusch DB, Committee YMPS, Members WG. 2013. The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. Frontiers in microbiology 4.

Jang Y-C, Lee C-W. 2009. Gotjawal Forest in Jeju Island as an Internationally Important Wetland. Journal of Wetlands Research 11.

Jansson M. 1993. Uptake, Exchange, and Excretion of Orthophosphate in Phosphate-Starved Scenedesmus quadricauda and Pseudomonas K7. Limnology and Oceanography 38:1162–1178.

Jost L. 2007. Partitioning Diversity into Independent Alpha and Beta Components. Ecology 88:2427–2439. doi:10.1890/06-1736.1.

Kamat SS, Williams HJ, Raushel FM. 2011. Intermediates in the Transformation of Phosphonates to Phosphate by Bacteria. Nature 480:570–573. doi:10.1038/nature10622.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28:27–30.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2015. KEGG as a reference resource for gene and protein annotation. Nucleic acids research:gkv1070.

Kim J-S, Jung M-Y, Lee KC, Kim D-S, Ko S-H, Lee J-S, Rhee S-K. 2014. The Archaea Community Associated with Lava-Formed Gotjawal Forest Soil in Jeju, Korea. Journal of Agricultural Chemistry and Environment 03:96. doi:10.4236/jacen.2014.33012.

Kim J-S, Kim D-S, Lee KC, Ko S-H, Lee J-S, King GM, Kang S. 2017. Microbial Community Structure and Functional Potential of Lava-Formed Gotjawal Soils in Jeju, Korea. In review.

Krom MD, Kress N, Brenner S, Gordon LI. 1991. Phosphorus limitation of primary productivity in the eastern Mediterranean Sea. Limnology and Oceanography 36:424–432.

Laliberté E, Lambers H, Burgess TI, Wright SJ. 2015. Phosphorus limitation, soil-borne pathogens and the coexistence of plant species in hyperdiverse forests and shrublands. New Phytol 206:507–521. doi:10.1111/nph.13203.

Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Thurber RLV, Knight R, et al. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature biotechnology 31:814–821.

Laureto LMO, Cianciaruso MV, Samia DSM. 2015. Functional diversity: an overview of its history and applicability. Natureza & Conservação 13:112–116. doi:10.1016/j.ncon.2015.11.001.

LeBrun ES, King RS, Back JA, Kang S. 2018. Microbial Community Structure and Function Decoupling Across a Phosphorus Gradient in Streams. Microb Ecol 75:64–73. doi:10.1007/s00248-017-1039-2.

LeBrun ES, Taylor DL, King RS, Back JA, Kang S. 2018. Rivers may constitute an overlooked avenue of dispersal for terrestrial fungi. Fungal Ecol In press.

Li D-W. 2016. Biology of Microfungi. Springer.

Lloyd DS, Koenings JP, Laperriere JD. 1987. Effects of turbidity in fresh waters of Alaska. North American Journal of Fisheries Management 7:18–33.

Logue JB, Stedmon CA, Kellerman AM, Nielsen NJ, Andersson AF, Laudon H, Lindström ES, Kritzberg ES. 2016. Experimental insights into the importance of aquatic bacterial community composition to the degradation of dissolved organic matter. The ISME journal 10:533–545.

Louca S, Parfrey LW, Doebeli M. 2016. Decoupling function and taxonomy in the global ocean microbiome. Science 353:1272–1277.

Maguire B. 1963. The Passive Dispersal of Small Aquatic Organisms and Their Colonization of Isolated Bodies of Water. Ecological Monographs 33:161–185. doi:10.2307/1948560.

Marra G, Wood SN. 2011. Practical variable selection for generalized additive models. Computational Statistics & Data Analysis 55:2372–2387.

McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one 8:e61217.

Micheli F, Halpern BS. 2005. Low functional redundancy in coastal marine assemblages. Ecology Letters 8:391–400.

Miki T, Yokokawa T, Matsui K. 2014. Biodiversity and multifunctionality in a microbial community: a novel theoretical approach to quantify functional redundancy. Proc. R. Soc. B 281:20132498. doi:10.1098/rspb.2013.2498.

Miller MP, Blinn DW, Keim P. 2002. Correlations between observed dispersal capabilities and patterns of genetic differentiation in populations of four aquatic insect species from the Arizona White Mountains, U.S.A. Freshwater Biology 47:1660–1673. doi:10.1046/j.1365-2427.2002.00911.x.

Miltner RJ, others. 1998. Primary nutrients and the biotic integrity of rivers and streams. Freshwater Biology 40:145–158.

Mukhamedova KS, Glushenkova AI. 2000. Natural Phosphonolipids. Chemistry of Natural Compounds 36:329–341. doi:10.1023/A:1002804409503.

Nakazawa M. 2017. fmsb: Functions for medical statistics book with some demographic data. R package version 0.6.1. CRAN. R-project. org/package= fmsb.

NCBI Resource Coordinators. 2017. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 45:D12–D17. doi:10.1093/nar/gkw1071.

Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS, Kennedy PG. 2016. FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecology 20:241–248. doi:10.1016/j.funeco.2015.06.006.

Nikolcheva LG, Bärlocher F. 2005. Seasonal and substrate preferences of fungi colonizing leaves in streams: traditional versus molecular evidence. Environmental microbiology 7:270–280.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2016. vegan: Community Ecology Package. R package version 2.4-1.

Pagaling E, Strathdee F, Spears BM, Cates ME, Allen RJ, Free A. 2014. Community history affects the predictability of microbial ecosystem development. ISME J 8:19–30. doi:10.1038/ismej.2013.150.

Pashley CH, Fairs A, Free RC, Wardlaw AJ. 2012. DNA analysis of outdoor air reveals a high degree of fungal diversity, temporal variability, and genera not seen by spore morphology. Fungal biology 116:214–224.

Peay KG, Bidartondo MI, Arnold AE. 2010. Not every fungus is everywhere: scaling to the biogeography of fungal–plant interactions across roots, shoots and ecosystems. The New Phytologist 185:878–882.

Peay KG, Bruns TD. 2014. Spore dispersal of basidiomycete fungi at the landscape scale is driven by stochastic and deterministic processes and generates variability in plant–fungal interactions. New Phytol 204:180–191. doi:10.1111/nph.12906.

Peay KG, Schubert MG, Nguyen NH, Bruns TD. 2012. Measuring ectomycorrhizal fungal dispersal: macroecological patterns driven by microscopic propagules. Molecular Ecology 21:4122–4136.

Pedersen JA, Soliman M, Suffet IH (Mel). 2005. Human Pharmaceuticals, Hormones, and Personal Care Product Ingredients in Runoff from Agricultural Fields Irrigated with Treated Wastewater. J. Agric. Food Chem. 53:1625–1632. doi:10.1021/jf049228m.

Pellegrino KCM, Rodrigues MT, Waite AN, Morando M, Yassuda YY, Sites JW. 2005. Phylogeography and species limits in the Gymnodactylus darwinii complex (Gekkonidae, Squamata): genetic structure coincides with river systems in the Brazilian Atlantic Forest. Biological Journal of the Linnean Society 85:13–26. doi:10.1111/j.1095-8312.2005.00472.x.

Pérez GL, Torremorell A, Mugni H, Rodríguez P, Vera MS, Nascimento M do, Allende L, Bustingorry J, Escaray R, Ferraro M, et al. 2007. Effects of the Herbicide Roundup on Freshwater Microbial Communities: A Mesocosm Study. Ecological Applications 17:2310–2322. doi:10.1890/07-0499.1.

Petersen I, Masters Z, Hildrew AG, Ormerod SJ. 2004. Dispersal of adult aquatic insects in catchments of differing land use. Journal of Applied Ecology 41:934–950. doi:10.1111/j.0021-8901.2004.00942.x.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. PLOS Biology 9:e1000602. doi:10.1371/journal.pbio.1000602.

Prágai Z, Harwood CR. 2002. Regulatory interactions between the Pho and σB-dependent general stress regulons of Bacillus subtilis. Microbiology 148:1593–1602. doi:10.1099/00221287-148-5-1593.

Qi Y, Kobayashi Y, Hulett FM. 1997. The pst operon of Bacillus subtilis has a phosphate-regulated promoter and is involved in phosphate transport but not in regulation of the pho regulon. J. Bacteriol. 179:2534–2539. doi:10.1128/jb.179.8.2534-2539.1997.

Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, et al. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490:nature11450. doi:10.1038/nature11450.

R Core Team. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Read DS, Gweon HS, Bowes MJ, Newbold LK, Field D, Bailey MJ, Griffiths RI. 2015. Catchment-scale biogeography of riverine bacterioplankton. The ISME journal 9:516–526.

Rhee S-K, Liu X, Wu L, Chong SC, Wan X, Zhou J. 2004. Detection of Genes Involved in Biodegradation and Biotransformation in Microbial Communities by Using 50-Mer Oligonucleotide Microarrays. Appl. Environ. Microbiol. 70:4303–4317. doi:10.1128/AEM.70.7.4303-4317.2004.

Richardson AE. 2001. Prospects for using soil microorganisms to improve the acquisition of phosphorus by plants. Functional Plant Biology 28:897–906.

Richardson AE, Simpson RJ. 2011. Soil microorganisms mediating phosphorus availability update on microbial phosphorus. Plant physiology 156:989–996.

Rier ST, Kinek KC, Hay SE, Francoeur SN. 2016. Polyphosphate plays a vital role in the phosphorus dynamics of stream periphyton. Freshwater Science 35:490–502.

Roberts DW. 2013. labdsv: Ordination and Multivariate Analysis for Ecology. R package version 1.6-1. h ttp. CRAN. R-project. org/package= labdsv.

Rodríguez H, Fraga R. 1999. Phosphate solubilizing bacteria and their role in plant growth promotion. Biotechnology Advances 17:319–339. doi:10.1016/S0734-9750(99)00014-2.

Santamaría L. 2002. Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. Acta Oecologica 23:137–154. doi:10.1016/S1146-609X(02)01146-3.

Saralov A, Mol'kov D, Bannikova O, Solomennyĭ A, Chikin S. 2000. Intracellular accumulation of monomer precursors of polyphosphates and polyhydroxyalkanoates from Acinetobacter calcoaceticus and Escherichia coli. Mikrobiologiia 70:737–744.

Savage D, Barbetti MJ, MacLeod WJ, Salam MU, Renton M. 2012. Seasonal and diurnal patterns of spore release can significantly affect the proportion of spores expected to undergo long-distance dispersal. Microbial ecology 63:578–585.

Schachtman DP, Reid RJ, Ayling SM. 1998. Phosphorus uptake by plants: from soil to cell. Plant physiology 116:447–453.

Scheu S. 1993. Analysis of the microbial nutrient status in soil microcompartments: earthworm faeces from a basalt–limestone gradient. Geoderma 56:575–586. doi:10.1016/0016-7061(93)90136-9.

Schimel J, Balser TC, Wallenstein M. 2007. Microbial Stress-Response Physiology and Its Implications for Ecosystem Function. Ecology 88:1386–1394. doi:10.1890/06-0219.

Schindler DW. 1977. Evolution of phosphorus limitation in lakes. Science 195:260–262.

Schlesinger WH, Bernhardt ES, editors. 2013. Front matter. In: Biogeochemistry (Third Edition). Boston: Academic Press. p. i–ii. [accessed 2017 Feb 13]. http://www.sciencedirect.com/science/article/pii/B9780123858740099854.

Schlesinger WH, Melack JM. 1981. Transport of organic carbon in the world's rivers. Tellus 33:172–187. doi:10.3402/tellusa.v33i2.10706.

Scott T, Cotner J, LaPara T. 2012. Variable Stoichiometry and Homeostatic Regulation of Bacterial Biomass Elemental Composition. Front. Microbiol. 3. doi:10.3389/fmicb.2012.00042. [accessed 2017 Feb 13]. http://journal.frontiersin.org/article/10.3389/fmicb.2012.00042/abstract.

Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, Huber DH, Langenheder S, Lennon JT, Martiny JB, et al. 2012. Fundamentals of microbial community resistance and resilience. [accessed 2017 Apr 24]. https://books.google.com/books?hl=en&lr=&id=p8FpBgAAQBAJ&oi=fnd&pg=PA166&dq=Fundamentals+of+microbial+community+resistance+and+resilience&ots=Qamsk2HJii&sig=7PnZPOfduVNr-dZrpP2HdIoM4o4.

Sharma SB, Sayyed RZ, Trivedi MH, Gobi TA. 2013. Phosphate solubilizing microbes: sustainable approach for managing phosphorus deficiency in agricultural soils. SpringerPlus 2:1.

Sharpley AN, Chapra SC, Wedepohl R, Sims JT, Daniel TC, Reddy KR. 1994. Managing agricultural phosphorus for protection of surface waters: Issues and options. Journal of environmental quality 23:437–451.

Smith SV. 1984. Phosphorus versus nitrogen limitation in the marine environment. Limnology and oceanography 29:1149–1160.

Smouse PE, Long JC, Sokal RR. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic zoology 35:627–632.

Søndergaard M, Jeppesen E. 2007. Anthropogenic impacts on lake and stream ecosystems, and approaches to restoration. Journal of applied ecology 44:1089–1094.

Sridhar KR, Bärlocher F. 1994. Viability of aquatic hyphomycete conidia in foam. Canadian journal of botany 72:106–110.

Sridhar KR, Bärlocher F. 2000. Initial colonization, nutrient supply, and fungal activity on leaves decaying in streams. Applied and environmental microbiology 66:1114–1119.

Stapleton RD, Singh VP. 2002. Biotransformations: Bioremediation Technology for Health and Environmental Protection. Elsevier.

Su J-Q, Ding L-J, Xue K, Yao H-Y, Quensen J, Bai S-J, Wei W-X, Wu J-S, Zhou J, Tiedje JM, et al. 2015. Long-term balanced fertilization increases the soil microbial functional diversity in a phosphorus-limited paddy soil. Mol Ecol 24:136–150. doi:10.1111/mec.13010.

Suberkropp K, Klug MJ. 1976. Fungi and bacteria associated with leaves during processing in a woodland stream. Ecology 57:707–719.

Suberkropp K, Wallace JB. 1992. Aquatic Hyphomycetes in Insecticide-Treated and Untreated Streams. Journal of the North American Benthological Society 11:165–171. doi:10.2307/1467382.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, et al. 2015. Structure and function of the global ocean microbiome. Science 348:1261359. doi:10.1126/science.1261359.

Sutton W, Hansen EM, Reeser PW, Kanaskie A. 2009. Stream monitoring for detection of Phytophthora ramorum in Oregon tanoak forests. Plant Disease 93:1182–1186.

Tapia-Torres Y, Rodríguez-Torres MD, Elser JJ, Islas A, Souza V, García-Oliva F, Olmedo-Álvarez G. 2016. How To Live with Phosphorus Scarcity in Soil and Sediment: Lessons from Bacteria. Appl. Environ. Microbiol. 82:4652–4662. doi:10.1128/AEM.00160-16.

Taylor DL, Bruns TD. 1999. Community structure of ectomycorrhizal fungi in a Pinus muricata forest: minimal overlap between the mature forest and resistant propagule communities. Molecular Ecology 8:1837–1850.

Taylor DL, Walters WA, Lennon NJ, Bochicchio J, Krohn A, Caporaso JG, Pennanen T. 2016. Accurate Estimation of Fungal Diversity and Abundance through Improved Lineage-Specific Primers Optimized for Illumina Amplicon Sequencing. Appl. Environ. Microbiol. 82:7217–7226. doi:10.1128/AEM.02576-16.

Taylor JM, King RS, Pease AA, Winemiller KO. 2014. Nonlinear response of stream ecosystem structure to low-level phosphorus enrichment. Freshwater Biology 59:969–984.

Ternan NG, Grath JWM, Mullan GM, Quinn JP. 1998. Review: Organophosphonates: occurrence, synthesis and biodegradation by microorganisms. World Journal of Microbiology and Biotechnology 14:635–647. doi:10.1023/A:1008848401799.

Thomas K, Chilvers GA, Norris RH. 1991. A dynamic model of fungal spora in a freshwater stream. Mycological Research 95:184–188.

Tilman D, Fargione J, Wolff B, D'Antonio C, Dobson A, Howarth R, Schindler D, Schlesinger WH, Simberloff D, Swackhamer D. 2001. Forecasting Agriculturally Driven Global Environmental Change. Science 292:281–284. doi:10.1126/science.1057544.

Tilman D, Knops J, Wedin D, Reich P, Ritchie M, Siemann E. 1997. The influence of functional diversity and composition on ecosystem processes. Science 277:1300–1302.

Toplis MJ, Libourel G, Carroll MR. 1994. The role of phosphorus in crystallisation processes of basalt: An experimental study. Geochimica et Cosmochimica Acta 58:797–810. doi:10.1016/0016-7037(94)90506-1.

Torsvik V, Øvreås L. 2002. Microbial diversity and function in soil: from genes to ecosystems. Current Opinion in Microbiology 5:240–245. doi:10.1016/S1369-5274(02)00324-7.

Torsvik V, Øvreås L, Thingstad TF. 2002. Prokaryotic Diversity--Magnitude, Dynamics, and Controlling Factors. Science 296:1064–1066. doi:10.1126/science.1071698.

Tu Q, Yu H, He Z, Deng Y, Wu L, Van Nostrand JD, Zhou A, Voordeckers J, Lee Y-J, Qin Y, et al. 2014. GeoChip 4: a functional gene-array-based high-throughput environmental technology for microbial community analysis. Mol Ecol Resour 14:914–928. doi:10.1111/1755-0998.12239.

Uchiyama T, Abe T, Ikemura T, Watanabe K. 2004. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nature Biotechnology 23:nbt1048. doi:10.1038/nbt1048.

Vadstein O, Olsen Y, Reinertsen H, Jensen A. 1993. The Role of Planktonic Bacteria in Phosphorus Cycling in Lakes-Sink and Link. Limnology and Oceanography 38:1539–1544.

Van Der Heijden MG, Bardgett RD, Van Straalen NM. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecology letters 11:296–310.

Van Der Heijden MGA, Horton TR. 2009. Socialism in soil? The importance of mycorrhizal fungal networks for facilitation in natural ecosystems. Journal of Ecology 97:1139–1150. doi:10.1111/j.1365-2745.2009.01570.x.

Vanschoenwinkel B, Gielen S, Vandewaerde H, Seaman M, Brendonck L. 2008. Relative importance of different dispersal vectors for small aquatic invertebrates in a rock pool metacommunity. Ecography 31:567–577. doi:10.1111/j.0906-7590.2008.05442.x.

Vazquez P, Holguin G, Puente ME, Lopez-Cortes A, Bashan Y. 2000. Phosphate-solubilizing microorganisms associated with the rhizosphere of mangroves in a semiarid coastal lagoon. Biol Fertil Soils 30:460–468. doi:10.1007/s003740050024.

Vitousek PM, Porder S, Houlton BZ, Chadwick OA. 2010. Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen–phosphorus interactions. Ecological applications 20:5–15.

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, et al. 2016. gplots: Various R programming tools for plotting data. R package version 3.0.1 2.

Waters TF. 1995. Sediment in streams. American Fisheries Society, Monograph 7.

Wells N, Saunders WMH. 1960. Soil studies using sweet vernal to assess element availability IV. Phosphorus. New Zealand Journal of Agricultural Research 3:279–299. doi:10.1080/00288233.1960.10418084.

Wickham H. 2006. ggplot: An Implementation of the Grammar of Graphics. R package version 2.1.0. [accessed 2016 Oct 27]. http://ftp.auckland.ac.nz/software/CRAN/src/contrib/Descriptions/ggplot.html.

Wood SN. 2001. mgcv: GAMs and generalized ridge regression for R. R news 1:20–25.

Wu L, Thompson DK, Li G, Hurt RA, Tiedje JM, Zhou J. 2001. Development and Evaluation of Functional Gene Arrays for Detection of Selected Genes in the Environment. Appl. Environ. Microbiol. 67:5780–5790. doi:10.1128/AEM.67.12.5780-5790.2001.

Xu X, Thornton PE, Post WM. 2013. A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. Global Ecology and Biogeography 22:737–749. doi:10.1111/geb.12029.

Xu Z, Malmer D, Langille MG, Way SF, Knight R. 2014. Which is more important for classifying microbial communities: who's there or what they can do. ISME J 8:2357–2359.

Zeglin LH. 2015. Stream microbial diversity in response to environmental changes: review and synthesis of existing research. [accessed 2017 Feb 13]. https://books.google.com/books?hl=en&lr=&id=CwO8DAAAQBAJ&oi=fnd&pg=PA247&dq=Stream+microbial+diversity+in+response+to+environmental+changes:+review+and+synthesis+of+existing+research&ots=SuYriN6_MV&sig=btSumHtJdtsirf3bdetxBW2CioU.

Zhou J, He Z, Yang Y, Deng Y, Tringe SG, Alvarez-Cohen L. 2015. High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats. mBio 6:e02288-14. doi:10.1128/mBio.02288-14.

Zinger L, Gobet A, Pommier T. 2012. Two decades of describing the unseen majority of aquatic microbial diversity. Molecular Ecology 21:1878–1896. doi:10.1111/j.1365-294X.2011.05362.x.