WILLIAMS, WILLIAM T., M.S. Biogeography: Do Microbes Have It? (2011)
Directed by Dr. Parke A. Rublee. 43 pp.

The purpose of this study was to use an empirical approach to assess microbial biogeography in freshwater lakes. The targeted, empirical approach was pursued in this work to 1) understand if the census-based methods are under-representing microbes in environmental samples, and 2) to determine if certain microbes are more "transportable" than others, and thus more widespread. Lake samples were collected from 7 clusters of lakes around the world and real-time quantitative PCR was used with 20 microbe-specific primers to assess presence and abundance of each target. The results showed that, for the target microbes, presence in lakes exceeded previous census based estimates that only 15% of microbes are detectable from more than one location. Target microbes were found at an average of 81.7% of sites worldwide, with two present at all lakes. No statistically significant correlation between either the Sørensen Similarity Index or the Bray-Curtis Similarity Index and distance between microbial communities was found in the overall data set. A *post hoc* analysis did find a statistically significant decrease in the Sørensen Index over distances up to 7500 km. However, the regression coefficient was low (1.40×10^{-5}) indicated that this would only reduce the number of species by about 10% over this distance. The results of this work provide support for the "everything is everywhere" hypothesis, suggesting little geographic limitation of microbial distribution. Based on the results, further study with additional targets and an increased number of geographically dispersed lakes is warranted.

BIOGEOGRAPHY: DO MICROBES HAVE IT?

By

William T. Williams

A Thesis Submitted to
The Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2011

Approved by

_____

Committee Chair

To my wife, Kate, without whom this would not have been possible. I love you.

APPROVAL PAGE

This thesis has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____

Date of Acceptance by Committee

ACKNOWLEDGMENTS

I would like to acknowledge the assistance and guidance of all members of my advisory committee: Dr. Parke Rublee (advisor), Dr. Vincent Henrich, and Dr. Scott Richter. I would like to thank Dr. Stan Faeth for allowing the use of his laboratory for a portion of this work and Mike Marshall for his assistance with location and identification of reagents. I would finally like to recognize the University of North Carolina at Greensboro's Department of Biology for providing funding for this work. Funding was also provided by the UNC Competitiveness Research Fund Award and EPA Star Grant RD83162701 Drs. Rublee and Henrich.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

BACKGROUND

Biogeography is a field with considerable historical work and data with respect to macroorganisms. However, it is still an emerging field from a microbial perspective. This is in part due to the fact that appropriate technology was not available until recently to assess the presence of specific microbes beyond easily culturable strains, which comprise only a small percentage of the total microbial community (*cf.* Rappe and Giovannoni 2003). The emerging nature of the field can also be attributed to the long-held paradigm regarding microbes and how they are distributed - namely that their small size and high abundance allow them to be transported by passive means easily and often, thus making them ubiquitous.

Increased study of microbial biogeography over the past decade and the resultant data do not clearly support the ubiquity model nor do they clearly reject it. This has led to a renewed interest in how we view microbes and their biogeography. One recent study suggests that most bacterial species have constrained geographic ranges (Nemergut *et al.* 2011). However, there is also work supporting cosmopolitanism in microorganisms, in which the authors suggest that organisms less than 1 mm in size are freely dispersed across the globe (Finlay and Fenchel 2004).

A major reason for the lack of consensus in microbial biogeography, in addition to limited data, is that a variety of methods are being used, none of which provide an ideal approach. Two recent studies have combined multiple clone libraries of microbial

DNA sequences to look for major patterns over broad geographic scales (Nemergut *et al.* 2011; Newton *et al.* 2011). Newton and colleagues (2011) found a widespread distribution of microbial taxa at the genus level and higher. They also noted that descriptive data is important as the field is still emerging. Nemergut and colleagues (2011) analyzed data from multiple studies consisting of238 bacterial assemblages from terrestrial, aquatic, and symbiotic habitats and found that the majority of species (85%) were only detected in one assemblage. More recently, pyrosequencing has been utilized and it provides a deeper analysis of microbial communities but the number of studies using it have been limited (Sogin *et al.* 2006; Lauber *et al.* 2009). Under-sampling is an issue with clone library approaches, and may be an issue with pyrosequencing, as the number of unique taxa at given site has been shown to be much larger than the libraries represent (Sogin *et al.* 2006). The absence of a universal methodology necessitates consideration of data from multiple techniques when investigating microbial biogeography, at least until better technology is available.

Currently there seems to be two generally accepted models of microbial biogeography. One is the cosmopolitanism theory, whereby microbes may exist anywhere that their environmental requirements are met due to high distribution rates. The other is the moderate endemicity model, which allows for variable distribution of different microbial organisms and places an emphasis on adaptation to the local environment. These two models are the middle ground on a spectrum of potential microbial biogeography characterization, with ubiquity and endemicity models at the extreme ends (Figure 1). Where on this spectrum a microbe lies depends on the relative

importance of characteristics that govern dispersal, survivability, adaptation, and growth, among several factors.

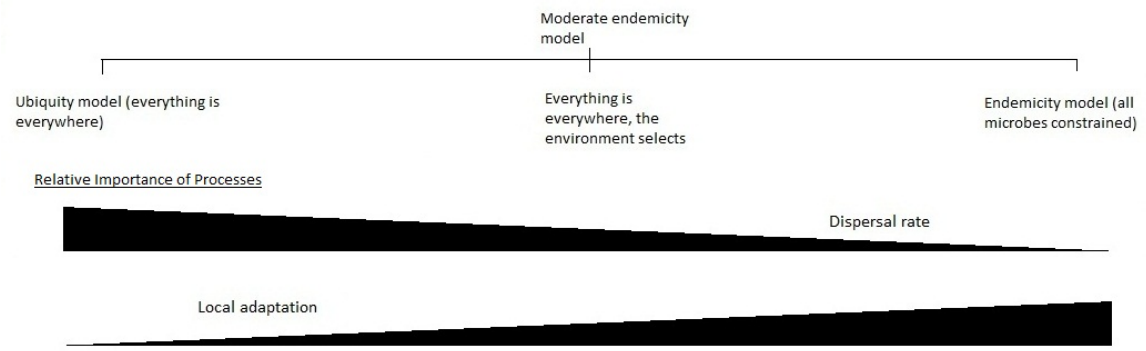

Figure 1: An ideological spectrum of possible models of microbial biogeography. The moderate endemicity model is displayed the above the rest as it allows for application of different models to unique microbes based on characteristics of the organism. The ubiquity model assumes that distribution of microbes is essential universal, and the local adaptation is not a major factor due to continued input of dispersed microbes. The "environment selects" represents the middle ground of the spectrum where dispersal is still significant but local conditions control what microbes exist at a given site. The far right of the spectrum constitutes the endemicity model, which represents a model similar to that of macroorganisms whereby local adaptation is the major controlling factor of species composition and dispersal has much lower importance.

The ubiquity model is a long held paradigm of microbiology. It was proposed first by M.W. Beijerinck and widely accepted throughout the 20th century prior to the advent of molecular methods (Brock 1961). This theory declares that any bacterial species may be found at any location on earth and is fundamentally based on the small size and high population counts of microbes. The crux of this theory is the suggestion that, unlike macro-organisms whose evolutionary and geographic histories define their current distribution, microbes are small and abundant enough that ubiquitous distribution is assumed to be possible in all organisms below a threshold of size, based on high dispersal rates. There is support for this theory in current literature (Fenchel and Finlay

2004). The theory can be neatly summed-up in the first part of L. Baas-Becking's 1934 hypothesis "everything [microbial] is everywhere" (Whitfield, 2005). Protist species present from two sites in Europe was the focus of a study that supports cosmopolitanism of microbes (Finlay and Fenchel 2004). Species present were tabulated by taxonomic identification from samples from the both sites in order to determine if body size provided a correlation for ubiquity. The data analysis suggests that distribution of species is related to body size, such that the smallest species (<1 mm) are near ubiquitous and larger organisms are constrained in distribution. This highlights one of the primary unknowns of biogeographical modeling in microbes which is the broad use of the term "microbe" to include a wide variety of organisms, both prokaryotic and eukaryotic spanning a sizes over three orders of magnitude (from 1 µm to 1 mm). Few studies have examined biogeography of species across the range of microbial sizes.

The endemicity model is in direct contrast with the ubiquity model. This model, although mentioned in literature (Martiny *et al.* 2006) has little support. Research has shown that total endemicity for microbes is highly unlikely (Fenchel and Finlay 2004). However, some microbial species have environmental requirements that are not as common across the globe (e.g. extremophiles, obligate symbionts). The potential for dispersal between sites of hot spring bacteria, for example, is expected to be much lower than for non-extremophiles. This limit in emigration input would, in theory, promote local adaptation and create distinct populations. Research in extremophiles has suggested this to be the case (Whitaker *et al.* 2003). Thus, endemicity may be a reality for some microbes.

A conservative middle-ground model is summarized in the second half of Baas-Becking's statement: "everything is everywhere; *the environment selects*" (Whitfield 2005). The qualifier "the environment selects" importantly suggests that microbes can and do exist ubiquitously so long as the necessary environmental conditions are met. This caveat would imply that the biogeographic range of a microbe is only limited by the availability of suitable habitat and physical conditions necessary for the organism to survive. In other words, similar habitats around the globe may have highly similar communities of those organisms, at least below the aforementioned size threshold.

The other middle-ground model has been "moderate endemicity model," which claims that while the majority of microbes may be ubiquitously distributed, some microbes are not, due to low transportability or low environmental tolerance (Foissner 2007). This model is unique from other models in that it does not lump all microorganisms under a single model of biogeography. Instead, the moderate endemicity model allows for variation among the distribution capabilities of microorganisms due to their wide range of sizes and survivability characteristics. The model also suggests that though transport is important, early colonizers of habitats are at an advantage due to short life cycles and rapid growth. Essentially, this model suggests microbial ecologists look at microbial species on an individual level when describing their distribution rather than assuming all microbes are transported equally, as well as considering how significant an impact immigration of organisms has on a habitat.

Ultimately, which model best describes how a microbe is distributed will be determined by several factors that make microorganisms unique from macroorganisms.

The most significant and obvious differences in microorganisms are the small size rapid growth rates and high abundance compared to macroorganisms. These intrinsic properties of microbial life make the scale of studies incomparable to those of macro-organisms and renders complete censuses of organisms logistically impossible. The properties shared by all microbes (size and abundance) suggest that these organisms can be distributed by passive means (e.g. water flow, wind, macro-organism movement, human activities, etc.) and with great frequency. However, differences among different groups of microbes may affect their relative transportability in a measurable way (Litchman 2010). Whitaker (2006) suggests that organisms which have active motility structures would be more likely to be cosmopolitan and not face distance constraints. Similarly, those organisms that generate resistant endospores or have resting stages would be better suited for passive transportation over distance and thus able to disperse effectively. Additionally, organisms that are closely tied to humans (disease) or human activities (agriculture) could be expected to have cosmopolitan distributions (Whitaker 2006). A recent study looking at migration effects in lakes shows that up to 20% of bacterial variation in highly-linked seepage lakes may be due to migration effects, suggesting that dispersal has some role to play in microbial biogeography (Jones and McMahon 2009).

Our current understanding of microbial distribution stems from our knowledge of macro-organisms and how gene flow works between and within populations. Whitaker (2006) provides a review outlining what is known about microbes and the extent of studies in microbial biogeography. Whitaker frames the topic of microbial biogeography as a topic of allopatric speciation, with the contrast of geographic isolation versus

unlimited dispersal capability shaping the debate. This is one of the major unknowns of microbial biogeography: whether geographic isolation and thus speciation occurs due to continued separation of gene pools in microbes. Whitaker (2006) suggests that an important challenge in determining if geographic distance and genetic distance are related in microbes is to control for other variables that can make comparisons among communities more complex.

The structure of a microbe's habitat and the microbe's relative abundance may control its distribution (Whitaker, 2006). Microbes which are in high abundance and form a significant component of their habitat would be more likely to have a more widespread biogeography due to random chance of a dispersal event. Additionally, greater relative size and frequency of habitat that a microbe may inhabit would also contribute to a wider distribution through chance of an immigration event. If this were to be the case, isolated habitats or those that are rare (*e.g.* hot springs) would be sites of potential speciation.

Recent studies have shown that some microbes face geographic constraints and that immigration is not great enough to yield homogeneous populations among study sites. *Perdinium limbatum* in northern Wisconsin lakes consists of two genetically distinct populations that the authors suggest have resulted from divergent evolution (Kim *et al.* 2004). Genetic distance between the two populations was statistically significant; however, hybrid sequences were detected suggesting that the populations were still experiencing some gene flow. These findings are intriguing considering the geographic scale was relatively small. A study of soil microbes in China using PLFA (phospholipid fatty acid) analysis found that climatologic history may play a role in determining microbial composition (Wu *et al.* 2010). The authors collected samples from three

unique climates and then incubated each at a range of temperatures, analyzing community structure at specific intervals. The results showed that some of the microbial samples incubated at temperatures distant from the samples' respective sites had noticeable changes in PLFA profile compared with those incubated at their native temperature. These findings suggest that some microbial communities may indeed be adapted to their geographic locations and thus not uniform over space.

Several studies have shown that some extremophiles experience some geographic isolation, possibly due to the small and sparse nature of their habitats. Hyperthermophilic archaea of the genus *Sulfolobus* that live in geothermal springs have been shown to be limited in geographic distribution, albeit with very small relative genetic distances (Whitaker *et al.* 2003). Additionally, populations were shown to be similar based on geographic distance between compared populations. A similar study with cyanobacteria from hot springs in Australia, Japan, North America, and Italy found genetically distinct populations (Papke *et al.* 2003). Since the pattern of phylogeny was not fully explainable by physical conditions measured, the authors concluded that as least some of the variation can be attributed to geographic isolation. The prevalence of such distinct microbial populations in hot spring environments can likely be attributed to the small and low frequency of the habitat across the globe, limiting random chance distribution from one population to another. However, there is data that suggest cosmopolitanism exists in some extremophiles. Flagellate diversity in deep-sea hydrothermal vents determined by SSU rDNA sequences suggest the primary taxonomic groups present at these sites are ubiquitous and common in a variety of marine and terrestrial environments worldwide (Atkins *et al.* 2000). It was suggested that broad

environmental tolerance of these species allows for such a distribution. It could be that the contiguous nature and global scale currents of oceans facilitate passive transport of the organisms as well.

There are also recent studies that suggest microbes are not limited in geographic distribution, suggesting cosmopolitanism through high dispersal. *Aspergillus fumigatus* was found to have little genetic variation in and among populations worldwide (Rydholm *et al.* 2006). The lack of detectable genetic variation from this study is likely due to the organism's ubiquity and habitat prevalence (soils). Human infectious strains of the same organism found a common global distribution as well (Pringle *et al.* 2005). A study looking at prokaryotic diversity in two distinct oceanic locations (Mediterranean Sea and Greenland sea) found similarities in community sequence libraries between the two despite geographic distance and environmental differences (Zaballos *et al.* 2006).

It is clear that there is a high potential for dispersal in microbes but whether this potential is reflected in microbial biogeography is not readily obvious. Because of this, several scenarios can be imagined that depict what the distribution of specific microbes might be. Considering dispersal capability and environmental tolerance as major factors controlling microbial distribution, there are four potential distribution patterns for a microbe (Figure 2). In scenarios where dispersal capability is strong, a pattern of presence at most sites should appear (Figure 2: A, B). In scenarios where dispersal capability is low, a pattern of presence at a few clustered sites would be expected depending on the microbe's environmental tolerance (Figure 2: C, D).

Molecular methods for identification of microbial species have become the standard to comprehensively identify the presence of microbes. It has been known for

some time that culturable prokaryotes constitute a small minority of all microbial species (Rappe and Giovannoni 2003). GenBank entry history shows a steady increase of known sequences, highlighting the significant growth in our knowledge of the variety of microbes in (Rappe and Giovannoni 2003; Newton *et al.* 2011). The most common DNA sequence of identification has become the rDNA gene, specifically 16S for prokaryotes and 18S for eukaryotes (Newton *et al.* 2011; Atkins *et al.* 2000). Quantitative PCR approaches can give data about presence or absence, as well as abundance, of a specific OTU in genomic DNA extracted from an environmental sample (Marshall *et al.* 2008). An OTU for rDNA sequences is a general term used to describe a group of organisms with a specific level sequence similarity. This grouping mechanism can be used to roughly describe the microbial presence and/or diversity from a sample are considered to correspond to species, genus, family, and order at 98%, 95%, 92%, and 89% sequence similarity respectively (Nemergut *et al.* 2011).

Research in microbial biogeography has not yielded a consensus model despite increased precision of molecular methods but a clearer understanding can be achieved through more work. This is because there is not a single methodology which reveals the presence of every single organism from a sample. Each sampling technique has weaknesses, and until new technology allows for comprehensive sampling there may be discrepancies among methods. One of the primary difficulties in addressing microbial biogeography is detection of rare taxa in the environment. The number and variety of taxa present in low abundance from a specific site is large (Sogin *et al.* 2006; Ashby *et al.* 2007; Figure 3). This poses a significant hurdle in understanding the nature of

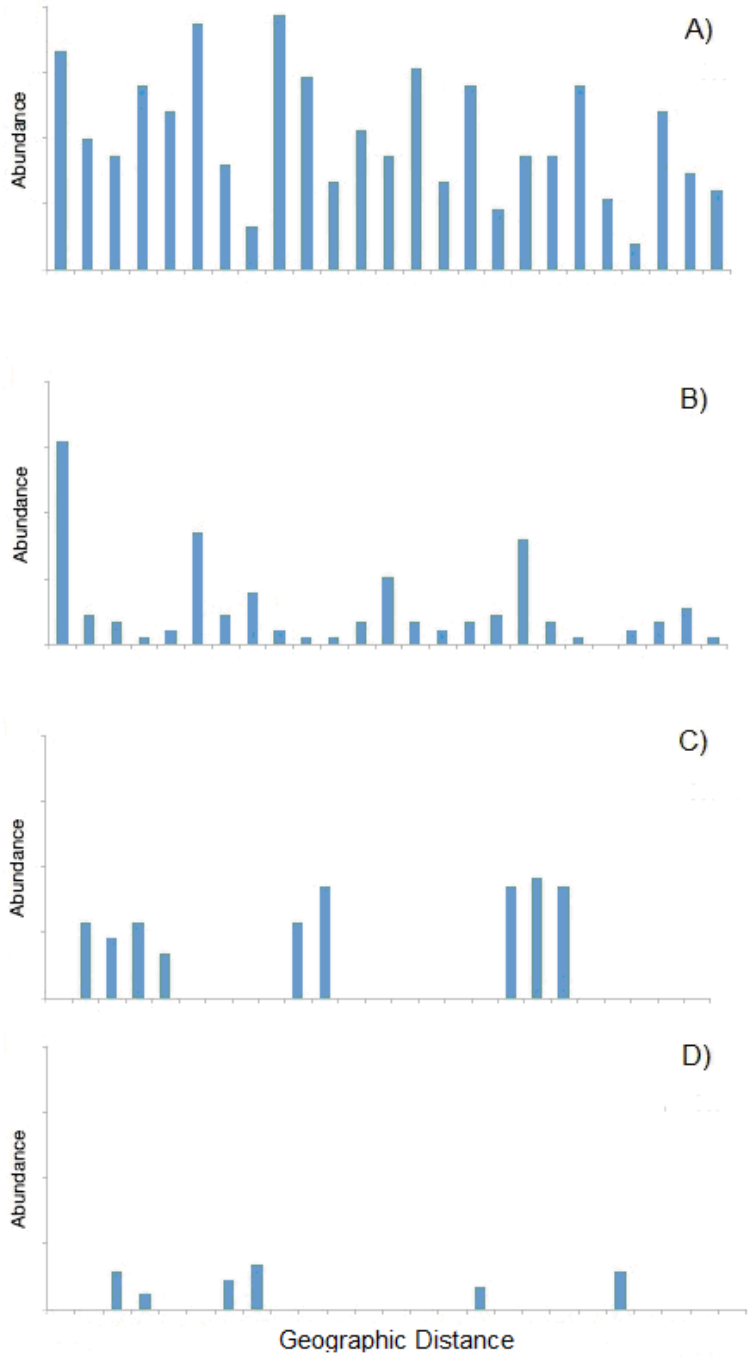

Figure 2: Theoretical distributions of a microbe. Bars represent unique sites and adjacent bars are geographically closer. Scenarios represent: an organism with high transportability and high environment tolerance; high transportability and low environmental tolerance; low transportability and high environmental tolerance; low transportability and low environmental tolerance.

microbial biogeography on a global scale as accurate characterization of numerous sites across a broad scale is resource intensive.

Census methods are commonly used in microbial biogeography studies to characterize community composition (Sogin *et al.* 2006; Nemergut *et al.* 2011). Results from census-based studies have revealed microbial community composition typically follows a negative logarithmic pattern, where few taxa are present in high abundance but many more taxa are present at low abundances. The region of low abundance has been shown to contain large numbers of unique taxa and has been termed the "rare biosphere" (Sogin *et al.* 2006). In some communities it appears that rare biosphere taxa make up the majority of the microbial biomass (Ashby *et al.* 2007). Censusing methods detect the presence of common species accurately. However, as only a limited number of unique sequences are captured from each sample, detection of each rare species is not possible. Thus, census methods must be used with caution in community comparisons between samples, particularly when species of interest are in the rare biosphere in one or more of the samples.

Recent work seeking to directly address biogeographic patterns of bacterial species on a global scale used clone libraries of bacterial 16S rRNA sequences to census from multiple habitat types across the globe (Nemergut *et al.* 2011). The authors analyzed data from multiple studies and habitat types and found that 85% of sequences were found in only one of 238 tested sites, where an average of 118 sequences per site was obtained. Additionally, the 10 most abundant OTUs from each site were detected at an average of 28% of all sites (compared to 2% of sites for all OTUs), suggesting more abundant taxa are more widely distributed or are more commonly detected by the

methods used. The results suggest that most bacterial taxa exhibit a limited distribution within habitat types. The authors state that newer technologies may reveal different patterns of dispersal in microbes (Nemergut *et al.* 2011).

There are other techniques used in recent studies that seek to characterize community structure through census approaches but are likely not fully representing the rare organisms. High throughput sequencing is useful technique for characterizing a community but is based on a limited number of sequences (Sogin *et al.* 2006). Deep sequencing of a sample would be the gold standard for microbial studies if the ability to identify all organisms was possible. However, this is currently cost-prohibitive. Another such approach is ARISA which, again gives strong community data but is not ideal for detecting organisms in low abundance (Fuhrman and Steele, 2008). All of these "shotgun" approaches leave the question of global distribution unanswered due to their limited unique organism counts.

The use of real-time quantitative-PCR is an effective method for detection of specific microbial rDNA sequences from environmental samples. Many microbial studies use PCR amplification of rDNA sequences to identify microbes present at sample sites (*e.g.* Zaballos *et al.* 2006; Pringle *et al.* 2005; Kim *et al.* 2004). as well as its relative abundance (Marshall *et al.* 2008). This allows for the selection of a set of primers for known rDNA sequences and then a comparison of the presence or absence and relative abundance of each organism of interest. Real-time Q-PCR is unique in that

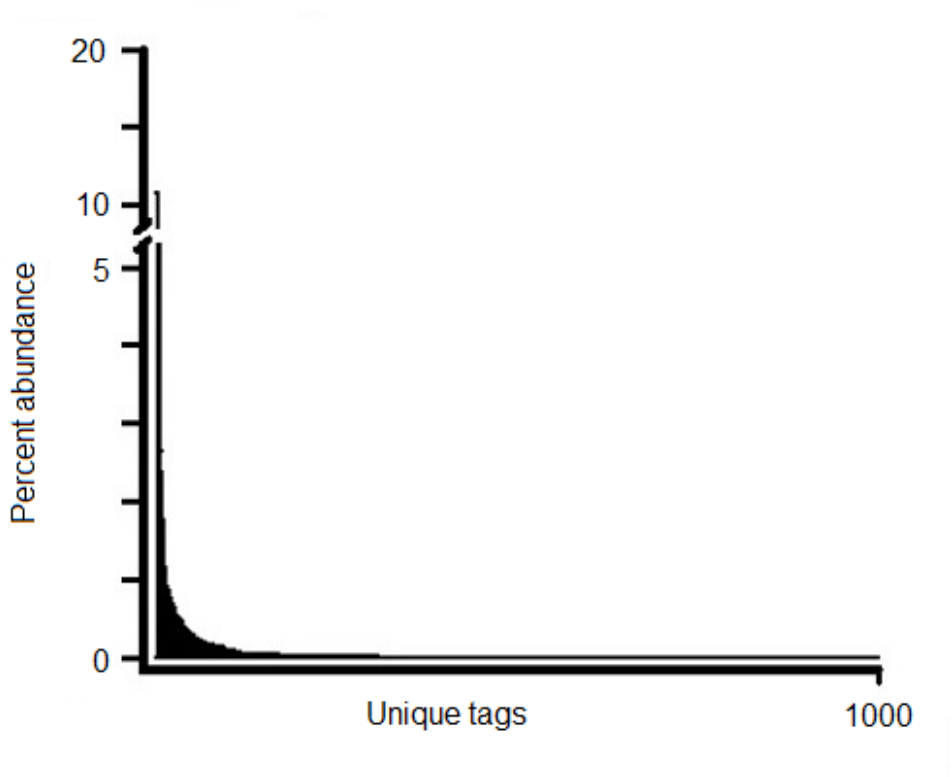

Figure 3: Rank abundance curve (Ashby *et al.* 2007). A rank abundance curve showing typical microbial community structure. One difficulty in addressing microbial biogeography is the detection of the large number of unique taxa in the long tail of a rank-abundance curve, or the "rare biosphere".

it detects the presence of a targeted sequence from samples quantitatively. The sensitivity for targets in low abundance (*i.e.* the rare biosphere) is not diminished compared to abundant targets, as is the case with clone library censusing. The weakness with this technique is the limited number of taxa that can be used to compare community composition among sites.

Objectives and Hypotheses

This study used real-time quantitative PCR to investigate microbial distribution over a broad scale. This study directly addressed a major shortcoming of census techniques which is the ability to consistently detect the presence of taxa in low abundance; thus the advantage of this approach is that if a target is either a common or rare member of the community it should be detected. This allows for a more accurate assessment of the presence of organisms over a global scale. Specifically this study investigated lake samples from locations across the globe and used a selection of microbial target organisms that span a spectrum of size and transportability characteristics to determine:

1) Are target microbial species more broadly distributed than suggested by a recent study (Nemergut *et al.* 2011), who only found 15% of taxa represented in multiple lakes?

Hypothesis 1:

H_o: Most target organisms (≥85%) will be detected in only one sample.

H_a: Most target organisms will be detected in more than one sample.

2) Does distance effect microbial community similarity and species distribution?

Hypothesis 2:

H_o: A correlation will be seen between of community similarity or abundance and distance.

H_a: No correlations will be found between pairwise community similarity indices or abundance and distance.

If the alternative for hypothesis 1 is supported by data, this would suggest that the quantitative PCR detection used in this study is a more accurate molecular method for studying microbial biogeography than current census methods.

If the alternative for hypothesis 2 is supported by data, this would suggest that no distance effect exists for the microbes examined in this study.

CHAPTER II

OUTLINE OF PROCEDURES

Materials and Methods

Sampling of lakes was performed in a non-random manner based on availability (Table 1). Seven clusters of lakes were included so that presence could be measured over both short and long distances. Within each cluster 5 lakes were chosen to span a spectrum of distances and lake types. All samples were collected during summer months. Samples were taken by surface water collection and 100 mL of collected water was passed through GF/F glass fiber filters to retain the microbial community.

Filters were then stored in 1 ml of CTAB (cetyltrimethylammonium bromide) buffer in a 2ml vial at room temperature until extraction. Genomic DNA was extracted by CTAB DNA isolation technique (Schaefer, 1997). Briefly, the glass fiber filters and CTAB were transferred to a 15 mL tube and an additional 1 ml of CTAB buffer was added. The tube was then heated at 65°C for 1 hour . After incubation and slight cooling 2 mL of chloroform/isoamyl alcohol mixture (24:1 ratio) was added and mixed for one minute. The tubes were then centrifuged for 15 minutes at 910 x g. Two aliquots of 850 µL were transferred from the aqueous layer into microcentrifuge tubes and 600 µL of 2-propanol was added to each. The tubes were centrifuged at high speed (20000 x g) at4°C for 25 minutes. Finally, the supernatant was p oured off and the sample was air dried at room temperature followed by resuspension in 20 µL of TE pH 8.0 buffer.

Table 1. Lake samples used in this study

Lake	Cluster	Country	Trophic state
S. China bot. garden pond	1	China	Eutrophic
Shatian SE GZ fish pond	1	China	Eutrophic
GZ 3rd pond	1	China	Eutrophic
Pond by small house	1	China	Eutrophic
Eutrophic farm pond	1	China	Eutrophic
Zellersee	2	Germany	Meso-oligotrophic
Lake Constance	2	Germany	Mesotrophic
Kochelsee	2	Germany	Meso-oligotrophic
Lake Wiritoa	3	New Zealand	Eutrophic
Lake Rotorua	3	New Zealand	Eutrophic
Lake Tarawera	3	New Zealand	Mesotrophic
Karori Lower Lake	3	New Zealand	Eutrophic
Lake Waikare	3	New Zealand	Eutrophic
GTH 65	4	USA – AK	Oligotrophic
GTH 98	4	USA – AK	Oligotrophic
S4	4	USA – AK	Oligotrophic
Lake Tahoe	5	USA - CA	Oligotrophic
Hemet Lake	5	USA - CA	Mesotrophic
Prado Lake	5	USA - CA	Eutrophic
Lake Balboa	5	USA - CA	Eutrophic
Lake Mission Viejo	5	USA - CA	Mesotrophic
Chebacco Lake	6	USA - MA	Mesotrophic
Hamilton Pond	6	USA - ME	Mesotrophic
Upper Hadlock	6	USA - ME	Mesotrophic
Echo Lake	6	USA - ME	Oligotrophic
Lake Marion	7	USA - SC	Mesotrophic
Falls Lake	7	USA - NC	Mesotrophic
Lake Brandt	7	USA - NC	Mesotrophic
Thagard's Lake	7	USA - NC	Mesotrophic
Lake Mackintosh	7	USA - NC	Mesotrophic

Extracted DNA was then quantified using a NanoDrop Spectrophotometer (Thermo Scientific) to determine DNA concentration of each extracted sample. These data were then used to create dilutions to a uniform concentration of 5 ηg/μL for each sample for Q-PCR.

Targets include both prokaryotic and eukaryotic lake microbes (Table 2). Targets were selected to cover a range of transportability characteristics, including size and resistance stages (Litchman *et al.* 2010). Non-native (of lakes) microbial targets are also included in the analysis as they represent high transportability (due to wind transport on dust from surrounding landscapes). Targets were chosen based on the availability of primers from previous work (e.g. Marshall, *et al.* 2008) and known organisms that represented a range of sizes and types with published primer sequences.

Q-PCR was run using and Applied Biosystems StepOne™ real-time PCR system in 48-well or 96-well runs. Reaction mixtures contained 10 μL Power Sybr® Green PCR master mix, 1 μL each of forward and reverse primer, 8 μL of DI water, and 1 μL template at a DNA concentration of ≈ 5 ηg μL^{-1}. Each run included 3 negative controls and 3 concentrations of standards run in triplicate. Samples will be run in duplicate in the remaining wells. The PCR temperature cycle consisted of: 1) an activation step at 95°C for 15 minutes; 2) 40 replication cycles of: 95° C for 15 seconds; an annealing step of 30 seconds to 1 minute at 55-60°C (temperatur e set for specific primer); an extension step 72°C for 1 minute; and fluorescence dete ction at 80°C for 15 seconds; 3) a melt curve step at the end of the PCR run. The relative abundance of targets was computed using standard curves. The melt curves of samples were examined to assure that no amplification was due to a false positive.

Table 2: Target microbial organisms

Organism	Size (µm)	Motility	Resting Stage	Primer Sequence	Source
Characterized microbes					
Burkholderia multivorans	1-3	Yes	No	F 5' AGG CGG TCT GTT AAG ACA - 3' 5'-AGC ACT CCC GAA TCT CTT - 3'	LiPuma 1999
Acinetobacter (genus)	1-3	No	No	F 5' - TTT AAG CGA GGA GGA GG -3' R 5' - ATT CTA CCA TCC TCT CCC -3'	Newton 2011
Zooglea ramigera	1-3	No	No	F 5' - AAC GTA CCC AAG AGT GGG -3' R 5' - AAG GAT ATT AGC CTC TAC CG -3'	Lab data
Bacillus subtilis	3-5	Yes	Spore	F 5'-AAGTCGAGCGGACAGATGG-3' R 5'-CCAGTTTCCAATGACCCTCCCC-3'	Wattiau 2001
Pseudomonas aeruginosa	1-5	Yes	No	F 5 ' - TTCCCTCGCAGAGAAAACATC - 3' R 5 - 'CCTGGTTGATCAGGTCGATCT - 3'	Da Silva 1999
Sulfolobus solfataricus	1	No	No	F 5' - GCT ATC GGG GTG GGG CTA A-3' R 5' - TGG TCG GGA CTC TTA CTG G -3'	Whitaker 2003
Aphanizomenon issatchenkoi	7 x 20-70	No	No	F 5' - GTG GCT AAT ACC GAA TGT GCC GA -3' R 5' - CCC TTW ACG CCC AAT CAT TCC GGA TAA -3'	Lab data
Cryptomonas ovata	13-41	Yes	Cyst	F 5' - TTC AAA CCG GCC TCG TTC TG -3' R 5' - CCC ATA ACC AAC GAA ATA GC -3'	Lab data
Asterionella formosa	60-80	No	Yes	F 5' - ATC GAG TAT CAA TTG GAG GG -3' R 5' - GAC GGG GTC AAT ACA ACG AC -3'	Lab data
Vorticella campanula	100-200	Yes	No	F 5' - AAG ATT AAG CCA TGC ATG TG -3' R 5' - TCC TTG CGG AAT TAG TTT AG -3' and 5' - AAT CAC CTA CCA GGA ATA CC -3'	Lab data
Dileptus	20-200	Yes	No	F 5' - TTA GCG AAT CGT GGC ACG TC -3' R 5' - AAT GTA TTC CTG CAA ACG CC -3'	Lab data
Prochlorothrix hollandica	1 x 3-10	No	No	F 5' - GGA AAC GAC TGC TAA TAC CCG ATG T -3' R 5' - GCC TAC GAA CGC TTT ACG CCC AA -3'	Lab data
Cylindrospermopsis raciborskii	3 x 10-120	No	Akinete	F 5' - GGT GAA AGA TTT ATC GCC TGG AGA TGA - 3' R 5' -GAC TAC WGG GGT ATC TAA TCC CWT T - 3'	Lab data
Microcystis aeruginosa	10-1,000	No	No	F 5' - HCT AAT TGG CCT GRA GAA GAG C -3' R 5' -GAC TAC WGG GGT ATC TAA TCC CWT T - 3'	Lab data

Table 2 continued

Nodularia spumigena	3 x 220	No	Akinete	F 5’ - GTG AAA GGT TAA TCG CCT GAA GGT -3‘ R 5’ -GAC TAC WGG GGT ATC TAA TCC CWT T - 3’	Lab data
Lyngbya wollei	5 x 40	No	No	F 5’ - GAT TAA TTG CCA GAA GAT GAG C - 3’ R 5’ -GAC TAC WGG GGT ATC TAA TCC CWT T - 3’	Lab data
Uncharacterized microbes					
Cyanobacteria LD27				F 5’ - ACA TGC AAG TCG TAC GAG AG -3’ R 5’ - ACA CGT CAT TTA TTC CTC CC -3’	Lab data
Hg 2				F 5’ - AGT CAT CGG CCA CAC CGT GG -3’ R 5’ - AAC TCT AAG GAG ACT GAA GG -3’	Lab data
Hg 7				F 5’ - GTA CAC ACT CTA GCA AAG TG -3’ R 5’ - ACC ATG GTA GGC ATA TCA CC -3’	Lab data
MTBE degrader				F 5’ - GGT AAC AGG TTA AGC TGA CG -3’ R 5’ - CAG AGT ATT AAT CCG AAG CG -3’	Lab data

Data Analysis

The first objective of this study was to determine the geographic distribution of targeted microbes and compare it with distributions from previous work. Through a clone census approach, the majority of sequences detected (85%) were found in only one of 238 assemblages (microbial communities) (Nemergut *et al.* 2011). The hypothesis is that the targeted probing methodology used in this research is more sensitive to the presence of less common (in a relative community assemblage) microbes.

The second objective of this study was to determine if there is a difference in transportability among targets. Two community comparison indices, Sørensen's similarity and Bray-Curtis similarity were utilized to compare lake communities (Bloom 1981). Sørensen's similarity index compares the presence or absence of taxa at two sites and follows the formula:

$$C = \frac{2N}{(S_A + S_B)}$$

where the Sorensen Similarity Index of two sites, C, = 2 times the number of taxa in common at both sites (N) divided by S_A = the total number of taxa at site A, plus S_B = the total number of taxa found at site B. The index ranges from 0 to 1, with 0 representing no commonality and 1 representing identical taxa at each site.

The Bray-Curtis similarity index is based on the abundance of taxa at two sites and follows the formula:

$$BC_{ij} = \sum_{k=1}^{n} \frac{|n_{ik} - n_{jk}|}{(n_{ik} + n_{jk})}$$

where the Bray-Curtis Similarity Index of two sites, i and j, is equal to the sum of the absolute difference in the abundance of each species, k, at the two sites, divided by the sum of the abundances of the species at the two sites. The Bray-Curtis index also ranges from 0 to 1, with 0 representing no commonality and 1 representing identical abundance of taxa at each site.

Additionally, residual target abundances between sites were plotted against geographic distance between sites to determine if abundances of targets show a distance effect.

The experimental setup of this study has statistical limitations that should be mentioned. There was no randomization of sampling or target selection, so p-values of a normal population are not necessarily appropriate. The format of this study is a non-randomized, observational study, and therefore no causation or greater population inferences can be drawn from the results.

CHAPTER III

RESULTS

Of the 20 target organisms in this study, all (100%) were identified by rDNA sequences from more than one lake (Table 3). A binomial test, assuming 0.15 (15%, Nemergut *et al.* 2011) as the predicted rate of targets at multiple lakes, yielded a one-sided p-value less than 0.0001. The binomial test assumes that each target is present or absent independent of the other targets, thus hypothesis 1 was supported. Targets were found at an average of 24.5 of the 30 lakes (81.6%) at detectable levels. Two of the organisms were detectable at every site and the least common organism was found at 10 of the 30 sites. With respect to lakes, two sites (both in the California cluster) displayed presence of all 20 organisms; the lowest number of targets detected in a lake was 9 of 20.

In addition to presence, the abundance was computed for the 17 organisms for which there were available standards. The abundance measures computed were in ng/ml of "target" DNA but must be considered relative values since the standards in some cases were cloned amplicons and in some cases were genomic DNA form pure cultures. Thus, between-organism comparisons were not appropriate (Table 4). There was broad observed variability in each organism's abundance among sites, ranging up to 7 orders of magnitude difference from the lowest detectable presence to the highest (Table 5), suggesting both broad dispersal and broad range of environmental tolerances as in Figure 2B.

Table 3: Target Presence

	Burkholderia multivorans	AclB1	*Zooglea ramigera*	*Sulfolobus solfataricus*	*Aphanizomenon issatchenkoi OTU 8*	*Cryptomonas ovata*	*Asterionella formosa*	*Vorticella campanula*	*Dileptus*	*Prochlorothrix hollandica OTU 1*	*Cylindrospermopsis raciborskii OTU 2*	*Microcystis aruginosa*	*Nodularia spumigena OTU 90*	*Lyngbya wollei*	LD27	Hg 2	Hg 7	B. subtilis	P. aeruginosa	MTBE	Total	Percent of Targets Present
S. China bot. garden pond	1	1	1	1	0	1	0	1	1	1	1	1	0	1	0	1	0	1	1	1	15	0.75
Shatian SE GZ fish pond	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	18	0.9
GZ 3rd pond	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	18	0.9
Pond by small house	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	19	0.95
Eutrophic farm pond	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19	0.95
Zellersee	0	1	1	0	1	0	1	1	1	1	0	1	0	1	1	1	0	1	0	1	13	0.65
Lake Constance	1	1	1	0	1	1	0	0	1	1	1	1	0	1	1	0	0	1	1	0	13	0.65
Kochelsee	1	1	1	0	1	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	9	0.45
Lake Wiritoa	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	19	0.95
Lake Rotorua	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	18	0.9
Lake Tarawera	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	18	0.9
Karori Lower Lake	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	18	0.9
Lake Waikare	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	19	0.95
GTH-98	1	1	0	1	1	1	1	1	0	1	1	1	0	0	1	1	0	1	0	0	13	0.65
S-4	1	1	1	1	1	1	0	1	1	1	0	1	0	1	1	1	0	1	0	1	15	0.75

Table 3 continued

	Burkholderia multivorans	AclB1	*Zooglea ramigera*	*Sulfolobus solfataricus*	*Aphanizomenon issatchenkoi OTU 8*	*Cryptomonas ovata*	*Asterionella formosa*	*Vorticella campanula*	*Dileptus*	*Prochlorothrix hollandica OTU 1*	*Cylindrospermopsis raciborskii OTU 2*	*Microcystis aruginosa*	*Nodularia spumigena OTU 90*	*Lyngbya wollei*	LD27	Hg 2	Hg 7	B. subtilis	P. aeruginosa	MTBE	Total	Percent of Targets Present
GTH-65	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	18	0.9
Lake Tahoe	1	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1	0	1	0	0	14	0.7
Hemet Lake	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	1
Lake Balboa	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	1
Lake Mission Viejo	1	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	0	14	0.7
Prado (CA-9)	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	18	0.9
Chebacco Lake	0	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	15	0.75
Upper Hadlock	1	1	0	0	1	1	0	0	1	1	1	1	0	1	1	0	0	1	0	0	11	0.55
Echo Lake	1	1	1	0	1	1	0	1	1	1	1	1	0	1	1	0	0	0	0	0	12	0.6
Hamilton Pond	1	1	1	1	1	1	0	1	1	1	0	1	0	1	1	1	1	1	0	0	15	0.75
Lake Marion	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	17	0.85
Falls Lake	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	19	0.95
Lake Brandt	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	18	0.9
Thagard's Lake	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	17	0.85
Lake Mackintosh	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	18	0.9
Total	*28*	*30*	*26*	*17*	*29*	*29*	*19*	*27*	*29*	*29*	*26*	*30*	*11*	*28*	*29*	*26*	*10*	*29*	*18*	*20*	*16.33*	*28*

Table 4: Target Abundances (ng/ml)

	P. hollandica (o.t.u. 1)	C. raciborskii (o.t.u. 2)	Aphanizomenon (o.t.u. 8)	Nodularia (o.t.u. 90)	Zooglea ramigera	LD27	C. ovata	Dileptus	Microcystis	B. subtilis	P. aeruginosa	Hg 7	MTBE	Vorticella	Hg 2	Asterionella
S. China bot. garden pond (CN 1)	3.63E-05	5.70E-08	0	0	3.11E-03	0	2.98E+00	1.55E-02	2.85E-03	3.35E-02	2.21E-03	0	4.48E-05	1.39E+00	9.21E-02	0
Shatian SE GZ fish pond (CN 4)	5.11E-03	4.99E-05	1.92E-01	0	2.31E-03	2.67E-01	9.85E+01	4.91E-01	5.31E-01	1.34E+00	9.11E-03	0	7.61E-03	1.00E+01	9.11E-01	2.37E-03
GZ 3rd pond (CN 5)	6.75E-02	6.86E-04	8.79E+00	2.31E+01	4.54E-03	2.34E-01	3.15E+02	9.54E-01	1.99E+00	1.23E+00	1.25E-02	0	2.74E-03	5.82E+00	2.00E-01	1.87E-04
Pond by small house (CN 6)	7.90E-04	6.26E-06	3.65E-02	2.90E-01	5.91E-04	3.71E-02	5.43E+01	6.00E-02	1.91E-02	8.35E-02	5.08E-04	4.85E-02	2.28E-04	8.97E-01	5.13E-01	0
Eutrophic farm pond (CN 7)	8.28E-04	7.31E-06	9.50E-01	8.58E-02	2.05E-05	2.23E-03	3.06E+01	3.20E-01	1.23E+00	5.29E-01	3.50E-03	1.35E-01	1.05E-03	1.05E+01	3.41E+00	9.59E-04
Zellersee	5.24E-06	0	4.58E-04	0	1.25E-05	3.06E-05	0	9.21E-05	2.00E-01	2.79E-02	0	0	3.45E-04	1.99E-01	2.30E-02	2.37E-03
Lake Constance	2.02E-05	8.94E-07	2.90E-02	0	4.23E-04	1.39E-03	3.64E-02	6.37E-04	7.15E-04	2.70E-03	2.81E-04	0	0	0	0	0
Kochelsee	0	0	4.02E-04	0	1.45E-03	1.80E-05	6.41E-03	1.16E-04	3.45E-05	2.23E-04	0	0	0	0	0	0
Lake Wiritoa	1.54E-02	5.94E-06	1.69E+01	0	3.51E-01	1.38E-02	2.87E+01	5.99E-02	8.43E-01	7.52E-02	1.15E-03	5.43E-02	4.69E-05	1.07E+00	5.95E-02	2.97E-03
Lake Rotorua	8.09E-04	1.61E-05	6.67E-01	2.84E-02	9.52E-02	7.39E-02	2.47E+01	2.33E-01	2.57E-01	2.99E-02	0	9.28E-04	2.27E-04	1.43E+00	6.60E-03	4.07E-03
Lake Tarawera	4.62E-04	2.25E-06	1.11E+00	0	5.13E-03	5.47E-04	2.01E-01	2.56E-01	1.65E-01	4.11E-03	1.83E-04	0	1.61E-05	1.95E-01	1.50E-02	3.88E-04
Karori Lower Lake	4.69E-03	1.07E-04	1.00E+03	0	5.33E-03	4.07E-02	4.86E+01	1.45E-01	4.81E-01	2.68E-01	8.41E-03	0	5.21E-04	6.67E-01	4.91E-03	1.27E-03
Lake Waikare	8.66E-02	1.64E-02	8.73E+00	9.55E-01	1.19E-02	5.56E-03	2.93E+01	7.09E-01	2.69E+01	5.25E-01	7.01E-03	0	4.63E-04	4.96E-01	1.11E-02	2.54E-03

Table 4 continued

	P. hollandica (o.t.u. 1)	C. raciborskii (o.t.u. 2)	Aphanizomenon (o.t.u. 8)	Nodularia (o.t.u. 90)	Zooglea ramigera	LD27	C. ovata	Dileptus	Microcystis	B. subtilis	P. aeruginosa	Hg 7	MTBE	Vorticella	Hg 2	Asterionella
GTH-65	9.98E-06	9.33E-06	1.10E-02	0	2.76E-05	6.26E-05	2.64E-01	1.14E-02	5.59E-03	1.10E-02	5.10E-04	5.43E-04	1.34E-03	3.23E-01	1.20E-04	1.24E-05
GTH-98	4.08E-06	2.01E-05	1.07E+00	0	0	1.75E-04	5.68E-02	0	3.88E-02	1.11E-02	0	0	0	4.98E-01	3.22E-04	1.29E-04
S-4	1.01E-07	0	9.89E-02	0	5.86E-05	8.52E-06	3.47E-03	9.74E-05	6.68E-03	3.62E-04	0	0	1.66E-04	1.93E-02	3.53E-05	0
Lake Tahoe	4.65E-06	1.72E-08	1.73E-03	0	9.98E-06	1.26E-04	1.14E-04	4.75E-03	2.06E-04	8.63E-05	0	0	0	1.66E-03	1.07E-06	0
Hemet Lake	6.04E-01	9.52E-04	8.54E+02	8.43E-01	6.16E-04	8.62E-02	1.37E+03	2.59E-01	1.49E+02	2.68E-01	3.53E-03	2.13E+00	2.17E-03	5.81E+00	1.85E+00	4.66E-03
Lake Balboa	9.16E-04	1.86E-05	3.63E-02	1.51E+00	9.32E-04	1.26E-01	4.21E+02	2.86E-01	2.33E-03	4.05E-01	2.31E-03	9.35E-04	5.91E-04	7.49E+00	5.34E-02	9.92E-04
Lake Mission Viejo	1.14E-04	5.23E-07	5.61E-03	0	0	2.81E-04	9.83E+00	6.84E-03	1.34E-03	2.83E-02	0	0	0	5.70E-01	4.74E-01	6.85E-04
Prado (CA-9)	2.91E-02	1.23E-03	7.94E+00	1.97E+01	3.74E-02	1.50E+00	1.94E+05	4.93E+00	1.39E+01	4.51E+00	9.41E-02	0	9.90E-03	2.17E+01	9.71E-02	0
Chebacco Lake	1.39E-04	2.22E-06	3.32E-01	0	2.01E-05	1.70E-04	1.88E+00	4.90E-02	5.16E-02	1.60E-02	0	9.37E-04	0	5.87E-02	6.52E-04	4.39E-04
Upper Hadlock	1.38E-05	4.07E-06	1.73E-02	0	0	1.51E-03	1.81E+00	5.49E-03	1.64E-03	8.27E-03	0	0	0	0	0	0
Echo Lake	7.95E-06	2.00E-06	1.68E-02	0	2.14E-04	4.26E-05	3.47E-01	4.03E-03	2.31E-02	0	0	0	0	6.26E-05	0	0
Hamilton Pond	4.26E-05	0	4.78E-03	0	2.34E-03	2.90E-04	3.03E+01	2.83E-02	2.69E-03	8.88E-02	0	5.47E-02	2.02E-04	5.51E-01	2.62E-03	0
Lake Marion	4.10E-05	6.09E-06	1.12E-02	3.36E-02	0	1.98E-03	1.54E+01	2.67E-01	4.53E-02	2.04E-02	0	8.62E-03	0	4.26E-01	2.87E-01	5.63E-05
Falls Lake	5.20E-02	2.35E-02	1.94E+01	3.22E+01	4.63E-05	8.45E-03	3.91E+01	1.31E-01	8.78E-01	5.24E-01	6.28E-03	0	9.69E-04	4.62E+00	9.21E-02	1.17E-01
Lake Brandt	1.16E-02	5.55E-03	7.22E+00	5.32E+00	9.95E-05	1.16E-01	2.56E+01	3.18E-01	7.02E-01	5.11E-01	3.98E-03	0	1.52E-03	1.07E+01	1.38E-01	3.07E-01
Thagard's Lake	2.49E-04	3.23E-06	4.55E-01	0	4.37E-04	4.03E-04	1.09E+01	1.49E-01	4.30E-01	8.00E-02	2.53E-03	0	1.28E-03	1.19E+00	1.20E-02	5.06E-05
Lake Mackintosh	3.80E-03	8.97E-02	5.11E+01	7.13E-01	8.99E-04	2.21E-03	1.22E+01	6.48E-02	2.21E-01	2.06E-01	3.94E-03	0	4.15E-04	2.47E+00	6.32E-03	5.41E-03

Table 5: Abundance Statistics (ng/mL of target DNA)

	Mean (SE)	Range
Zooglea ramigera	0.0175±0.012	0-0.3508
Bacillus subtilis	0.3609±0.156	0-4.5069
Pseudomonas aeruginosa	0.0054±0.0031	0-0.0941
Aphanizomenon issatchenkoi	66.0472±42.9696	0-1001.8846
Cryptomonas ovata	6564.5799±6476.1997	0-194369
Asterionella formosa	0.0151±0.0108	0-0.3067
Vorticella campanula	2.9701±0.8914	0-21.6974
Dileptus	0.3252±0.1639	0-4.9279
Prochlorothrix hollandica	0.0295±0.0202	0-0.6042
Cylindrospermopsis raciborskii	0.0046±0.0031	0-0.0897
Microcystis aeruginosa	6.5939±5.0049	0-148.9049
Nodularia spumigena	2.8241±1.4184	0-32.2154
LD27	0.0842±0.0505	0-1.5043
Hg 2	0.2753±0.1281	0-3.4131
Hg 7	0.081±0.0707	0-2.1256
MTBE Degrader	0.0011±0.0004	0-0.0099

Using organism presence in a binary matrix (Table 3), the data were then used to calculate Sørensen's similarity coefficient (Bloom 1981) for every pair-wise comparison among lakes. These coefficients, ranging from 0 (no species in common) to 1 (all species present in both sites) representing a presence-based measure of community similarity, were plotted against pair-wise geographic distances between equivalent sites (Figure 4). A linear regression found no statistically significant correlation between

Sørensen Similarity and distance. Log transformations of both variables individually, and then together, yielded no statistically significant correlations, nor did converting either or both variables to non-linear ranks.

The Bray-Curtis Similarity value for each pair of samples was also calculated and plotted vs. geographic distance between samples (Figure 5). The Bray-Curtis index incorporates both presence and abundance of organisms at each site (Bloom 1981). As with the Sørensen Index, no statistically significant correlation was found between the Bray-Curtis index and distance. Log transformations and rank calculations were performed as with the Sørensen data, but no statistically significant correlations were found.

In order to compare dispersal of individual targets, residual abundances (the difference in abundance of a microbial taxon between two lakes) were calculated with all pair-wise site comparisons for each unique target, for a total of 17 residual matrices. Each of these sets of residuals was plotted against the pair-wise distance data between lakes. Regression fits were created for each plot and these plots generally had near-zero or slightly negative slopes, with no statistically significant correlations (Figure 6A). Perhaps surprisingly, the remaining regressions with non-statistically non-significant trends showed slightly negative slopes (Figure 6B).

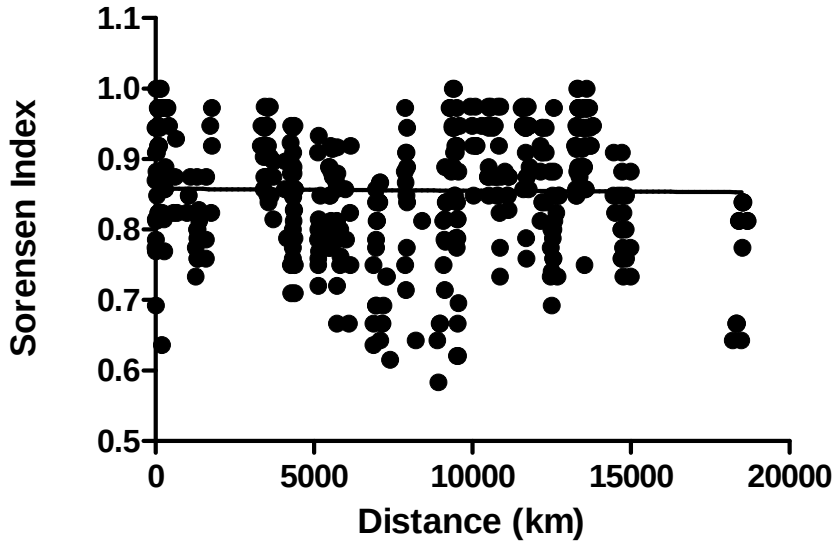

Figure 4: Sørensen index vs. Geographic distance. All pair-wise comparisons between lakes are present. The regression shows non-statistically non-significant correlation (p-value = 0.75; r-value =0.00023).

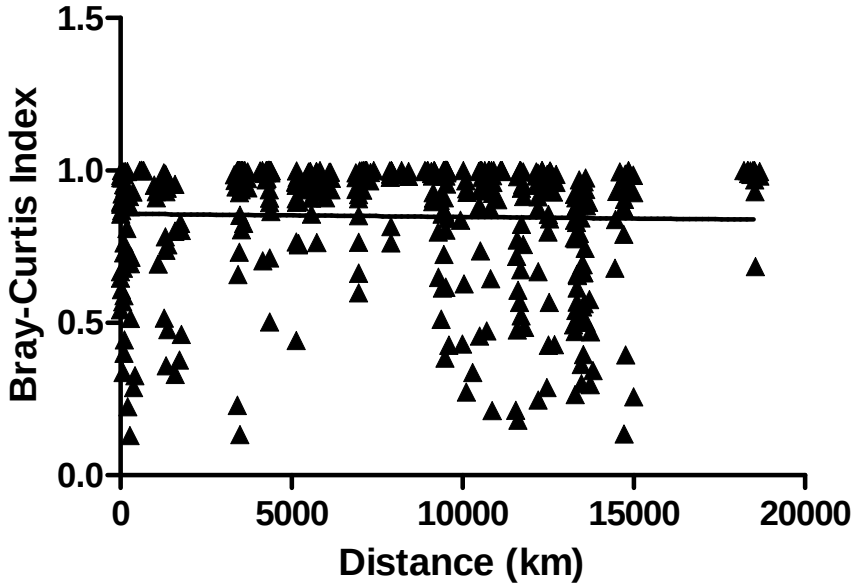

Figure 5: Bray-Curtis index vs. Geographic distance. All pair-wise comparisons between lakes are present. The regression shows non-statistically non-significant correlation (p-value = 0.622, r-value = 0.00056).

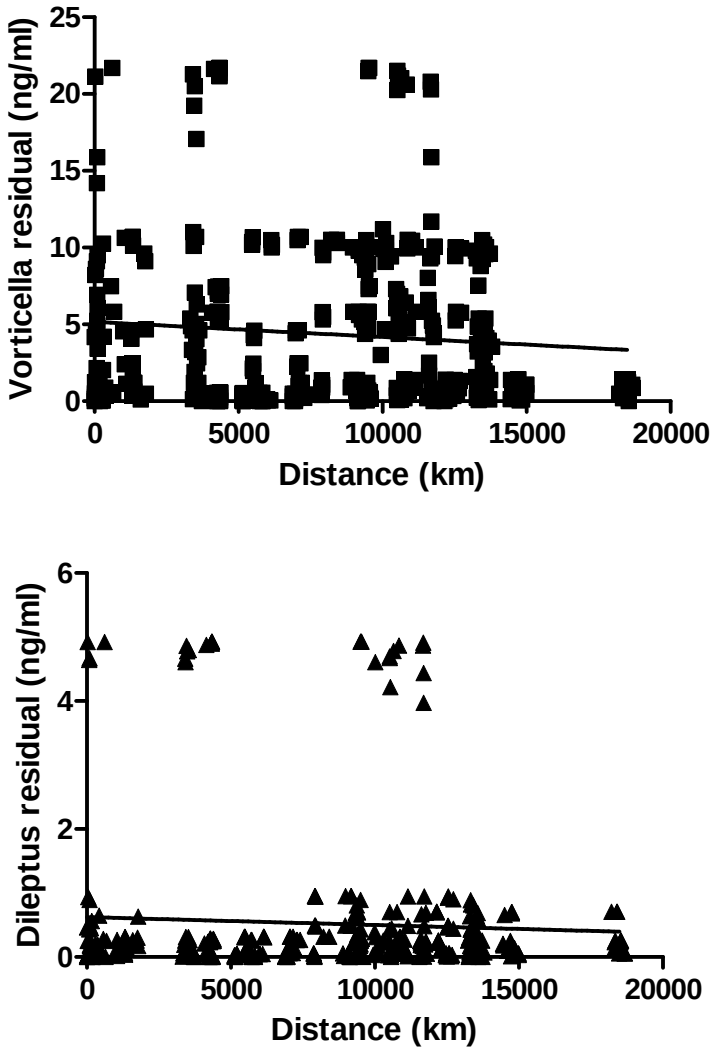

Figure 6 Typical residual abundance (ng/ml) vs. distance plots. No regressions of residuals vs. distance were statistically significant. The upper plot is an example of a suggestive but non-statistically non-significant p-value (0.059; r-value = 0.008). The lower plot is an example of a high, non-statistically non-significant p-value fit (p-value = 0.276; r-value = 0.003).

To investigate potential patterns of regional grouping, both similarity indices were plotted with only within-cluster comparisons. These comparisons appeared to have no detectable difference in variation from inter-cluster comparisons (Figure 7), except for the Carolinas cluster, which showed an apparent linear decrease. A linear regression

was fit onto these data and the only statistically significant correlation was a negative slope for Sørensen index versus distance for the Carolinas cluster (slope = -0.001, p-value = 0.001). All other clusters yielded non-statistically non-significant correlations and were generally scattered similar to the overall data.

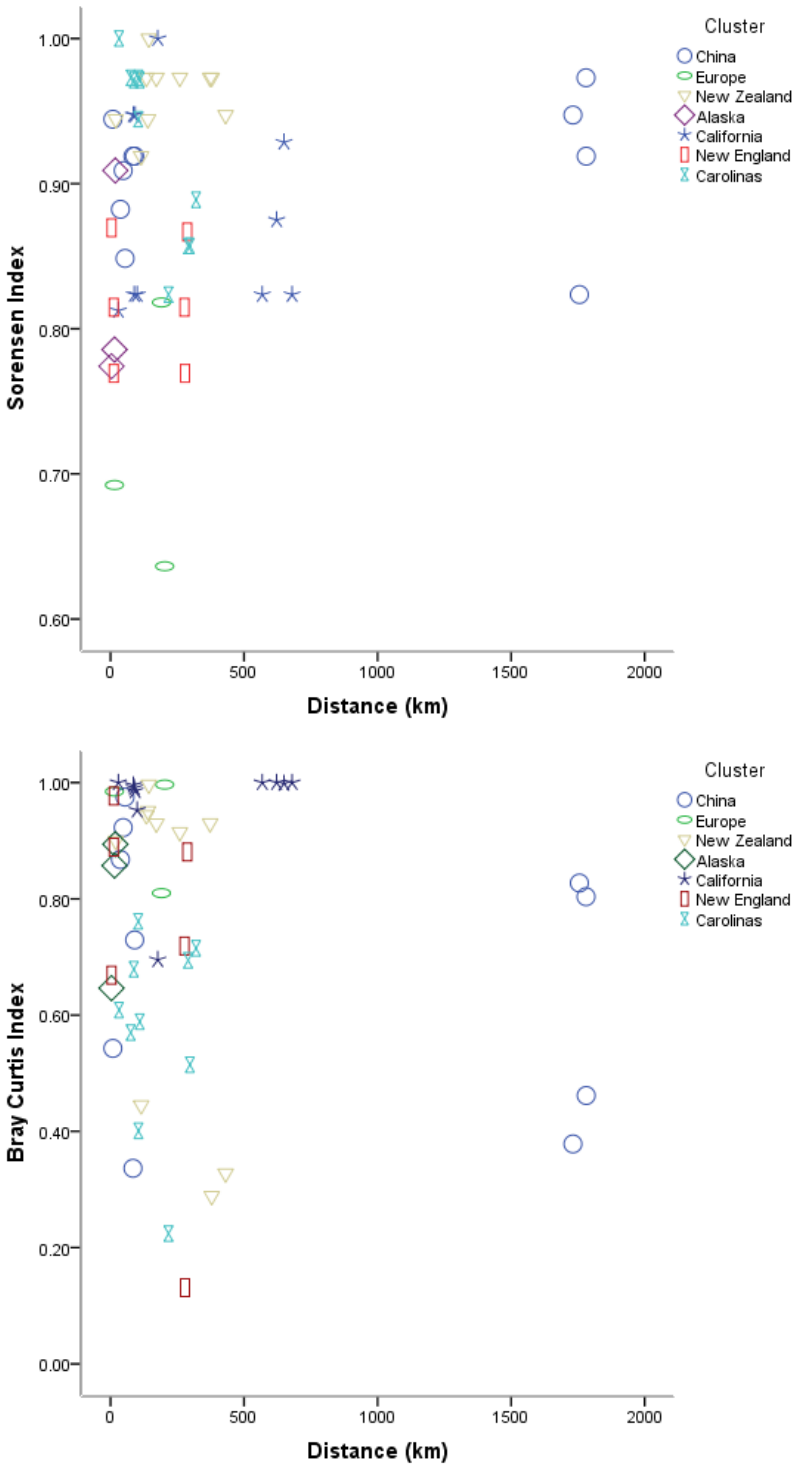

Figure 7: Indices vs. distance, within cluster comparisons. The plots show only within cluster comparisons of Similarity Indices versus distance.

CHAPTER IV

DISCUSSION

This study explored the topic of microbial biogeography with the novel approach of targeting species in lakes across a large geographic scale. The resulting data suggest that, at least for the targets chosen in this study, microbial organisms are broadly distributed and show little to no dispersal limitation. Experimental Hypothesis 1 was supported by result that 100% of targets were found in multiple lakes (81.67% of lakes on average), supporting the alternative hypothesis. These results contrast with data obtained from community profiling (census) studies, which have suggested individual taxa are restricted to one or a few sites and that only 15% of taxa were found in multiple sites (Nemergut *et al.* 2011).

Microbial target used in this study were characterized for organism size, as well as motility (flagella, cilia) and resistance stages (spore, cyst, akinete) (Table 2). These qualities were used to determine if they can be correlated to a certain organisms' distribution. However, no correlation with size could be found. Some of the most abundant targets had larger cell sizes (*Cryptomonas ovata*, *Vorticella campanula*, *Microcystis aeruginosa*). Additionally, some of the smallest were not widely distributed (*Pseudomonas aeruginosa*, *Zooglea ramigera*). These data provide no evidence of cellular size affecting an organism's distribution. One suggestive piece of evidence, however, is that two organisms with cyst or spore stages (*C. ovata* and *Bacillus subtilis*, respectively) were two of the more common targets (Table 3). As these were the only

targets that utilize these resistance stages little can be concluded from their prevalence. Nevertheless, it might prove a prevalent trend if more similar microbes were assayed for.

Overall, abundances of targets that might be expected to be cosmopolitan or limited generally were seen to be so. For example, *Acinetobacter* AcIB1, a common lakewater microbe and that has been suggested by other studies to be widely distributed (Newton 2011), was found in 100% of lakes in this study. Cyanobacteria LD27 was also expected to be widely distributed (29 of 30 lakes), as it was an organism found in multiple lakes from a clone library study (Zwart et al. 1998; GenBank accession number AJ007876; Mike Marshall, UNCG, personal communication). Microbes expected to be limited in range were the potential bioindicators Hg 2 (found in 26 of 30 lakes), MTBE degrader (found in 20 of 30 lakes), and Hg 7 (found in 10 of 30 lakes), as well as the thermophilic Archaea *Sulfolobus solfataricus* (found in 17 of 30 lakes). The bioindicators were expected to be limited based on the presence of their associated contaminant, while *Sulfolobus* was expected to be limited due to its primary habitat being hot springs. Only Hg 2 showed above-average distribution, but was present in relatively low abundances.

The overall results of this study provide no evidence for a distance effect on the similarity of microbial communities. Experimental hypothesis 2 was not supported by data, which showed no linear or non-linear correlations with distance. This suggests that some microbes may fit into Baas-Becking's "everything is everywhere" model. Although no statistically significant correlations were found in the planned analyses, a *post hoc* analysis of Sørensen's index versus distance using 7500 kilometers as a maximum cut-off yielded a statistically significant correlation (slope = $-1.395\ e^{-5} \pm 2.265\ e^{-6}$; p-value

<0.0001; Figure 8). Considering the inherent biases and as this was not a planned cut-off criterion, the relevance of this should not be given too much weight. It does, however, provide a scale for future analyses to explore. The distance of 7500 kilometers suggests that at continental scales and possibly even between continents there may be a slight distance effect. However, the magnitude of this perceived effect is quite small; only around a 10% change in similarity would be expected over a continental scale. Plots of both similarity index values limited to within cluster comparisons (Figure 7) suggest slight within-cluster grouping effects on similarity. For example, California lakes and some New Zealand lakes showed high levels of similarity. The variability within other clusters was very broad. A linear decrease in Sørensen index with distance was seen in the Carolinas cluster. It is unclear why only this cluster and no others showed such a trend.

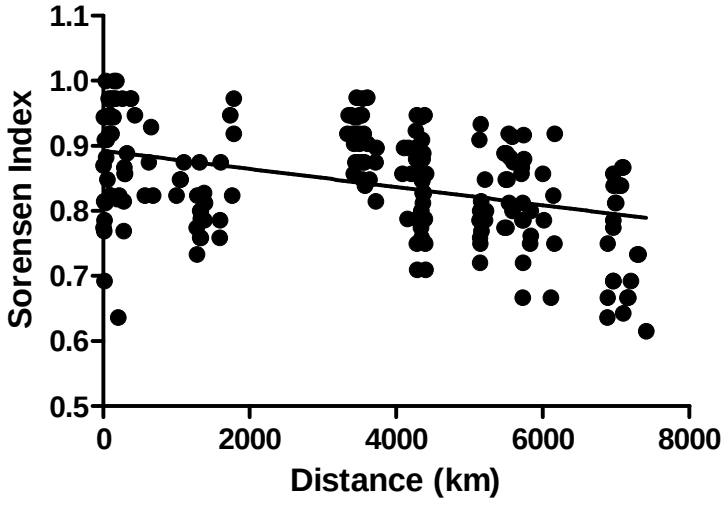

Figure 8: Sørensen index vs. Geographic distance 0-7500 km. All pair-wise comparisons between lakes are present. The regression shows a statistically significant correlation with a slope of -1.395×10^{-5} (p-value <0.0001; r-value =0.1621).

Several limitations of this study preclude broad generalizations. First, the selection of lakes was done by convenience, so there is no implication that these lakes represent a random sampling of world lakes. Samples were obtained based on availability of access (local lakes) or ability to obtain samples (distant lakes) from willing colleagues. Additionally, all lake samples were taken during the summer season (respective to locale) at all sites but not during the same year, as sampling dates span several years. It was not feasible to sample all lakes during the same year, but given the data, it would have been unlikely that this would have shown correlations beyond what was found. Further, the limited availability of lake samples from some clusters, and the fact that not all samples had intact microbial rDNA after extraction left some clusters with fewer than five representative lakes.

Second, a limited range of microbial species were selected as targets for this study. Since many of the targets were identified from census-based studies of local lakes (Marshall 2008) and several common cyanobacteria (unpublished data), the "community" of chosen targets could have been biased toward more common lake microbes, similar to clonal census studies. These include *Asterionella formosa, Vorticella campanula, Dileptus,* LD27, *Zooglea ramigera, Cryptomonas ovata*; and the cyanobacteria: *Aphanizomenon issatchenkoi, Prochlorothrix hollandica, Cyllindrospermopsis raciborskii, Microcystis aeruginosa, Nodularia spumigena,* and *Lyngbya*. The remaining targets were derived from a study looking at bioindicators: Hg 2, Hg 7, and MTBE; (Mike Marshall, UNCG, personal communication); and targets from literature which were selected for various characteristics: AcIB1 (Acinetobacter genus, cosmopolitan microbe); *Burkholderia multivorans* and *Pseudomonas aeruginosa*

(common aquatic microbes); *Sulfolobus* (thermophile, found in hot springs); and *Bacillus subtilis* (common soil microbe). The selection of targets was, as with the lakes, based on opportunity and convenience: the availability of primers as well as quantifiable standards weighed heavily in target selection. A broader selection of targets that would include some less-abundant components of lake communities might show some microbes to be more restricted than most (as with Hg 7 in this study).

Finally, the sampling method of this study was limited. As previously stated, all sampling was done during summer months, which potentially biased the data toward prevalence of the cyanobacterial species, which represented 30% of the target organisms. Sampling from year-round time points might show some of these targets to be more or less common at different seasonal time points. Second, the water samples consisted of only 100-200 mL of epilimnetic water from near-shore sites. It is unclear whether increasing or decreasing volume filtered would significantly change the results. However, the fact that more than one site contained all target species rDNA supports the suitability of the volumes this study used.

Statistically, this study is limited by the method of non-random sampling or assignment. However, randomized selection of lakes was not feasible because samples were “samples of opportunity”. Additionally, while some analyses suggested correlations of decreased similarity with distance, the biological importance of these statistically significant correlations is unclear.

This study used quantitative PCR to compare microbial assemblages across geographic distance rather than censusing methods as used in most other recent studies. The results show microbes are statistically more widespread than census data

suggest (p-value <0.05). There was no substantial evidence of a distance effect in the targets selected for this study. The results suggests that the "everything is everywhere" model of microbial dispersal might apply to at least some microbes, a result which contrasts with the results of other recent studies (Whitaker *et al.* 2003; Papke *et al.* 2003, Nemergut, *et al.* 2011). In fact, recent work using a method with a known bias towards common species supports the lack of a distance relationship over short distance (hundreds of meters) (Bell 2010).

Although this study cannot be viewed as definitive (large sample sizes of targets and lakes would be needed), nevertheless, the results are compelling and strongly suggest that microbes are widely distributed. Future studies should attempt to assay for a wider range of microbial targets and lake characteristics that may provide more statistical power to argue that results apply more broadly. A random selection of targets should be utilized, which would serve to better represent rare taxa. Given the *post hoc* analysis correlation of Sørensen similarity at distances up to 7500 kilometers, a larger selection of lakes in this range of distances would potentially reveal more about this continental-scale phenomenon.

REFERENCES

Ashby MN, Rine J, Mongodin EF, Nelson KE, Dimster-Denk.. 2007. Serial analysis of rRNA gens and the unexpected dominance of rare members of microbial communities. Applied and Environmental Microbiology 73: 4532-4542.

Atkins MS, Teske AP, Anderson OR. 2000. A survey of flagellate diversity at four deep-sea hydrothermal vents in the Eastern Pacific Ocean using structural and molecular approaches. Journal of Eukaryotic Microbiology 47: 400-411.

Bell T. 2010. Experimental tests of the bacterial distance-decay relationship. International Society for Microbial Ecology Journal 4: 1357-1365.

Bloom SA. 1981. Similarity indexes in community studies - potential pitfalls. Marine Ecology-Progress Series 5: 125-128.

Brock TD. 1961. Milestones in Microbiology. Englewood Cliffs (NJ): Prentice-Hall.

Countway PD, Gast RJ, Savai P, Caron DA. 2005. Protistan diversity estimates based on 18S rDNA from seawater incubations in the western North Atlantic. Journal of Eukaryotic Microbiology 52: 95-106.

Da Silva Filho, et al. PCR identification of Pseudomonas aeruginosa and direct detection in clinical samples from cystic fibrosis patients J Med Microbiol April 1999 vol. 48 no. 4 357-361

Dolan JR. 2005. An introduction to the biogeography of aquatic microbes. Aquatic Microbial Ecology 41: 39-48.

Fenchel T, Finlay BJ. 2004. The ubiquity of small species: Patterns of local and global diversity. Bioscience 54: 777-784.

Finlay BJ, Fenchel T. 2004. Cosmopolitan metapopulations of free-living microbial eukaryotes. Protist 155: 237-244.

Fuhrman JA, Steele JA. 2008. Community structure of marine bacterioplankton: patterns, networks, and relationships to function. Aquatic Microbial Ecology 53: 69-81.

Jones SE, McMahon KD. 2009. Species-sorting may explain an apparent minimal effect of immigration on freshwater bacterial community dynamics. Environmental Microbiology 11: 905-913.

Katz LA, McManus GB, Snoeyenbos-West OLO, Griffin A, Pirog K, Costas B, Foissner W. 2005. Reframing the 'Everything is everywhere' debate: evidence for high gene flow and diversity in ciliate morphospecies. Aquatic Microbial Ecology 41: 55-65.

Kim E, Wilcox L, Graham L, Graham J. 2004. Genetically distinct populations of the dinoflagellate Peridinium limbatum in neighboring Northern Wisconsin lakes. Microbial Ecology 48: 521-527.

Lauber CL, Hamady M, Knight R, Fierer N. 2009. Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. Applied and Environmental Microbiology 75: 5111-5120.

LiPuma JJ, Dulany BJ, McMenamin JD, Whitby PW, Stull TL, Coenye T, Vandamme P. 1999. Development of rRNA-Based PCR Assays for Identification of Burkholderia cepacia Complex Isolates Recovered from Cystic Fibrosis Patients. Journal of Clinical Microbiology 37(10): 3167-3170.

Litchman E. 2010. Invisible invaders: non-pathogenic invasive microbes in aquatic and terrestrial ecosystems. Ecology Letters 13: 1560-1572.

Liu WT, Mirzabekov AD, Stahl DA. 2001. Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach. Environmental Microbiology 3(10): 619-629.

Marshall MM, Amos RN, Henrich VC, Rublee PA. 2008. Developing SSU rDNA metagenomic profiles of aquatic microbial communities for environmental assessments. Ecological Indicators 8: 442-453.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreas L, Reysenbach AL, Smith VH, Staley JT. 2006. Microbial biogeography: putting microorganisms on the map. Nature Reviews Microbiology 4: 102-112.

Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, Fierer N, Townsend AR, Cleveland CC, Stanish L, Knight R. 2011. Global patterns in the biogeography of bacterial taxa. Environmental Microbiology 13: 135-144.

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. 2011. A guide to the natural history of freshwater lake bacteria. Microbiology and Molecular Biology Reviews 75: 14-49.

Papke RT, Ramsing NB, Bateson MM, Ward DM. 2003. Geographical isolation in hot spring cyanobacteria. Environmental Microbiology 5: 650-659.

Pringle A, Baker DM, Platt JL, Wares JP, Latge JP, Taylor JW. 2005. Cryptic speciation in the cosmopolitan and clonal human pathogenic fungus Aspergillus fumigatus. Evolution 59: 1886-1899.

Rappe MS, Giovannoni SJ. 2003. The uncultured microbial majority. Annual Review of Microbiology 57: 369-394.

Rydholm C, Szakacs G, Lutzoni F. 2006. Low genetic variation and no detectable population structure in Aspergillus fumigatus compared to closely related Neosartorya species. Eukaryotic Cell 5: 650-657.

Schaefer E.F. 1997. A DNA assay to detect the toxic dinoflagellate Pfiesteria piscicida, and the application of a PCR based probe. MS Thesis. University of North Carolina, Greensboro.

Schiaffino MR, Unrein F, Gasol JM, Massana R, Balague V, Izaguirre I. 2011. Bacterial community structure in a latitudinal gradient of lakes: the roles of spatial versus environmental factors. Freshwater Biology 56: 1973-1991.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences of the United States of America 103: 12115-12120.

Wattiau, et al. 2001. A PCR test to identify Bacillus subtilis and closely related species and its application to the monitoring of wastewater biotreatmen. Applied Microbiology and Biotechnology 56(5-6):816-819.

Whitaker RJ. 2006. Allopatric origins of microbial species. Philosophical Transactions of the Royal Society B-Biological Sciences 361: 1975-1984.

Whitaker RJ, Grogan DW, Taylor JW. 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science 301: 976-978.

Whitfield J. 2005. Biogeography: Is everything everywhere? Science 310: 960-961.

Wu YP, Yu XS, Wang HZ, Ding N, Xu JM. 2010. Does history matter? Temperature effects on soil microbial biomass and community structure based on the phospholipid fatty acid (PLFA) analysis. Journal of Soils and Sediments 10: 223-230.

Zaballos M, Lopez-Lopez A, Ovreas L, Bartual SG, D'Auria G, Alba JC, Legault B, Pushker R, Daae FL, Rodriguez-Valera F. 2006. Comparison of prokaryotic diversity at offshore oceanic locations reveals a different microbiota in the Mediterranean Sea. Fems Microbiology Ecology 56: 389-4

Zwart,G., Hiorns,W.D., Methe,B.A., van Agterveld,M.P., Huismans,R., Nold,S.C., Zehr,J.P. and Laanbroek,H.J. 1998. Nearly identical 16S rRNA sequences recovered from lakes in North America and Europe indicate the existence of clades of globally distributed freshwater bacteria. Syst. Appl. Microbiol. 21 (4), 546-556