Title
Microbial Activity & Diversity in High-Altitude Oligotrophic Lakes of the Sierra Nevada, California

Permalink
https://escholarship.org/uc/item/5682s67z

Author
Hayden, Curtis J.

Publication Date
2014

Peer reviewed|Thesis/dissertation
Microbial Activity & Diversity in High-Altitude Oligotrophic Lakes of the Sierra Nevada, California

A Thesis submitted in partial satisfaction of the requirements for the degree of Master of Science

in

Environmental Systems

by

Curtis Janes Hayden

Committee in charge:

Chair Professor J. Michael Beman
Professor Asmeret Asefaw Berhe
Professor Carolin Frank

2014
The Thesis of Curtis Janes Hayden is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

____________________________________________

J. Michael Beman, Chair

____________________________________________

Asmeret Asefaw Berhe

____________________________________________

Carolin Frank

University of California, Merced

2014
Acknowledgements

I would like to acknowledge and thank the numerous people who made this thesis possible. First, I owe a great deal of gratitude to my advisor, Mike Beman. I cannot thank Mike enough for the countless hours spent enhancing my abilities as a researcher. I feel very fortunate to have had such a great and talented mentor in this endeavor. Also, thank you to my thesis committee members, Prof. Asmeret Asefaw Berhe and Prof. Carolin Frank, for support and guidance for this project.

A huge thanks to all the current and former members of the Beman Lab for their camaraderie and contributions during the numerous memorable expeditions to Yosemite, Palau, and Bodega Bay. Dr. Kelly Marie Henry, Susan Alford, Jesse Wilson, and Molly Carolan were instrumental in the success of this project and provided ample help and guidance throughout. I am also very grateful for the Berhe, Ghezzai, Fogel, and Blois Labs for their insight and input during our numerous Castle Meetings.

Finally, many thanks to my family and friends whose support have made this thesis possible, including my parents Bob and Michele Hayden and my girlfriend Emily Lisker. Thanks for handling my numerous days spent incommunicado in the mountains and working in Merced.
TABLE OF CONTENTS

Signature Page...........................................................................................................iii
Acknowledgements.................................................................................................iv
Table of Contents.....................................................................................................v
List of Figures and Tables...........................................................................................vi

Abstract......................................................................................................................1
1. Introduction: Microbial communities and freshwater ecosystems.......................2
2. Background – Sierra Nevada aquatic ecosystems..................................................2
  2.1 Study Location......................................................................................................2
  2.2 Microbial Diversity of Sierra Nevada Lakes......................................................3
  2.2 Effects of atmospheric nitrogen on Sierra Nevada aquatic ecosystems..............4
3. High Abundances of Active Ammonia Oxidizing Bacteria and Archaea in Oligotrophic
   High-altitude Lakes of the Sierra Nevada, California, USA..................................5
  3.1 Abstract...............................................................................................................5
  3.2 Introduction..........................................................................................................6
3.3 Materials and Methods........................................................................................8
  3.3.1 Study Site.......................................................................................................8
  3.3.2 Sampling .......................................................................................................8
  3.3.3 DNA extraction and quantification...............................................................8
  3.3.4 Real time qPCR analysis .............................................................................8
  3.3.5 Nutrient analyses .........................................................................................9
  3.3.6 $^{15}$NH$_4^+$ oxidation rate measurements ...............................................9
  3.2.7 Statistical Analyses ....................................................................................9
3.4 Results and Discussion.........................................................................................10
4. Microbial Community Composition, Diversity and Abundance in High-Altitude
   Oligotrophic Lakes of the Sierra Nevada, California.............................................14
  4.1 Abstract..............................................................................................................15
  4.2 Introduction.........................................................................................................15
4.3 Materials and Methods.......................................................................................16
  4.3.1 Study Site ....................................................................................................16
  4.3.2 Sampling ....................................................................................................16
  4.3.3 DNA extraction and quantification.............................................................17
  4.2.4 Nutrient Analyses ......................................................................................17
  4.2.5 Illumina Sequencing ..................................................................................18
  4.2.6 Diversity Calculations ..............................................................................18
  4.2.7 Statistical Analyses ....................................................................................19
4.4 Results & discussion.........................................................................................19
  4.5 Conclusion.......................................................................................................24

References...............................................................................................................26
LIST OF FIGURES AND TABLES

FIGURES

Figure 1.1.1: Map of sampling locations in Yosemite NP……………………………………33
Figure 1.1.2: Plot of sampling site elevation versus longitude………………………35
Figure 1.2: Environmental data for individual sites……………………………………..36
Figure 1.3: Boxplot comparison of AOA/AOB abundance versus elevation……………37
Figure 1.4: Bar graph of observed ammonia oxidation rates…………………………38
Figures 2.1.1 through 2.1.4: Bar graphs of bacterial community composition ………39
Figures 2.2.1 through 2.2.6: Area graphs of bacterial community composition………42
Figure 2.3.1: Temporal multivariate redundancy analysis (RDA) plot…………………48
Figure 2.3.2: Spatial multivariate redundancy analysis (RDA) plot…………………..49
Figures 2.4.1 through 2.4.6: Alpha Diversity calculation plots…………………………50
Figure 2.5.1: Combined Beta Diversity calculation………………..……………………55
Figures 2.6.1 through 2.6.3: non-metric multidimensional scaling (NMDS) plots………56

TABLES

Table 1.1: List of lake names, elevation, location and depth…………………………….34
Table 1.2: List of ammonia oxidation rates by lake……………………………………38
Table 2.1: Percent community composition for most abundant bacterial classes……46
Table 2.2: Monthly $R^2$ for Beta diversity calculation versus inter-lake distance………..53
ABSTRACT

Title: Microbial Activity and Diversity in High-Altitude Oligotrophic Lakes of the Sierra Nevada, California
Student: Curtis Janes Hayden
Degree: Master of Science, Environmental Systems
School: University of California, Merced, 2014

Lakes traditionally have been considered unimportant in global and regional biogeochemical cycling, however these systems receive and process large volumes of material within their watersheds and thus function as integrators and processors of relatively large amounts of material compared with the surrounding landscape. Due to the relative concentration of biogeochemical cycling, these systems disproportionately influence global biogeochemical cycling (e.g. carbon and nitrogen cycling). This is critical as these cycling processes have the potential to be significantly influenced by anthropogenic activities — i.e., lakes are considered ‘sentinels’ of global change. Microbial communities are integral components of these systems as archaea and bacteria — ubiquitous organisms that we still know very little about — almost exclusively perform these processes. To investigate the controls governing the abundance, activity and distribution of these microbial communities, I sampled multiple lakes (n=15) at multiple time points along an elevational transect located in Yosemite National Park, California, USA. At these sites I examined the effects of spatio-temporal and environmental variability on the abundance, activity, and composition of archaeal and bacterial communities. This was accomplished using a combination of methods that included quantification of functional nitrification genes (ammonia monooxygenase) using quantitative polymerase chain reaction, stable isotope based in-situ measurements of nitrification, chemical analyses of critical micronutrients, and next-generation high-throughput sequencing to determine community composition and diversity among microbial communities. Active nitrogen cycling and high abundances of ammonia-oxidizing bacteria (AOB) were captured by my sampling, as well as the presence of a large and diverse microbial community that exhibits a strong distance-decay relationship with increasing inter-lake distance. In my thesis I address questions regarding the dominance of ammonia-oxidizing archaea (AOA) versus AOB in nitrogen cycling processes and the effects of spatio-temporal and environmental variability on microbial community composition and diversity in these systems.
Introduction: Microbial communities and freshwater ecosystems

Microbial communities play a dominant role in influencing the biogeochemical cycling of essential elements within ecosystems. The organisms that comprise these communities have evolved a multitude of metabolic lifestyles that utilize essential elements for their energy requirements, including Hydrogen, Carbon, Nitrogen, Oxygen, Sulfur, and many other organic and inorganic materials. As a whole, these metabolic processes drive global biogeochemical cycles through the transformation of elements (Falkowski et al. 2008). In lakes, microbes determine the chemical species present and in turn control water quality and drive primary productivity (Newton et al. 2011). Aside from the important influence that these microbial processes have on the biogeochemistry of freshwater environments, they are understudied relative to other ecosystems (e.g. marine environments). Few publications have looked at the activity of microbial communities in freshwater ecosystems let alone the effects that environmental factors (i.e. temperature, nutrient availability, light attenuation) have on controlling microbial processes. Considering the importance of these processes and the potential for anthropogenic activities to increase temperature and pollution of synthetically produced nutrients (i.e. nitrogen), it is essential to understand the susceptibility of these processes to changing environmental conditions. The goal of this research is to determine how spatio-temporal and environmental variability affect the abundance, activity, and diversity of microbial communities in the high-altitude oligotrophic systems of the Sierra Nevada. To explore these questions I surveyed multiple lakes along an elevational gradient ranging from 2,289 to 3,160 meters in Yosemite National Park, USA. Samples collected from sites were analyzed in order to determine microbial community composition and diversity, quantify specific functional genes and nutrient cycling rates, and explore what factors were most significant in governing microbial community composition, abundance, and activity in these systems.

Study Location

The Sierra Nevada is a 400-mile long north-south running mountain range located between the U.S. states of California and Nevada. The range gradually rises west to east from California’s Central Valley and reaches an apex of 3,000 to 4,200 meter peaks on its eastern edge, abruptly dropping off into the deep valleys of the Great Basin. The range has multiple deep west-facing river drainages and vegetation along them consists of grasslands and foothill woodlands at lower elevations, transitioning to mixed conifer forests, alpine meadows, and lakes at increasing elevations (Cisneros et al. 2010). These river drainages are partitioned into 24 discernable watersheds that contain over 4,000 lakes above 2,500 meters (Nelson et al. 2008). The majority of these drainages are located in federally protected wilderness areas and are an essential watershed for California, serving as the public’s preeminent drinking water source and also providing agricultural, hydroelectric, and industrial benefits. As these lakes are located in the alpine and sub-alpine zone, they are generally low in productivity due to low levels of nitrogen and phosphorous inputs, exhibiting high degrees of seasonal variation in temperature due to high-elevation (Wetzel 2001). Due to the relatively low levels of productivity and isolation from anthropogenic influences, these systems are appropriate lenses with which to look at spatio-temporal and environmental controls over
biogeochemical cycling. In addition, elevational variability and distance between selected sites (see Fig 1.1.1 and 1.1.2) allow for the consideration of spatial effects and temperature differences on cycling processes and microbial community structure.

Microbial Diversity of Sierra Nevada Lakes

In the high-altitude lakes of the Sierra Nevada, microbial communities provide important ecosystem services through their metabolic processing and transformation of essential nutrients that influence the concentrations, stoichiometric ratio, and bioavailability of nutrients in these systems. These processes include decomposition of organic matter, mineralization of nitrogen, storage of inorganic compounds, as well as roles in other significant biogeochemical cycles. The structure of these microbial communities is important as it can influence the rates and magnitude of biogeochemical cycling processes (Schimel and Gulledge 1998, Fuhrman 2009). Community structure formation of microbial communities in lakes has been credited to temporal and spatial variability, though a universal pattern for predicting community responses based on spatio-temporal and physicochemical parameters is not available (Fuhrman et al. 2006). Past surveys and comparisons of multiple lakes systems demonstrated that environmental, temporal, regional, and landscape-level factors combine to influence the composition of lake microbial communities (Yannarell and Triplett 2005) and the distribution of specific freshwater bacterial groups can be determined by lake acidity, temperature, and retention time (Lindstrom et al. 2005).

Due to the seasonality experienced in high-altitude systems and related physicochemical influences associated with temperature, acidity, lake-turnover, and allochthonous nutrient enrichment events from snowmelt, aquatic ecosystems in the Sierra Nevada experience a high degree of inter- and intra-annual variability within individual lake systems. Past work in Sierra Nevada lakes (i.e. Emerald Lake, California) found consistent phenological patterns of community composition in lakes related to dynamic events, including snowmelt, ice-off, summer stratification, and fall-overturn, where over multiple years distinctive microbial communities arose subsequent to the onset of changes in the physical structure of the water column (Nelson 2009).

Effects of Nitrogen Pollution

The atmospheric deposition of foreign chemical species into lakes can subject these systems to pollutants and lead to modifications of existing biogeochemical conditions through subsequent nutrient enrichment (Ballantyne et al. 2011). In the Sierra Nevada, where ~22% of lakes are strongly nitrogen limited (Eilers JM 1987), this increased nitrogen input can have particularly pronounced effects. Owing to the geologic history of the Sierra Nevada, the watersheds that feed into its alpine and subalpine lakes are high in granitic parent material and generally have thin soils. Due to these characteristics, these watersheds have a limited buffering capacity in terms of their ability to neutralize foreign chemical species (Meixner et al. 2004). Moreover, nutrient conditions of alpine and subalpine lakes of the Sierra Nevada indicate that they are generally oligotrophic, meaning they have low nutrient concentrations, low supply of nutrients, and low primary
productivity (Goldman and Horne 1983). The limited buffering capacity of these systems, coupled with the oligotrophic conditions, means minor modifications of the existing nutrient conditions in a watershed (e.g. through atmospheric dust deposition) could increase biological activity, leading to more eutrophic conditions (Sickman et al. 2003). These modifications could have especially pronounced effects for microbial communities as these populations may rapidly respond to and be altered by nutrient enrichment due to increased sensitivity relative to other organisms (i.e. phytoplankton) (Nelson et al. 2008).

In light of the absence of detailed information regarding microbial community composition and diversity in the systems, as well as the nature of microbial mediated biogeochemical cycling of nitrogen in high-altitude lakes of the Sierra Nevada, the goal of this work is to:

1. Assess the presence, abundance, diversity and activity of nitrifying microorganisms and relationships with spatial, temporal, and geochemical variability in high-altitude lakes of the Sierra Nevada

2. Use next generation high-throughput sequencing to explore spatio-temporal and geochemical effects on the presence, absence, abundance, diversity, and composition of the microbial communities in high-altitude lakes of the Sierra Nevada
High Abundances of Active Ammonia Oxidizing Bacteria and Archaea in Oligotrophic High-altitude Lakes of the Sierra Nevada, California, USA

ABSTRACT

Nitrification plays a central role in the nitrogen cycle by determining the oxidation state of nitrogen and its subsequent bioavailability and cycling. However, relatively little is known about the underlying ecology of the microbial communities that carry out nitrification in freshwater ecosystems, and particularly in high-altitude oligotrophic lakes, where nitrogen is frequently a limiting nutrient. We quantified ammonia-oxidizing archaea (AOA) and bacteria (AOB) in 9 high-altitude lakes (2289-3160 m) in the Sierra Nevada, California, USA, in relation to spatial and biogeochemical data. Based on their ammonia monooxygenase (amoA) genes, AOB and AOA were frequently detected. AOB were present in 88% of samples and were more abundant than AOA in all samples. Both groups showed >100 fold variation in abundance between different lakes, and were also variable through time within individual lakes. Nutrient concentrations (ammonium, nitrite, nitrate, and phosphate) were generally low and varied across and within lakes; AOB abundance was significantly correlated with phosphate ($r^2 = 0.32$, $p < 0.1$), whereas AOA abundance was most strongly correlated with lake elevation ($r^2 = 0.43$, $p < 0.05$). We also measured low rates of ammonia oxidation—indicating that AOB, AOA, or both, are biogeochemically active in these oligotrophic ecosystems. Our data indicate that dynamic microbial communities and internal N cycling occur in oligotrophic, high-altitude, freshwater lakes.
INTRODUCTION

Nitrogen (N) is an essential nutrient for all life, and its availability serves as a critical factor for the growth of individual organisms, community composition, and ecosystem primary productivity in freshwater lakes (Elser et al. 2007) (Vrede et al. 2009). In many ecosystems, N availability—both quantity and chemical form—is largely dictated by microbial communities, which transform inorganic nitrogen into bioavailable forms (N$_2$ fixation), and actively cycle N through oxidation-reduction (redox) processes. Phosphorus (P) typically limits primary production in freshwater (Schindler 1977), but both absolute amounts, and relative ratios, of N and P are highly variable due to variations in lake nutrient sources, as well as internal cycling by phytoplankton, zooplankton, and microbes (Elser et al. 2007, Elser et al. 2009, Harpole et al. 2011). In oligotrophic aquatic systems, in particular, differences in size, growth rate, and chemical form of available nutrients may favor microorganisms in competition with phytoplankton (Cotner and Biddanda 2002). Microbial control of both N quantity and chemical form has important implications for the degree of eutrophication in these ecosystems, and the degree to which allochthonous N inputs (i.e. atmospheric pollutants) may affect oligotrophic lakes (Fenn et al. 2003).

Within the microbial N cycle, nitrification is a two-step process that involves the aerobic oxidation of reduced inorganic nitrogen compounds (i.e. NH$_3$/NH$_4^+$) to nitrite (NO$_2^-$), and the subsequent oxidation of NO$_2^-$ to nitrate (NO$_3^-$). Nitrification links the mineralization of nitrogen to its eventual removal as dinitrogen gas (N$_2$) via either denitrification or anaerobic ammonium oxidation (anammox). The first step of nitrification is carried out by a few bacterial lineages within the Beta- and Gamma-proteobacteria and also by the archaeal phylum *Thaumarchaeota* (previously know as the group 1 Crenarchaeota) (Brochier-Armanet et al. 2008). These ammonia-oxidizing bacteria (AOB) and ammonia-oxidizing archaea (AOA) use ammonia monoxygenase (AMO) to catalyze the oxidation of NH$_4^+$ to NO$_2^-$. As AOA were confirmed to be capable of ammonia oxidation only recently (Konneke et al. 2005), the physical and chemical factors that control the abundance and function of these organisms, and their relative influence on nitrification rates, are not entirely understood—particularly in freshwater environments (Francis et al. 2007, Erguder et al. 2009, Hatzenpichler 2012, Small et al. 2013).

AOA, AOB, and nitrification have been examined within few freshwater lakes, yet AOA appear to be important and dynamic components of lake plankton and biogeochemical cycles: *Thaumarchaeota* are abundant in lakes (Auguet and Casamayor 2008), most appear to be AOA (Auguet et al. 2011), and their populations fluctuate over time (Auguet et al. 2011). AOB are found in lakes ranging from temperate eutrophic, to high-altitude oligotrophic, but in contrast to AOA, how AOB abundance varies in lakes through space and time is not well known (Whitby et al. 2001, Cebron et al. 2004, Hu et al. 2010). The abundances of AOA and AOB can be controlled by differential sensitivities to temperature (Tourney et al. 2008), pH (Nicol et al. 2008, Auguet and Casamayor 2013), ammonium concentrations (Martens-Habbena et al. 2009), and light (French et al. 2012)—all of which may be relevant in high-altitude lakes, but have not been examined.
Nitrification varies with depth and time, and is quantitatively important within lake water columns (Hall and Jeffries 1984, Rudd et al. 1988, Finlay et al. 2007, Small et al. 2013)—for example, Finlay et al. 2007 showed that within-lake production of NO$_3^-$ through nitrification is the predominant source of NO$_3^-$ in Lake Superior. However, a lone study has measured both AOA and ammonia oxidation rates in freshwater lakes (Small et al. 2013), and AOB have rarely been quantified in lakes (Hugoni et al. 2013). We therefore know little about variations in AOA and AOB abundance and activity over time and across different lakes—let alone how AOA, AOB, and ammonia oxidation rates respond to changes in temperature, N availability, and other environmental factors within freshwater systems.

Of particular relevance are the potential inhibitory effects of light on nitrification: while these have been known for some time (reviewed by (Lomas and Lipschultz 2006), the relative effects of different wavelengths of light on AOA versus AOB, and in the field versus lab, are mixed. AOA appear highly sensitive to light in controlled experiments (French et al. 2012, Merbt et al. 2012): the AOA *Nitrosopumilus maritimus* and *Nitrosotalea devanaterra* were inhibited by lower light levels than AOB, and showed little recovery of ammonia oxidation over 8/16 hour light/dark cycles (Merbt et al. 2012). French et al. (2012) likewise found that ammonia oxidation by three freshwater AOA isolates was strongly inhibited by white and blue light, whereas an AOB isolate was inhibited only by blue light and recovered partial oxidation ability in the dark. Notably, all of the AOA isolates used in these studies have been recovered from sediments or soil, and it is possible that pelagic AOA are less light-sensitive—for example, Auguet and Casamayor (2008) proposed that surface waters of mountain lakes are an archaeal ‘hotspot’ based on high crenarchaeal abundance in the neuston. AOA also actively express *amoA*, and nitrification is known to occur at least transiently, in the upper ocean (Church et al. 2010, Santoro et al. 2010, Beman et al. 2012).

We quantified the abundance of AOA and AOB across a high-altitude lake transect in Yosemite National Park, in the Sierra Nevada mountain range, California, USA (Figure 1.1.1). In Sierra Nevada lakes, the differing susceptibility of AOA and AOB to photoinhibition could be a crucial factor in N cycling, as there is a strong natural increase in ultraviolet radiation with increasing elevation (Hall & Jeffries, 1984). Moreover, these high-altitude lakes have relatively low light attenuation due to high transparency typical of oligotrophic aquatic ecosystems found at high elevations (Hall and Jeffries 1984, Sommaruga 2001). Contrary to the paradigm that freshwater lakes are traditionally limited by phosphorous availability (Sterner 2008), the availability of nitrogen is also a critical factor for primary productivity in aquatic ecosystems of the Sierra Nevada: biological activity in many lakes is strictly limited by nitrogen availability, with ~22% of lakes being limited by nitrogen (Eilers JM 1987). Internal nitrogen cycling may therefore play an important role in the overall productivity and structure of these freshwater ecosystems. We used natural variations in temperature, radiation, and N deposition, based on elevational and temporal variability between sampling sites, to examine the prevalence and abundance of AOA and AOB in high-altitude lakes.
MATERIALS & METHODS

**Study Site.** The Sierra Nevada (California) is a 400-mile long mountain range that gradually rises from the valley floor from west to east and reaches an apex of 3,000 to 4,200 meter peaks on its eastern edge. Vegetation along the mountain range is composed of grasslands and foothill woodlands at lower elevations, with a transition to mixed conifer forests, and then alpine meadows and lakes at higher elevations. Aquatic ecosystems in the Sierra Nevada are located downwind of large urban or agricultural areas that emit high levels of nitrogen (Clarisse et al. 2010) and can experience elevated levels of nitrogen deposition (Bytnerowicz et al. 2002). This N deposition is known to increase nitrogen concentrations in lakes (Fenn et al. 2003), and in the case of the Sierra Nevada, represents a large fraction of the nitrogen input to high-elevation lakes: (Baron 2006) established a critical load threshold of 1.5 kg N ha$^{-1}$ year$^{-1}$ for high-elevation lakes located in Rocky Mountain National Park, yet current annual nitrogen loading in the Sierra Nevada (i.e. Emerald Lake Watershed, Sequoia National Park) ranges from 2.0 to 4.9 kg N ha$^{-1}$ year$^{-1}$ (Sickman et al. 2001). Moreover, the watersheds are high in granitic parent material and generally have thin soils; both characteristics cause aquatic ecosystems in this Sierra Nevada to have limited buffering capacity in terms of their ability to neutralize foreign chemical species (Meixner et al. 2004). Clow et al. 2010 suggested that this property, coupled with high precipitation at high elevations, leads to high N loading at high elevations despite greater distances from emission sources. Our data are relevant to this as our selected sites range from 2300 m (Harden Lake) to 3160 m (Upper Gaylor Lake) and our transect terminus is adjacent to the steep Sierra escarpment (See Fig 1.1.2).

**Sampling.** Water samples were collected from lakes located in Yosemite National Park (YNP) during the 2012 summer (for list of sampled sites see Table 1.1 and map see Figure 1.1.1). Nine lakes were visited a total of 41 times, which ranged in elevation from 2289 m to 3160 m and in depth from 2.5 m to 10 m. Water samples were collected from the middle of each lake at 1 m below the surface using a Van Dorn water sampler (Lamotte). Duplicate water samples were filtered in the field onto 0.22 µm PVDF Membrane Filters (Millipore) using sterile 60 mL Polycarbonate syringes (Cole-Parmer). After filtration, filter membranes were packed into bead tubes (MP Biomedicals) containing 800 µL of Sucrose-Tris-EDTA. Filtrate was collected in 60 mL HDPE Bottles (Nalgene) for subsequent nutrient analysis. Samples were transported on ice to UC Merced’s main campus or Yosemite Field Station (within hours of collection) and stored at -80° or -20°C until extraction of DNA and nutrient analysis.

**DNA Extraction and quantification and real-time qPCR Analysis.** DNA extraction followed Beman et al. (2012) using Sucrose-Tris-EDTA (STE) lysis buffer, sodium dodecyl sulfate (SDS), and proteinase K with bead-beating. DNA was further purified using a DNeasy Blood & Tissue Extraction Kit (Qiagen) and resolubilized in 50 uL of ultra-pure DNA free water (Qiagen). DNA was quantified using a PicoGreen dsDNA quantification kit (Invitrogen) and an Mx3005P real-time thermocycler (Agilent Technologies).
Quantitative Polymerase Chain Reaction (qPCR) was used to quantify the abundance of *amoA* genes in lake samples. Primers, reaction chemistry, thermocycling, QPCR standards, and data analysis exactly followed Beman et al. (2012). In brief, we used SYBR Green chemistry and the primers crenamoAF and crenamoAR for archaeal *amoA* (Francis et al. 2005) and beta-amoA1F and beta-amoA1R for betaproteobacterial *amoA* (Rotthauwe et al. 1997) (Primer sequences listed in Table 1.3).

**Nutrient Analyses.** Filtered lake water samples were analyzed for orthophosphate (μmol l⁻¹) nitrite (μmol l⁻¹) and nitrate (μmol l⁻¹) using flow-injection analysis on a QuikChem 8000 (Zellweger Analytics, Inc.) at the University of California, Santa Barbara Marine Sciences Institute Analytical Laboratory (standard curve r² = 0.999 for all assays). Filtered lake water samples were analyzed for ammonium (NH₄⁺) following Holmes et al. 1999 (Holmes et al. 1999). 8 mL of filtered water was combined with 2 mL of reagent that consisted of 95%: 0.1M sodium tetraborate, 0.015M O-pthalaldialdehyde, 0.03mM sodium sulfite and 5%: Ethanol. After aging, samples were measured in triplicate for fluorescent intensity using a fluorometer (Trilogy® Laboratory Fluorometer, Turner Designs) Standards ranged from 31.2 to 186.8 nM and for different runs, standard curve r² = 0.998–0.999.

**¹⁵NH₄⁺ oxidation rate measurements.** Ammonia oxidation rates were measured by adding 99 atom percent (at%) ¹⁵NH₄⁺ to a concentration of 200 nmol L⁻¹, and measuring the accumulation of ¹⁵N label in the oxidized NO₂⁻ + NO₃⁻ pool after incubation for ~24 hours (Ward 2008). All samples were incubated within lakes at 1m depth to mimic in situ conditions as accurately as possible. δ¹⁵N of NO₂⁻ + NO₃⁻ was measured at the UC Davis Stable Isotope Facility using the ‘denitrifier method’ (Sigman et al. 2001), which produces N₂O that can be analyzed on mass spectrometer. Isotopic reference materials bracketed every 3-4 samples and coefficients of variation were 0.6%.

Initial at% enrichment of the substrate at the beginning of the experiment (n₀NH₄⁺, see Eq. 1) was calculated by isotope mass balance based on NH₄⁺ concentrations determined fluorometrically (Holmes et al. 1999) assuming that the ¹⁵N activity of unlabeled NH₄⁺ was 0.3663 at% ¹⁵N. Rates of ammonia oxidation (¹⁵Rox) were calculated using an equation modified from Ward et al. (1989a):

\[
¹⁵R_{ox} = \frac{(n_t - n_{NOx}) \times [NO_3^- + NO_2^-]}{(n_{NH_4^+} - n_{NH_4^+}) \times t}
\]

where nₜ is the at% ¹⁵N in the NO₃⁻ + NO₂⁻ pool measured at time t, n₀NOx⁻, is the measured at% ¹⁵N of unlabeled NO₂⁻ + NO₃⁻, n₀NH₄⁺ is the initial at% enrichment of NH₄⁺ at the beginning of the experiment, nₜNH₄⁺ is at% ¹⁵N of NH₄⁺ at time t, and [NO₃⁻ + NO₂⁻] is the concentration of the NO₃⁻ pool.

**Statistical analyses** were performed using the R statistical environment (http://www.r-project.org/) and the vegan package.
RESULTS AND DISCUSSION

AOB amoA genes were detected in all lakes from June to September, and in 54 out of 61 total samples, whereas AOA were found in 28 of 61 total samples. AOB amoA gene copies ranged from $3.04 \times 10^2$ to $2.07 \times 10^5$ genes mL$^{-1}$, however the majority (73%) of AOB values fell below $2 \times 10^4$ genes mL$^{-1}$, with values exceeding this occurring in June, August and September at high elevations (Figure 1.3). AOA ranged from 0 to $4.58 \times 10^3$ genes mL$^{-1}$ (See Fig 1.3) and there were zero instances where AOA outnumbered AOB. When AOA were detected, they were outnumbered by AOB by $2.8 \times 10^{80}$-fold.

Highest average abundances of both AOB and AOA were found in Middle Sunrise Lake (2826 m), while lowest average AOB abundances were found in Lukens Lake (2506 m), Lower Cathedral Lake (2832 m) and Lake Elizabeth (3019 m) (See Fig 1.3). For AOA, multiple lakes had low average values, including Lukens, Lower Cathedral, Elizabeth, and Lower Gaylor (3064 m) (See Fig 1.3). Over time, the highest numbers of amoA genes were present in September for AOB (4 out of 6 lakes above 2800 m) (See Fig 1.3), and in June for AOA (5 out of 6 lakes above 2800 m)(See Fig 1.3).

Previous studies that have quantified AOA abundance in high-altitude oligotrophic lakes reported AOA amoA gene abundances as high as $3 \times 10^4$ genes mL$^{-1}$ (Auguet et al. 2011), while AOB were below detection limits in the same high-altitude lakes (Auguet et al. 2012). Our results indicate that AOB are more prevalent and abundant than AOA within oligotrophic lakes of the Sierra Nevada, CA, and AOB dominance occurs regardless of lake nutrient concentrations (see below), date of sampling, or lake elevation. This contrasts with earlier work that found AOA were more prevalent (detected via PCR but not quantified by QPCR) than AOB in oligotrophic lakes of the Tibetan plateau (Hu et al. 2010), and dominant in oligotrophic lakes of the Spanish Pyrenees, Spain (Auguet et al. 2012). In two contrasting, lower elevation (231-825 m), lakes in France, Hugoni et al. 2013 found a dominance of AOA under low ammonia concentrations and oligotrophic conditions, and dominance of AOB in nutrient-rich waters (Hugoni et al. 2013). The lakes sampled in Yosemite are uniformly nutrient-poor and have low ammonium conditions (<75 nM), however our data indicate that AOB are dominant over AOA. One explanation for AOB dominance is that the amoA primers used to detect AOB and AOA may over- or under-estimate their abundances, respectively. However, the AOB primers used here are specific for betaproteobacterial AOB and are widely used (Junier et al. 2008), while the AOA primers amplify a wide range of AOA groups from water, sediments, and soils (Francis et al. 2005). Nor does this explain the prevalence of AOB, which were detected in 88.5% of samples. AOA were also detected in nearly 50% of samples despite high light levels and oligotrophic conditions that are presumably hostile to nitrifiers in general. AOB and AOA therefore appear to be present—and sometimes abundant—within the water columns of high-altitude oligotrophic lakes in Yosemite National Park.

These AOB and AOA populations could be sustained by relatively high N fluxes that are not reflected in nutrient pools. That is, ammonium may be rapidly regenerated, oxidized, and assimilated, but because of high demand, does not accumulate. For example, lakes are watershed ‘integrators’ (Williamson et al. 2009) that may function as hotspots of N-
cycling—including nitrification (Finlay et al. 2007, August and Casamayor 2008, Clow et al., 2010, Baron et al. 2013). In Yosemite, lake N loading also increases with elevation (Clow et al. 2010), and NO$_3^-$ concentrations are correlated with modeled N deposition rates (Clow et al. 2010). N deposition to Sierra watersheds occurs primarily as NH$_4^+$ (Bytnerowicz et al. 2002) and this flux of N—which can be comparatively large in these oligotrophic ecosystems (Murphy and Knopp 2000)—must have been nitrified at least once if it accumulates as NO$_3^-$. Ammonium concentrations therefore may not be the sole predictor for AOA and AOB abundances—elevation and nitrate concentrations may also be relevant—and we analyzed relationships between AOA and AOB and several types of spatial and nutrient concentration data, using a variety of statistical approaches.

During the sampling period NO$_3^-$ concentrations in all lakes (1m depth) ranged from 150 nM to 990 nM, PO$_4^{3-}$ concentrations from 60 nM to 120 nM, NO$_2^-$ concentrations from 20 nM to 120 nM, and NH$_4^+$ concentrations from 2.8 nM to 72.2 nM (See Fig 1.2). All of these concentrations where generally low. As expected in aquatic ecosystems, NO$_3^-$ levels were higher than either PO$_4^{3-}$ or NO$_2^-$. In almost all lakes, the molar concentration of PO$_4^{3-}$ was higher than that of nitrite, which is typical as NO$_2^-$ is quickly oxidized to NO$_3^-$ in the presence of oxygen (Goldman and Horne 1983) (See Fig 1.2). Across our samples, AOB and AOA were significantly correlated with a few individual variables. For example, AOB abundance was most strongly correlated with PO$_4^{3-}$ concentrations (Pearson’s $r^2=0.32, p<0.1$), while AOA abundance was most strongly correlated with elevation (Pearson’s $r^2=0.43, p<0.05$). AOB displayed an opposing trend with elevation—increasing with altitude—but this relationship was not significant (Pearson’s $r^2=0.21, p=0.28$). Both groups showed wide variation in Middle Sunrise Lake at 2826 m. These patterns are evident in Figure 1.3, where AOB were notably more abundant—but also variable—in the Gaylor Lakes at >3000 m elevation, whereas AOA were more abundant and variable at lower elevations. For all samples collected in the three lakes >3000 m elevation, AOA were only detected four times.

Analysis of sites within specific months (e.g., statistically comparing only sites only sampled in the month in June) yielded stronger correlations between environmental and spatial factors and AOA/AOB abundance. In June, the most significant factor for both AOA and AOB abundance was PO$_4^{3-}$ (AOA Pearson’s $r^2=0.84, p<0.005$, AOB Pearson’s $r^2=0.76, p<0.05$). In July, AOB also had a strong positive correlation with PO$_4^{3-}$ (Pearson’s $r^2=0.84, p<0.01$), and there was a strong negative relationship with elevation for AOA (Pearson’s $r^2=0.91, p<0.01$). In August and September, there were no significant correlations between AOA/AOB abundance and environmental and spatial factors. These data suggest that PO$_4^{3-}$ may be an important variable governing N-cycling organisms and are consistent with work by Sundaweshar et al. (Sundareshwar et al. 2003), which showed P-limitation of N-cycling. However, differences in slope indicate that the form of this relationship varies from month to month. This may reflect competition with other organisms for P, or changing P demands among AOB—e.g., through differences in community composition or growth rates (Elser et al. 2000). The inverse relationship between AOA abundance and increasing elevation may also reflect the effect of increased (UV) radiation at higher elevations. Our interpretation of these data is that P limits AOA and AOB growth in June, whereas high light levels inhibit
AOA in July, especially at higher elevations. We used redundancy analysis (RDA) to analyze these multivariate relationships between AOA and AOB abundance and spatial (elevation and longitude), temporal (sampling date) and environmental (NO$_3^-$, NO$_2^-$, NH$_4^+$ and phosphate concentrations) data. 33% of the variability in AOA and AOB abundance was explained by these data, with nutrient concentrations accounting for 24% of the variability in AOA and AOB abundance, and site location and sampling date accounting for 9% of the variability. Collectively, nutrients explain nearly a quarter of the variation in AOA and AOB abundance in these lakes, and this includes ammonium, as well as nitrate, nitrite, and phosphate, concentrations. The modest percentage of constrained variance overall indicates that AOA and AOB populations are affected by other, unmeasured factors, or this may reflect stochastic variation in populations. In lakes, AOB and AOA abundance may be affected by active growth, but also by transport of cells into lakes via air, water, or suspended particles; trophic interactions, such as grazing and viral infection and lysis; and competition with other organisms (especially heterotrophic microbes) for ammonium.

The prevalence and abundances of AOA and AOB suggest that active N cycling and nitrification may be occurring in high-altitude lakes, but is surprising given that high-altitude oligotrophic lakes have low N concentrations and experience high light levels. To determine whether AOA and AOB may be active under these conditions, we measured ammonia oxidation rates in three lakes (Lukens Lake, Lower Cathedral Lake and Lower Gaylor Lake) using $^{15}$NH$_4^+$ incubations. These were conducted in both light and dark bottles that were incubated within the lakes under in situ conditions. We detected ammonia oxidation in all of these lakes and in both light and dark bottles (See Table 1.2 and Figure 1.4). While the rates were low, this is due to the fact that the total pool of measured nitrate + nitrite is small. In other words, the $^{15}$N values were demonstrably enriched (8.4-20.5 per mil) compared with in situ $^{15}$N, but a small pool has been enriched and we compute a low rate. On average dark bottles had 36% higher rates than bottles incubated in the light, which is consistent with light limitation of ammonia oxidation during daylight. Overall, these data indicate that during the night—and possibly during daylight—ammonia is actively oxidized within Yosemite’s oligotrophic lakes. This expands the range of habitats in which ammonia oxidation is known to occur to include the water columns of high-altitude, freshwater lakes.

The results of this study demonstrate that AOB and AOA are frequently present in high-altitude, oligotrophic, freshwater lakes, and that the abundance of both varies between lakes and over time within individual lakes. However, AOB were more commonly detected and were more abundant than AOA under all sampled conditions. The presence of AOA is consistent with the idea that AOA are adapted to the low-nutrient conditions that are characteristic of these lakes—for example, the marine AOA *Nitrosopumilus maritimus* has a remarkably high affinity for ammonia and appears to be adapted to life under extreme nutrient limitation (Martens-Habbena et al. 2009). The fact that AOB are abundant at high altitudes fits with the recent idea that they are more light-tolerant than AOA (Merbt et al., French et al. 2012), and high altitude lakes may also experience higher fluxes of N that would periodically favor AOB (Clow et al. 2010). Both AOA and AOB may contribute to nitrification in freshwater oligotrophic lakes: we detected low
rates of ammonia oxidation in three lakes, and in both light and dark bottle treatments. Our data add to the limited information available on microbial contributions to N cycling in lakes—particularly for AOA/AOB and nitrification—and additional work should expand these approaches to additional lakes and additional sampling periods. Nitrification under ice could be important during winter, and nitrification may be especially significant following spring snowmelt, when pulses of N likely enter these lakes, and may be metabolized by microbes. Our findings are indicative of dynamic microbial communities and internal N cycling in high altitude lakes.
Microbial Community Composition, Diversity, and Abundance in High-Altitude Oligotrophic Lakes of the Sierra Nevada, California

ABSTRACT

High-altitude lakes receive and process large volumes of material from their watersheds, functioning as biogeochemical integrators of the surrounding landscape. Lakes therefore have a disproportionate influence on global carbon and nitrogen cycling, yet they are also ‘sentinels’ of global change that may be significantly altered by anthropogenic activities. Despite the fact that microbial communities largely control biogeochemical cycling in lakes and provide critical ecosystems with services through processing and transformation of elements, our understanding of freshwater microbial ecology and biogeochemistry is still limited, particularly within high-altitude lakes. In light of this, we used Illumina sequencing of 16S ribosomal ribonucleic acid (rRNA) to examine microbial community composition, abundance, and diversity in relation to spatial, temporal, and geochemical data within lakes located along an elevation gradient in Yosemite National Park, California, USA. Elevation, temperature, atmospheric nitrogen deposition, and exposure to UV all vary along this transect, allowing us to examine what variables may influence microbial communities in these systems. Our sampling captured a diverse and sizable microbial community that is dominated by a handful of bacterial classes. A redundancy analysis (RDA) of microbial operational taxonomic units (OTUs) showed elevation and date to be the two most significant factors in determining the composition of microbial communities in these systems. In addition, we found a strong distance-decay relationship between microbial community diversity and inter-lake distance. Collectively, this data demonstrates variability in microbial abundance, diversity, and community composition within lakes and among different lakes based on spatial, temporal, and environmental heterogeneity.
INTRODUCTION
Lakes receive and process large volumes of material within their watersheds and therefore function as integrators and processors of material from the surrounding landscape (Williamson et al. 2009). Due to their relatively high concentration of biogeochemical cycling activity, lakes have a disproportionate influence on principal global biogeochemical cycles (e.g. carbon and nitrogen) (Cole et al. 2007, Finlay et al. 2013). This is critical as the rates and nature of these processes in lakes have the potential to be significantly influenced by anthropogenic activities — i.e., they are considered ‘sentinels’ of global change (Tranvik et al. 2009). Microbes largely control biogeochemical cycling in lakes, yet the diversity, composition, and activity of freshwater microbial communities are not well understood (Newton et al. 2011). ‘Typical’ freshwater bacteria and their phenology were identified only in the past decade, and it is not fully understood how increased nutrient concentrations and fluctuating temperatures may influence microbial community composition and activity (Comte and del Giorgio 2010, Newton et al. 2011). Understanding the dynamics of these microbial communities and the environmental factors that govern their activity and distribution is essential in order to determine the response of these systems to future perturbations (e.g., increasing nutrient pollution and global climate change).

Determining the relationship of microbial community composition to biogeochemical cycling requires understanding the factors that influence community structure in these systems. Although spatial patterns of macrofauna have been documented in many types of habitats, these patterns of diversity are not well explored in microbial communities (Martiny et al. 2006). There is evidence that suggests that microbes do follow some classic ecological patterns (i.e. latitudinal gradients, taxa-area relationships, distance-decay relationships, and co-occurrence patterns) (Fuhrman 2009). However, studies show that environmental heterogeneity is a primary factor in determining community structure (Martiny et al. 2006). In the Sierra Nevada’s chain lakes, there is support for microbial community structure being determined by within-lake environmental selection (Nelson et al. 2009), however another regional scale study of high-elevation lake systems the Spanish Sierra Nevada Mountains found that geographic distance between lakes—rather than environmental variation—was the primary determinant of bacterial community composition (Reche et al. 2005).

This study applied recently developed high-throughput sequencing technologies to describe the microbial communities of multiple Sierra Nevada lakes and explored how spatio-temporal and environmental variation affect microbial community composition. In oligotrophic high-altitude lakes — of which there are roughly 4,000 above 2,500 m in the Sierra Nevada— microbial communities have a significant influence on productivity as they determine the eventual form and fate of multiple ecologically important compounds including carbon, nitrogen, and phosphorous. Although past studies have characterized the community composition and diversity of microbial communities along elevational gradients (Bryant et al. 2008, Wang et al. 2012) and within multiple high-elevation lakes in the Sierra Nevada (Nelson et al. 2009), these studies focused primarily on the
composition of the bacterial fraction of the microbial communities (Bryant et al. 2008, Wang et al. 2012) and used molecular methods for characterizing community composition that are dated relative to current technologies. The recent development of new high-throughput sequencing technologies for metagenomic analysis of microbial communities allows for a finer-scale examination of microbial diversity and community composition relative to previous approaches. Moreover, the Illumina Mi-seq sequencing used in this study provides a sizable increase in the number of sequences yielded compared with all other existing sequencing technologies (e.g. 454 pyrosequencing), allowing for a more precise description of microbial communities than previously achieved (Williamson et al. 2009).

MATERIALS & METHODS

Study Site. The Sierra Nevada (California) is a 400-mile long mountain range that gradually rises from the valley floor from west to east and reaches an apex of 3,000 to 4,200 meter peaks on its eastern edge. Vegetation along the mountain range is composed of grasslands and foothill woodlands at lower elevations, with a transition to mixed conifer forests and then alpine meadows and lakes at higher elevations. Aquatic ecosystems in the Sierra Nevada are located downwind of large urban or agricultural areas that emit high levels of nitrogen (Clarisse et al. 2010) and can experience elevated levels of nitrogen deposition (Bytnerowicz et al. 2002). This deposition is known to increase nitrogen concentrations in lakes (Fenn et al. 2003), and in the case of the Sierra Nevada represents a large fraction of the nitrogen input to high-elevation lakes: (Baron 2006) established a critical load threshold of 1.5 kg N ha\(^{-1}\) year\(^{-1}\) for high-elevation lakes located in Rocky Mountain National Park, yet current annual nitrogen loading in the Sierra Nevada (i.e. Emerald Lake Watershed, Sequoia National Park) ranges from 2.0 to 4.9 kg N ha\(^{-1}\) year\(^{-1}\) (Sickman et al. 2001). Moreover, the watersheds are high in granitic parent material and generally have thin soils; both characteristics cause aquatic ecosystems in the Sierra Nevada to have limited buffering capacity in terms of their ability to neutralize foreign chemical species (Meixner et al. 2004). Clow et al. (2010) suggest that this property, coupled with high precipitation at high elevations, leads to high nitrogen loading at high elevations despite greater distances from emission sources. Our data are relevant to this as our selected sites range from 2,300 m (Harden Lake) to 3,160 m (Upper Gaylor Lake) and our transect terminus is adjacent to the steep Sierra escarpment (see Fig. 1.1.2).

Sampling. Water samples were collected from lakes located in Yosemite National Park during the 2012 summer (for a list of sampled sites see Table 1.1 and for a map see Fig. 1.1.1). Water samples were collected from the middle of each lake at 1 m below the surface and 1 m above the lake bottom using a Van Dorn water sampler (Lamotte). Duplicate water samples were filtered in the field onto 0.22 µm PVDF Membrane Filters (Millipore) using sterile 60 mL polycarbonate syringes (Cole-Parmer). After filtration, filter membranes were packed into bead tubes (MP Biomedicals) containing 800 µL of Sucrose-Tris-EDTA. Filtrate was collected in 60 mL HDPE Bottles (Nalgene) for subsequent nutrient analysis. Samples were transported on ice to the University of
DNA Extraction and quantification. DNA extraction followed Beman et al. (2012), using Sucrose-Tris-EDTA (STE) lysis buffer, sodium dodecyl sulfate (SDS), and proteinase K with bead-beating. DNA was further purified using a DNeasy Blood and Tissue Extraction Kit (Qiagen) and resolubilized in 50 uL of ultra-pure DNA free water (Qiagen). DNA was quantified using a PicoGreen dsDNA quantification kit (Invitrogen) and an Mx3005P real-time thermocycler (Agilent Technologies).

Nutrient Analyses. Filtered lake water samples were analyzed for orthophosphate (μmol l⁻¹) nitrite (μmol l⁻¹) and nitrate (μmol l⁻¹) using flow-injection analysis on a QuikChem 8000 (Zellweger Analytics, Inc.) at the University of California, Santa Barbara Marine Sciences Institute Analytical Laboratory (standard curve r² = 0.999 for all assays). Filtered lake water samples were analyzed for ammonium (NH₄⁺) following Holmes et al. 1999 (Holmes et al. 1999). An 8 mL of filtered water was combined with 2 mL of reagent that consisted of 95% 0.1M sodium tetraborate, 0.015M O-phtaldialdehyde, 0.03mM sodium sulfite, and 5% ethanol. After aging, samples were measured in triplicate for fluorescent intensity using a fluorometer (Trilogy® Laboratory Fluorometer, Turner Designs). Standards ranged from 31.2 to 186.8 nM and for different runs, standard curve r² = 0.998–0.999.

Illumina Sequencing. Thirty-four DNA samples representing extracted DNA from Yosemite lake-water column samples were sequenced using the Illumina Mi-seq platform located at the Argonne National Laboratory’s (ANL) Institute for Genomics and Systems Biology (Chicago, IL; http://www.igsb.anl.gov/). These samples were chosen based on temporal and spatial characteristics in order to maximize spatio-temporal diversity of sequenced samples and were diluted to a DNA concentration of 1 ng/uL before shipment to ANL. Prior to sequencing, 16S rRNA was amplified in triplicate via Polymerase Chain Reaction (PCR) using primers 515F (forward primer) 5′-GTGCCAGCMGCCGCGGTAA -3′ and 806R (reverse primer) 5′-GGACTACHVGGGTWTCTAAT -3′. The reverse primer (806R) was followed by a sample-identifying 12-base Golay barcode so that sequencing results could be extracted on a sample-by-sample basis. The 515F and 806R are universal archaeal and bacterial primers used to amplify sections of the 16s rRNA (Caporaso et al. 2012). The PCR reaction mixture consisted of PCR Grade H₂O (13.0 µL), 5 Prime Hot Master Mix (10.0 µL) (5 Prime Inc., Gaithersburg, MD), Forward Primer (0.5 µL at 10µM Concentration), Reverse Primer (0.5 µL at 10µM Concentration), and DNA (1.0 µL) for a total reaction volume of 25.0 µL. PCR conditions included 35 cycles consisting of 94°C for 3 min 45 sec, 50°C for 60 seconds, followed by a final cycle of 72°C for 10 min and then a 4°C hold. The barcode-tagged 16S rRNA PCR products were then pooled. In preparation for sequencing, 15-30% denatured PhiX Control V3 (Illumina Inc.) was added to all samples. PhiX Control V3 is a control library that is generated from the PhiX virus and is used as a control for Illumina sequencing runs. Subsequently, sequencing was performed over a 750 base-pair length region covering V3-V6 of 16S rRNA variable regions. MiSeq
Control Software (Illumina Inc.) was used to capture and process images of clusters on the flow cell.

An analysis of Illumina results was performed using the Mothur bioinformatics program (Schloss et al. 2009). Forward and reverse sets of reads were combined into single contigs and any sequences with ambiguous bases and reads longer than 275 base pairs (75% of average) were removed. Sequences were then screened to the majority start and end positions, and homopolymers with a maximum length of eight were removed. Subsequently, sequences were aligned with the Greengenes 16s rRNA gene database (DeSantis et al. 2006) to ensure quality control across all data; this database was chosen as it is a universal alignment that contains both archaea and bacteria. Sequences were then pre-clustered using the algorithm described by Huse et al. (2010) and chimeras were removed.

After pre-processing, sequences were classified using the SILVA archaeal, bacterial, and eukaryotic database (Quast et al. 2013). This database was selected as it is the longest (Schloss 2009) and yields the highest-quality results. A 65% cutoff was used as the probability that particular sequences fell under particular taxonomic classifications. Sequences were then subsampled based on the size of the smallest library (i.e. 40,000) in order to eliminate bias based on sequencing effort. This yielded a total of 1,735,535 bacterial sequences, 349,244 archaeal sequences, and 16,624 eukaryotic sequences in the samples.

**Diversity Calculations.** In order to explore alpha diversity within individual samples, I first calculated uncorrected pairwise distances between aligned DNA sequences with a cutoff for OTUs formation at 0.10. The cluster command was then used to assign sequences to OTUs using the furthest neighbor method, which clustered all sequences with an OTU that were at most 3% distant from all other sequences within the same OTU. Subsequently multiple statistical methods, including the ACE richness estimator (Chao et al. 1993), the Chao1 richness estimator (Chao 1984), the inverse of the Simpson diversity estimator (Simpson 1949), and the Shannon index (Shannon 1948), were applied to explore the species richness of individual samples relative to the number of OTUs found.

To examine beta diversity among separate samples, all samples were clustered using the furthest neighbor method. Afterward, statistical methods, including the Bray-Curtis Dissimilarity index (Bray and Curtis 1957), the abundance-based Jaccard dissimilarity index (Chao et al. 2005), the abundance-based Sorensen dissimilarity index (Chao et al. 2005), the Yue and Clayton measure of dissimilarity (Yue and Clayton 2005), the non-metric multidimensional scaling function (NMDS) (Borg 1997), and principal coordinates analysis (PCoA), were applied to analyze beta diversity among separate lakes. Results were then compared between sampling sites based on the pairwise distances, which were calculated using sampling site geographic coordinates and geographic distance matrix generating software (Ersts 2013).
Statistical analyses were performed using the R statistical environment (http://www.r-project.org/) using the vegan package.

RESULTS & DISCUSSION

Bacterial Community Composition
Examination of classified bacterial sequences (based on a ≥ 65% similarity level) revealed a total of 51 bacterial phyla in all lakes. However, lakes were dominated by a few groups of widespread freshwater bacteria, including Proteobacteria (>730,000 sequences), Actinobacteria (>715,000 sequences), Bacteroidetes (>324,600 sequences), Verrucomicrobia (>176,000 sequences), Cyanobacteria (>55,800 sequences), and Planctomycetes (>15,800 sequences). The remaining phyla made up <0.7% of the total number of identified bacterial sequences. These observations are inline with Newton et al. (2011) and Zwart et al. (2002) who both reported that within an extensive ecological and geographic range of freshwater lakes that Actinobacteria and Proteobacteria are the two most common bacterial phylum found in water-columns, and that Bacteroidetes and Verrucomicrobia commonly comprise a large percentage of the remaining community.

Bacterial sequences were further classified by bacterial class and separated based on sampling site and location (see Table 2.1 and Figs. 2.1.1-2.1.4 and 2.2.1-2.2.6). When all sequences were combined, the most abundant classes were Actinobacteria representing 35.1% of all bacterial sequences, followed by Betaproteobacteria (27.9%), Sphingobacteria (14.2%), Alphaproteobacteria (6.1%), Opitutae (5.0%), and Spartobacteria (3.3%). Of the 31 different samples, Actinobacteria was the most abundant in 24 of them, ranging from 17.2% to 47.1% of total bacterial sequences. Betaproteobacteria represented the most abundant class in six lakes and was the second most abundant bacterial class in 24, ranging from 19.0% to 40.8% of total bacterial sequences. In 26 out of 31 samples, Sphingobacteria was the third-most abundant bacterial class, ranging from 7.0% to 23.6% of total bacterial sequences. This was followed by Alphaproteobacteria, the fourth-most abundant on average and ranging from 1.1% to 12.7% of total bacterial sequences, and Opitutae (on average the fifth most abundant), ranging from 0.01% to 20.1% of total bacterial sequences. Interestingly, Spartobacteria, which ranged from 0.01% to 30.53% (mean = 3.26%) of total bacterial sequences, was the most abundant bacterial species in Lower Sunrise Lake during the September 2012 sampling period, comprising 30.5% of the total bacterial sequences. Remaining bacterial classes included Gammaproteobacteria (range 0.1% to 8.8%, mean=1.37%), Flavobacteria (range 0.01% to 6.5%, mean=1.27%), and Planctomycetaceae (range 0.1% to 8.25%, mean=0.77%). All other classes were never greater than ~2% of any individual sample and had mean percentages of <0.3% of total bacterial sequences (for listing of remaining bacterial classes and percentages see Table 2.1). On the family taxonomic level the three most abundant bacterial classes were primarily represented by the Sporichthyaceae (97% of all Actinobacteria sequences), Burkholderiales (90% of Betaproteobacteria sequences) and Chitinophagaceae and Cytophagaceae (68% and 24% of Sphingobacteria sequences, respectively).
Spatio-Temporal & Environmental Effects on Bacterial Community Composition

Across the 16 most abundant bacterial classes (all listed in Fig. 2.1), Pearson’s product-moment correlations were applied to examine spatio-temporal (elevation and date) and environmental (NO$_3^-$, NO$_2^-$, NH$_4^+$ and phosphate concentrations) effects on bacterial class abundance. For Actinobacteria there were negligible relationships with all factors. Date of sampling showed a moderately strong positive relationship with Chloroflexi ($r = 0.37$, $p < 0.05$) and unclassified-Proteobacteria ($r = 0.35$, $p < 0.05$) while having a strong negative relationship with Gammaproteobacteria ($r = -0.39$, $p < 0.05$). Elevation had a moderately negative relationship with Alphaproteobacteria ($r = -0.30$, $p < 0.1$) and Planctomycetacia ($r = -0.33$, $p < 0.1$) while having a strong positive relationship with Chloroflexi ($r = 0.40$, $p < 0.05$) and a moderately positive relationship with Flavobacteria ($r = 0.31$, $p < 0.1$). In regards to environmental effects on bacterial class abundance, of all environmental parameters measured, NH$_4^+$ concentration was most commonly correlated with abundance, having a strong positive relationship with OPB35 ($r = 0.25$, $p = 0.17$), Opitutae ($r = 0.25$, $p = 0.18$), and Cyanobacteria-SubsectionI ($r = 0.28$, $p = 0.14$). NH$_4^+$ exhibited moderately strong negative relationships with unclassified-Proteobacteria ($r = -0.36$, $p < 0.05$) and Spartobacteria ($r = -0.30$, $p < 0.1$). Phosphate concentration was very strongly correlated with unclassified-Proteobacteria abundance ($r = 0.91$, $p < 0.001$), though showed no significant correlation with other bacterial classes. NO$_3^-$ exhibited a strong positive relationship with Sphingobacteria ($r = 0.42$, $p < 0.05$), while NO$_2^-$ showed a strong positive relationship with unclassified-Proteobacteria ($r = 0.39$, $p < 0.05$).

UV stress resistance was suggested to be one of the reasons for the success of Actinobacteria in the upper waters of lakes, especially lakes with high UV transparency (Newton et al. 2011). In addition, one study demonstrated that there is an established positive relationship between Actinobacteria abundance and UV transparency in high-elevation lakes (Warnecke et al. 2005), though in a second study there was no observed correlation between these factors (Wu et al. 2006). Our samples showed a very weak correlation between Actinobacteria abundance and elevation ($r = -0.04$) and depth ($r = -0.1$), indicating that UV stress resistance is not a significant factor over the elevational gradient we sampled.

Temporal Effects on Lake Species Richness and Abundance

Across all sampled lakes, the number of OTUs (based on 97% similarity) varied from a low of 3,284 (Lower Sunrise Lake, 7/24/12) to a high of 5,976 (Ten Lakes #2, 6/30/12). Over the course of the sampling period (June to September 2012) there were no clear positive or negative trends in the number of OTUs. For example, in Harden Lake the highest number of OTUs was observed in June (5,316 OTUs), in July the number fell to 3,695 OTUs, and in August it rose to 4,457 OTUs (see Figs. 2.4.1 through 2.4.6). The highest average number of OTUs was found in Upper Gaylor Lake (43,418 OTUs) whereas the lowest average number was found in Lower Sunrise Lake (28,715 OTUs). Multiple diversity indexes were applied to determine the variability in species richness
and evenness within lakes over the four-month sampling period. These indexes included the ace-community richness estimator, the Shannon diversity index, the Simpson diversity index, and Smith and Wilson’s metric of community evenness (see Figs. 2.4.1 through 2.4.6). The ace-community richness values ranged from a low of 18,919 (Lower Sunrise Lake, 7/24/12) to a high of 77,598 (Ten Lakes #2, 6/30/12). Highest and lowest averages were observed in Upper Gaylor Lake (43,418) and Lower Sunrise Lake (28,715), respectively. The most variable ace-richness values were observed in Harden Lake and Lower Sunrise Lake (Coefficient of Variability (CV) = 0.42 and 0.41, respectively), whereas Upper Cathedral Lake was the least variable (CV = 0.16). For the Shannon diversity index, a measurement of overall biodiversity (Shannon 1948), values ranged from a low of 4.74 (Harden Lake, 6/25/12) to a high of 5.95 (Lower Gaylor Lake 6/4/12), with highest and lowest averages found in Harden Lake (5.32) and Lower Sunrise Lake (5.0). In Harden Lake, the Shannon Index was also the most variable (CV=0.08), whereas Upper Gaylor Lake was the least variable (CV=0.01). For the Simpson diversity values, which in this application is the probability that two randomly selected samples would belong the same OTU (Simpson 1949), the highest index level was Lower Sunrise Lake (0.04) and the lowest index level was Lower Cathedral Lake (0.02). The Simpson index was most variable in Lower Cathedral Lake (CV=0.36), whereas it was least variable in Upper Gaylor Lake (CV=0.02). For Smith and Wilson’s metric of community evenness, based on the variance of log abundances (Smith and Wilson 1996), the lowest and highest average values were found in Upper Gaylor Lake (0.15) and Lower Cathedral Lake (0.24), respectively. Upper Gaylor Lake also exhibited the highest variability (CV=-.39), whereas Upper Cathedral Lake was the least variable (CV=0.1).

Analysis of associations among spatio-temporal and environmental variables and alpha diversity indexes yielded no strong correlations, although we did find a moderate negative correlation between depth and Simpson diversity index \( (r = -0.37, p < 0.05) \). Altogether, there was relatively little temporal variability in the range of sequences observed or the applied indexes (i.e. ace-community richness, Shannon, Simpson and Smith-Wilson). It can be understood that the influence of temporal and environmental variability on both species’ richness and evenness is relatively small. However, more extensive observations (i.e. longer sampling period and more extensive geographic area) would clarify the important factors and long-term trends for alpha diversity in these systems.

**Beta-Diversity and Sierra Nevada Lakes**

Beta diversity among lakes was measured through application of multiple calculators (available in the Mothur bioinformatics program (Schloss et al. 2009)) that use various metrics to determine similarity or dissimilarity between lakes based on the presence, absence, and abundance of OTUs and similarity and dissimilarity was evaluated based on the inter-spatial geodesic distance between individual lakes (Ersts 2013) (Fig. 2.5.1 and Table 2.2).
Shared community richness between lakes was measured using two metrics; the *two-sample shared ACE richness estimator*, which calculates the shared ACE richness estimate for an OTU definition, and the *two or more sample shared Chao1 richness estimator*, which calculates the shared Chao1 richness estimate for an OTU definition. When plotted against inter-lake geodesic distance, both these metrics returned moderately strong negative correlations ($r = -0.35$ and $r = -0.38$ respectively, $p < 0.001$ for both) (see Fig. 2.5.1). These results indicate that there is a moderate negative relationship between increased inter-lake distance and the similarity in the richness of the community.

To measure the similarity in community structure between lakes I used four metrics: 1) the *abundance-based Jaccard dissimilarity coefficient* (Chao et al. 2005), which is the fraction of sequences that don’t belong to shared OTUs based on the Jaccard index, 2) the *abundance-based Sorenson dissimilarity coefficient* (Chao et al. 2005), which also is the fraction of sequences that don’t belong to shared OTUs, though based on the Sorenson index, 3) the *Smith theta similarity coefficient* calculator, which returns the Jaccard index Smith—a description of the differences between the structures of two communities, and 4) the *Yue and Clayton theta similarity coefficient* (Yue & Clayton 2005), which calculates the Yue and Clayton measure of dissimilarity between the structures of two communities. When plotted against inter-lake distance these metrics all returned variable correlations ($r = -0.33$, $r = 0.54$, $r = 0.29$, and $r = 0.28$ respectively, $p < 0.001$ for all) (see Fig. 2.6.1), though the abundance-based Sorenson dissimilarity coefficient showed a very strong correlation ($r = 0.54$) with distance.

The similarity in community membership between lakes was measured using the *traditional Jaccard similarity coefficient based on the observed richness*, which calculates the traditional Jaccard index describing the dissimilarity between two communities, the *Jaccard similarity coefficient based on the Chao1 estimated richness*, which estimates the traditional Jaccard index using Chao1 richness estimates for the richness of individual samples and the richness of the shared OTUs, and the *Sorenson similarity coefficient based on the observed richness*, which calculates the traditional Sorenson index describing the dissimilarity between two communities. When plotted against inter-lake distance, these metrics all returned strong correlations ($r = 0.53$, $r = 0.37$, $r = 0.37$ respectively, $p < 0.001$ for all) (see Fig. 2.5.1), suggesting a strong positive relationship between inter-lake distance and microbial community dissimilarity. This correlation is indicative of a strong distance decay relationship between microbial communities in lakes over the study-site landscape. These findings are in line with (Reche et al. 2005), who found that geographic location was an important factor in regards to formation of microbial community structure in similar ecosystems to the Sierra Nevada in California (i.e. the Sierra Nevada, Spain).

**Redundancy Analysis.** To determine in greater detail how environmental factors and spatio-temporal variability collectively influenced microbial community composition, I used multivariate redundancy analysis (RDA) to analyze variation in community composition (number of OTUs per lake) (total number of unique OTUs for all sites = 48,839 at 97% similarity cutoff) and spatial (elevation), temporal (sampling date), and environmental ($\text{NO}_3^-$, $\text{NO}_2^-$, $\text{NH}_4^+$, N:P Ratio and phosphorous concentrations) factors.
Cumulatively, spatio-temporal and environmental variability explained 29% of the presence, absence, and abundance of OTUs (constrained variance=29%, $F=1.45$, $P<0.05$; Figs. 2.3.1 and 2.3.2). Lake elevation and sampling date accounted for ~14% of the constrained variability, while environmental conditions accounted for 15% of constrained variability. As the spatio-temporal variability was limited to only two factors as opposed to six for the environmental variability, these results suggest that microbial community composition is more strongly related to elevation and date of sampling relative to lake environmental conditions. In the RDA this stronger influence is depicted in Figures 2.3.1 and 2.3.2. On the horizontal axis (RDA1, 14% of constrained variability) the most influential constraining variable was elevation (biplot score = -0.71) followed by date (biplot score = -0.59), $\text{NO}_2^-$ (biplot score = 0.46), N:P Ratio (biplot score = -0.30), $\text{NO}_3^-$ (biplot score = -0.26), $\text{NH}_4^+$ (biplot score = -0.18), and phosphorous (biplot score = 0.07). On the vertical axis (RDA2, 7% of constrained variability) the most influential constraining variable was $\text{NO}_3^-$ (biplot score = -0.46), followed by the N:P ratio (biplot score = -0.21), $\text{NO}_2^-$ (biplot score = 0.19), $\text{NH}_4^+$ (biplot score = -0.15), elevation (biplot score = 0.13), phosphorous (biplot score = -0.10), and date (biplot score = 0.04). These results demonstrate that multiple factors interact to determine which specific bacterial populations are most abundant, although within these selected systems elevation and date appear to be the most influential contributory factors. These results are in line with previous findings that described seasonality (i.e. date of sampling) (Nelson 2009) and geographic location (Reche et al. 2005) as the most important factors for determining lake microbial community composition.

Non-Metric Multidimensional Scaling. In order to illustrate the influence of spatio-temporal variability on beta-diversity in as few dimensions as possible, I used non-metric multidimensional scaling (NMDS) to evaluate the association between Sorenson abundance index calculations and the presence, absence, and abundance of OTUs. These results were then plotted to visualize differences between lakes based on elevation (Fig. 2.6.1), date (Fig. 2.6.2), and individual lake microbial community (Fig. 2.6.3). In regards to elevation, visually there is a noticeable gradient from top to bottom and right to left representing declining similarity with elevation. This relationship was analyzed for statistical significance using an analysis of molecular variance test (amova). The amova tests whether the centers of the clouds representing OTUs are more separated than the variation among samples of the same elevation using distance matrixes (Schloss et al. 2009). For elevation (Fig. 2.6.1), this showed that the observed separations between different elevations are statistically significant ($p$-value: <0.001). In addition, time of sampling (across all months) (Fig. 2.6.2) was also statistically significant ($p$-value <0.01), as was separation based on individual lakes ($p$-value<0.001) (Fig. 2.6.3).

CONCLUSIONS
High-altitude lakes of the Sierra Nevada harbor dynamic microbial communities that are collectively dominated by a few bacterial classes. However, this is not surprising in these lakes given that this is typical of other freshwater systems where Actinobacteria and Betaproteobacteria are consistently dominant (Newton et al. 2011). On a more precise
taxonomic scale, our results show that there is a strong distance-decay relationship with increasing inter-lake distance for microbial community diversity in Sierra Nevada lakes. These results are in line with previous findings showing that lake location is an important factor in regards to formation of microbial community structure. Based on the RDA analysis of sequencing results, the two most significant factors for determining within-lake microbial community composition were lake elevation and sampling date. These results suggest that the eventual structure of microbial communities is determined through a complex interplay between multiple spatio-temporal and environmental variables. Alpha diversity was essentially unchanging within individual lakes during our sampling period, however the period of sampling (i.e. June to September) did not cover a large degree of the seasonal events (e.g. fall overturn, spring melting) that are strong drivers for changes in alpha diversity. Given this sequence of diversity and ecological diversity within these systems and the inter-lake differences, our findings prove that these systems bear diverse and active microbial communities whose structures are based on multiple variable factors.
References:


Ersts, P. J. 2013. Geographic Distance Matrix Generator(version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation.


Sickman, J. O., A. Leydecker, and J. M. Melack. 2001. Nitrogen mass balances and abiotic controls on N retention and yield in high-elevation catchments of the


LIST OF FIGURES AND TABLES

FIGURES

Figure 1.1.1: Map of sampling locations in Yosemite NP…………………….…………33
Figure 1.1.2: Plot of sampling site elevation versus longitude………………………35
Figure 1.2: Environmental data for individual sites……………………………………36
Figure 1.3: Boxplot comparison of AOA/AOB abundance versus elevation………………37
Figure 1.4: Bar graph of observed ammonia oxidation rates……………………….……38
Figures 2.1.1 through 2.1.4: Bar graphs of bacterial community composition ………39
Figures 2.2.1 through 2.2.6: Area graphs of bacterial community composition………..42
Figure 2.3.1: Temporal multivariate redundancy analysis (RDA) plot………………….48
Figure 2.3.2: Spatial multivariate redundancy analysis (RDA) plot………………….…49
Figures 2.4.1 through 2.4.6: Alpha Diversity calculation plots………………………50
Figure 2.5.1: Combined Beta Diversity calculation………………………………………55
Figures 2.6.1 through 2.6.3: non-metric multidimensional scaling (NMDS) plots……56

TABLES

Table 1.1: List of lake names, elevation, location and depth…………………………34
Table 1.2: List of ammonia oxidation rates by lake……………………………………38
Table 2.1: Percent community composition for most abundant bacterial classes……..46
Table 2.2: Monthly R^2 for Beta diversity calculation versus inter-lake distance……….53
Figure 1.1.1: Sampling Locations (yellow circles) in Yosemite National Park (green shaded area of inset map). Sample locations are listed below in order of farthest west to east.
Table 1.1: table listing lake names, lake elevation, lake longitude, latitude, and maximum observed depth. Lakes are listed in geographic order from west to east.

<table>
<thead>
<tr>
<th>Lake Name</th>
<th>Elevation (m)</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Max Depth (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harden Lake</td>
<td>2289</td>
<td>-119.6764</td>
<td>37.8957</td>
<td>4.5</td>
</tr>
<tr>
<td>Lukens Lake</td>
<td>2506</td>
<td>-119.6160</td>
<td>37.8598</td>
<td>6</td>
</tr>
<tr>
<td>Grant Lakes</td>
<td>2822</td>
<td>-119.5392</td>
<td>37.8867</td>
<td>4.5</td>
</tr>
<tr>
<td>Ten Lakes #1</td>
<td>2727</td>
<td>-119.5260</td>
<td>37.9045</td>
<td>10</td>
</tr>
<tr>
<td>Ten Lakes #2</td>
<td>2866</td>
<td>-119.5143</td>
<td>37.9004</td>
<td>16</td>
</tr>
<tr>
<td>Lower Sunrise</td>
<td>2801</td>
<td>-119.4533</td>
<td>37.8047</td>
<td>5.5</td>
</tr>
<tr>
<td>Middle Sunrise</td>
<td>2826</td>
<td>-119.4497</td>
<td>37.8087</td>
<td>5</td>
</tr>
<tr>
<td>Lower Cathedral</td>
<td>2832</td>
<td>-119.4233</td>
<td>37.8453</td>
<td>10</td>
</tr>
<tr>
<td>Upper Cathedral</td>
<td>2923</td>
<td>-119.4147</td>
<td>37.8395</td>
<td>3.5</td>
</tr>
<tr>
<td>Elizabeth Lake</td>
<td>3050</td>
<td>-119.3697</td>
<td>37.8453</td>
<td>9</td>
</tr>
<tr>
<td>Vogelsang Lake</td>
<td>3203</td>
<td>-119.3436</td>
<td>37.7871</td>
<td>-</td>
</tr>
<tr>
<td>Fletcher Lake</td>
<td>3101</td>
<td>-119.3406</td>
<td>37.7970</td>
<td>5</td>
</tr>
<tr>
<td>Evelyn Lake</td>
<td>3155</td>
<td>-119.3257</td>
<td>37.8061</td>
<td>7</td>
</tr>
<tr>
<td>Lower Gaylor</td>
<td>3064</td>
<td>-119.2861</td>
<td>37.9086</td>
<td>11</td>
</tr>
<tr>
<td>Upper Gaylor</td>
<td>3160</td>
<td>-119.2691</td>
<td>37.9124</td>
<td>7</td>
</tr>
</tbody>
</table>
Figure 1.1.2: Lake elevation plotted against lake longitude. Sampling site elevations and longitudes are listed in table 1.1.
Figure 1.2: Lake nutrient concentrations for phosphorous, nitrite (NO$_2^-$), ammonium (NH$_4^+$) and nitrate (NO$_3^-$) plotted against elevation. Colors denote which month sample was collected [June (red), July (blue), August (black) and September (green)]. The vertical axes represent nutrient concentrations in nanomolar (nM) and plotted on a base 10 log scale. The horizontal axes represent lake elevation in meters.
Figure 1.3: Boxplot comparison of number of gene copies per milliliter (genes mL$^{-1}$) of ammonia oxidizing archaea (AOA) and ammonia oxidizing bacteria (AOB) found in lake samples at 1 meter below the lake surface during June, July, August and September 2012. Lakes include (from left to right) Harden Lake, Lukens Lake, Lower Cathedral Lake, Upper Cathedral Lake, Lower Sunrise Lake, Middle Sunrise Lake, Lake Elizabeth Lower Gaylor Lake and Upper Gaylor Lake. The vertical axes (log 10 base scale) denotes the number of genes mL$^{-1}$ based on quantitative polymerase chain reaction (qPCR) of the ammonia monooxygenase subunit A gene in AOB (bottom plot) and AOA (top plot). Outliers (represented by blue and red circles) are samples that were more than 3/2 times larger than the upper or lower quartile. The horizontal axis represents the elevation of the sampling site, and is ordered from lowest elevation (Harden Lake) to highest elevation (Upper Gaylor Lake).
Table 1.2: calculated ammonia/ammonium (NH$_3$/NH$_4^+$) oxidation rates measured in Lukens, Lower Cathedral, and Lower Gaylor Lakes. Triplicate dark/triplicate light samples were incubated in situ for 24-hours under zero-light/light conditions at 1 m depth.

<table>
<thead>
<tr>
<th>Lake</th>
<th>Treatment</th>
<th>Avg nmol/L*d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lukens</td>
<td>dark</td>
<td>0.011</td>
</tr>
<tr>
<td>Lukens</td>
<td>light</td>
<td>0.009</td>
</tr>
<tr>
<td>Lower Cathedral</td>
<td>dark</td>
<td>0.013</td>
</tr>
<tr>
<td>Lower Cathedral</td>
<td>light</td>
<td>0.011</td>
</tr>
<tr>
<td>Lower Gaylor</td>
<td>dark</td>
<td>0.015</td>
</tr>
<tr>
<td>Lower Gaylor</td>
<td>light</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Figure 1.4: Boxplot comparison of calculated ammonia/ammonium (NH$_3$/NH$_4^+$) oxidation rates measured in Lukens, Lower Cathedral, and Lower Gaylor Lakes. The vertical axis represents the number of nanomoles of ammonia/ammonium oxidized per liter of lake water per day (nmol L$^{-1}$D$^{-1}$). The blue colored bars (dark) represent samples that were incubated in zero-light conditions during sample incubation, while the green colored bars represent samples that were exposed to ambient light.
Figures 2.1.1 through 2.1.4: 100% stacked bar graphs depicting bacterial community composition at the class taxonomic level based on illumina sequencing data. Sequences were classified using the SILVA bacterial database and a 65% cutoff was used as the probability that particular sequences fell under particular taxonomic classifications. The vertical-axes show the percentage of the total number of bacterial sequences that are represented by a specific bacterial genus. The specific bacterial genus is represented by the color of the bar graph (e.g. *Actinobacteria* is represented by the red portion of the bar graph and during July 2012 in Harden lake represented ~26% of total bacterial sequences within the sample). The legend displays bacterial groups that represented >1% of total bacterial sequences for any site (See Table 2.1 for specific values). The horizontal axes represent the lake that was sampled and are ordered from lowest elevation to highest elevation. Each different graph represents an individual month during which the samples depicted were collected.
Figure 2.1.1

Figure 2.1.2
Figures 2.2.1 through 2.1.6: 100% stacked area graphs depicting bacterial community composition at the **class** taxonomic level for individual sampling sites across the months of June, July, August, and September 2012. Community composition is based on illumina sequencing data. Sequences were classified using the SILVA bacterial database and a 65% cutoff was used as the probability that particular sequences fell under particular taxonomic classifications. The legend displays bacterial groups that represented >1% of total bacterial sequences for any site (See Table 2.1 for values). The horizontal axes represent the month that the sample was collected. Each different graph represents a different lake and are listed in order of elevation from lowest (Harden Lake) to highest (Upper Gaylor Lake). Lakes depicted had a minimum of three separate sampling timepoints.
Figure 2.2.1
Harden Lake

Figure 2.2.2
L. Cathedral Lake
Figure 2.2.1

Harden Lake

Figure 2.2.2

L. Cathedral Lake
Figure 2.2.1

Figure 2.2.2
Table 2.1: calculated % abundances of most prevalent bacterial classes. Blue colored cells represent most abundant bacterial classes while red cells represent least abundant bacterial classes.
Figures 2.3.1 through 2.3.2: Biplots of multivariate redundancy analysis (RDA) showing collective relationship between the presence, absence and abundance of unique OTUs observed at sampling sites (Total number of unique OTUs for all sites = 48839 at 97% similarity cutoff) and environmental (i.e. NO$_3^-$, NO$_2^-$, NH$_4^+$, N:P Ratio and phosphorous (P)) and spatio-temporal (i.e. date and elevation) parameters (constrained variance=29%, F= 1.45, P<0.05). The horizontal axis (RDA1) represents 14% of constrained variability while the vertical axis (RDA2) represents 7% of constrained variability. Arrows pointing towards spatio-temporal and environmental variables denote biplot scores for the constraining variables, with the longest independent variable vectors being most important in explaining variation in OTU presence, absence or abundance along that axis. The circles represent individual sampling sites (listed in table 2.1) and the numbers adjacent represent which site. The color key denotes which month the sample was collected (Figure 2.3.1) or the elevation of the sampling site (Figure 2.3.2). Constraining variable biplot scores are scaled to 100 times their value in the RDA analysis.
Fig 2.3.1
Figure 2.3.2
Figures 2.4.1 through 2.4.6: Bar graphs depicting lake alpha diversity calculations based on illumina 16S rRNA sequencing (vertical axes) and date sampled (horizontal axes). Lakes represented were all sampled a minimum of three times during the summer of 2012 and include Harden Lake (Fig 2.4.1), Lower Cathedral Lake (Fig 2.4.2), Upper Cathedral Lake (Fig 2.4.3), Lower Sunrise Lake (Fig 2.4.4), Lower Gaylor Lake (Fig 2.4.5) and Upper Gaylor Lake (Fig 2.4.6) and are listed in order of elevation from lowest (Harden Lake) to highest (Upper Gaylor Lake). Each separate figure includes 5 bar graphs that depict (listed from top to bottom) the number of sequences observed and four diversity calculations including the ACE richness estimator, the Shannon index, the Simpson diversity index, Smith and Wilson's metric of community evenness, and the log series which tests whether observed data follow the log series distribution.
Table 2.2: $R^2$ values for beta diversity calculations based on illumina 16S rRNA sequencing and inter-lake geodesic distance. Each separate column represents a different month during which the samples were collected. Beta diversity calculations included the abundance-based Sorensen similarity coefficient (Bray and Curtis 1957) (Bray and Curtis 1957), shared Chao1 richness estimate, the shared ACE-richness estimate, the abundance-based Jaccard dissimilarity index, the traditional Jaccard similarity coefficient based on the observed richness, the Sorensen similarity coefficient based on the observed richness, the Jaccard similarity coefficient based on the Chao1 estimated richness, the Yue & Clayton theta similarity coefficient, the Sorensen similarity coefficient based on the Chao1 estimated richness, and the Smith theta similarity coefficient.

<table>
<thead>
<tr>
<th>Diversity Calculator</th>
<th>June ($R^2$)</th>
<th>July ($R^2$)</th>
<th>August ($R^2$)</th>
<th>September ($R^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorensen similarity coefficient</td>
<td>0.38</td>
<td>0.18</td>
<td>0.79</td>
<td>0.13</td>
</tr>
<tr>
<td>Shared Chao1 richness estimate</td>
<td>0.15</td>
<td>0.03</td>
<td>0.8</td>
<td>0.11</td>
</tr>
<tr>
<td>Shared ACE-richness estimate</td>
<td>0.08</td>
<td>0.05</td>
<td>0.8</td>
<td>0.12</td>
</tr>
<tr>
<td>Abundance-based Jaccard dissimilarity index</td>
<td>0.1</td>
<td>0.04</td>
<td>0.74</td>
<td>0.09</td>
</tr>
<tr>
<td>Traditional Jaccard similarity coefficient based on the observed richness</td>
<td>0.36</td>
<td>0.18</td>
<td>0.81</td>
<td>0.11</td>
</tr>
<tr>
<td>Sorensen similarity coefficient based on the observed richness</td>
<td>0.21</td>
<td>0.01</td>
<td>0.79</td>
<td>0.11</td>
</tr>
<tr>
<td>Jaccard similarity coefficient based on the Chao1 estimated richness</td>
<td>0.21</td>
<td>0.01</td>
<td>0.81</td>
<td>0.1</td>
</tr>
<tr>
<td>Yue &amp; Clayton theta similarity coefficient</td>
<td>0.13</td>
<td>0.00001</td>
<td>0.88</td>
<td>0.08</td>
</tr>
<tr>
<td>Sorensen similarity coefficient based on the Chao1 estimated richness</td>
<td>0.14</td>
<td>0.000004</td>
<td>0.8</td>
<td>0.08</td>
</tr>
</tbody>
</table>

*Bold numbers indicate a p-value < 0.1
Figure 2.5.1: Scatter plots depicting beta diversity calculations based on illumina 16S rRNA sequencing (vertical axes) and distance (km) between lakes (horizontal axes). Each separate graph represents a different diversity calculator. Beta diversity calculations included the abundance-based Sorenson similarity coefficient (Bray and Curtis 1957) shared Chao1 richness estimate, the shared ACE-richness estimate, the abundance-based Jaccard dissimilarity index, the traditional Jaccard similarity coefficient based on the observed richness, the Sorenson similarity coefficient based on the observed richness, the Jaccard similarity coefficient based on the Chao1 estimated richness, the Yue & Clayton theta similarity coefficient, the Sorenson similarity coefficient based on the Chao1 estimated richness and the Smith theta similarity coefficient.
Figure 2.5.1
Figures 2.6.1 through 2.6.3: Plots of non-metric multidimensional scaling (NMDS) (Stress = 0.16, $R^2 = 0.89$) results showing collective relationship between the presence, absence and abundance of unique OTUs observed at sampling sites (Total number of unique OTUs for all sites = 48839 at 97% similarity cutoff) and environmental (i.e. NO$_3^-$, NO$_2^-$, NH$_4^+$, N:P Ratio and phosphorous (P)) and spatio-temporal (i.e. date and elevation) parameters
Figure 2.6.1

Figure 2.6.2
Figure 2.6.3