MOLECULAR DIVERSITY OF BACTERIA FROM THREE DISTINCT ECOSYSTEMS WITHIN GREAT SMOKY MOUNTAINS NATIONAL PARK

By

Melissa Brooke Collins

A Thesis
Submitted to the
Faculty of the Graduate School
of
Western Carolina University
In Partial Fulfillment of
the Requirements for the Degree
of
Master of Science

Committee:

Seán O'Connell _____ Director

Debra Lindle

Katherine Mathews

_____ Dean of the Graduate School

Date: July 31, 2006

Summer 2006
Western Carolina University
Cullowhee, North Carolina

MOLECULAR DIVERSITY OF BACTERIA FROM THREE DISTINCT ECOSYSTEMS WITHIN GREAT SMOKY MOUNTAINS NATIONAL PARK

A thesis presented to the faculty of the Graduate School of Western Carolina University in partial fulfillment of the requirements for the degree of Master of Science

By

Melissa Brooke Collins

Director: Dr. Seán O'Connell
Assistant Professor of Biology
Department of Biology

July 2006

Acknowledgements

I would like to thank my advisor Seán O'Connell for all of his hard work, help, and dedication to not only myself and this project, but also any student in need. His passion for science is inspiring and makes working with him exciting and rewarding. I would also like to thank the Biology department and the Office of Graduate Studies at WCU for providing equipment and funding without which this research would not have been possible. Thank you to Kristina Reid, Heather Sink, Darby Harris, Barclay Taylor, and especially Philip Drummond for all of the time and energy that they donated to helping inside the lab. Finally, I would like to thank my family and Peter Gut for all of their love, support, and help with Aubrie during this time.

Table of Contents

List of Tables

List of Figures

Abstract

MOLECULAR DIVERSITY OF BACTERIA FROM THREE DISTINCT ECOSYSTEMS WITHIN GREAT SMOKY MOUNTAINS NATIONAL PARK

Melissa B. Collins M.S.

Western Carolina University (August 2006)

Director: Dr. Seán O'Connell

The number of microbial species in nature may be in the millions, but most have never been observed or detected (Hong et al. 2006). For over 100 years, studies have focused primarily on culturing species from environmental samples in order to examine diversity of the community. With advancements in molecular techniques, a shift has occurred in both the approaches used to create community profiles and to explain what these profiles look like. This knowledge of microbial diversity is crucial for our understanding of the structure, function, and evolution of biological communities.

The biodiversity of several thousand organisms has been catalogued throughout Great Smoky Mountains National Park (GSMNP) as part of a long term study called the All Taxa Biodiversity Inventory (ATBI). Recently, prokaryotes have become important within this study as well, and early work was focused on collecting data through culture-dependent techniques.

Here, I implemented a protocol, based completely on molecular techniques to create a library of species in order to describe the community of bacteria within ATBI plots. Through the use of Polymerase Chain Reaction (PCR), molecular cloning,

Restriction Fragment Length Polymorphism (RFLP), and DNA sequencing I have been able to compare the diversity of bacteria among three different ATBI plots. Identifications were made for 177 bacterial species representing eleven different phylum including *Acidobacteria, Firmicutes, Actinobacteria, Bacteroidetes, Verrucomicrobia, Planctomycetes, Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, Gammaproteobacteria,* and OP10. The community profiles detected via these methods provided a new outlook on what bacterial species were dominating these three plots compared to what the previous culture-dependent methods had suggested. Overall, the *Acidobacteria* and *Fimricutes* divisions dominated the entire community profile. Albright Grove had nine different divisions represented with the *Acidobacteria* dominating this site. Cataloochee and Purchase Knob both had eight different divisions represented with the *Acidobacteria* dominating at Cataloochee and the *Proteobacteria* dominating at Purchase Knob.

Microorganisms are extremely important and essential for all ecosystems; yet prokaryotes are the least understood of all organisms and the least defined taxonomically. Analyzing, comparing, and identifying these different bacterial species in GSMNP provides a better understanding of microbial distribution in soil environments. This allows for a better development of bacterial taxonomy and ultimately will help in understanding bacterial niches.

Introduction

"The key to taking the measure of biodiversity lies in a downward adjustment of scale. The smaller the organisms, the broader the frontier and the deeper the unmapped terrain" (Wilson 1994).

When looking at biodiversity on a global scale, rRNA phylogenetic trees have shown that the main extent of the Earth's biodiversity is microbial (Hugenholtz et al. 1998). One can easily observe this overwhelming trend via the universal tree of life. At one point the tree of life was divided into five major kingdoms, *Animalia, Plantae, Fungi, Protista and Monera*. In 1990 Carl Woese split the *Monera* Kingdom into two domains, *Archaea* and *Bacteria*, and combined the other four kingdoms into one domain, *Eukarya* (DeLong and Pace 2001; Figure 1).

It is now known that everywhere one finds life one also finds bacteria. This is because free-living bacteria are able to survive every environment that supports eukaryotes and even those that cannot (Cohan 2001). Hence, one can only imagine the tremendous amount of ecological diversity within the prokaryotic world. This in part could be due to the enormous potential for speciation within the bacterial domain. Large population sizes for bacteria and a rapid rate of reproduction contribute to the increase in opportunity for speciation compared to more highly sexual plants and animals. Plus, bacteria are highly adaptable (i.e., mutations are manifested more quickly due to higher

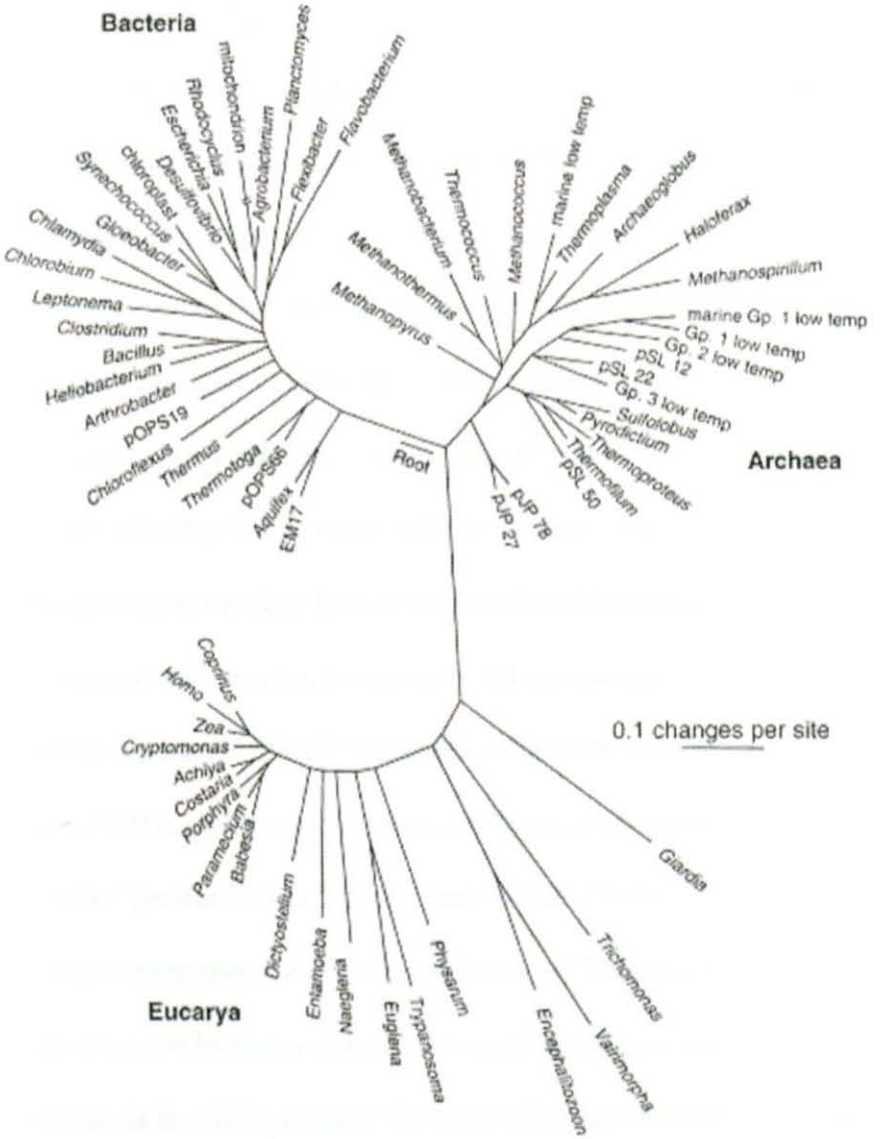

Figure 1. Universal phylogenetic tree based on comparison of small subunit rRNA sequences. Sixty-four rRNA sequences representative of all known phylogenetic domains were aligned, and a tree was produced with fastDNAml (Olsen et al. 1994). That tree was modified, resulting in the composite one shown, by trimming lineages and adjusting branch points to incorporate results of other analyses. The scale bar corresponds to 0.1 changes per nucleotide (Pace 1997). Figure obtained from Jurgens (2002).

reproduction rates) making them better able to adapt and to thrive in just about any environment (Madigan et al. 2003).

Soils sustain an immense diversity of microbes, which, to a large extent, remains unexplored (Curtis et al. 2002; Gans et al. 2005; Torsvik et al. 1990 a and b). In fact, assessing the diversity of bacteria in soil has been an ongoing issue for several years. This is due to the fact that the ability to measure diversity is a prerequisite for any systematic study of biogeography and community assembly (Curtis et al. 2002). Unfortunately, the extent of prokaryotic diversity is widely held to be beyond simple calculation and is left to more complex models. As knowledge of the microbial world is expanded, it seems that the estimation of total bacterial diversity grows. It is known that one gram of soil may harbor up to 10 billion microorganisms, and, it was thought, possibly thousands of different species (Torsvik and Ovreas 2002). In 1990, Torsvik et al. used DNA-DNA reassociation to estimate approximately 4,000-10,000 different bacterial "genomic units" in one gram of soil (Torsvik et al. 1990 a and b). These estimates were concluded through the use of DNA melting/reannealing data, which is likely the least biased molecular diversity technique used. The downside to this method, however, is that it is possibly the least informative method, only measuring total diversity (i.e., it is very sensitive to DNA heterogeneity but cannot be used to identify species). Torsvik et al.'s estimate is now thought to be low because they used a mathematical model that assumes all bacterial species in a sample are equally abundant (Gans et al. 2005). Gans et al. (2005) instead used quantitative comparisons of different species-

abundance models to increase the estimate to 10^7 different bacterial species in 10 grams of pristine soil.

Some think that microbial diversity cannot be estimated because many microbial accumulation curves are linear or close to linear because of high diversity and small sample size (Hughes et al. 2001). These accumulation curves are important because knowledge of the extent of phylogenetic diversity can indicate how many functional groups have not yet been accounted for (Schloss and Handelsman 2004). As of today there are 52 different bacterial phyla, and half of them are composed entirely of uncultured bacteria. (Figure 2). Also, three phyla contain less then 10% cultured members and six phyla contain more than 90% cultured members. Thus, it is apparent how much information is actually missing regarding bacterial species.

Even with this high diversity in soil, many of the organisms belong to groups for which no cultivated representatives are known. In fact, it is estimated that 1% or less of soil bacteria have been cultured (Hugenholtz et al. 1998). This means that DNA sequence data obtained by direct PCR amplification from the environment provides most of the information available for up to 99% of the prokaryotes in natural communities (Schloss and Handelsman 2004). Staley and Konopka (1985) coined the term "the great plate count anomaly" to describe the discrepancy between the number of countable and culturable cells present in any given environmental sample. This discrepancy has limited our understanding of the species diversity of soil bacterial communities (Joseph et al. 2003), but has been partially overcome through the application of molecular techniques.

Molecular techniques to assess diversity include guanine plus cytosine (G + C) content, nucleic acid reassociation, DNA microarrays, DNA hybridization, denaturing and temperature gradient gel electrophoresis (DGGE and TGGE), single strand

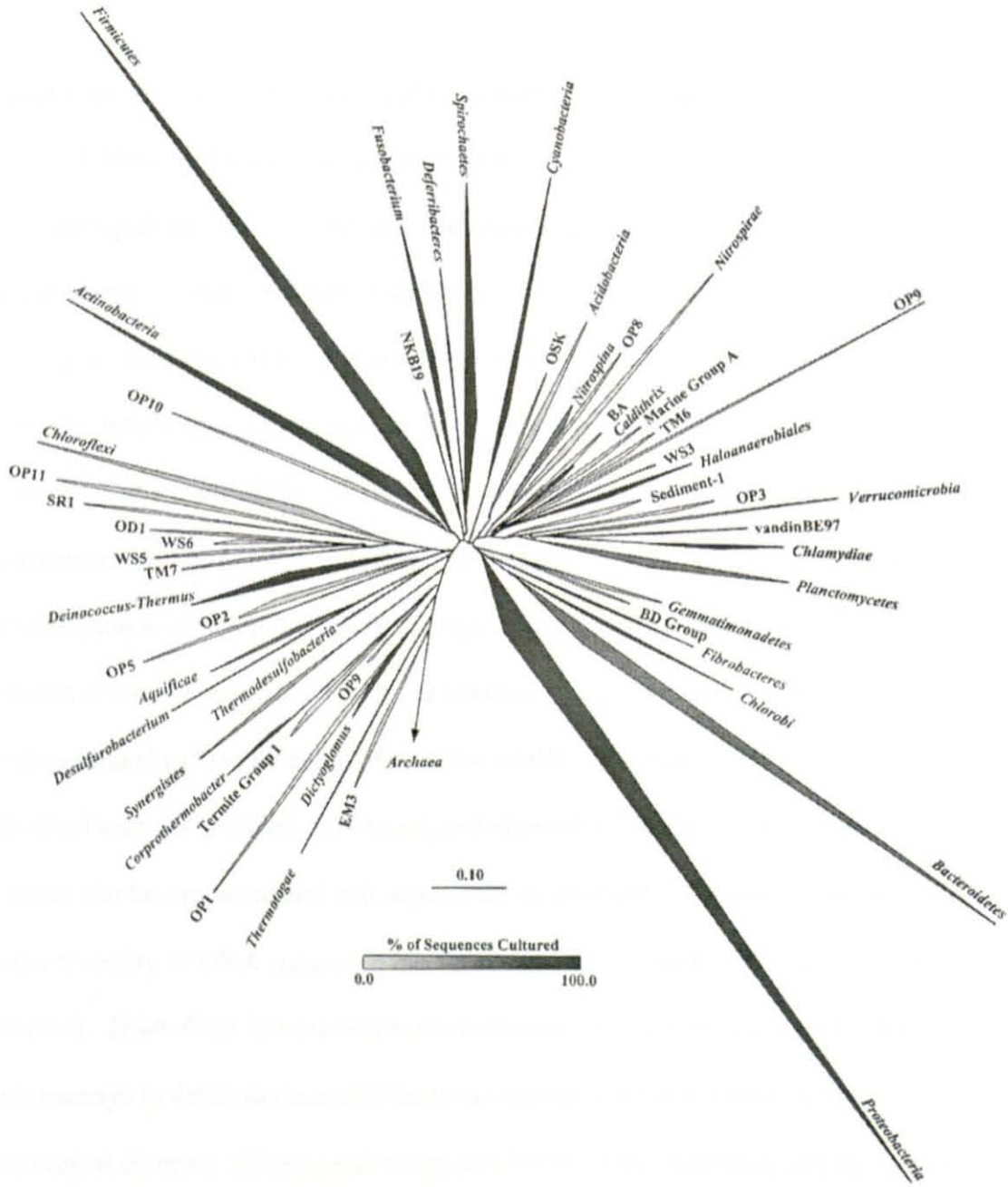

Figure 2. Phylogenetic tree of *Bacteria* showing established phyla (italicized Latin names) and candidate phyla. The vertex angle of each wedge indicates the relative abundance of sequences in each phylum; the length of each side of the wedge indicates the range of branching depth found in that phylum; the darkness of each wedge corresponds to the proportion of sequences in that phylum obtained from cultured representatives. Candidate phyla do not contain any cultured members (from Schloss and Handelsman 2004).

conformation polymorphism (SSCP), amplified ribosomal DNA restriction analysis (ARDRA) or restriction fragment length polymorphism (RFLP), terminal restriction fragment length polymorphism (T-RFLP), and ribosomal intergenic spacer analysis (RISA)/automated ribosomal intergenic spacer analysis (ARISA) (Kirk et al. 2004). G+C methodology is based on the knowledge that bacteria differ in their G+C molar content. Therefore, this information can be used to study the bacterial diversity of soil communities (Tiedje et al. 1999). There are some disadvantages to this methodology in that taxonomically related groups only differ between 3% and 5% which leads to a coarse level of resolution as different taxonomic groups might share the same G+C range. DNA reassociation is used to estimate diversity as a measure of genetic complexity of the microbial community (Torsvik et al. 1990). The total DNA is extracted from environmental samples, purified, denatured, and allowed to reanneal. The rate of reassociation can be measured and will depend on the similarity of sequences present (i.e., as the diversity of DNA sequences increases, the rate at which DNA reassociates will decrease). DNA-DNA hybridization has been more recently used together with DNA microarrays to detect and identify bacterial species (Cho and Tiedje 2001) or to assess microbial diversity (Greene and Voordouw 2003). The microarray can then either contain specific target genes to provide functional diversity information or can contain a sample of environmental DNA fragments representing different species found in the environmental sample. While the above approaches to assessing diversity have not been based on PCR, there are several that are PCR based.

The first of these PCR-based methods is DGGE and TGGE. They are very similar with the only differences being the method of species separation. During denaturation, DNA melts in "domains", which are sequence specific causing differential migration through a polyacrylamide gel. DGGE has a gel with a gradient of increasing concentrations of formamide and urea that causes different melting behaviors of the double-stranded DNA (Muyzer 1999). TGGE uses the same principle as DGGE except the gradient is temperature rather than chemical denaturants. By examining these gels, different community analyses can be made and species identified by band sequencing. SSCP is another technique that relies on separation of DNA based on differences in sequences. Here, single-stranded DNA molecules are separated on a polyacrylamide gel based on differences in mobility caused by their folded secondary structure (Lee et al. 1996). RFLP or ARDRA is yet another tool used to study microbial diversity that relies on DNA polymorphisms. PCR amplified rDNA is digested with restriction enzyme(s) that cut DNA at a particular sequence segment. This causes different fragment lengths which can be detected using agarose gels. These banding patterns can then be analyzed to assess diversity and unique species sequenced (Pace 1996). T-RFLP uses a similar technique as RFLP except that one of the PCR primers is labeled with a fluorescent dye. This allows detection of only the labeled terminal restriction fragment, which is detected in a capillary sequencer and yields community patterns but rarely species identifications (Liu et al. 1997). Finally, RISA and ARISA also provide ribosomal-based fingerprinting of the microbial community. In RISA and ARISA, the intergenic spacer region between the 16S and 23S ribosomal subunits is amplified by PCR, denatured, and separated on a

polyacrlyamide gel under denaturing conditions. In RISA, the sequence polymorphisms are detected using a silver stain while in ARISA the forward primer is fluorescently labeled and automatically detected with the use of an automated sequencer with laser detection (Fisher and Triplett 1999).

Each of the molecular techniques has its advantages and disadvantages. One of the most important advantages is that most molecular-based techniques do not require culturing and allow for detection of many different phyla and may give a more accurate account of the most numerically dominant organisms (Janssen 2006). Generating ribosomal sequence data is also an advantage to ultimately describing species. Sequences obtained through direct amplification from the environment provide the only information available for 99% of the prokaryotes in most natural communities (Schloss and Handelsman 2004). Some analyses can be made for the community as a whole without using molecular techniques, but it is necessary to acquire sequence data to determine diversity on a species or even phylum level.

There are several biases involved in using molecular microbial ecology methods including lysis efficiency of cells (Kirk et al. 2004). Since bacteria exist in or on the surface of soil aggregates, the ability to separate these cells from soil components is vital for studying biodiversity. The method of DNA or RNA extraction used can also bias diversity studies. If the method used is too harsh, nucleic acids can be sheared, which might cause problems with PCR. It is important to remove humic acids which can be coextracted and interfere with PCR analysis as well (Kirk et al. 2004). PCR, in general, which is used in most molecular techniques can also cause biases. Some of these issues

include different affinities of primers to templates, different copy numbers of target genes, and primer specificity (von Wintzingerode et al. 1997). Other issues mostly stemming from PCR include that sequence artifacts may arise due to the formation of chimerical molecules (Acinas et al. 1997; Hugenholtz and Huber 2003; Qui et al. 2001; Wang and Wang 1997), the formation of heteroduplex molecules (Speksnijder et al. 2001; Qui et al. 2001), *Taq* DNA polymerase error (Eckert and Kunkel 1991; Qui et al. 2001), and heterogeneity of 16S rDNA sequences (von Wintzingerode et al. 1997).

Phylogenetic studies using RNA, and eventually DNA, extracted directly from the environment have played a key role in exposing the gap in our knowledge about microbial species diversity (Handelsman 2004). This new ability to uncover taxonomic relationships for large numbers of species based on extracted DNA, combined with creative culture-based techniques designed to identify novel species, will provide insight into the biology, physiology and ecology of many presently unknown organisms living on this planet. This could lead to countless applications in biotechnology, medicine, bioremediation and environmental monitoring.

For this experiment, I decided to use molecular-based techniques. This is because pure culture techniques alone are inadequate for describing all naturally-occurring microbial assemblages, because appropriate media and conditions for growth are simply not well-developed, available, or practically feasible for microorganisms to be representative of their actual ecological niches (DeLong and Pace 2001). New developments from the 1980's and forward have allowed for more accurate descriptions of natural microbial diversity. The cultivation-independent approach involves the

recovery of phylogenetically informative gene sequences, usually from 16S rDNA nucleic acids extracted directly from microbial biomass. These informative gene sequences extracted from mixed microbial populations can be isolated as DNA clones and then sorted and sequenced to allow for this biodiversity to be better understood (DeLong and Pace 2001). Even with some biases in these methods, these culture-independent methods should allow for detection of numerous bacterial species, including the detection of unculturable species. In comparison to culturing, they also allow for a faster assessment of diversity and a less biased assessment when considering bacterial communities.

Great Smoky Mountains National Park (GSMNP) is a 2,200 km^2 reserve that lies on the mountainous divide between the states of North Carolina and Tennessee ("Discover Life", 2004). Some 95% of this area is forested, with much of it subjected to disturbance (e.g., logging, road building, air pollution, etc.) at some point in the past. GSMNP is known for its temperate forest richness, old-growth forests, and its diversity of species. As a result, an extensive study of the biodiversity inside the park is being conducted. This study, started in 1997, is known as the All Taxa Biodiversity Inventory (ATBI) (Sharkey 2001). This study concentrates on three questions: what is it, where is it, and what does it do? Therefore, the value of the ATBI is not just placed on what is found, but also on discovering the organisms' park-wide distribution, relative abundance, seasonality and ecological relationships. Even though insects, arachnids, and vertebrates have been the main focus of the inventory, recently a new interest has also developed in the study of prokaryotes and their diversity and importance throughout the park

(O'Connell 2002, 2003). Previously, about 250 bacterial species have been cultured and categorized in the park and three have been categorized that are uncultured (O'Connell, personal communication).

Some preliminary data have already been collected from GSMNP using culture-dependent studies (Figure 3). From 80 isolates sequenced six different phyla were observed of bacteria grown on solid media (O'Connell, submitted for publication). These included *Firmicutes, Actinobacteia, Bacteroidetes, Alphaproteobacteria, Betaproteobacteria*, and *Gammaproteobacteria*. When looking at Albright Grove the predominant phylum was *Firmicutes* (~80% of isolates), but in Cataloochee the predominant phyla were the *Betaproteobacteria* and *Firmicutes* with *Bacteroidetes* and *Gammaproteobacteria* of secondary dominance. In Purchase Knob, *Firmicutes* was dominant and the *Betaproteobacteria* and *Actinobacteria* codominated, secondarily. There was a much higher diversity at the genus level at the two second-growth forest sites compared with the old-growth forest site. It was also interesting to see that while Cataloochee and Purchase Knob both contained all six divisions, Albright Grove only had four and *Alphaproteobacteria* and *Gammaproteobacteria* were not observed there. Via these culture-dependent methods, differences between sites was observed, and I hypothesized differences would also be seen through culture-independent methods.

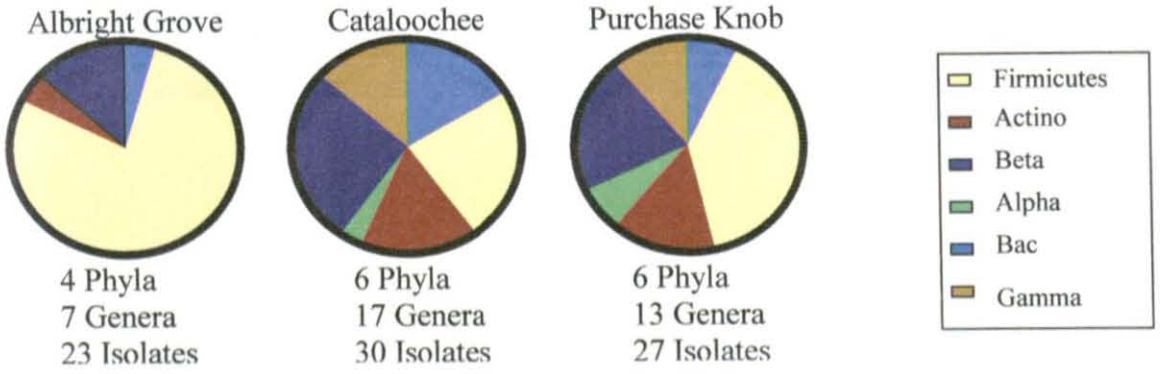

Figure 3. Phylum level diversity (in percent total for each group) for bacteria cultivated from soil from the three sites in this study, showing predominance of the *Firmicutes* at the old growth forest site and higher diversity for the two second growth sites; samples were obtained from bulk soils near hemlock. (Bac is *Bacteroidetes*; Actino is *Actinomycetes*; Alpha, Beta, and Gamma are subphyla within the Proteobacteria.

Due to the diversity of ecosystems throughout the park, there is much to learn about the relationships and differences between bacterial species within each location. The three sites that I explored within GSMNP were the Albright Grove, Cataloochee, and Purchase Knob long-term ATBI study plots (Figure 4). Each of these sites differs by forest type, soil chemistry, and elevation (Sharkey 2001; Table 1). It has been shown that diversity of soil microorganisms is determined primarily by the vegetative cover but also by the climatic and soil conditions (Campbell et al. 1999). Changes in land use will affect microbial diversity and also the balance between different microbial processes. Studies observing microbial community changes after forest impacts such as ash treatment, clear-cutting, and prescribed burning found that r-strategists predominated the community directly after the forest disturbance, taking advantage of the lack of competition and readily decomposable substrates (Staddon et al. 1998). After time, K-

strategists increased in numbers as the community became more complex. Such complexity should already exist in forests that have not experienced any major impacts.

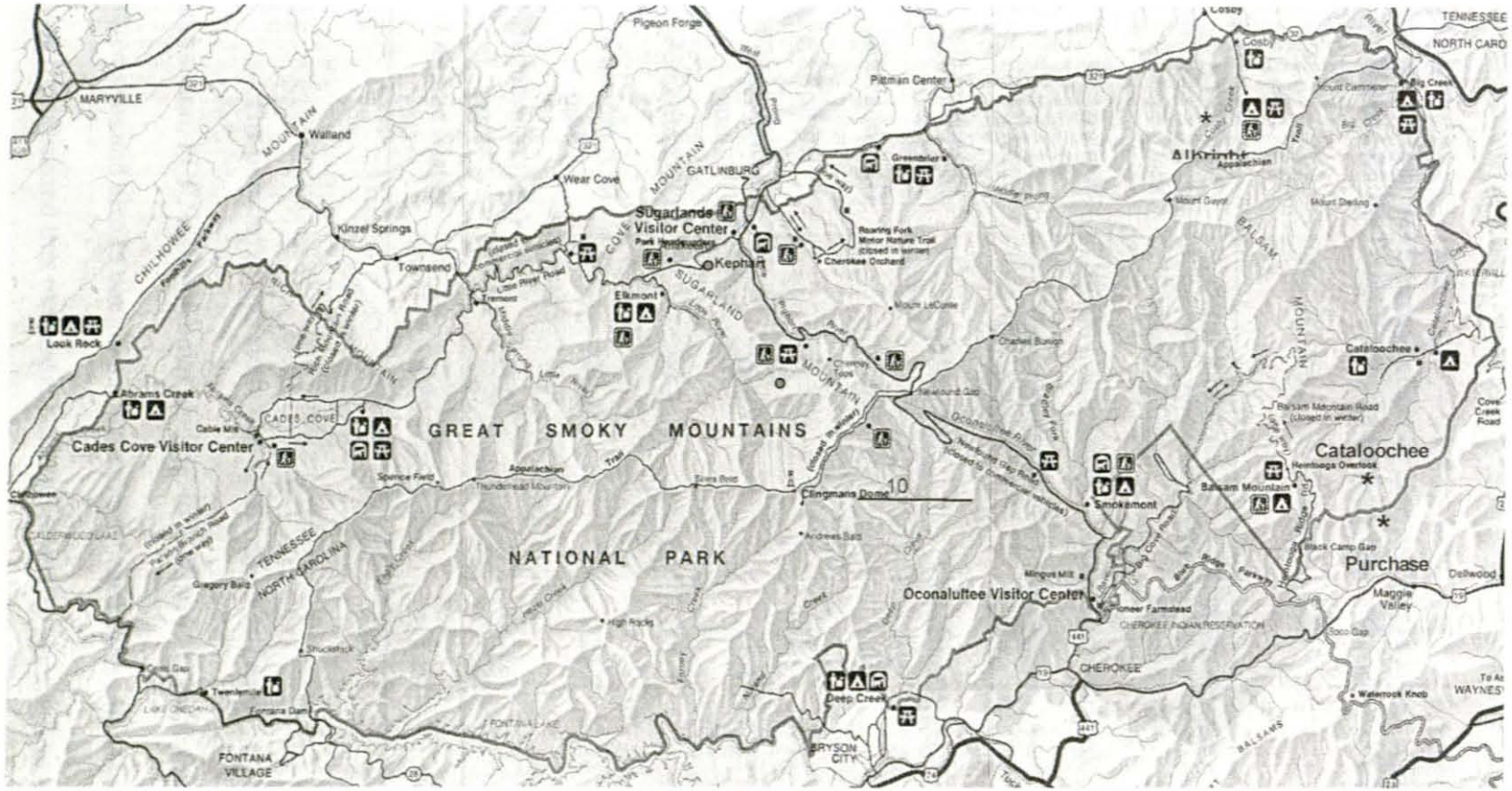

Figure 4. Map of Great Smoky Mountains National Park and ATBI sample sites Albright Grove, Cataloochee, and Purchase Knob, on the eastern side of the Park.

Table 1. Three biodiversity reference plots examined in this study and previously established for the Great Smoky Mountains National Park All Taxa Biodiversity Inventory.

ATBI plot	Albright Grove	Cataloochee	Purchase Knob
Forest Class	Montane Cove	Mesic Oak	Northern Hardwood
Watershed	Indian Camp Creek	Cataloochee Creek	Cove Creek
Geology	Thunderhead Sandstone	Thunderhead Sandstone	Biofite Augen Greiss
Disturbance History	Undisturbed	Chestnut Blight	Logged
Elevation (ft)	3,390	4,530	5,020
Soil pH	4.3	4.3	4.8
Phosphorus (P) ppm	18.7	13.3	12.0.
Potassium (K)	93.3	81.7	85.7
Calcium (Ca) ppm	224.8	222.8	274.3
Magnesium (Mg) ppm	35.3	35.2	42.7
Organic Matter (%)	3.9	3.8	3.5

The purpose of this study was to use direct molecular-techniques, i.e., DNA extraction, PCR, and molecular cloning, to compare bacterial communities among the three sites and also to compare communities based on previous culture-dependent data. It was hypothesized that molecular bacterial diversity from soil would differ among the three forested sites because of chemical, vegetational, and land history differences.

Molecular techniques should also select for different bacterial species to be identified when compared to cultured bacteria from the same sites (based on previous work by O'Connell). This would presumably be due to the differences between easily-cultivated bacteria versus rare and/or culture-resistant species.

Methods and Materials

Soil Sample Collection

Soil samples were collected from three ATBI plots in GSMNP (Albright Grove, Cataloochee, and Purchase Knob) on February 13th, 2005 and placed on dry ice. The soil samples were collected using aseptic techniques by removing the leaf litter and any roots with EtOH rinsed and flame sterilized tools (small shovel and garden trowel). The soil was then homogenized in the upper 4-5 inches of the ground and an aliquot transferred to a sterile 50 mL centrifuge tube. Three replicates were taken at each site from near Eastern Hemlock (*Tsuga Canadensis*) stands and were within 100 feet of each other. Soil pH measurements were also taken at each site (Table 1).

Methods Development

DNA Extraction. Comparisons were made using the maximum yield protocol from the Mo Bio UltraClean Soil DNA Isolation Kit and the Mo Bio PowerSoil DNA Isolation Kit with the alternative lysis method (Mo Bio Industries, Inc., Solana Beach, CA). DNA was extracted directly from the soil of Albright Grove replicate 1 and Purchase Knob replicate 1. Comparisons of the kits were made in an attempt to minimize humic acid content in samples in order to maximize PCR amplification. The PowerSoil DNA Isolation Kit has an extra proprietary chemical added to help remove humics. A 1% agarose gel stained with ethidium bromide was run at 45V for 90 minutes and viewed with UV illumination to compare the DNA from the extraction kits.

PCR Amplification. PCR of 16S rDNA was performed to amplify total bacterial community DNA and further compare the DNA isolation kits. Final DNA extracts from both Albright Grove replicate 1 and Purchase Knob replicate 1 were amplified at both 100% and 10% concentration. PCR was conducted using a "touchdown" approach using primers 341F and 907R (based on *Escherichia coli* numbering; Casamayer et al. 2000). PCR conditions that were used to amplify the 16S rDNA gene fragment were as follows (volumes are per reaction): Master Mix = Eppendorf nuclease free water, 1% Igepal, Eppendorf Buffer (10X), 341F primer (25pmol/μL), 907R primer (25pmol/μL), 2.5U Eppendorf Taq, and Eppendorf dNTPs (10mM each). To 49.5μL of master mix was added 1.0μL of DNA. Thermal cycler (Eppendorf Corporation, Westbury, NY) conditions for "touchdown" PCR were as follows: Initial Denaturation: 5 minutes at 94°C; 30X PCR cycles→ Denaturation: 1 minute at 94°C, Annealing: 1 minute at * °C -- *Start at 65°C (2X), drop 1°C each cycle (10X), end at 55°C (18X), Elongation: 3 minutes at 72°C; Final Elongation: 7 minutes at 72°C; and Sample Hold: ∞ at 4°C. A 1% agarose gel stained with ethidium bromide was run at 90V for 30 minutes to compare amplified products from all samples.

Denaturing Gradient Gel Electrophoresis (DGGE). DGGE methods (adapted from Muyzer et al. 1998) consisted of a polyacrylamide gel impregnated with a gradient of 20% (urea/formamide) to 60% (urea/formamide) to which 20μL of community PCR products were added. A Bio-Rad DCode Universal Mutation Detection system (Bio-Rad Laboratories, Hercules, CA) was used to electrophorese samples at 65V for 15 hours at 60°C. Gels were stained with ethidium bromide for thirty minutes, destained for ten

minutes, and photographed with UV illumination using an EDAS 290 gel imaging system (Eastman Kodak Company, Rochester, NY). Band locations correspond to unique species, with each sequence becoming immobilized at its mimicked melting temperature in the urea/formamide gradient. Bands in the same vertical position hypothetically represent the same species, while those that are staggered likely represent different species.

Methods Used in the Full Study

DNA Extraction. The PowerSoil DNA Isolation Kit was used with the alternative lysis method. DNA was extracted directly from the soil of each replicate from all three sites, screened using agarose gel electrophoresis, and stored at -20°C for later PCR amplification.

PCR Amplification (1500 bp). Approximately 1500 base pair fragments of the 16S rDNA from the mixed bacterial species were amplified using bacterial primers 27F and 1492R (based on *Escherichia coli* numbering; Corinaldes et al. 2005). Albright Grove, Cataloochee, and Purchase Knob replicates were all diluted to 10% and amplified using the same PCR chemical conditions as before (substituting 27F/1492R primers for 341F/907R primers). Thermal cycler conditions were as follows (Corinaldes et al. 2005): Initial Denaturation: 3 minutes at 94°C; 30X PCR cycles→ Denaturation: 1 minute at 94°C, Annealing: 1 minute at 55°C, Elongation: 2 minutes at 72°C; Final Elongation: 10 minutes at 72°C; and Sample Hold: ∞ at 4°C. PCR products were screened as before in an agarose gel.

PCR Clean-Up. Montage PCR Centrifugal Filter Devices (Millipore Corporation, Bedford, MA) were used for PCR product purification. This step allowed for high quality nucleic acids for use in molecular cloning, RFLP, and sequencing reactions.

Molecular Cloning. Approximately 1500bp PCR fragments were cloned into *Escherichia coli* using the pGEM-T-Easy Vector System (Promega Corporation, Madison, WI) using a three step approach (protocol shared by R. Lehman, unpublished). First, ligation was performed from the products obtained through PCR and PCR clean-up. Ligation reactions were set up in PCR tubes for all products as follows and refrigerated overnight:

Reagent	3:1	1:1	1:3
2X Rapid Buffer	5μL	5μL	5μL
Vector	1μL	1μL	1μL
T4 DNA Ligase	1μL	1μL	1μL
H_2O	2μL	2μL	2μL
PCR Product*	1μL	1μL	1μL

*PCR products were used at different concentrations
in an effort to maximize the number of transformed cultures
3:1 Sample = Straight PCR products
1:1 Sample = 3μL PCR products + 9μL water
1:3 Sample = 1μL of 1:1 Sample + 2μL water

Transformation was performed by first withdrawing 2μL of the ligation reaction and placing it into new, sterile PCR tubes. The next step was to transfer 50μL of JM109 *E.coli* cells into each tube, mix gently, and incubate in an ice bath for 20 minutes. These tubes were then placed into a 42°C water bath for 45-50 seconds and then returned to the ice bath for another 2 minutes. These contents were placed into 950μL of room temperature SOC Media [per 100mL; 2.0g tryptone, 0.5g yeast extract, 1mL of 1M NaCl,

0.25mL of 1M KCl, 1mL of 2M Mg^{2+} stock (20.33g $MgCl_2.6H_20$, 24.65g $MgSO_4.7H_2O$ in 100mL water; filter sterilized), 1mL of 2M glucose stock (filter sterilized) Mg^{2+} and glucose added after autoclaving the other ingredients] in sterile 15mL tubes and incubated at 1.5 hours at 37°C shaking at 150 RPM. Ten microliters of 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal) and 50µL of isopropyl-beta-D-thiogalactopyranoside (IPTG) were spread onto fresh Luria-Bertani (LB)/Ampicillin (AMP) plates and warmed in a 37°C incubator (according to manufacturer's recommendations). Contents of each tube were plated at 100µL per culture onto these warmed LB/AMP/IPTG/X-Gal plates and incubated upside down for 20 hours at 37°C. The final step for molecular cloning was blue/white screening. After the 37°C incubation, the plates were refrigerated for 1-2 hours and then the plate with the PCR dilutions that produced the greatest number of white colonies was chosen for clone selection. Colonies (150 per site) were collected using sterile toothpicks, placed into numbered LB/glycerol tubes (100 µL of 15% glycerol in 200 µL PCR tubes), and stored at -70°C until further processing could occur. Numbering of colonies was as follows: Albright Grove replicate 1 = numbers 1-50, Albright Grove replicate 2 = numbers 51-100, Albright Grove replicate 3 = 101-150, Cataloochee replicate 1 = 151-200, Cataloochee replicate 2 = 201-250, Cataloochee replicate 3 = 251-300, Purchase Knob replicate 1 = 301-350, Purchase Knob replicate 2 = 351-400, and Purchase Knob replicate 3 = 401-450.

Whole Cell PCR. Colonies from molecular cloning were plated out onto fresh LB/AMP/IPTG/X-Gal plates. Colonies that still grew up white were used during whole

cell PCR. Protocols for whole cell PCR required two separate reactions to be set up. The pre-master mix required a 10µL/reaction solution while the post-master mix required a 39.5µL/reaction solution to be set up. The pre- master mix consisted of 9µL/reaction volume of nuclease free water and 1.0µL/reaction volume of PCR buffer (10X). This solution was mixed and dispensed as 10µL/reaction into labeled PCR tubes. White colonies were collected off of the plates using toothpicks and mixed into this 10µL solution. Tubes were placed into a thermal cycler and the cell lysis accomplished at 99°C for 15 minutes. After this step hot start was run at 80°C for 5 minutes (this step was used to place the post-master mix solution into tubes). The post-master mix consisted of nuclease free water, PCR Buffer (10X), 1% IgePal , M13 Forward primer (25pmol/µL), M13 Reverse primer (25pmol/µL), 2.5U DNA Polymerase *Taq*, and dNTPs (10mM ea.) for a total volume of 39.5µL. Thermal cycler conditions were then continued with an Initial Denaturation: 4 minutes at 94°C; 30X PCR cycles→ Denaturation: 1 minute at 94°C, Annealing: 1 minute at 55°C, Elongation: 1 minute at 72°C; Final Elongation: 4 minutes at 72°C; and Sample Hold: ∞ at 4°C. Products were screened as previously. Montage PCR clean-up was also performed for each working product to be used in RFLP.

Restriction Fragment Length Polymorphism (RFLP). RFLP digestions were performed for each PCR product (protocol shared by R. Lehman, unpublished). Master mix solution volumes were made as follows: Eppendorf nuclease free water, Buffer B (10X), BSA (10mg/µL), RsaI restriction enzyme (10U/µL), and MspI restriction enzyme (10U/µL) (Promega, Inc., Madison, WI). The master mix was mixed well and 10µL/reaction was

dispensed into labeled PCR tubes. 10μL of each whole cell PCR product was then added to each PCR tube, spun down, and placed into a thermal cycler. Restriction digest conditions were as follows: 3 hours at 37°C, 15 minutes at 65°C, and then held for ∞ at 4°C. These products were run on a RFLP gel prepared as follows: 125mL of cold 1X TBE was placed into a container along with a Teflon coated stir bar and stirred rapidly on a magnetic stir plate; 5.0g of Metaphor agarose (Cambrex Bio Science Rockland, Inc., Rockland, ME) was slowly added and allowed to stir for 15 minutes until all clumps were gone; this mixture was then placed into a microwave and heated to the point of boiling; it was then placed back onto the stir plate for another 15 minutes (this time stirring slowly); the mixture was placed into the microwave again and heated until all granules had been dissolved; it was then placed back on the stir plate until the solution had reached 50-60°C. Ethidium bromide was added to the agarose solution and poured into gel casts; gels were then allowed to solidify (10-15 minutes) and TBE buffer was then added to the top of the gels. Finally, each gel was placed in the refrigerator for 10-15 minutes. 10μL of each PCR product digest along with 1.5μL of loading dye was added to each well. This was run at 210V for three, 1 minute intervals with 10 second pauses in between and then at 68V for 180 minutes. Afterwards, images were captured using UV transillumination and banding patterns analyzed to detect unique DNA sequences.

Sequencing. PCR using primers 341F/907R was performed on the clone inserts from unique banding patterns to amplify ~550bp of the product. The PCR products were cleaned using AutoSeq Sephadex G-50 spin columns (Amersham Biosciences, Piscataway, NJ). After cleanup, sequencing PCR products were dried using a speed

vacuum. Samples were resuspended in 10μL HiDi formamide (Applied Biosystems) and then sequenced using the BigDye Terminator Version 3.0 Cycle Sequencing Kit and a 3130 Automated DNA Sequencer (Applied Biosystems, Foster City, CA).

Sequences were compared to previously identified clones and isolates that were in the Ribosomal Database Project II (RDP II) using both the "Classifier" and "Sequence Match" programs (Maidak, 2001). All sequences were checked for chimeras by first aligning them with ClustalW (Vector NTI, Invitrogen, Inc., Carlsbad, CA) and then using the Belephron (Huber et al. 2004), Mallard and Pintail computer programs (Ashelford et al. 2005).

DNA sequence similarity matrices and simple phylogenetic trees were generated by Vector NTI following alignment using ClustalW in order to better compare the sequences from each clone with other clones identified to the same phylum. Trees were created using the neighbor-joining algorithm. A similarity matrix was generated for each phylum containing more than three sequences in order to make comparisons regarding how similar these clones actually were (based on percentages).

All sequence data will be deposited in the ATBI and GenBank (and directly into the RDP II) databases.

Results

Methods Development

The Mo Bio Ultra Clean Soil kit yielded the highest amount of genomic DNA (Figure 5). However, the PowerSoil kit yielded the strongest bands of PCR products (at 10% strength solution of extracts (Figure 6) and was used for the full study. When comparing molecular techniques for community analysis, DGGE did not yield adequate banding patterns (results not shown) for further analysis, and molecular cloning was used instead.

Methods Used In Soil Study

Molecular Cloning Results. All three sites from which samples were collected showed amplification of bacterial 16S rDNA. Molecular cloning of samples from Albright Grove, Cataloochee, and Purchase Knob yielded 450 clones (50 clones from each soil sample replicate, 150 clones from each site). These numbers decreased as clones were re-streaked and then again as whole cell PCR was performed (Table 2).

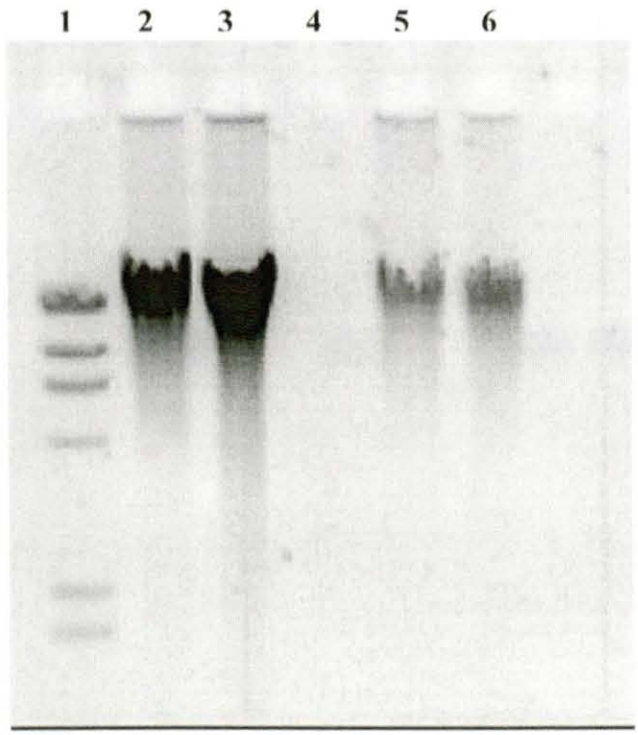

Figure 5. Comparison of Mo Bio PowerSoil DNA Isolation Kit and Mo Bio UltraClean Soil DNA Isolation Kit for yield of genomic DNA from soils from Great Smoky Mountains National Park. This gel shows that there is more genomic DNA in the UltraClean Kit. Lane 1, λ/Hind III ladder; lanes 2 and 3, AG-Ultra and PK-Ultra, respectively; lanes 5 and 6, AG-Power and PK-Power, respectively. (AG= Albright Grove; PK= Purchase Knob; Ultra=UltraClean Kit; Power= PowerSoil Kit).

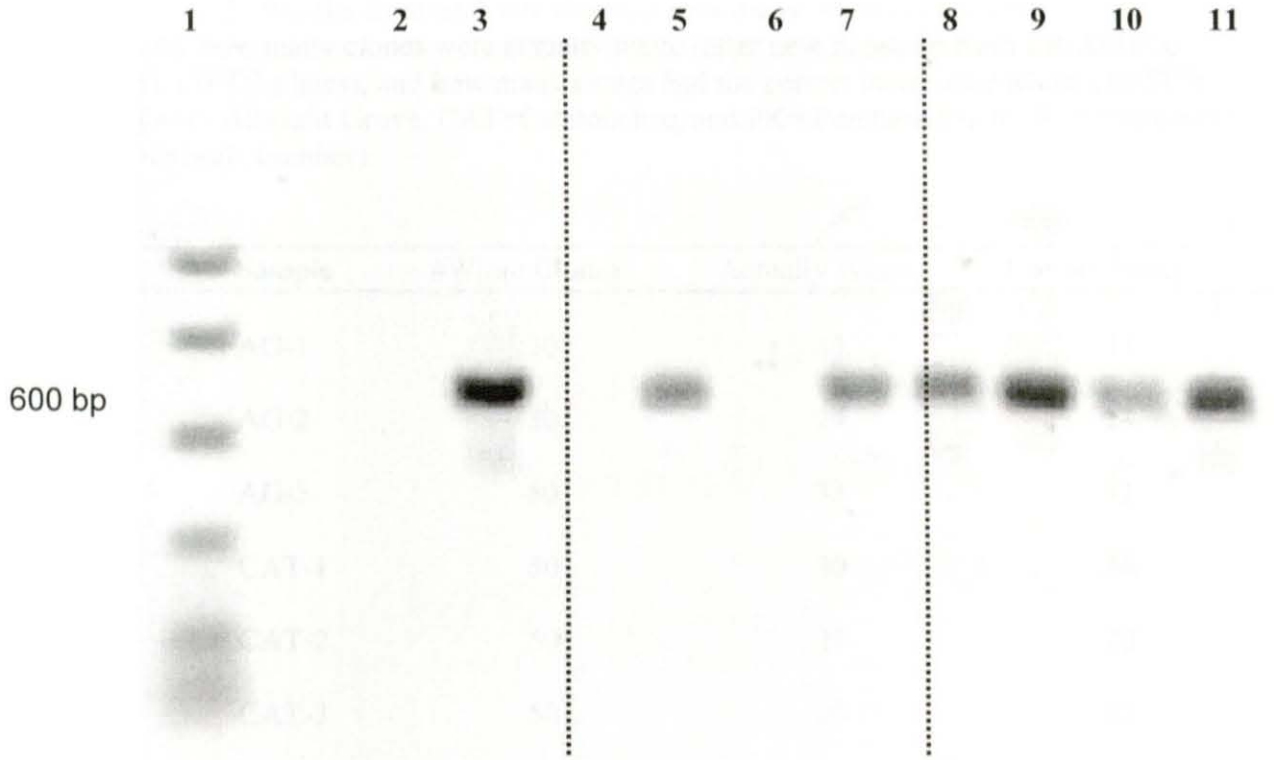

Figure 6. Agarose gel of PCR amplification products comparing the UltraClean DNA Isolation Kit with the PowerSoil DNA Isolation Kit for diluted and undiluted genomic DNA stocks, showing that the PowerSoil Kit with DNA diluted to 10% had the best amplification. Lane 1, PCR ladder; lanes 2 and 3, (-) Ctrl and (+) Ctrl, respectively; lanes 4 to 7, AG-Ultra 100%, AG-Ultra 10%, PK-Ultra 100%, and PK-Ultra 10%, respectively; lanes 8 to 11, AG-Power 100%, AG-Power 10%, PK-Power 100%, and PK-Power 10%, respectively. (AG= Albright Grove; PK= Purchase Knob; Ultra=UltraClean Kit; Power= PowerSoil Kit).

Table 2. Results from each site showing how many white clones were chosen form each site, how many clones were actually white (after new streak on fresh LB/AMP/X-Gal/IPTG plates), and how many clones had the correct insert after whole cell PCR. (AG= Albright Grove, CAT=Cataloochee, and PK= Purchase Knob; -X corresponds to replicate number).

Sample	#White Clones	Actually White	Correct Insert
AG-1	50	45	43
AG-2	50	24	23
AG-3	50	33	32
CAT-1	50	30	26
CAT-2	50	21	20
CAT-3	50	29	22
PK-1	50	3	7
PK-2	50	11	11
PK-3	50	10	8

RFLP Results. Banding patterns from RFLP resulted in 180 unique banding patterns from the three sites (Figure 7). Only four clones shared banding patterns within Albright Grove and only two clones shared banding patterns between Albright Grove and Cataloochee.

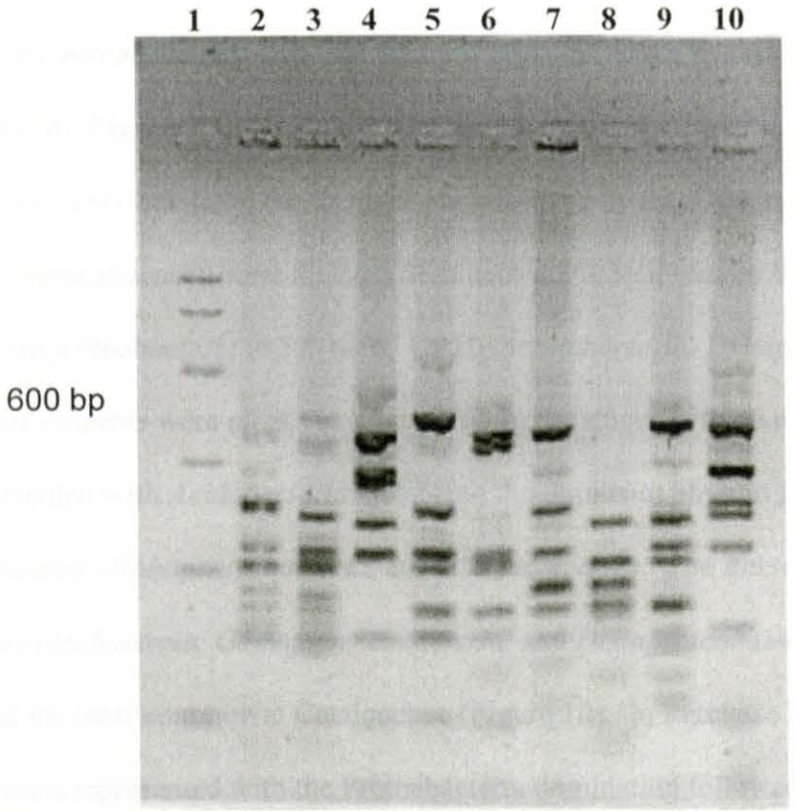

Figure 7. Example of a restriction fragment length polymorphism (RFLP) gel used to compare banding patterns between clones. Each lane represents a different clone from Albright Grove. Lane 1, PCR ladder; lanes 2 to 10, A12, A52, A55, A57, A58, A59, A102, A104, and A105, respectively. (A = Albright Grove; -X corresponds to replicate number)

Sequencing and RDP II Results. Sequences were acquired for 177 out of 192 clones corresponding to unique banding patterns. Three sequences of the 180 unique banding pattern sequences had ten or more bases that were indecisive and were not used. Sequences were then entered into the RDP II "Classifier" and "Sequence Match"

programs, and a complete list of species classifications is given in Appendix A. RDP II analyses resulted in 11 total phyla for all sites. These 11 phyla were *Acidobacteria, Firmicutes, Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, Gammaproteobacteria, Actinobacteria, Verrucomicrobia, Planctomycetes, Bacteroidetes*, and OP10 (Figure 8; Table 3). In Albright Grove there were nine phyla represented with *Acidobacteria* being the dominant phylum follwed by the *Firmicutes*. The *Alphaproteobacteria* were the next dominant at 9% followed by the *Verrucomicrobia* and *Gammaproteobacteria* at 3% each. OP10, *Actinobacteria, Deltaproteobacteria,* and *Planctomycetes* were all at 1% (Figure 9). In Cataloochee there were eight phyla represented with *Acidobacteria* also being the dominant phylum followed by the *Firmicutes, Alphaproteobacteria,* and *Planctomycetes*. The *Bacteroidetes, Betaproteobacteria, Gammaproteobacteria,* and *Deltaproteobacteria* all followed as being the least common at Cataloochee (Figure 10). In Purchase Knob there were eight divisions represented with the Proteobacteria dominating followed by the *Acidobacteria* and *Firmicutes*. The least common were the OP10 and *Betaproteobacteria* (Figure 11).

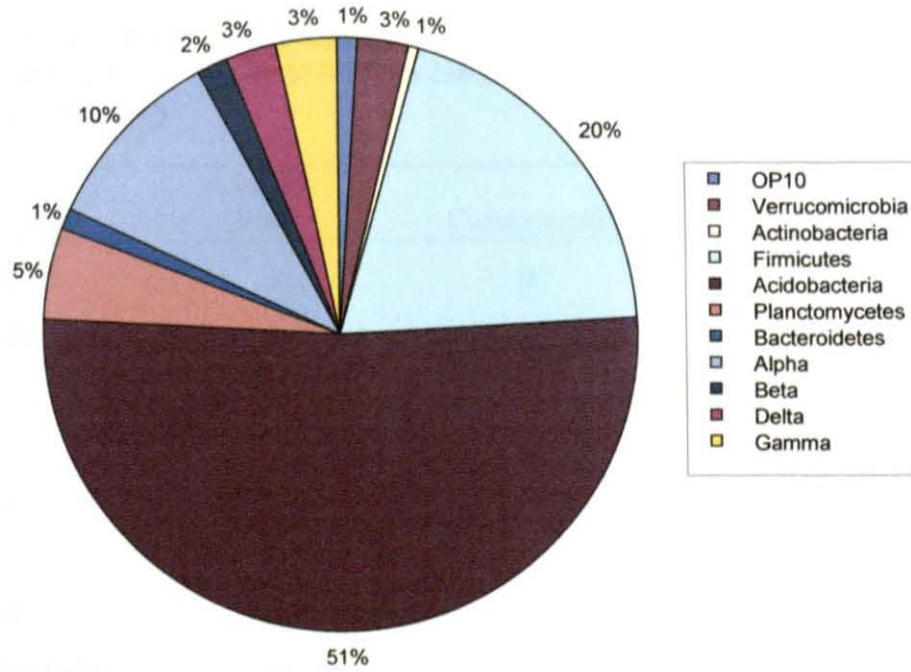

Figure 8. RDP II "Classifier" results for phylum level diversity among all three sites from this study. Dominance is shown by the *Acidobacteria* followed by the *Firmicutes*. (Alpha, Beta, Gamma, and Delta are subphyla within the *Proteobacteria* division).

Table 3. Phylum level diversity for bacterial 16S rDNA sequences cloned from soil from the three study sites in Great Smoky Mountains National park (number is the total number of clones obtained).

	Albright Grove	Cataloochee	Purchase Knob
Verrucomicrobia	3	0	2
Planctomycetes	1	8	0
OP10	1	0	1
Firmicutes	21	9	5
Acidobacteria	52	33	6
Bacteroidetes	0	2	0
Actinobacteria	1	0	0
Alphaproteobacteria	8	8	2
Betaproteobacteria	0	2	1
Deltaproteobacteria	1	2	2
Gammaproteobacteria	3	1	2

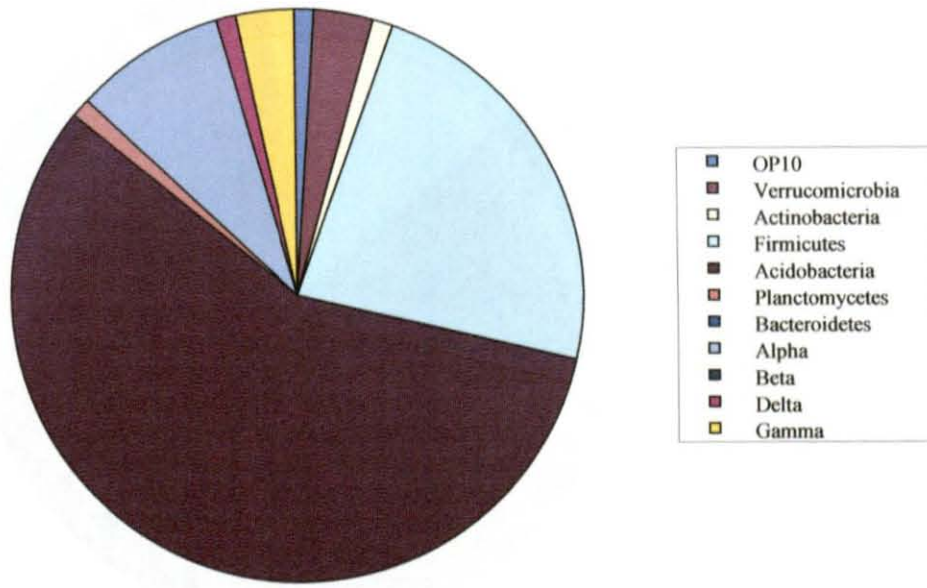

Figure 9. RDP II "Classifier" results for phylum level diversity in the Albright Grove ATBI site, showing the dominance of the *Acidobacteria*. (Alpha, Gamma, Delta are subphyla within the *Proteobacteria* division). *Acidobacteria* = 58%; *Firmicutes* = 23%; *Alphaproteobacteria* = 9%; *Gammaproteobacteria* = 3%, *Verrucomicrobia* = 3%; *Actinobacteria* = 1%; OP10 = 1%; *Deltaproteobacteria* = 1%; and *Planctomycetes* = 1%.

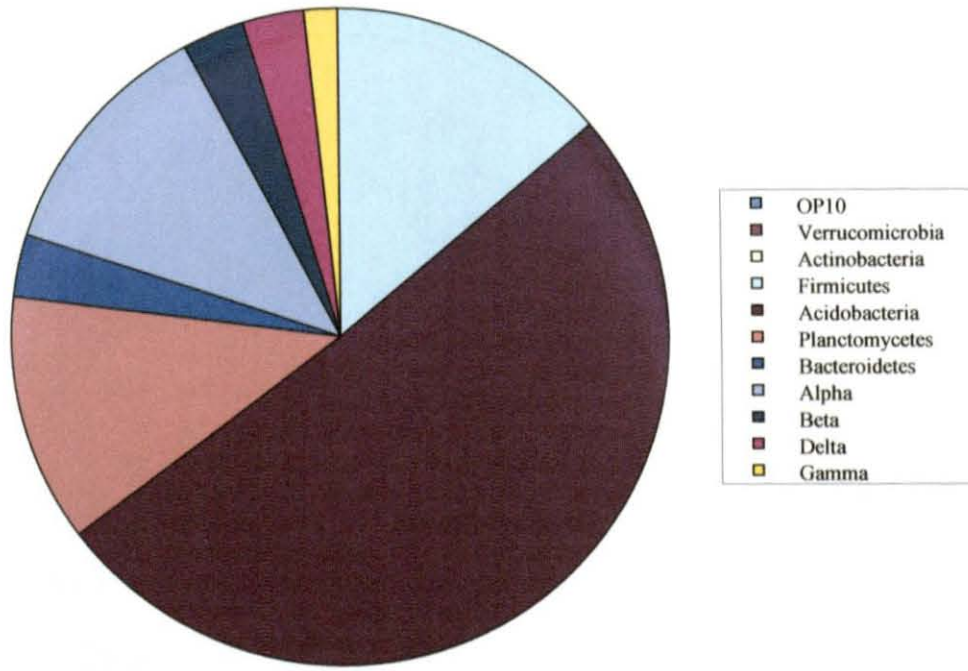

Figure 10. RDP II "Classifier" results for phylum level diversity at the Cataloochee ATBI site, showing dominance by the *Acidobacteria*. (Alpha, Beta, Gamma, Delta are subphyla within the *Proteobacteria* division). *Acidobacteria* = 51%; *Firmicutes* = 14%; *Alphaproteobacteria* = 12%; *Planctomycetes* = 12%; *Bacteroidetes* = 3%; *Betaproteobacteria* = 3%; *Deltaproteobacteria* = 3%; and *Gammaproteobacteria* = 2%.

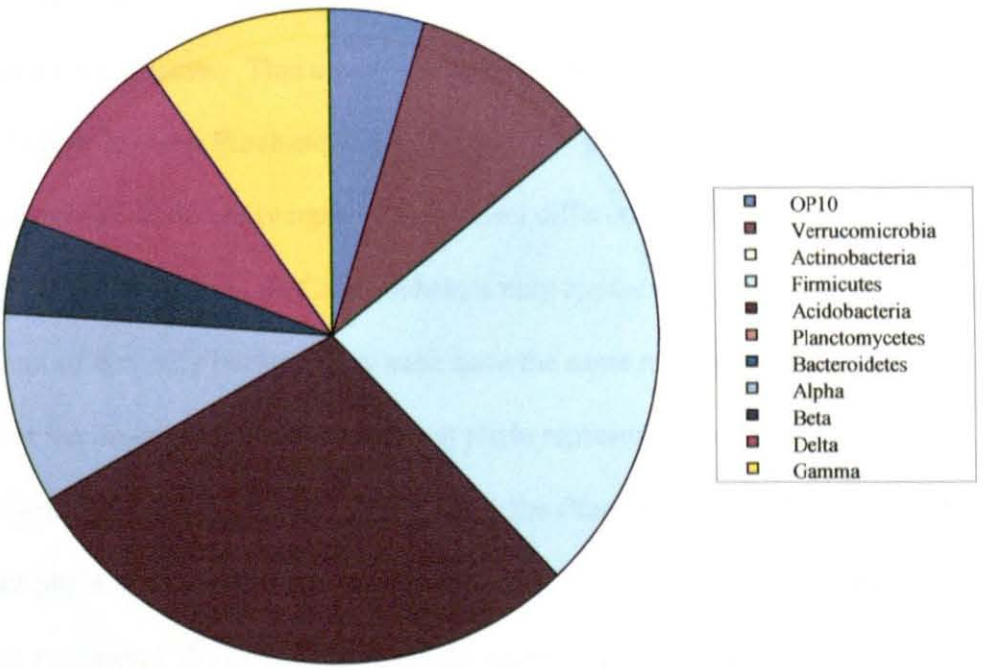

Figure 11. RDP II "Classifier" results for phylum level diversity at the Puchase Knob ATBI site, showing codominance by the *Acidobacteria* and *Firmicutes*. (Alpha, Beta, Gamma, Delta are subphyla within the *Proteobacteria* division). *Acidobacteria* = 28%; *Firmicutes* = 23%; *Alphaproteobacteria* = 10%; *Deltaproteobacteria* = 10%; *Gammaproteobacteria* = 10%; *Verrucomicrobia* = 9%; OP10 = 5%; and *Betaproteobacteria* = 5%.

It was interesting to see how diversity not only differed between sites, but also within sites between replicates. This could be viewed in Albright Grove (Figure 12), Cataloochee (Figure 13), and Purchase Knob (Figure 14). In Albright Grove, replicate 1 has the highest amount of diversity representing seven different phyla, while replicate 2 has four and replicate 3 has five. In Cataloochee, it may appear that each replicate has the same amount of diversity because they each have the same number of phyla represented, but the diversity lies in the different phyla represented and the proportion that each is represented. For example, in replicate 1 the *Planctomycetes* are the second most dominant phylum, in replicate 2 the *Firmicutes* are the second most dominant phylum, and in replicate 3 the *Firmicutes* and *Alphaproteobacteria* are the second most dominant phyla. Also, the *Bacteroidetes* are only found in replicate 2, but the *Betaproteobacteria* are found in every replicate except replicate 2. Similarly, the *Gammaproteobacteria* are only found in replicate 3, but the *Deltaproteobacteria* are found in every replicate except replicate 3. In Purchase Knob, the same trend is in effect and can not see the overall diversity of the site by only looking at one replicate. For example, the *Firmicutes* are not even seen in replicate 3, but the *Betaproteobacteria* are only found in replicate 3. OP10 is only found in replicate 1, but both the *Verrucomicrobia* and *Gammaproteobacteria* are found in every replicate except replicate 1.

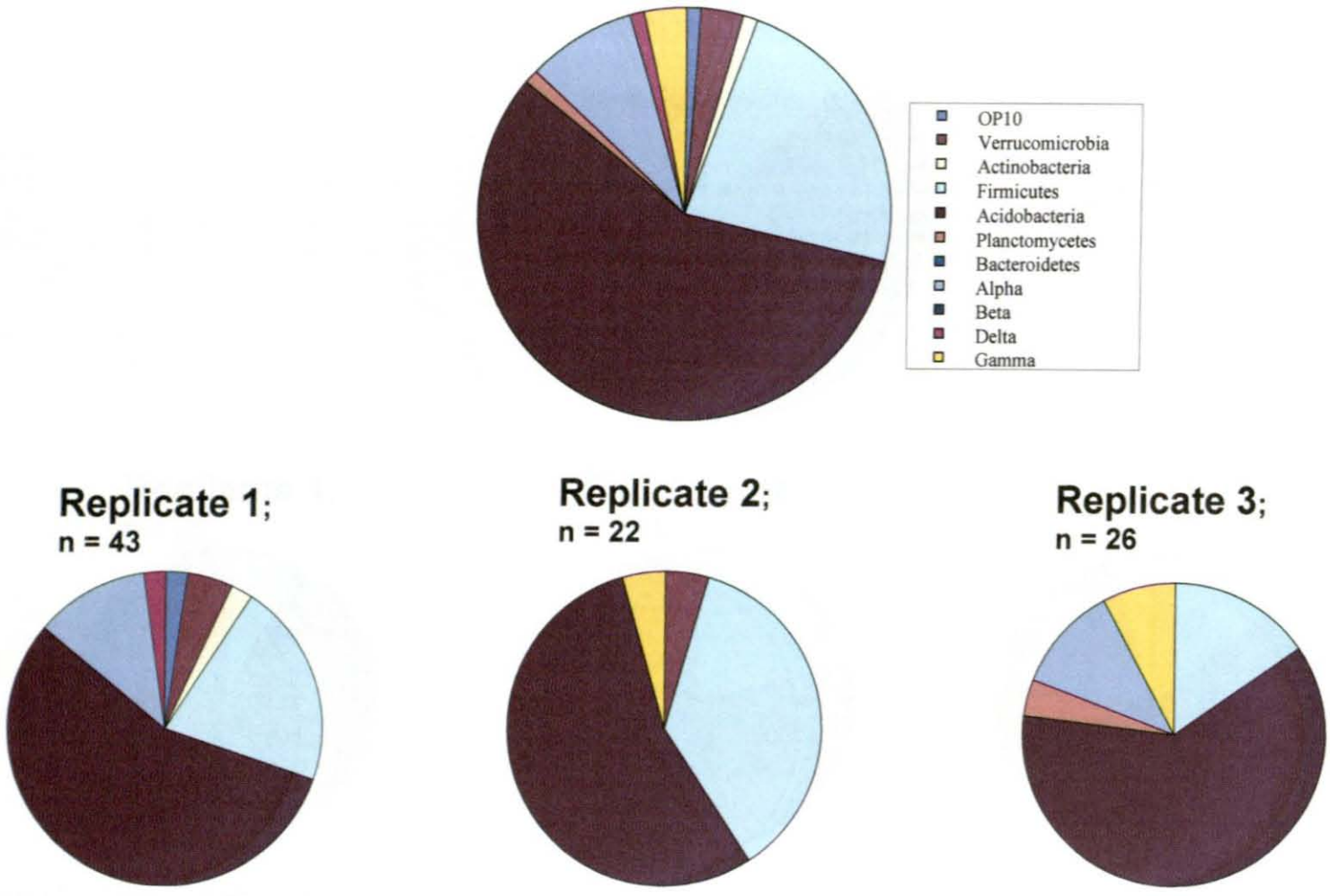

Figure 12. Phylum diversity patterns for Albright Grove broken down into the phylum diversity patterns for each of the three replicates.

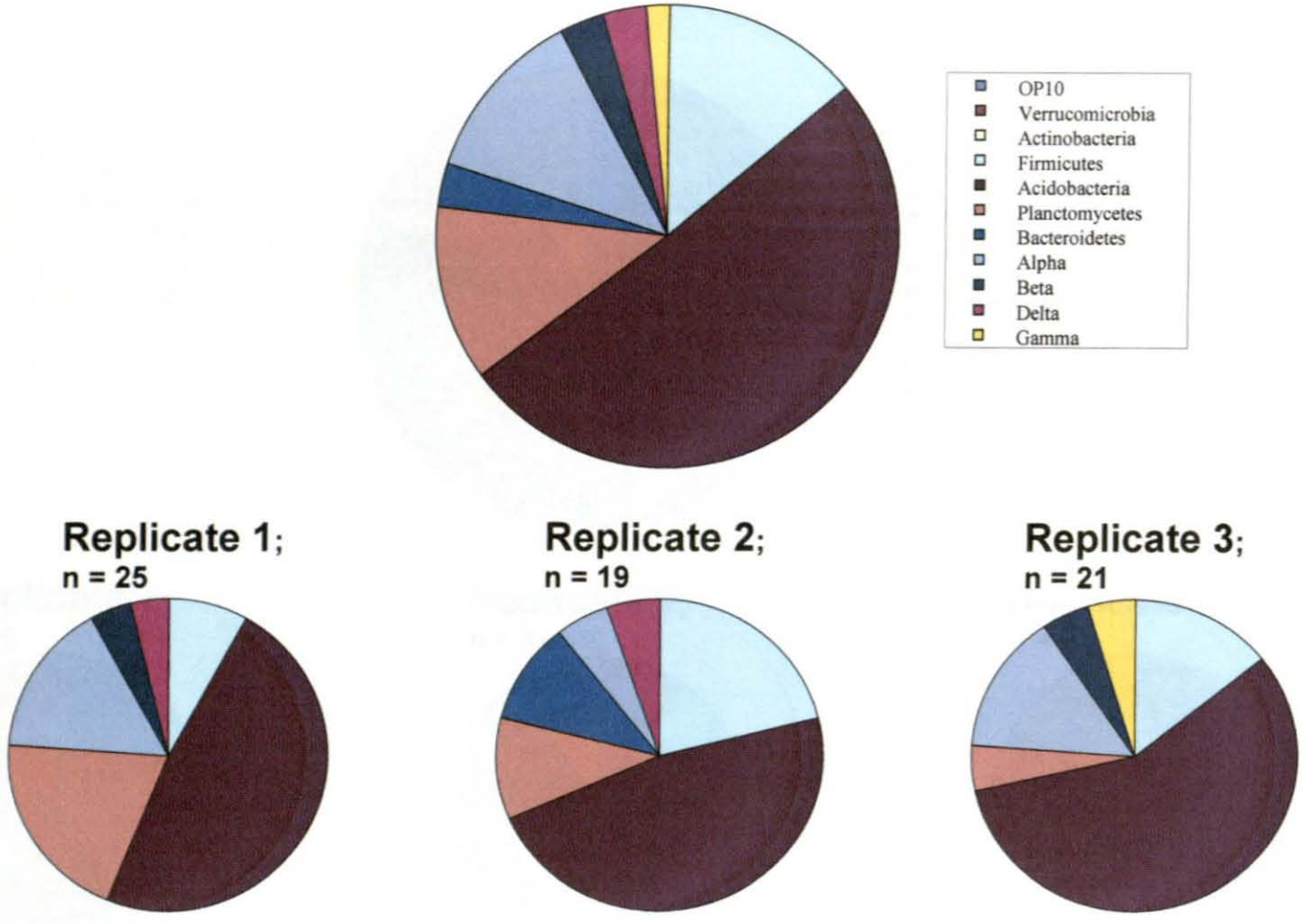

Figure 13. Phylum diversity patterns for Cataloochee broken down into the phylum diversity patterns for each of the three replicates.

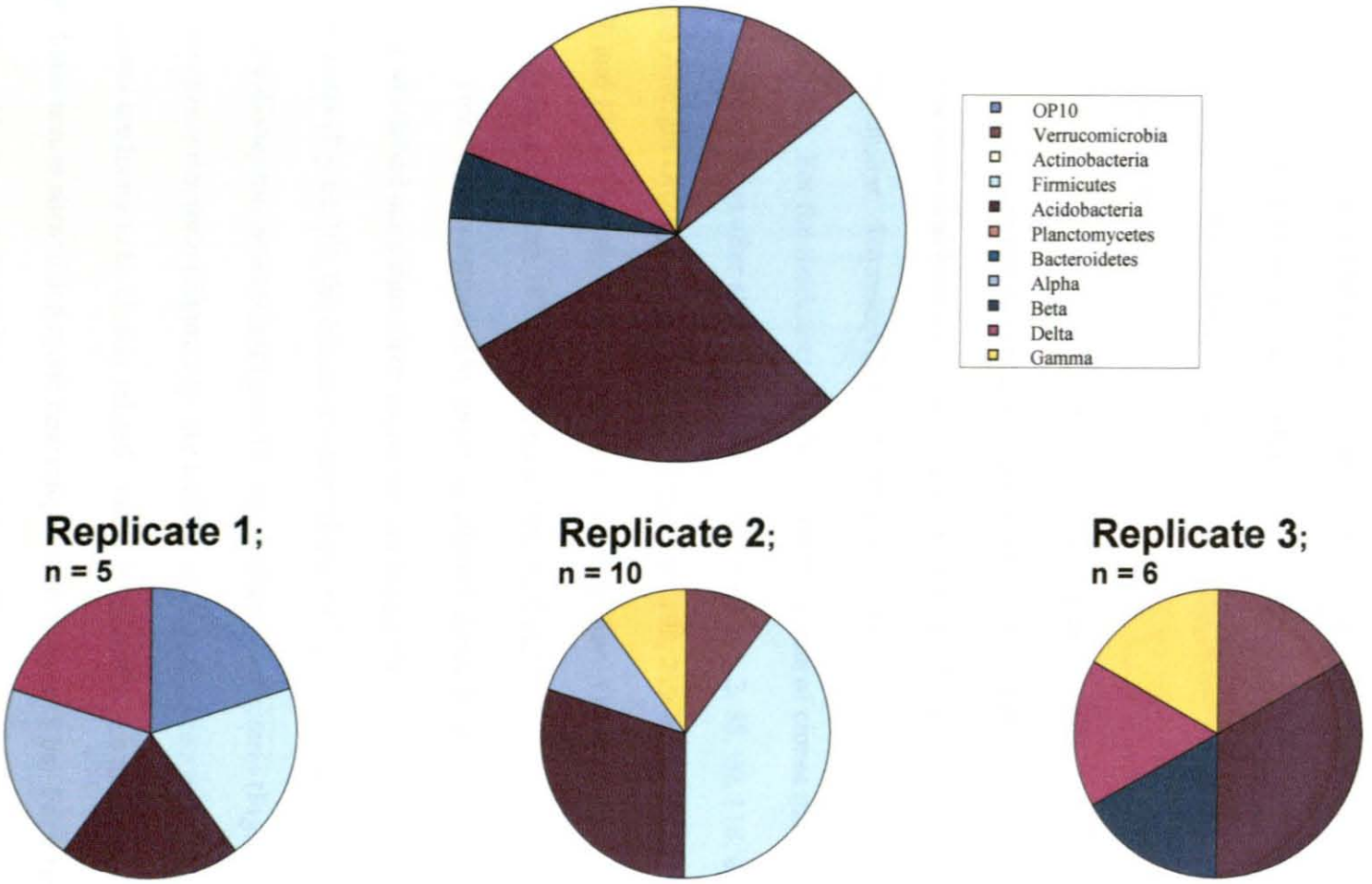

Figure 14. Phylum diversity patterns for Purchase Knob broken down into the phylum diversity patterns for each of the three replicates.

Another interesting result found when looking at the RDP II "Classifier" data was the number of genera detected in the eleven phyla (Table 4). These data suggest that the *Firmicutes* represents the phylum with the most genera detected (20). It is also noteworthy that within the most predominant division, *Acidobacteria*, there is only one genus, *Acidobacterium*.

The proposed phyla OP10, *Bacteroidetes, Actinobacteria*, and *Betaproteobacteria* were not included in treeing since few clones were obtained from these groups, however, differences were seen between sites. Sequences 19 and 307 were of the phyla OP10 and were 94% similar. Sequences 205 and 249 were of the phyla *Bacteroidetes* and were 91% similar. For the *Acidobacteria*, there were eight sets of clones that had 100% sequence identities after alignment. These were clones 1, 2, 38; 90, 118; and 104, 140 from Albright Grove; 176, 197; 209, 213; and 168, 192, 219 from Cataloochee; and 102, 257, and 127, 275 from Albright Grove and Cataloochee. One *Firmicutes* clone overlap was also seen, between 183 and 241, both from the Cataloochee site.

Trees were constructed by grouping aligned clones by phylum or division for any group which had more than three sequences, including the *Acidobacteria* (Figure 15), the *Firmicutes* (Figure 16), the *Planctomycetes* (Figure 17), the *Verrucomicrobia* (Figure 18), the *Alphaproteobacteria* (Figure 19), the *Deltaproteobacteria* (Figure 20), and the *Gammaproteobacteria* (Figure 21). By looking at these trees one can assume that clades of clones are likely to be closely related. These trees enable one to compare relatedness of clones across sites to determine how unique each clade may be. For a simple example, in Figure 18, three clades of *Verrucomicrobia* are illustrated, indicating clones 23 and 31

Table 4. A complete list of all the different genera represented by the eleven phyla found within the entire clone library. (Each genus is color coded to correspond with its correct phylum).

Division	Albright Grove	Cataloochee	Purchase Knob
Firmicutes	*Acetanaerobacterium* *Anaeroglobus* *Bryantella* *Caminicella* *Faecalibacterium* *Gelria* *Pelotomaculum* *Shuttleworthia* *Subdoligranulum* *Syntrophothermus* *Thermacetogenium* *Thermobrachium* *Thermodesulfobium* *Thermohalobacter*	*Acidaminobacter* *Anaeroglobus* *Faecalibacterium* *Johnsonella* *Quinella* *Subdoligranulum* *Thermacetogenium* *Thermanaeromonas* *Thermobrachium*	*Anaerobaculum* *Faecalibacterium* *Soehngenia* *Subdoligranulum* *Thermacetogenium*
Alphaproteobacteria	*Blastochloris* *Bradyrhizobium* *Odyssella* *Methylosinus*	*Acidisphaera* *Bradyrhizobium* *Magnetospirillum* *Methylosinus* *Phenylobacterium* *Tistrella*	*Rhodoplanes* *Roseomonas*
Betaproteobacteria		*Burkholderia* *Tepidiphilus*	*Caenibacterium*
Deltaproteobacteria	*Desulfomonile*	*Desulforegula*	*Hippea*
Gammaproteobacteria	*Alkalispirillum* *Rickettsiella* *Thiorhodospira*	*Isochromatium*	*Alkalispirillum*
Acidobacteria	*Acidobacterium*	*Acidobacterium*	*Acidobacterium*
Actinobacteria	*Acidimicrobineae*		
Bacteroidetes		*Chitinophaga*	
OP10	*OP10*		*OP10*
Planctomycetes	*Isosphaera*	*Isosphaera* *Planctomyces*	
Verrucomicrobia	*Verrucomicrobium*		*Verrucomicrobium*

from Albright Grove cluster together but separately from clones 380 and 401, which form distinct clades from Purchase Knob. For *Planctomycetes* one clade indicated that clones 137 and 190 were from Albright Grove and Cataloochee, respectively. The other clades consisted of six clones that were all from Cataloochee. For the *Alphaproteobacteria* clone 14 from Albright Grove is by itself while other clades have clones representing all three sites. The *Deltaproteobacteria* has three clades inidicating clone 2 from Albright Grove separate from clones 171 and 240 from Cataloochee and clones 330 and 415 from Purchase Knob. The *Gammaproteobacteria* has four clades with clone 83 from Albright Grove separate from clones 116 and 424 from Albright Grove and Purchase Knob, respectively. Also, clone 261 from Cataloochee is separate from clones 150 and 377 from Cataloochee and Purchase Knob, respectively. For the *Acidobacteria,* 42 clades had clones represented in only one site and 16 clades had clones represented in two or more sites. For the *Firmicutes*, 20 clades had clones represented in only one site and only 2 clades had clones represented in two sites.

Finally, the clones can be classified by grouping them within sequence similarity boundaries, a technique for simplifying data for 16S rDNA sequences from clone libraries (Hong et al. 2006; Table 5). For each division, clones were grouped by sequence identities of 100% (same sequence), 99%, 98%, 97% (same species), 96%, 95% (same genus), 90-94% (same family/class), 80-89% (same phylum), and 70-79% and 60-69% (deep differences between clones). Interestingly, for all divisions the majority of sequence identities fell in the 80-89% category. Therefore, over 50% of the clones were 89% or less similar to the entire clone library. In the division *Acidobacteria* the majority

were in the 80-89% and the 90-94% categories, indicating major subdivisions within this phylum. This deep diversion of sequence similarities was also true of the *Verrucomicrobia* division. The *Planctomycetes, Alphaproteobacteria,* and *Gammaproteobacteria* clones mostly fell into the 80-89% category. The *Firmicutes, Betaproteobacteria,* and *Deltaproteobacteria* were largely in the 70-79% category. It was also noteworthy that the *Acidobacteria* and *Firmicutes* divisions were the only divisions where a 100% similarity was found with ten identical sequences being found in *Acidobacteria* and one in the *Firmicutes*. For the OP10 and *Bacteroidetes* there were only two representatives, which were found to be 94% and 91% similar, respectively. Also, since there was only one representative for the division *Actinobacteria*, it was left out of Table 5; however, its closest relative in RDP II was 78% similar to it.

Table 5. Comparison of sequence identities to all clones within a phylum as defined by operational taxonomic units (OTU) with 100% representing clones of identical sequence, 97% representing clones likely of the same species, 95% representing clones from the same genus, 90% representing clones of the same family or class, and 80% representing clones within the same phylum. (*Alpha, Beta, Delta, and Gamma* are subphyla of the *Proteobacteria* division).

	Acidobacteria		*Firmicutes*		*Verrucomicrobia*		*Planctomycetes*	
OTU Boundary	Clones Pertaining to Each Boundary	% of Phylum Found at Boundary	Clones Pertaining to Each Boundary	% of Phylum Found at Boundary	Clones Pertaining to Each Boundary	% of Phylum Found at Boundary	Clones Pertaining to Each Boundary	% of Phylum Found at Boundary
100%	10	0.24%	1	0.16%	0	0%	0	0%
99%	35	0.83%	3	0.48%	0	0%	0	0%
98%	105	2.49%	15	2.38%	0	0%	0	0%
97%	125	2.96%	14	2.22%	0	0%	1	2.78%
96%	138	3.27%	9	1.43%	0	0%	2	5.56%
95%	176	4.17%	5	0.79%	1	10.00%	0	0%
90-94%	1417	33.59%	94	14.92%	2	20.00%	6	16.67%
80-89%	2045	48.47%	68	10.79%	3	30.00%	18	50.00%
70-79%	47	1.11%	361	57.30%	0	0%	1	2.78%
60-69%	30	0.71%	1	0.16%	0	0%	0	0%
	Alpha		*Beta*		*Delta*		*Gamma*	
OTU Boundary	Clones Pertaining to Each Boundary	% of Phylum Found at Boundary	Clones Pertaining to Each Boundary	% of Phylum Found at Boundary	Clones Pertaining to Each Boundary	% of Phylum Found at Boundary	Clones Pertaining to Each Boundary	% of Phylum Found at Boundary
100%	0	0%	0	0%	0	0%	0	0%
99%	3	1.75%	0	0%	0	0%	1	4.76%
98%	9	5.26%	0	0%	0	0%	0	0%
97%	0	0%	0	0%	0	0%	0	0%
96%	2	1.17%	0	0%	0	0%	0	0%
95%	0	0%	0	0%	0	0%	0	0%
90-94%	16	9.36%	0	0%	0	0%	2	9.52%
80-89%	68	39.77%	1	16.67%	1	6.67%	11	52.38%
70-79%	53	30.99%	2	33.33%	9	60.00%	1	4.76%
60-69%	2	1.17%	0	0%	0	0%	0	0%

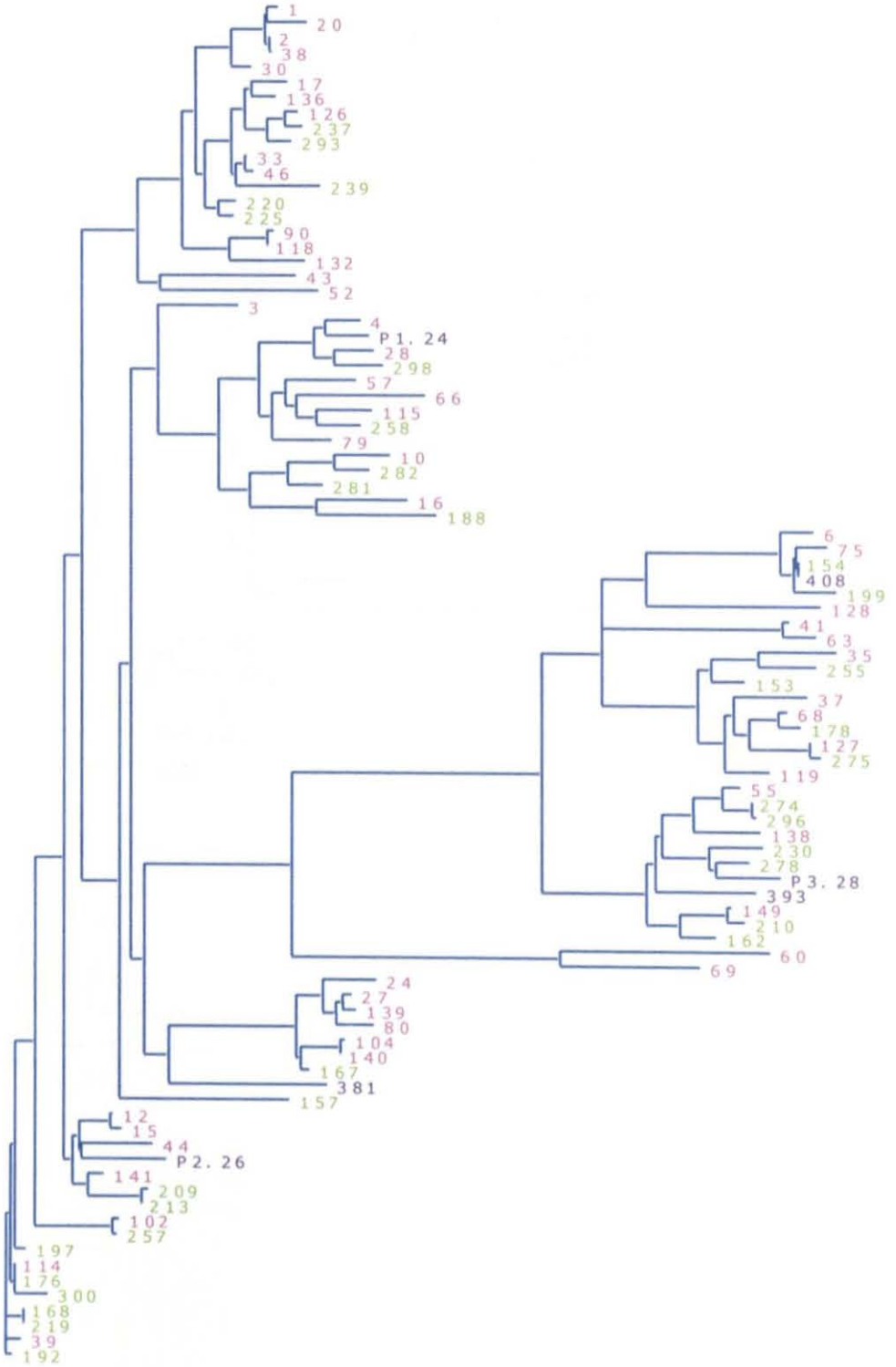

Figure 15. Tree formation for the clones represented in the *Acidobacteria* phylum. (Pink=Albright Grove, Green=Cataloochee, and Purple=Purchase Knob.)

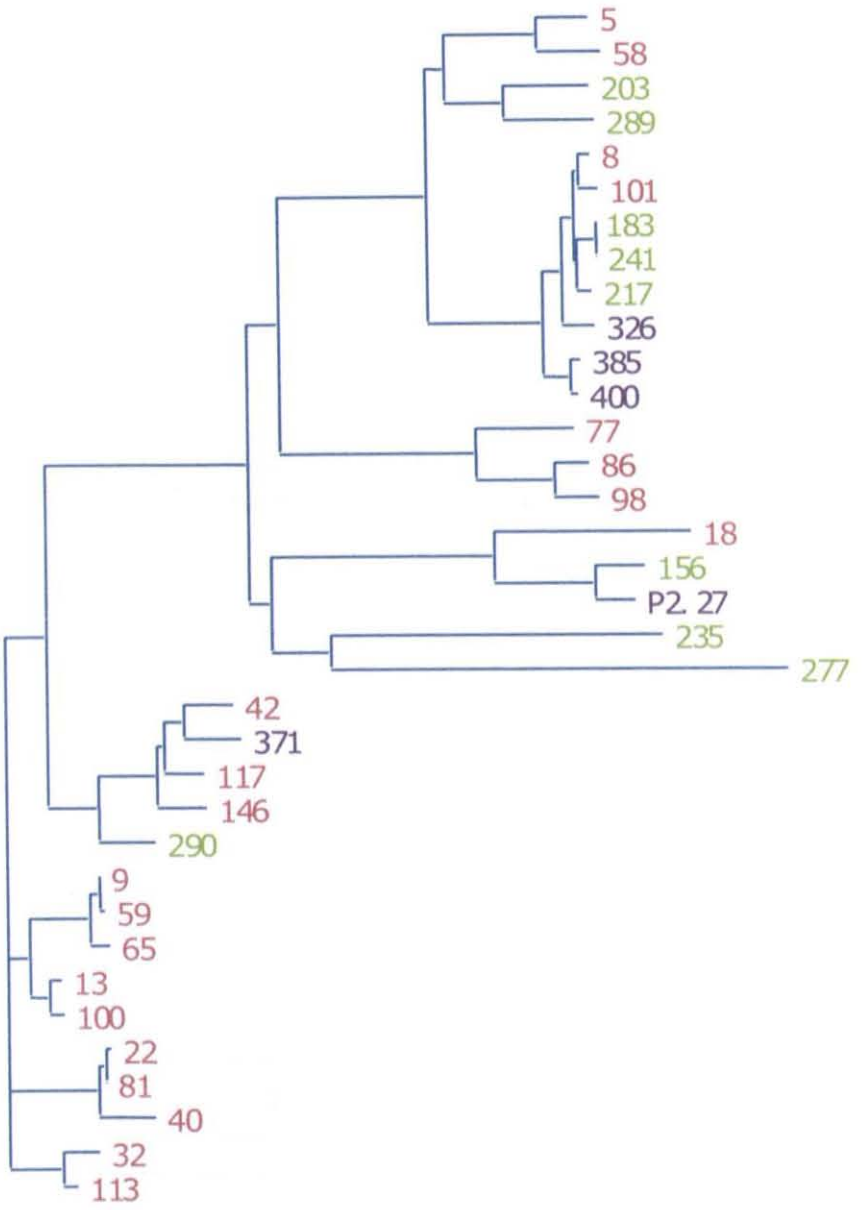

Figure 16. Tree formation for the clones represented in the *Firmicutes* phylum. (Pink=Albright Grove, Green=Cataloochee, and Purple=Purchase Knob.)

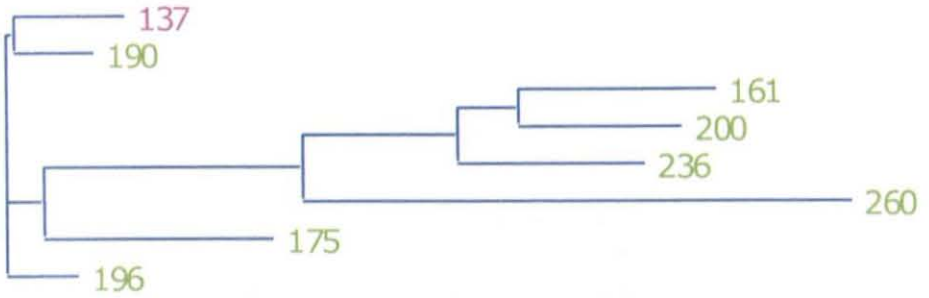

Figure 17. Tree formation for the clones represented in the *Planctomycetes* phylum. (Pink=Albright Grove, Green=Cataloochee, and Purple=Purchase Knob.)

Figure 18. Tree formation for the clones represented in the *Verrucomicrobia* phylum. (Pink=Albright Grove, Green=Cataloochee, and Purple=Purchase Knob.)

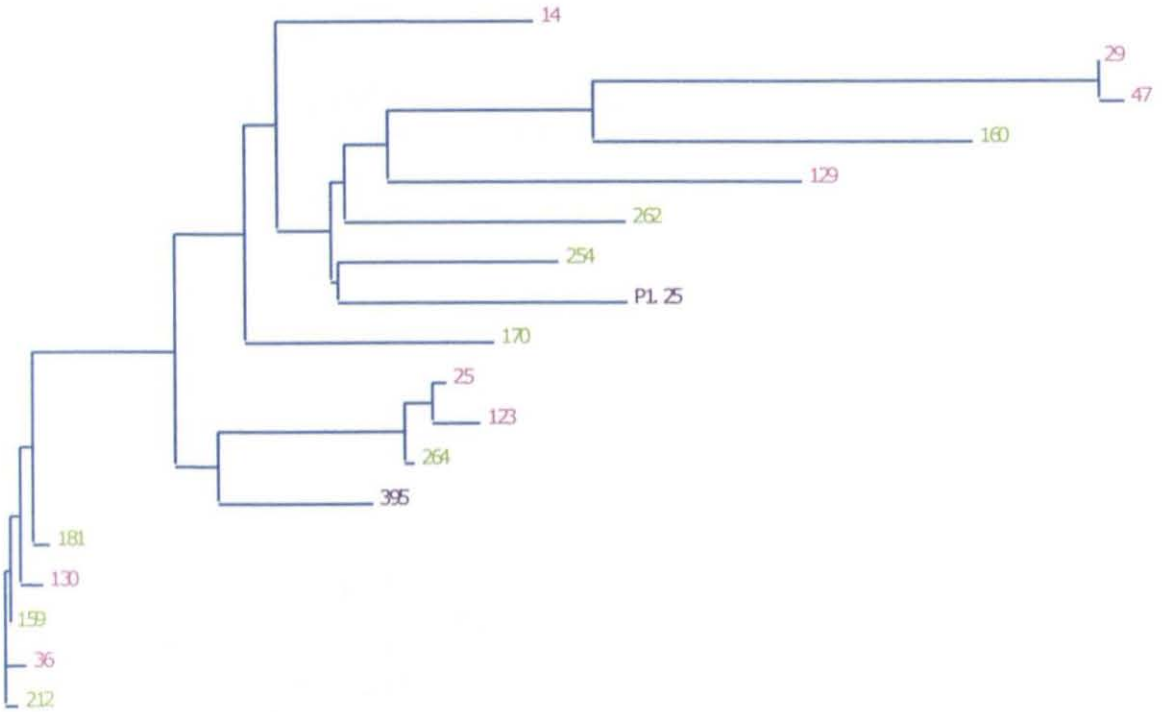

Figure 19. Tree formation for the clones represented in the *Alphaproteobacteria* sub-phylum. (Pink=Albright Grove, Green=Cataloochee, and Purple=Purchase Knob.)

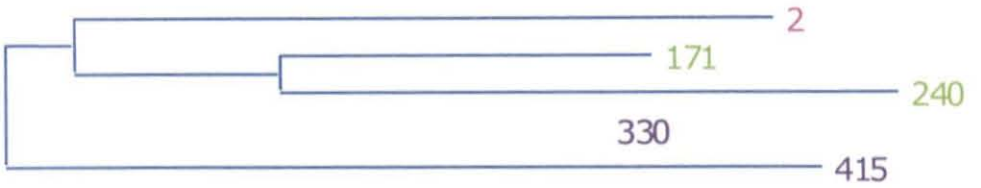

Figure 20. Tree formation for the clones represented in the *Deltaproteobacteria* sub-phylum. (Pink=Albright Grove, Green=Cataloochee, and Purple=Purchase Knob.)

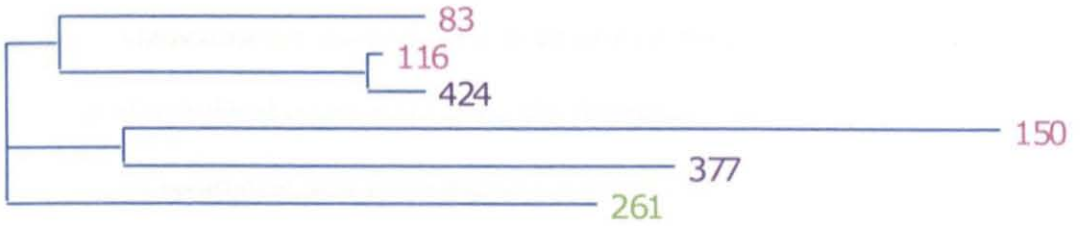

Figure 21. Tree formation for the clones represented in the *Gammaproteobacteria* sub-phylum. (Pink=Albright Grove, Green=Cataloochee, and Purple=Purchase Knob.)

Discussion

Bacteria are found in every environment that supports eukaryotes and even those that do not (Madigan et al. 2003). Due to the vast array of environments within which these bacteria are living, there is clearly tremendous ecological diversity within the prokaryotic world (Cohan 2001). In fact, the amount of diversity is so high that after years of characterizing the prokaryotic realm only a window to this diversity has been opened. Many different methods have been used to characterize prokaryotes resulting in patterns of individual organisms falling into discrete clusters on the basis of their phenotypic, ecological, and DNA sequence characteristics (Cohan 2001). Interestingly, when observing community patterns of bacteria in soil a shift in what were thought to be the prevalent taxonomic divisions has occurred due to the availability of modern molecular approaches to diversity (Janssen 2006).

Soil bacteria are an essential component of the community in forests, and they are largely responsible for ecosystem functioning because they participate in most nutrient transformations (Hackl et al. 2004). Although the bulk of the diversity of life has been proven to be microbial, the vast majority of soil bacteria still remain unknown because only a minor percentage of naturally occurring microorganisms can be cultured (Pace 1997). In 1977, Martin Alexander listed in the second edition of his book *Introduction to Soil Microbiology* what were at that time considered to be the most important genera of

soil bacteria based on cultivation studies. He suggested that there were nine genera that were significant in soils: *Agrobacterium*, *Alcaligenes*, *Arthrobacter*, *Bacillus*, *Flavobacterium*, *Micromonospora*, *Nocardia*, *Pseudomonas*, and *Streptomyces* (Alexander 1977). Through the years since there have been two major changes in microbiology that have caused this list to be called into question. First, many of these genera listed have undergone taxonomic changes causing them to be grouped into other taxonomic categories. Second, and possibly more important, new methodology using molecular approaches has allowed for surveying of 16S rRNA genes in soil permitting a more direct census of soil bacteria without the limitations of culturing. These new approaches now show that Alexander's list of nine genera actually only make up about 2.5 to 3.2% of soil bacteria (Janssen 2006).

I had proposed to see a different community of bacteria using culture-independent methods from previous culture-dependent work. This in fact was true. Previous culture-dependent work resulted in the findings of four different phyla and three subphyla (Figure 3). The four phyla found were *Firmicutes*, *Bacteroidetes*, *Actinomycetes*, and *Proteobacteria*. Within the *Proteobacteria* the subphyla *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria* were found. Although these phyla and subphyla were also found in my study, the overall patterns of these phyla were different. For instance, at Albright Grove, the *Firmicutes* accounted for 75% of the isolates. The *Betaproteobacteria* followed at about 15% and the *Actinomycetes* and *Bacteroidetes* were both at about 5% of the isolates. However, the molecular clone work from Albright Grove idicated that the *Acidobacteria* were dominant at about 58% of the community and

the *Firmicutes* followed at about 23%. Following this trend, I observed that there were seven more phyla occurring in Albright Grove. Therefore, I was able to obtain a much higher phylum-level diversity with my clone work than the culture data showed. I was also able to detect unique phyla that are difficult to culture, but are apparently wide-spread in soil such as *Verrucomicrobia* and OP10 (Janssen 2006).

Even more defining differences were observed between the two different approaches to detect diversity at Cataloochee. In the culture-dependent methods, the *Betaproteobacteria* were found to be the most common followed closely by the *Firmicutes*. The *Bacteroidetes, Gammaproteobacteria,* and *Actinomycetes* were all distributed almost equally through the site followed finally by the least common group, the *Alphaproteobacteria*. Yet, in the culture-independent methods (Figure 10), the *Acidobacteria* dominated again followed by the *Firmicutes, Alphaproteobacteria,* and *Planctomycetes*. The *Bacteroidetes, Betaproteobacteria, Gammaproteobacteria,* and *Deltaproteobacteria* all followed as being the least common at Cataloochee. Again, while the culture-dependent methods detected five different phyla, I was able to detect seven phyla through cloning with some of these phyla representing groups that to date have few or no culture representatives.

Finally, differences were also discovered between Purchase Knob culture-dependent and culture-independent diversity patterns. The culture-dependent methods showed that the *Firmicutes* dominated followed again by the *Betaproteobacteria*. This was followed by the *Actinomycetes,* then the *Gammaproteobacteria* and *Bacteroides*. The *Alphaproteobacteria* were the least commonly found. In the culture-independent

methods (Figure 11), the *Acidobacteria* and *Firmicutes* dominated, followed by the *Alphaproteobacteria, Gammaproteobactera, Deltaproteobactera,* and *Verrucomicrobia*. The least common were the OP10 and *Betaproteobacteria*. This same trend of finding a higher phylum-level diversity including groups that are difficult to culture through cloning instead of culturing was also found in Purchase Knob.

It is obvious through these data that the two types of methodology affected the outcome of the community profile for each of the three sites. Not only were the phyla and subphyla patterns different, but the extent of diversity at each site was also much larger using molecular techniques compared to culture-based techniques. This is mostly because when using culturing as a method of detection, one is selecting for a particular phenotype (i.e., heterotrophic bacteria) based on media conditions. Although it may appear that this would limit the importance of culture-based methods, they are still needed in developing our understanding of bacterial physiology, genetics, and ecology (Janssen 2006). In fact, parallel study of laboratory cultures would strongly complement molecular ecological investigations and enhance research into the roles of soil bacteria and their biotechnological potentials. Assigning functions to bacteria known only by their 16S rRNA genes is a difficult task, and detailed investigations of their physiologies and genomes are even more challenging. The availability of pure cultures would greatly simplify such studies (Joseph et al. 2003).

Using molecular techniques has been shown to advance our knowledge of bacterial diversity greatly. Just as my results have shown a shift in community profiles from culture-based work to molecular-based work, so have many others. It is now known

that members of the phyla *Acidobacteria* and *Proteobacteria* are the most common in soil (Hugenholtz et al. 1998; Janssen 2006; Janssen et al. 2002; Joseph et al. 2003; and Rappé and Giovannoni 2003). The *Acidobacteria* group is a newly recognized bacterial phylum with very few cultivated representatives (Hugenholtz et al. 1998). This limitation provides little information regarding biochemical and metabolic properties that might be generally distributed throughout this phylum. In fact, the majority of sequences that make up this phylum are from environmental clones. Yet, the widespread occurrence of environmental sequences that have been found to belong to the *Acidobacteria* suggests that members of this group are ecologically significant constituents of many ecosystems, particularly in soil communities (Hugenholtz et al. 1998; Joseph et al. 2003; and Rappé and Giovannoni 2003). Some authors suggest that the *Acidobacteria* may be nearly as diverse as the *Proteobacteria*, but currently only three genera are defined in the former (Hugenholtz et al. 1998; Janssen 2006). The *Proteobacteria*, on the other hand, is represented by a large number of described subtaxa, including at least 528 named genera in 72 named families (Janssen 2006). Even with this large amount of information, analysis of soil bacterial communities by directly surveying 16S rRNA has revealed the presence of many clades at the genus, family, and order levels that are not represented by named species (Joseph et al. 2003). Through "Classifier" in the RDP II program about 60% of my clones that were assigned to the *Proteobacteria* phylum had less than 50% confidence at the genus, family, and order levels. This indicates that many proteobacterial groups still remain to be described and named in environmental samples.

Other bacterial phyla that are found to be dominant in libraries of soil samples include *Actinobacteria, Verrucomicrobia, Bacteroidetes, Chloroflexi, Planctomycetes, Gemmatimonadetes,* and *Firmicutes* (Janssen 2006; Janssen et al. 2002; and Joseph et al. 2003). Most of these phyla are virtually unstudied and have few or no known pure culture representatives from soils. These trends found in many other studies of bacterial diversity in soil are reflected throughout my findings in Great Smoky Mountains National Park. It is also important to note that when looking at the "Sequence Match" program that is part of RDP II, the majority of my clones were matched with those that had also been directly amplified from forest soil, indicating that these groups are common soil inhabitants.

Another one of my hypotheses was that each site would have different bacterial diversity based on differences in vegetation, elevation, and soil chemistry. When looking at each site separately (Figures 9, 10, and 11), it is apparent that Albright Grove has nine different phyla while Cataloochee and Purchase Knob each have eight different phyla. Therefore, each site is relatively diverse. One important factor when considering the amount of diversity within a site is the number of clones available for that sample set. While Albright Grove had 91 different clones analyzed, Cataloochee had 67, and Purchase Knob only had 20. This means that while Purchase Knob had less than a quarter of the clones that Albright Grove had, Albright Grove still only had one more phylum than Purchase Knob. Similarly, Purchase Knob had less than a third of the clones compared with Cataloochee but had the same number of phyla. Therefore, when comparing the number of clones to the number of phyla, the ratios at the different sites

were Albright Grove = 0.09; Cataloochee = 0.12; and Purchase Knob = 0.40 phylum per clone, respectively. Puchase Knob displayed a much higher phylum diversity than the other two sites, and it would be interesting to see what patterns would emerge if a greater number of clones were able to be sequenced. There was some disparity in why we saw a drop in number of clones between sites. One possibility could have been that there was not enough X-Gal on the blue/white plates causing some clones to appear white that were actually blue. Another discrepancy could have been that different people picked the clones between sites. Yet, even with the appearance of a low phylum-level diversity, the percent of clones that were unique at each site according to both RFLP patterns and the similarity matrix showed just how unique the microflora at each site was. For Albright Grove, the RFLP banding patterns showed that 84.7% of the clones were unique while the similarity matrix showed that 89.0% of the sequences for the clones were unique. For Cataloochee, the RFLP banding patterns showed that 77.9% of the clones were unique while the similarity matrix showed that 81.5% were unique. For Purchase Knob, both the RFLP and similarity matrix showed that 100% of the clones were unique. Therefore, while the *Acidobacteria* and *Firmicutes* phyla dominated every site there was still a high amount of diversity according to the uniqueness of the clones.

The differences between sites based on the types of phyla found were masked based on the *Acidobacteria* domination at every site. However, even within this hugely diverse group, patterns could be seen that distinguished clones from this phylum between sites (Figure 15). One problem that is reoccurring in this phylum that causes the appearance of limited diversity with the *Acidobacteria* is just how little we know about

this group. The *Acidobacteria* have only three formally described genera in the phylum that have been cultivated (Hugenholtz et al. 1998; Janssen 2006). Therefore, the majority of sequences that make up this phylum are from environmental clones. By looking at Figure 15, it is apparent that while only one genus was found, there are sequence differences between each clone. Therefore, a possible explanation might be that there are not enough cultivated representatives available defining these observed differences.

In examining the remaining phyla at each site, there were clear patterns of differences. These could be seen taking the non-*Acidobacteria* clones for each site and recalculating the percent that each were found (Table 6). By interpreting these percentages, one can easily see how each remaining phylum differs between sites.

Table 6. Percentage of phyla found within a site excluding the *Acidobacteria* clone data.

	Albright Grove	Cataloochee	Purchase Knob
Verrucomicrobia	7.70%	0%	13.30%
Planctomycetes	2.60%	25.00%	0%
OP10	2.60%	0%	6.70%
Firmicutes	53.80%	28.10%	33.30%
Bacteroidetes	0%	6.30%	0%
Actinobacteria	2.60%	0%	0%
Alphaproteobacteria	20.50%	25.00%	13.30%
Betaproteobacteria	0%	6.30%	6.70%
Deltaproteobacteria	2.60%	6.30%	13.30%
Gammaproteobacteria	7.70%	3.10%	13.30%

necessarily targeted for reduction. Fragmented nucleic acids (results of harsh conditions during extraction methods) are sources of artifacts in PCR and may contribute to the formation of chimeric PCR products. Also, various biotic and abiotic components of environmental ecosystems, such as inorganic particles or organic matter, affect lysis efficiency and may interfere with subsequent DNA purification (Narang and Dunbar 2004; and von Wintzingerode et al. 1997). During my methods development I performed many tests comparing DNA extraction kits to see which kit provided the most amplifiable DNA during PCR. This resulted in a method that removed some of these biotic and abiotic components from the sample enough to not interfere with PCR amplifications.

PCR amplification of the 16S rDNA and molecular cloning are the most common methods associated with biases (Narang and Dunbar 2004). The most common biases include PCR artifacts such as chimeras and heteroduplexes, choosing primers that will amplify the majority of prokaryotes in a sample, and efficiency of primer binding (Acinas et al. 1997; Baker et al. 2003; Hugenholtz and Huber 2003; Narang and Dunbar 2004; Qui et al. 2001; and von Wintzingerode et al. 1997). The appearance of PCR artifacts is a potential risk in the PCR-mediated analysis of complex microbiota as it suggests the existence of organisms that do not actually exist in the sample investigated (von Wintzingerode et al. 1997). Chimeras can be generated during the PCR process as DNA strands compete with specific primers during the annealing process and two sequences from two different species anneal to make one sequence consisting of DNA from two species (Ashelford et al. 2005). This causes the sequence to appear to be "unclassified" according to the RDP II database (Maidak et al. 2001). Chimeric anomalies have long

been recognized and if left undetected can generate misleading impressions of environmental diversity. It is also known that these chimeric anomalies have been known to accumulate in public databases (Ashelford et al. 2005). Sequences in this study were checked with Belephron, Mallard, and Pintail and resulted in no sequences that could be claimed as a chimera. On the other hand, sequences in this study were not checked for heteroduplexes. When a heteroduplex molecule is cloned and transformed, two homoduplex molecules of 16S rRNA genes will be produced and segregated as a result of plasmid propagation (Qui et al. 2001). When these are then subjected to methods such as RFLP they result in artificial RFLP patterns. Heteroduplexes can be determined by comparing RFLP banding patterns to those of reference homoduplex molecules. If the clones show extra bands that migrate more slowly than the homoduplex molecules but faster than single-stranded DNA molecules they can be considered heteroduplexes (Qui et al. 2001). This can also lead to double bands in DGGE gels as well.

The other problem with PCR revolves around the primers used. For a study such as this one, "universal" primers are used in order to amplify as much of the prokaryotic community as possible. It is important to know that no primers in current use are truly universal and no single set of primers can be recommended that are guaranteed to amplify all prokaryotes (Baker et al. 2003). Consequently, many 16S rDNA libraries will not be totally representative of microbial communities, especially on a quantitative level. Samples would have to be amplified with several different primers in order to have a more complete community analysis and this would represent a significant increase in laboratory time and expense. Primers can also affect PCR when considering varying

quantities of template DNA. In large quantities, the primers will find the most common DNA strands more often than the rare, which will then dominate the reaction as they multiply exponentially (Baker et al. 2003). Finally, biases can occur when analyzing sequences. Not only are artifacts in PCR going to be a problem (as mentioned before), but the quality of results obtained by comparative 16S rRNA sequence analyses strongly depends on the available dataset (von Wintzingerode et al. 1997). Even though the dataset of RDP II contains hundreds of thousands of sequences, this number only reflects a minor part of the expected microbial diversity. As seen in this data set, a low sequence similarity to known sequences occurred quite often making their phylogenetic affiliation difficult. This leads to the question of whether environmental sequences represent uncultured, novel microorganisms or whether they cannot be assigned to known taxa due to the fact that for even many cultivated microorganisms, 16S rRNA and rDNA sequences are not available or are of low quality (von Wintzingerode et al. 1997).

Conclusions and Possible Future Work

Molecular methods used in this study produced 177 unique 16S rDNA sequences that did not match any previously found in Great Smoky Mountains National Park (GSMNP). Out of those 177 unique sequences only one common genus was found between the use of culture-dependent and culture-independent methods, *Burkholderia*, a member of the *Betaproteobacteria*. The phyla *Acidobacteria, Planctomycetes, Verrucomicrobia,* and OP10 were all phyla that were previously undetected in GSMNP through culture-dependent methods. Due to the high amounts of *Acidobacteria* found through these molecular methods and the findings of these new phyla, the community profiles of all three sites differed than the profiles from previous culture-dependent methods. Within the RDP II "Classifier" results, approximately 80 clones were determined to be "unclassifiable" due to low sequence matches to the database. Of these 80 clones, possible new species, new genera, or perhaps novel families or classes could be present.

There are several routes one could take to further investigate the diversity of bacteria within these clones. In regards to the 80 "unclassified" bacteria and the 77 clones that had less than 50% confidence according to "Classifier" the first step would be to sequence the entire 1500bp 16s rDNA region. It is hoped that, by sequencing the entire region instead of a ~500bp excerpt, confidence in the identification of a species would be more accurate. If low confidence rates still occurred or the sequence resulted in

the bacteria still unclassifiable, one could investigate the possibilities of a new species or new genera being found.

Another direction of future work could lead to cultivating some of these clones to learn more about their physiological and ecological roles in the environment. As of now, not much is known about the roles that these species play in their natural habitats due to discrepancies with culturing. It can be assumed that these high numbers of some phyla could only mean that these species are members of functionally dominant groups that may have a substantial impact on the environments they inhabit. Each clone, or ecotype, may play a vital role in carbon cycling (heterotrophy, chemolithotrophy), nitrogen cycling (fixation, ammonia oxidation, denitrification), sulfur cycling (sulfur oxidation, sulfate reduction), or any of the many other geochemical processes dominated by microorganisms. Also, by learning more about these roles that each species is playing, one might be able to hypothesize more on why these community profiles look the way they do and why they differ between sites based on forest history and environmental factors. The only way to truly understand the entire community is through long-term studies that utilize multiple culturing and molecular approaches.

Although many species were found in this clone library, all of these sequences were new to GSMNP (and probably to science), the community profiles from each site were found to differ completely from previous culturing approaches; however, it is suspected that diversity from each site is only a fraction of the complete bacterial community was discovered. Yet how would one completely assess a community with billions and perhaps millions of species? The best approach is to slowly piece together

the unknown as both molecular techniques and culturing take place; this will allow us to better understand and pose better hypotheses about the amazing realm of the bacterial world.

Literature Cited

Literature Cited

Acinas, S., R. Rodrigues-Valera, and C. Pedros-Alio. 1997. Spatial and temporal variation in marine bacterioplankton diversity as shown by RFLP fingerprinting of PCR amplified 16S rDNA. FEMS Microbial Ecology. 24: 27-40.

Alexander, M. 1977. Introduction to soil microbiology (2nd ed.). John Wiley and Sons, New York, NY.

Ashelford, K.E., N.A. Chuzhanova, J.C. Fry, A.J. Jones, and A.J. Weightman. 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Applied and Environmental Microbiology. 71: 7724-7734.

Baker, G.C., J.J. Smith, and D.A. Cowan. 2003. Review and re-analysis of domain-specific 16S primers. Journal of Microbiological Methods. 55: 541-555.

Campbell, C., S. Grayston, and S. Chapman. 1999. Soil health and microbial diversity. http://www.macaulay.ac.uk/annualreport/1999/soilhealth99.pdf

Casamayor, E.O., H. Schäfer, L. Bañeras, C. Pedrós-Alió, and G. Muyzer. 2000. Identification of and spatio-temporal differences between microbial assemblages from two neighboring sulfurous lakes: comparison by microscopy and denaturing gradient gel electrophoresis. Applied and Environmental Microbiology. 66: 499-508.

Cho, J.C. and J.M. Tiedje. 2001. Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. Applied and Environmental Microbiology. 67: 3677-3682.

Cohan, F.M. 2001. Bacterial species and speciation. Systems Biology. 50: 513-524.

Corinaldes, C., R. Danovaro, and A. Dell'Anno. 2005. Simultaneous recovery of extracellular and intracellular DNA suitable for molecular studies or studies from marine sediments. Applied and Environmental Microbiology. 71: 46-50.

Curtis, T.P., W.T. Sloan, and J.W. Scannell. 2002. Estimating prokaryotic diversity and its limits. PNAS Early Edition. www.pnas.org/cgi/doi/10.1073/pnas.142680199

DeLong, E.F. and N.R. Pace. 2001. Environmental diversity of bacteria and archaea. Systems Biology. 50: 470-478.

Discover life in America: Great Smoky Mountain National Park All Taxa Biodiversity Inventory. "*The role of ATBI's in the global biodiversity crisis: notes from the Great Smokies ATBI.*" Great Smoky Mountains Association. 15 November 2004. <discoverlifeinamerica.org/atbi.>

Dunbar, J., S.M. Barns, L.O. Ticknor, and C.R. Kuske. 2002. Empirical and theoretical bacterial diversity in four Arizona soils. Applied and Environmental Microbiology. 68: 3035-3045.

Eckert, K. A., and T. A. Kunkel. 1991. DNA polymerase fidelity and the polymerase chain reaction. PCR Methods and Applications. 1: 17-24.

Fisher, M.M. and E.W. Triplett. 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. Applied and Environmental Microbiology. 65: 4630-4636.

Gans, J., M. Wolinsky, and J. Dunbar. 2005. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science. 309: 1387-1390.

Greene, E.A. and G. Voordouw. 2003. Analysis of environmental microbial communities by reverse sample genome probing. Journal of Microbiological Methods. 53: 211-219.

Hackl, E., S. Zechmeister-Boltensern, L. Bodrossy, and A. Sessitsch. 2004. Comparison of diversities and compositions of bacterial populations inhabiting natural forest soils. Applied and Environmental Microbiology. 70: 5057-5065.

Handelsman, J. 2004. Metagenomics: application of genomes to uncultured microorganisms. Microbiology and Molecular Biology Reviews. 68: 669-685.

Hong, S-H., J. Bunge, S-O. Jeon, and S.S. Epstein. 2006. Predicting microbial species richness. Proceedings of the National Academy of Sciences. 103: 117-122.

Huber, T, G. Faulkner, and P. Hugenholtz. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20:2317-2319.

Hugenholtz, P. and T. Huber. 2003. Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. International Journal of Systematic and Evolutionary Microbiology. 53: 289-293.

Hugenholtz, P., B.M. Goebel, and N.R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. Journal of Bacteriology. 180: 4765-4774.

Hughes, J.B., J.J. Hellmann, T.H. Ricketts, and B.J.M. Bohannan. 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. Applied and Environmental Microbiology. 67: 4399-4406.

Janssen, P.H. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rDNA and 16S rRNA genes. Applied and Environmental Microbiology. 72: 1719-1728.

Janssen, P.H., P.S. Yates, B.E. Grinton, P.M. Taylor, and M. Sait. 2002. Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions *Acidobacteria, Actinobacteria, Proteobacteria,* and *Verrucomicrobia.* Applied and Environmental Microbiology. 68: 2391-2396.

Joseph, S.J., P. Hugenholtz, P. Sangwan, C.A. Osborne, and P.H. Janssen. 2003. Laboratory cultivation of widespread and previously uncultured soil bacteria. Applied and Environmental Microbiology. 69: 7210-7215.

Jurgens, G. 2002. Molecular phylogeny of *Archaea* in boreal forest soil, freshwater and temperate estuarine sediment. Academic Dissertation. Helsinki University Printing House.

Kirk, J.L., L.A. Beaudette, M. Hart, P. Moutoglis, J.N. Klironomos, H. Lee, and J.T. Trevors. 2004. Methods of studying soil microbial diversity. Journal of Microbiological Methods. 58: 169-188.

Lee, D.H., Y.G. Zo, and S.J. Kim. 1996. Nonradioactive method to study genetic profiles of natural bacterial communities by PCR-single strand conformation polymorphism. Applied and Environmental Microbiology. 62: 3112-3120.

Liu, W.T., T.L. Marsh, H. Cheng, and L.J. Forney. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphism of genes encoding 16S rRNA. Applied and Environmental Microbiology. 63: 4516-4522.

Madigan, M.T., J.M. Martinko, and J. Parker. 2003. Brock biology of microorganisms: tenth edition. Prentice-Hall Inc., Upper Saddle River, NJ.

Maidak, B. L., Cole J.R., T. G. Lilburn, C. T. Parker Jr., P. R. Saxman, R. J. Farris, G. M. Garrity, G. J. Olsen, T. M. Schmidt, J. M. and Tiedje. 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Research. 29:173-174.

Muyzer, G. 1999. DGGE/TGGE a method for identifying genes from natural ecosystems. Current Opinions in Microbiology. 2: 317-322.

Muyzer, G., T. Brinkhoff, Nübel, C. Santegoeds, H. Schäfer, and C. Wawer. 1998. Denaturing gradient gel electrophoresis (DGGE) in microbial ecology. In Molecular Microbial Ecology Manual. (Akkermans ADL, van Elsas JD, and de Bruijn FJ, eds.), 3.4.4:1-27. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Narang, R. and J. Dunbar. 2004. Modeling bacterial species abundance from small community surveys. Microbial Ecology. 47: 396-406.

O'Connell, S. 2002. Microorganisms as the cornerstone of the Smoky Mountain Ecosystem. All Taxa Biodiversity Inventory (ATBI) Quarterly. 3:1.

O'Connell, S. 2003. "Subvisibile" biodiversity in Great Smoky Mountains National Park. All Taxa Biodiversity Inventory (ATBI) Quarterly. 4:1.

Olsen, G.J., Matsuda, H., Hagstrom, R. and Overbeek, R. 1994. FastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. CABIOS. 10: 41-48.

Pace, N.R. 1996. New perspective on the natural microbial world: molecular and microbial ecology. ASM News. 62: 463-470.

Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. Science. 276: 734-740.

Qui, X.Y., L.Y. Wu, H.S. Huang, P.E. McDonel, A.V. Palumbo, J.M. Tiedje, and J.Z. Zhou. 2001. Evaluation of PCR-generated chimeras: mutations, and heteroduplexes with 16S rRNA gene-based cloning. Applied and Environmental Microbiology. 67: 880-887.

Rappé M.S. and S.J. Giovannoni. 2003. The uncultured microbial majority. Annual Review of Microbiology. 57: 369-394.

Schloss, P.D. and J. Handelsman. 2004. Status of the microbial census. Microbiology and Molecular Biology Reviews. 68: 686-691.

Sharkey, M.J. 2001. The all taxa biological inventory of the Great Smoky Mountains National Park. Florida Entomologist. 84: 556-564.

Speksnijder, A., G.A. Kowalchuk, S. DeJong, E. Kline, J.R. Stephen, and H.J. Laanbroek. 2001. Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. Applied and Environmental Microbiology. 67: 469-472.

Staddon, W.J., L.C. Duchesne, and J.T. Trevors. 1998. Impact of clear-cutting and prescribed burning on microbial diversity and community structure in a Jack pine (*Pinus banksiana* Lamb.) clear-cut using Biolog Gram-negative microplates. World Journal of Microbiology and Biotechnology. 14: 119-123.

Staley, J. T., and A. Konopka. 1985. Measurements of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annual Review of Microbiology. 39:321-346.

Tiedje, J.M., S. Asuming-Brempong, K. Nusslein, T.L. Marsh, and S.J. Flynn. 1999. Opening the black box of soil microbial diversity. Applied Soil Ecology. 13: 109-122.

Torsvik, V. and Ovreas, L. 2002. Microbial diversity and function in soil: from genes to ecosystems. Current Opinion in Microbiology. 5: 240-245.

Torsvik, V., J. Goksoyr, and F.L. Daae. 1990a. High diversity in DNA of soil bacteria. Applied and Environmental Microbiology. 56: 782-787.

Torsvik, V., K. Salte, R. Soerheim, and J. Goksoeyr. 1990b. Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. Applied and Environmental Microbiology. 56: 77-781.

von Wintzingerode, F., U.B. Gobel, and E. Stackebrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiology Review. 21: 213-229.

Wang, G.C-Y. and Y. Wang. 1997. Frequency of formation of chimeric molecules is a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. Applied and Environmental Microbiology. 63: 4645-4650.

Wilson, E.O. 1994. Naturalist. Warner Brookes, New York.

Appendix

Appendix

Ribosomal Database Project II "Classifier" results for 16S rDNA bacterial clones obtained from soils from Albright Grove (clones 1-150), Catalocchee (clones 151-300), and Purchase Knob (clones 301-450) in Great Smoky Mountains National Park.

Clone #	Phylum	Class	Order	Family	Genus
1	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
2	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
3	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
4	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
5	*Firmicutes*	*Clostridia*	*Thermoanaerobacteriales*	*Thermodesulfobiaceae*	*Thermodesulfobium*
6	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
8	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Parasporobacterium*
9	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Thermobrachium*
10	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
12	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
13	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Bryantella*
14	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Hyphomicrobiaceae*	*Blastochloris*
15	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
16	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
17	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
18	*Firmicutes*	*Clostridia*	*Clostridiales*	*Acidaminococcaceae*	*Anaeroglobus*
19	OP10				*Genera incertae sedis*
20	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
21	*Proteobacteria*	*Deltaproteobacteria*	*Syntrophobacterales*	*Syntrophaceae*	*Desulfomonile*
22	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Caminicella*
23	*Verrucomicrobia*	*Verrucomicrobiae*	*Verrucomicrobiales*	*Verrucomicrobiaceae*	*Verrucomicrobium*
24	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
25	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Bradyrhizobiaceae*	*Bradyrhizobium*
26	*Actinobacteria*	*Actinobacteria*	*Acidimicrobiales*	*Acidimicrobiaceae*	*Acidimicrobium*

27	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
28	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
29	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Subdoligranulum*
30	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
31	*Verrucomicrobia*	*Verrucomicrobiae*	*Verrucomicrobiales*	*Verrucomicrobiaceae*	*Verrucomicrobium*
32	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Bryantella*
33	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
35	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
36	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Methylocystaceae*	*Methylosinus*
37	*Actinobacteria*	*Rubrobacteridae*	*Rubrobacterales*	*Rubrobacteraceae*	*Conexibacter*
38	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
39	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
40	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Caminicella*
41	*Firmicutes*	*Clostridia*	*Thermoanaerobacteriales*	*Thermoanaerobacteriaceae*	*Thermacetogenium*
42	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Bryantella*
43	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
44	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
46	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
47	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
52	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
55	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
57	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
58	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
59	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Thermobrachium*
60	*Firmicutes*	*Clostridia*	*Thermoanaerobacteriales*	*Thermoanaerobacteriaceae*	*Gelria*
63	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Subdoligranulum*
65	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Caloranaerobacter*
66	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
68	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
69	*Firmicutes*	*Clostridia*	*Thermoanaerobacteriales*	*Thermoanaerobacteriaceae*	*Thermanaeromonas*
75	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
77	*Firmicutes*	*Clostridia*	*Clostridiales*	*Syntrophomonadaceae*	*Syntrophothermus*
79	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*

80	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
81	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Thermohalobacter*
83	*Proteobacteria*	*Gammaproteobacteria*	*Chromatiales*	*Ectothiorhodospiraceae*	*Thiorhodospira*
86	*Firmicutes*	*Clostridia*	*Thermoanaerobacteriales*	*Thermoanaerobacteriaceae*	*Gelria*
90	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
98	*Firmicutes*	*Clostridia*	*Thermoanaerobacteriales*	*Thermoanaerobacteriaceae*	*Gelria*
100	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Thermobrachium*
101	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
102	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
104	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
113	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Thermobrachium*
114	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
115	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
116	*Proteobacteria*	*Gammaproteobacteria*	*Chromatiales*	*Ectothiorhodospiraceae*	*Alkalispirillum*
117	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Soehngenia*
118	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
119	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
123	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Bradyrhizobiaceae*	*Bradyrhizobium*
126	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
127	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
128	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
129	*Proteobacteria*	*Alphaproteobacteria*	*Rickettsiales*	*Incertae sedis*	*Odyssella*
130	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Methylocystaceae*	*Methylosinus*
132	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
136	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
137	*Planctomycetes*	*Planctomycetacia*	*Planctomycetales*	*Planctomycetaceae*	*Isosphaera*
138	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
139	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
140	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
141	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
146	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Acetanaerobacterium*
149	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
150	*Proteobacteria*	*Gammaproteobacteria*	*Legionellales*	*Coxiellaceae*	*Rickettsiella*

153	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
154	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
156	*Firmicutes*	*Clostridia*	*Thermoanaerobacteriales*	*Thermoanaerobacteriaceae*	*Thermacetogenium*
157	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
159	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Methylocystaceae*	*Methylosinus*
160	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacteriales*	*Desulfobacteraceae*	*Desulforegula*
161	*Planctomycetes*	*Planctomycetacia*	*Planctomycetales*	*Planctomycetaceae*	*Isosphaera*
162	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
167	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
168	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
170	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacteriales*	*Caulobacteraceae*	*Phenylobacterium*
171	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacteriales*	*Desulfobacteraceae*	*Desulforegula*
175	*Planctomycetes*	*Planctomycetacia*	*Planctomycetales*	*Planctomycetaceae*	*Isosphaera*
176	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
178	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
181	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Methylocystaceae*	*Methylosinus*
183	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Parasporobacterium*
184	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiaceae*	*Burkholderia*
188	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
190	*Planctomycetes*	*Planctomycetacia*	*Planctomycetales*	*Planctomycetaceae*	*Isosphaera*
192	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
196	*Planctomycetes*	*Planctomycetacia*	*Planctomycetales*	*Planctomycetaceae*	*Isosphaera*
197	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
199	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
200	*Planctomycetes*	*Planctomycetacia*	*Planctomycetales*	*Planctomycetaceae*	*Isosphaera*
203	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
205	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Crenotrichaceae*	*Chitinophaga*
209	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
210	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
212	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Methylocystaceae*	*Methylosinus*
213	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
217	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Subdoligranulum*
219	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*

220	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
225	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
230	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
235	*Firmicutes*	*Clostridia*	*Clostridiales*	*Acidaminococcaceae*	*Quinella*
236	*Planctomycetes*	*Planctomycetacia*	*Planctomycetales*	*Planctomycetaceae*	*Isosphaera*
237	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
239	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
240	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
241	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Subdoligranulum*
242	*Planctomycetes*	*Planctomycetacia*	*Planctomycetales*	*Planctomycetaceae*	*Planctomyces*
249	*Bacteroidetes*	*Sphingobacteria*	*Sphingobacteriales*	*Crenotrichaceae*	*Chitinophaga*
254	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Rhodospirillaceae*	*Magnetospirillum*
255	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
257	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
258	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
260	*Planctomycetes*	*Planctomycetacia*	*Planctomycetales*	*Planctomycetaceae*	*Planctomyces*
261	*Proteobacteria*	*Gammaproteobacteria*	*Chromatiales*	*Ectothiorhodospiraceae*	*Nitrococcus*
262	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Acetobacteraceae*	*Rhodopila*
264	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Bradyrhizobiaceae*	*Bradyrhizobium*
273	*Proteobacteria*	*Betaproteobacteria*	*Hydrogenophilales*	*Hydrogenophilaceae*	*Tepidiphilus*
274	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
275	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
277	*Firmicutes*	*Clostridia*	*Clostridiales*	*Syntrophomonadaceae*	*Aminomonas*
278	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
279	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
281	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
282	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
289	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Faecalibacterium*
290	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Thermobrachium*
293	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
296	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
298	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
300	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*

307	OP10				*genera incertae sedis*
326	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Subdoligranulum*
330	*Proteobacteria*	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*	*Hippea*
371	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Acetanaerobacterium*
377	*Proteobacteria*	*Gammaproteobacteria*	*Chromatiales*	*Ectothiorhodospiraceae*	*Alkalispirillum*
380	*Verrucomicrobia*	*Verrucomicrobiae*	*Verrucomicrobiales*	*Verrucomicrobiaceae*	*Verrucomicrobium*
381	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
385	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Soehngenia*
393	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
395	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Hyphomicrobiaceae*	*Rhodoplanes*
400	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Acidaminobacter*
401	*Verrucomicrobia*	*Verrucomicrobiae*	*Verrucomicrobiales*	*Verrucomicrobiaceae*	*Verrucomicrobium*
408	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
415	*Proteobacteria*	*Deltaproteobacteria*	*Desulfovibrionales*	*Desulfovibrionaceae*	*Lawsonia*
424	*Proteobacteria*	*Gammaproteobacteria*	*Chromatiales*	*Chromatiaceae*	*Rhabdochromatium*
445	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiaceae*	*Chitinimonas*
P.1.24	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
P.1.25	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Rhodospirillaceae*	*Azospirillum*
P.2.26	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*
P.2.27	*Firmicutes*	*Clostridia*	*Thermoanaerobacteriales*	*Thermoanaerobacteriaceae*	*Thermacetogenium*
P.3.28	*Acidobacteria*	*Acidobacteria*	*Acidobacteriales*	*Acidobacteriaceae*	*Acidobacterium*