POSSIBLE DETECTION OF PATHOGENIC BACTERIAL SPECIES INHABITING STREAMS IN GREAT SMOKY MOUNTAINS NATIONAL PARK

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

By

Kwame Agyapong Brown

Director: Dr. Sean O'Connell

Associate Professor of Biology and

Biology Department Head

Committee members: Dr. Sabine Rundle and Dr. Timothy Driscoll

November 2016

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my adviser Dr. Sean O'Connell for his insightful mentoring and thoughtful contribution throughout this research. I would also like to acknowledge my laboratory teammates: Lisa Dye, Rob McKinnon, Kacie Fraser, and Tori Carlson for their diverse contributions to this project. This project would not have been possible without the generous support of the Western Carolina University Biology Department and stockroom; so I would like to thank the entire WCU biology faculty and Wesley W. Bintz for supporting me throughout my Masters program. Finally, I would like to thank my thesis committee members and reader: Dr. Sabine Rundle, Dr. Timothy Driscoll and Dr. Anjana Sharma for their contributions to this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

POSSIBLE DETECTION OF PATHOGENIC BACTERIAL SPECIES INHABITING STREAMS IN GREAT SMOKY MOUNTAINS NATIONAL PARK

Kwame Agyapong Brown, M.S.

Western Carolina University, November 2016

Director: Dr. Sean O'Connell

Numerous pathogenic bacterial species have been found in many freshwater systems around the world. These pathogens affect the overall water quality of these systems and may cause diseases in both aquatic and terrestrial animals which may lead to loss of species diversity and abundance in their environments. This study sought to identify and document pathogenic bacterial species that may inhabit the streams that flow through Great Smoky Mountains National Park. Bacterial cells were collected by filtering water from four streams (Oconaluftee River, Kephart Prong, Little Pigeon River and Hickory King Branch Stream) through separate capsule filters. The cells were later backflushed from the filters and cultured on various selective and differential media. Ten isolates were selected based on phenotypic characteristics such as colony color and growth on specific media type, and sample origin. The nearly full 16S rDNA was sequenced for all ten isolates and analyzed to determine their identity.

Out of the ten isolates, four isolates were from the phylum Firmicutes while the other six were in the phylum *Proteobacteria*. Phylogenetic analysis of these isolates showed eight out of the ten isolates were related to known opportunistic pathogens. The other two were related to a

ubiquitous *Bacillus* species that is considered to be a probiotic. Although none of the isolates had a 100% match to a known obligate or opportunistic pathogen, many isolates matched > 97% to opportunistically pathogenic species. Follow up molecular and metabolic tests need to be employed to determine the pathogenicity of each isolate.

INTRODUCTION

Pristine water resources, both surface and groundwater, are becoming more scarce because of global increases in population and the action of humans in the environment (Geldreich 1996). Water-borne pathogen contamination in water resources and related diseases are a major water quality concern throughout the world (Pandey et al. 2014).

Bacteria and other single-celled microbes play a vital role in freshwater ecosystems (Findlay 2010). They perform numerous processes such as organic matter breakdown and nitrogen fixation (Findlay 2010). Bacteria are essential to the health of all water systems and also serve as food for other organisms present in the water bodies (Cairns 1971). Conversely, some bacterial species (classified as pathogens) are harmful to aquatic and other organisms that share the same freshwater body, including terrestrial organisms that drink the water.

Numerous pathogenic bacterial species have been found in many freshwater systems around the world (Geldreich 1996). These pathogens affect the overall water quality of these systems (Pandey et al. 2014) and cause diseases in both aquatic and terrestrial animals which may lead to loss of species diversity or species abundances in their environments. For example *Clostridium botulinum* and *Vibrio cholerae* are common pathogenic bacteria that cause fatal illnesses in terrestrial animals that drink water that is contaminated with fecal matter (Cabral 2010). Other pathogenic bacteria such as *Legionella pneumophila* are opportunistic and cause diseases in both aquatic and terrestrial animals that are immunocompromised or have a break in their physical protective barrier such as a cut on their skin (Dey 2009).

Water quality is routinely evaluated by determining the concentration of coliforms such as *Escherichia coli* (American Public Health Association 1965). Although this method is an

efficient way of determining water quality based on a particular subset of potential pathogenic bacteria, it does not identify the different species of bacterial pathogens that are present in the water and misses many other types.

Great Smoky Mountains National Park (GSMNP) contains one of the most diverse assemblages of plants, animals, microbes and invertebrates species in North America (Jenkins 2007). According to the National Park Service (NPS) only 18,545 species have been documented in the park: Scientists believe an additional 30,000-80,000 species of plants and animals (excluding bacteria and other microorganisms) may live there. Also there are about 19 major streams that exist in the park and these streams serve as a drinking water source for many of the diverse animals found in the park.

In 1998, GSMNP embarked on a project to determine all life forms that resided or spent time in the Park (Nichols & Langdon 2007). By 2006 over 300 species of bacteria and archaea were identified using DNA sequencing technology; many of them are novel species that have never before been described (Nichols & Langdon 2007). There is still ongoing research by systematists to identify the estimated 100,000 species of organisms that are found in the GSMNP, both microbial and larger.

Previous research performed by O'Connell et al. (2007) to inventory bacteria found within the soils and streams of the GSMNP revealed there was a total of eleven phyla present including six which were found only via culture-independent techniques. Overall they found about 69 genera of bacteria with *Bacteroidetes* being the dominant phylum in water (O'Connell et al. 2007). Although this research focused on general bacteria diversity and did not select for known pathogenic bacterial species, results from the study alluded to the presence of pathogenic species in the streams sampled. Bacteria from the genera *Bacillus, Pseudomonas, Enterobacter*

and Staphylococcus were cultured from water sampled from GSMNP (O'Connell et al. 2007). Certain species from these genera such as *Bacillus anthracis* and *Pseudomonas aeruginosa* are clinically relevant to humans, livestock and wildlife.

There is little known about the diversity of water-borne pathogenic bacteria that exist in the streams in the park. Since there are hundreds of water-borne pathogenic bacteria that can exist in a stream (Geldreich 1996), this study focused on those that could cause disease in terrestrial mammalian species that inhabit the Great Smoky Mountains including but not limited to *Staphylococcus, Pseudomonas* and *Salmonella* species. With the aid of 16S ribosomal DNA gene sequencing techniques and phylogenetic analysis, pathogenic bacterial species were targeted to be identified (Emerson et al. 2008), and added to the All Taxa Biodiversity Inventory (ATBI; Nichols and Langdon 2007). For this study, a 100% DNA sequence match to a known type strain pathogenic bacteria that shared an immediate common ancestor with the same type strain species would be counted as a positive finding of a pathogen from GSMNP. This work will lay the groundwork for further studies that will confirm each isolate's pathogenicity via immunoassays and antibiotic resistance testing, for example.

Specific Aims

The purpose of this research was to sample, identify and document potential pathogenic bacterial species that may inhabit the Oconaluftee River, Kephart Prong, Little Pigeon River and Hickory King Branch.

Significance

This study will help assess the water quality in GSMNP streams. Poor water quality could lead to various diseases that could end up sickening or killing many of the animals in the park that rely on these streams. In addition, knowing the diversity of pathogenic bacteria in these

streams could help the NPS to develop better management systems to control proliferation of these species and maintain pristine water quality.

METHODS

Site Description

Four streams were selected from GSMNP; two of the streams were located in North Carolina while the other two were located in Tennessee (Table 1). In North Carolina the streams sampled were the Oconaluftee River and Kephart Prong; and in Tennessee the streams sampled were the Little Pigeon River and Hickory King Branch. The Oconaluftee and Little Pigeon rivers are large streams (>18 ft wide) with little plant cover over the middle, but the banks and edges are surrounded by trees and other vegetation. The Little Pigeon River is rockier and faster flowing than the Oconaluftee River which is slow moving. Kephart Prong and Hickory King Branch are at a higher elevation and smaller than the other two streams (<18 ft wide). Kephart Prong is rocky and fast flowing while Hickory King Branch flowed slower than the other three. Both Hickory King Branch and Kephart Prong had a greater degree of shade compared to the Little Pigeon River and the Oconaluftee River, with Hickory King Branch being the most shaded among the sites sampled.

Sample Collection and Storage

Water samples were collected from four different sites using aseptic techniques. A custom-made water pump filtration system (description in Appendix) was used to draw water from the steams through a 0.22 μm sterile capsule filter (Pall Gelman Sciences). A 6-gallon Jerry can was used to estimate the amount of water filtered from each sample site. Each capsule was clamped shut using a small length of tubing and sealed after the designated amount of water had been filtered and they were then immediately stored on ice in a cooler at approximately 4°C. The capsules were later refrigerated at 4°C in the laboratory for 12 days.

Table 1: List of four sites sampled for bacteria in Great Smoky Mountains National Park including GPS coordinates and the volume of water filtered (L) at each site.

State	Sample Site	GPS Coordinates	Volume of Water Filtered (L)
North Carolina	Kephart Prong (NCKP)	35.35.222 N 83.21.431 W	22.7
	Oconaluftee (NCOR)	35.30.710 N 83.81.150 W	34.1
Tennessee	Little Pigeon River (TLPR)	35.63.336 N 83.51.526 W	34.1
	Hickory King Branch (THKB)	35.68.700 N 85.53.484 W	22.7

Recovery of Cells from Capsules

Bacterial cells were backflushed from the filter capsules. The backflushing protocol used to recover cells from the 0.22 μm filter involved warming the filters gradually to room temperature. Once each filter was brought to room temperature, ~150 mL of 0.1% sodium pyrophosphate (NaPP) was added (~80 mL into the outlet end and ~70 mL to the inlet end of the filter capsule). The filter ends were sealed using sterile tubing attached tightly to the filter and clamped shut. The filters were shaken on a culture shaker for 5 minutes at 400 rpm and then the clamp was removed on one end and sterile plastic tubing (tubing was sterilized in 10% bleach solution for five minutes and rinsed with sterile deionized water and 0.1% NaPP) was attached. This was then repeated for the other end of the capsule. The free end of tubing attached to the outlet end of the filter housing was tightly connected to a pressurized air nozzle in the laboratory. Air was forced into the filter housing at low to medium pressure and discharge collected from the inlet end in order to remove cells from the filter material. A total of approximately 150mL of

discharge was collected from the capsule and placed in a 1000-ml Erlenmeyer flask. This protocol was repeated for the other three capsules.

Enrichment Culturing and Screening of Bacteria (Gram Negative vs. Gram Positive)

About 350 ml of trypticase soy broth (TSB) was added to the cells recovered from each capsule filter and incubated at 37°C for 48 hours. After 48 hours of culture enrichment, the enrichment broth was divided into two equal aliquots with each aliquot poured into separate 500-ml Erlenmeyer flasks. One ml of phenylethyl alcohol (PEA) was added to one of the aliquots to select for Gram positive bacterial species while 0.1 ml of concentrated crystal violet solution (3.5g/liter) was added to the other aliquot to select for Gram negative bacterial species. PEA inhibits the growth of Gram negative bacteria by altering membrane permeability of cell and increasing leakage of cellular potassium (Silver and Wendt 1967, Leboffe and Pierce 2010). Crystal violet inhibits the growth of Gram positives by forming covalent adducts with thioredoxin reductase 2 (a highly conserved protein that is essential for cellular activity) once it penetrates the cell (Zhang et al. 2011). The flasks were shaken until a uniform mixture of culture and selective agent solution was achieved. Aluminum foil and parafilm were used to seal the flasks. This procedure was repeated for the other three samples from the other sites. All eight flasks were then incubated at 37°C for 48 hours.

Pathogen Selection

After 48 hours of incubation, each flask was shaken very well to suspend the bacterial cells in solution. Different media were employed to select for target pathogenic species as below and in Table 2.

Table 2. List of media used for enrichment and pathogen selection from four streams in Great Smoky Mountains National Park.

Media	Classification	Type of Organism Used to Isolate
Trypticase Soy Broth	Enrichment medium	All high nutrient requiring bacteria
Trypticase Soy Broth + Phenylethyl Alcohol	Selective medium	Gram positive bacteria
Trypticase Soy broth + Crystal Violet	Selective medium	Gram negative bacteria
Trypticase Soy Broth + 6.8% NaCl	Selective medium	Halotolerant bacteria
Mannitol Salt Agar (MSA)	Selective medium	*Staphylococcus* species
Hektoen Enteric Agar (HE)	Selective medium	*Salmonella* and *Shigella* species
MacConkey Agar	Selective medium	Gram negative enteric bacilli
Xylose Lysine Deoxycholate Agar (XLDA)	Selective medium	*Salmonella, Shigella and Providencia* species
Salmonella Typhimurium Isolation Agar (STIA)	Selective medium	*Salmonella typhimurium*
Pseudomonas Isolation Agar (PIA)	Selective medium	*Pseudomonas* species

Selection for Pseudomonas *Species*

Pseudomonas isolation agar (PIA) was prepared using dry premixed media and glycerol as indicated in the HIMEDIA product manual. One milliliter of Gram negative culture solution (TSB + 0.1 ml crystal violet solution + bacterial cells) was pipetted into a 2ml collection tube under sterile conditions. The collection tube was then placed on a high speed vortex to mix the cell suspension and a 100 µl aliquot was pipetted onto the solidified PIA media and spread-plated to uniformly cover the media with culture. This was done in triplicate and was repeated for the other three Gram negative culture solutions.

Selection for Salmonella typhimurium

Salmonella typhimurium isolation agar (STIA) was prepared according to the recipe laid out in the Handbook of Microbiological Media (Atlas 2010). The same protocol was used to spread Gram negative culture solution on the STIA plates as for the PIA media.

Selection for Gram negative Enteric Bacilli

MacConkey agar was used to screen for Gram negative enteric bacilli such as *Escherichia coli, Salmonella* species*, Shigella* species, and *Providencia* species. The same protocol utilized for PIA and STIA was employed for the MacConkey selection media. All media plates were incubated for 48 hours at 37°C. Unique bacterial colonies were selected and streaked onto three different media types: Hektoen enteric agar, xylose lysine tergitol-4 (XLT4) agar and xylose lysine deoxycholate agar. This procedure was done in triplicate and was repeated for the other three Gram negative culture solutions from all sites.

Selection for Staphylococcal Species

One hundred and twenty-five ml of Gram positive culture solution (TSB + PEA + bacterial cells) were added to 125ml of 13.6% NaCl TSB solution. This step yielded a 6.8% NaCl culture solution which was then incubated at 37°C for 48 hours. After incubation, a spread plate technique was utilized to culture bacterial colonies on mannitol salt agar plates. All plates were incubated for 48 hours at 37°C.

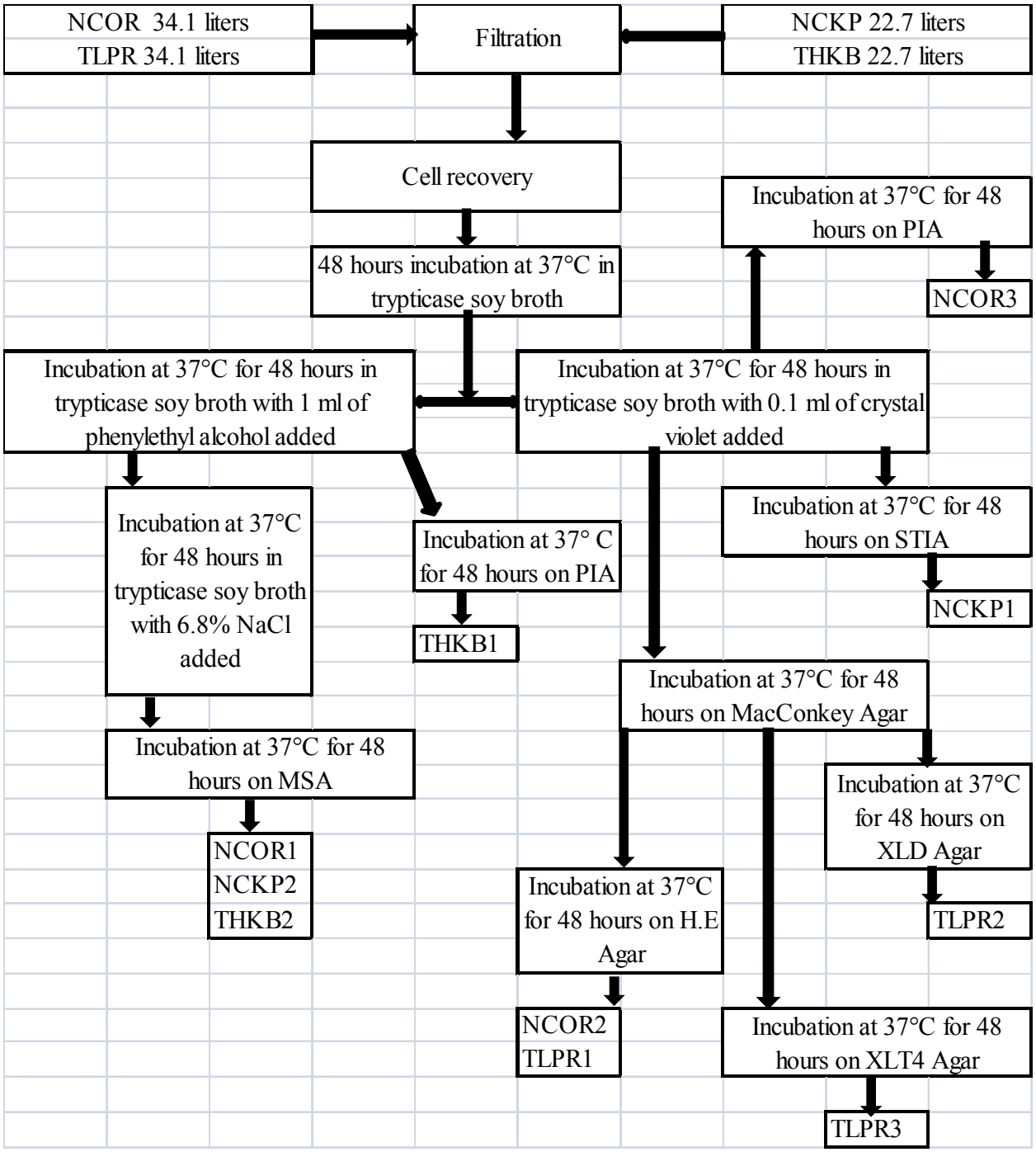


Figure 1: Schematic showing selective protocol sequence used to isolate potentially pathogenic bacterial species from water samples collected from four streams in Great Smoky Mountains National Park and the media type on which each isolate was isolated (key for isolates: NC =

North Carolina, T = Tennessee; KP = Kephart Prong, OR = Oconaluftee River, HKB = Hickory King Branch, LPR = Little Pigeon River; 1, 2, 3 = sequential isolates obtained from each site).

Isolate Designations

Isolates were initially labeled based on their location of origin and which media type they were isolated from, this made tracking each isolate easy but the names were too long. For simplification, each isolate was renamed based only on site of origin and culture number. For example, an isolate obtained from Oconaluftee River was labeled NCOR meaning North Carolina Oconaluftee River and a selection number, e.g. NCOR1.

DNA Sequencing and Culture Matching

Streak plating was employed to obtain pure cultures from different colonies with ten unique pure culture plates selected based on location of sample origin and growth on specific media types. These ten cultures (NCOR1, NCOR2, NCOR3, THKB1, THKB2, NCKP1, NCKP2, TLPR1, TLPR2 and TLPR3) were streaked onto separate trypticase soy agar plates and shipped to Genewiz Labs in South Plainville, New Jersey to have their 16S ribosomal DNA sequenced. Partial forward and reverse sequences were obtained for the 16S rRNA gene allowing for a nearly complete sequence to be assembled. Each sequence was assessed for quality using Finch TV version 1.4.0 (Geospiza, Inc.). Any base with a single distinct peak was defined as acceptable while those with multiple peaks of different colors were classified as unacceptable, if the peaks were of the same size. Where there was overlap between the two sequences, the corresponding DNA strand was assessed at the same position to determine the correct base for that position. The forward and reverse sequences were aligned using the National Center for Biotechnology Information (NCBI) Align Two Sequences Nucleotide BLAST tool (Altschul 1997). This alignment tool allowed for sequence proofreading of the overlapping areas of the

forward and reverse sequencing reads when paired with Finch TV's graphical analysis tool. Corrected forward and reverse sequencing data were then combined into one sequence with approximately 1400 nucleotide bases.

The DNA sequence for each isolate was compared against sequences in the Ribosomal Database Project (RDP) (Cole et al. 2009). DNA sequencing data were first run through the Classifier tool (Wang et al. 2007) on the RDP website to obtain the hierarchical classification of each isolate from phylum to genus. The Sequence Match tool (SeqMatch) was then employed to compare each isolate's DNA sequence to known species DNA sequences (Cole et al. 2014). To achieve this, the following parameters were set:

1. Strain option was limited to type strain only.
2. Only sequences from isolated cultures (not environmental DNA or RNA) were compared.
3. Only full length or close to full length (>1200 bases) sequences were selected to be compared from the database.
4. Sequence quality was set to good (excluded sequences of suspect or poor quality) and Taxonomy was set to nomenclatural (Cole et al. 2014).

This procedure produced a list of type bacterial species whose sequences were compared to each GSMNP isolate's 16 rDNA sequence so that any possible relatives to the isolate could be determined.

Phylogenetic Analysis

To perform phylogenetic analysis, Seqmatch was utilized but one parameter was changed (strain parameter was changed from only type to both type and non-type) because the results of the initial search were limited to a few species. DNA sequences from the result of this search that

had a percent match >95% were used to generate a phylogenetic tree with the aid of the MUSCLE multiple sequence alignment tool (Edgar 2004), Gblocks conserved block selective tool (Castresana 2000) and PhyML phylogeny tool (Guindon et al. 2010). Gblocks eliminated poorly aligned positions and divergent regions in the aligned 16S rDNA sequences (Castresana 2000). Only well aligned positions were used to determine phylogeny.Bootstrapping procedure was set to 100 bootstraps and the GTR substitution model was selected and rendered via TreeDyn tool (Dereeper et al. 2008, Timothy Driscoll personal communication). The phylogenetic tree was constructed to aid in the identification of each isolate and to determine whether it was related to a pathogenic species.

RESULTS

Classification of Isolates

The RDP Classifier tool was able to categorize the isolates into specific genera with 100% confidence except for isolates TLPR2 and TLPR3 which had 83% and 81% percent confidence, respectively (Table 3). Isolates THKB1, NCKP2 and THKB2 were found to belong to the genus *Bacillus*, isolates NCOR2 and TLPR1 belonged to the genus *Proteus,* TLPR2 and TLPR3 belonged to the genus *Enterobacter,* NCOR1, NCKP1 and NCOR3 were members of the genera *Staphylococcus, Serratia* and *Providencia,* respectively.

Table 3: RDP Classifier results showing the genus each isolate obtained from GSMNP belonged to, and the accuracy (confidence) of the result in percentage.

Isolate	Genus	% Confidence
NCKP1	*Serratia*	100
NCKP2	*Bacillus*	100
NCOR1	*Staphylococcus*	100
NCOR2	*Proteus*	100
NCOR3	*Providencia*	100
THKB1	*Bacillus*	100
THKB2	*Bacillus*	100
TLPR1	*Proteus*	100
TLPR2	*Enterobacter*	83
TLPR3	*Enterobacter*	81

SeqMatch Results & Phylogenetic Trees

NCKP1

The RDP SeqMatch tool revealed that the isolate NCKP1's 16S ribosomal DNA (rDNA) sequence had a 98.4% identity to a type strain of *Serratia marcescens* (KRED; accession number

AB061685) and 99.6% identity to a non-type strain *Serratia* sp. NT3. Phylogenetic analysis of NCKP1 (Figure 2a & b) showed *Serratia* sp. NT3 shares a common ancestor with the ancestor of isolate NCKP1.

Table 4: List of type strain bacterial species with ≥ 90% 16S rDNA sequence match to isolate NCKP1 and their percent identity to NCKP1.

Isolate	Best SeqMatch Type Strain Results	Accession Number	% Identity to Isolate
NCKP1	*Serratia marcescens; KRED*	AB061685	98.4
	Serratia nematodiphila	EU036987	98.3
	Serratia marcescens	AJ233431	96.4
	Serratia ureilytica	AJ854062	92.8

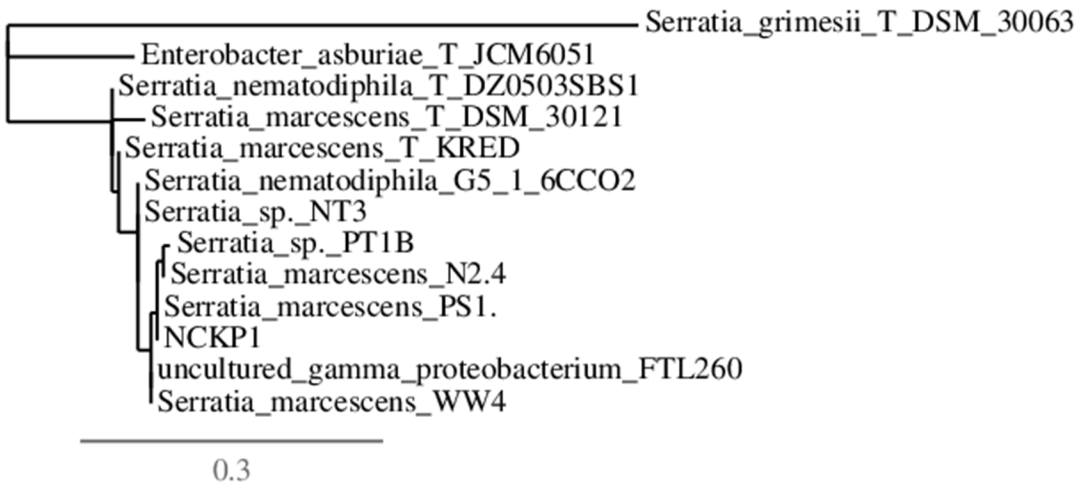


Figure 2a: Phylogenetic tree constructed using the top 12 sequence matches from the Ribosomal Database Project (type & non-type strains) analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tools and rendered via TreeDyn tool for isolate NCKP1 from Great Smoky Mountains National Park.

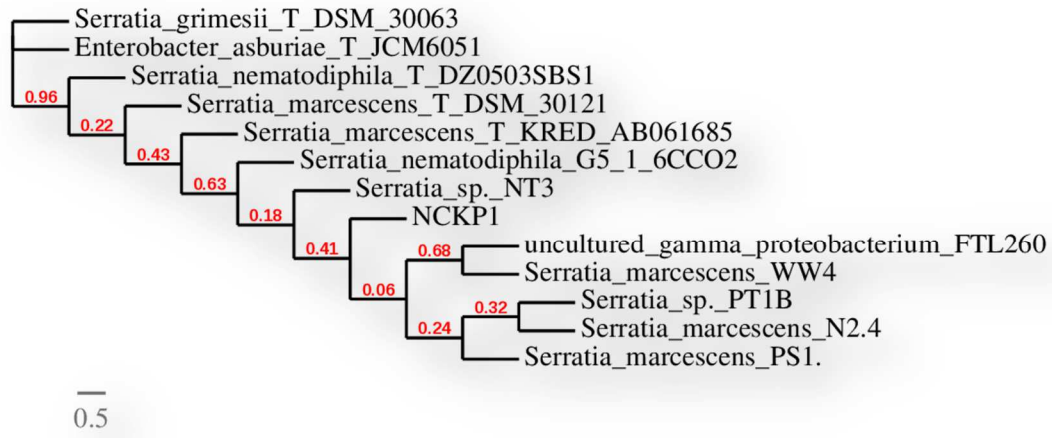


Figure 2b: Phylogenetic tree constructed using the top 12 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool with branch distance ignored for isolate NCKP1 from Great Smoky Mountains National Park Branch support values are out of 1.0.

TLPR1 & NCOR2

The 16S rDNA sequence for isolate TLPR1 had a 98.1% identity to type strain *Proteus mirabilis* (accession number DQ885256) while isolate NCOR2 had a 97.6 % identity to the same type strain when its 16S rDNA sequence was run through SeqMatch. Phylogenetic analysis revealed TLPR1 was related to the ancestor of non-type strain *Proteus penneri* (YCY34) while NCOR2 was shared a common ancestor with non-type strains *Proteus mirabilis* (SPC04) and *Proteus mirabilis* (FUA 1240 5b1).

Table 5: List of type strain bacterial species with ≥ 90% 16S rDNA sequence match to isolates TLPR1 and NCOR2 and their percent identity to TLPR1 and NCOR2.

Isolate	Best SeqMatch Type Strain Results	Accession Number	% Identity to Isolate
NCOR2	*Proteus mirabilis*	DQ885256	97.6
	Proteus vulgaris	DQ885257	93.3
	Proteus penneri	DQ885258	91.8
	Proteus hauseri	DSM14437	91.1
TLPR1	*Proteus mirabilis*	DQ885256	98.1
	Proteus vulgaris	DQ885257	93.8
	Proteus penneri	DQ885258	92.1
	Proteus hauseri	DSM14437	91.6

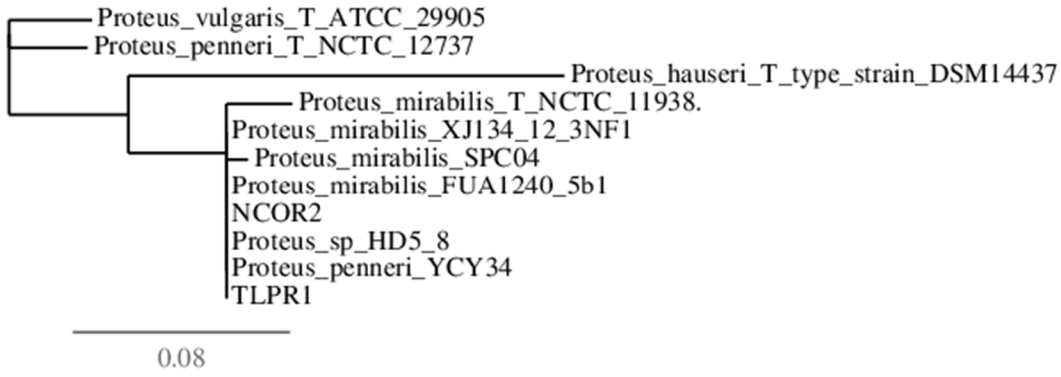


Figure 3a: Phylogenetic tree constructed using the top 9 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool for isolates NCOR2 and TLPR1 from Great Smoky Mountains National Park.

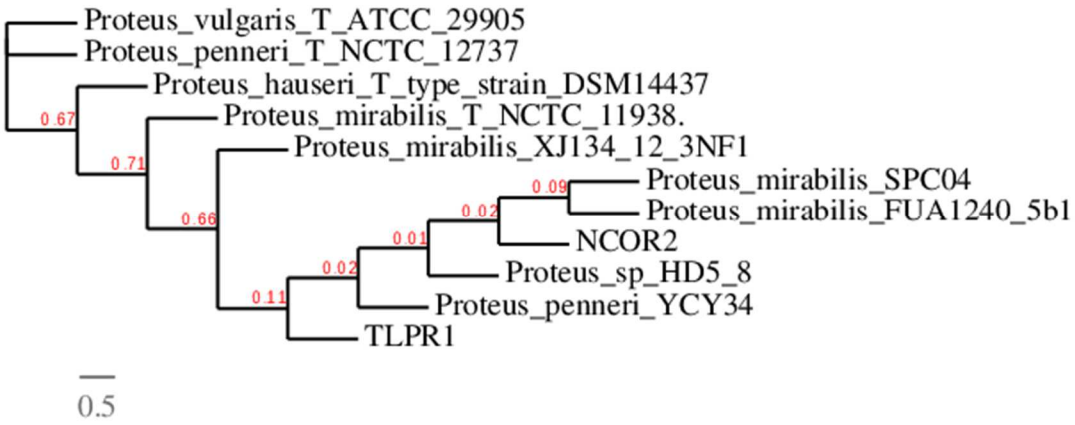


Figure 3b: Phylogenetic tree constructed using the top 9 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool with branch distance ignored for isolates NCOR2 & TLPR1 from Great Smoky Mountains National Park. Branch support values are out of 1.0.

NCOR1

The RDP SeqMatch tool revealed that the 16S rDNA sequence for isolate NCOR1 had a 99% identity to type strain *Staphylococcus sciuri* (accession number AJ421446) and a 99.8% identity to non-type strain *Staphylococcus* sp. K4-STE/2013. Phylogenetic analysis showed NCOR1 shares a common ancestor with the ancestor of type strain *Staphylococcus sciuri* (accession number AJ421446) and non-type strain *Staphylococcus sciuri* subsp. sciuri (P3-3-b-1).

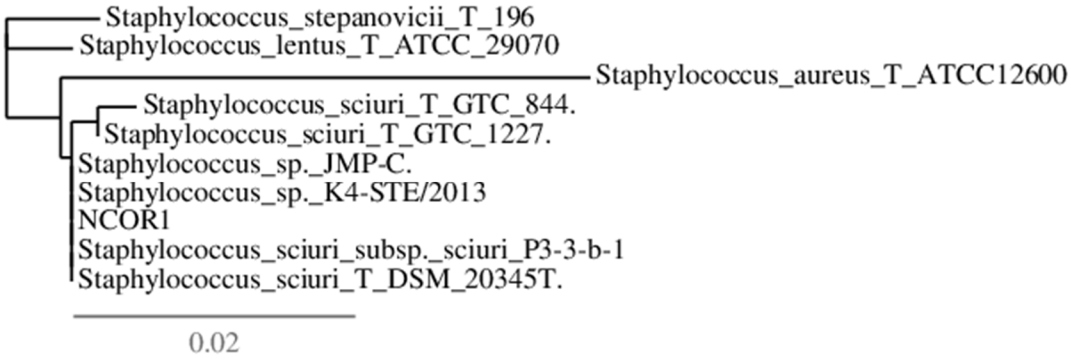


Figure 4a: Phylogenetic tree constructed using the top 9 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool for isolate NCOR1 from Great Smoky Mountains National Park.

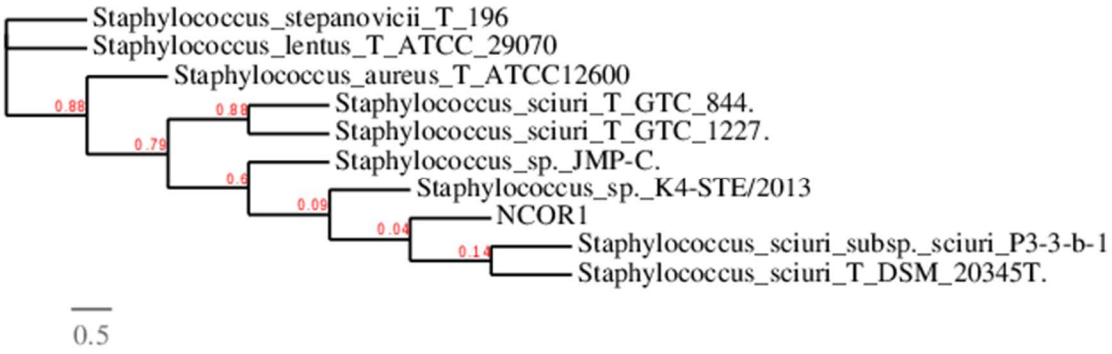


Figure 4b: Phylogenetic tree constructed using the top 9 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool with branch distance ignored for isolate NCOR1 from Great Smoky Mountains National Park. Branch support values are out of 1.0.

Table 6: List of type strain bacterial species with ≥ 90% 16S rDNA sequence match to isolate NCOR1 and their percent identity to NCOR1

Isolate	Best SeqMatch Type Strain Results	Accession Number	% Identity to Isolate
NCOR1	*Staphylococcus sciuri*	AJ421446	99
	Staphylococcus sciuri	AB233331	98.2
	Staphylococcus sciuri	AB233332	96.9
	Staphylococcus lentus	D83370	94.6
	Staphylococcus vitulinus	AB009946	93.4

NCOR3

The 16S rDNA sequence for isolate NCOR3 had a 98.3% identity to type strain *Providencia rettgeri* (accession number AM040492). Phylogenetic analysis revealed NCOR3 shares a common ancestor with non-type strain *Providencia rettgeri* VITJCSTT3.

Table 7: List of type strain bacterial species with ≥ 90% 16S rDNA sequence match to isolate NCOR3 and their percent identity to NCOR3

Isolate	Best SeqMatch Type Strain Results	Accession Number	% Identity to Isolate
NCOR3	*Providencia rettgeri*	AM040492	98.3
	Providencia vermicola	AM040495	96.8
	Providencia burhodogranariea	HM038004	94
	Providencia rustigianii	AM040489	93.1
	Providencia sneebia	HM038003	92.5

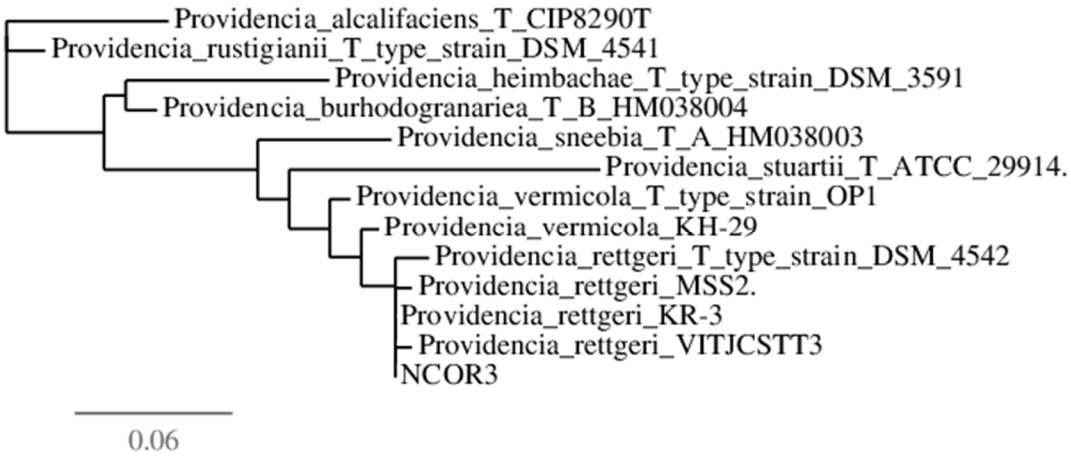


Figure 5a: Phylogenetic tree constructed using the top 8 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool for isolate NCOR3 from Great Smoky Mountains National Park.

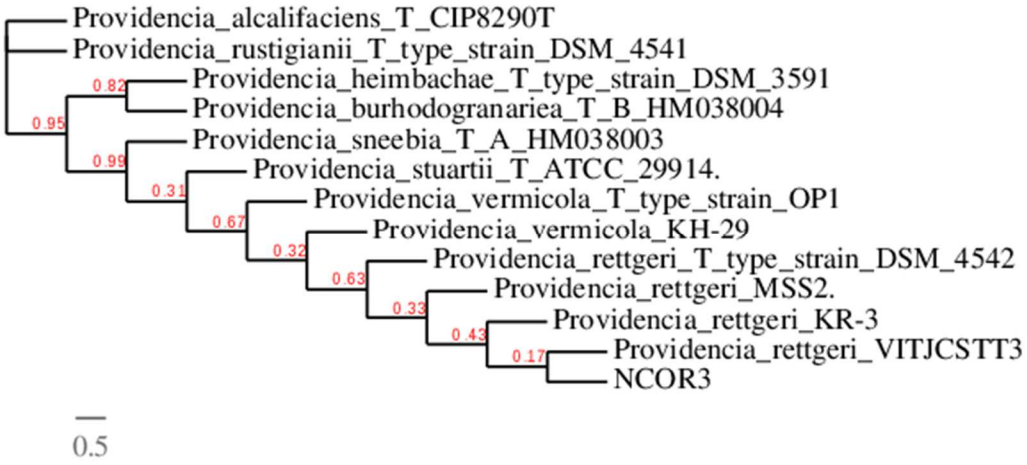


Figure 5b: Phylogenetic tree constructed using the top 8 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool for isolate NCOR3 from Great Smoky Mountains National Park. Branch support values are out of 1.0.

THKB2 & NCKP2

Both isolates THKB2 and NCKP2 16S rDNA sequences had a high percent identity to type strain *Bacillus subtilis* (accession number AJ276351). THKB2 had 99.4% identity to this bacteria strain while NCKP2 had a 99.5% identity with the strain. Phylogenetic analysis of both 16S rDNA sequences revealed THKB2 and NCKP2 share a common ancestor. This ancestor shares a common ancestor with *Bacillus subtilis* (accession number AJ276351).

Table 8: List of type strain bacterial species with ≥ 90% 16S rDNA sequence match to isolates THKB2 and NCKP2 and their percent identity to THKB2 and NCKP2.

Isolate	Best SeqMatch Type Strain Results	Accession Number	% identity to Isolate
NCKP2	*Bacillus subtilis*	AJ276351	99.5
	Bacillus mojavensis	AB021191	98
	Bacillus vallismortis	AB021198	97.4
	Bacillus subtilis	AF074970	97.3
	Bacillus methylotrophicus	EU194897	97.1
THKB2	*Bacillus subtilis*	AJ276351	99.4
	Bacillus mojavensis	AB021191	98.1
	Bacillus subtilis	AF074970	97.7
	Bacillus vallismortis	AB021198	97.2

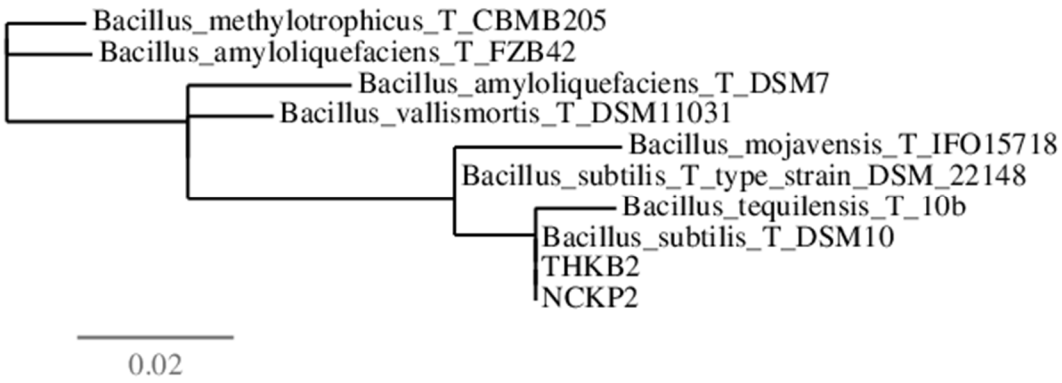


Figure 6a: Phylogenetic tree constructed using top the 8 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool for isolates THKB2 & NCKP2 from Great Smoky Mountains National Park.

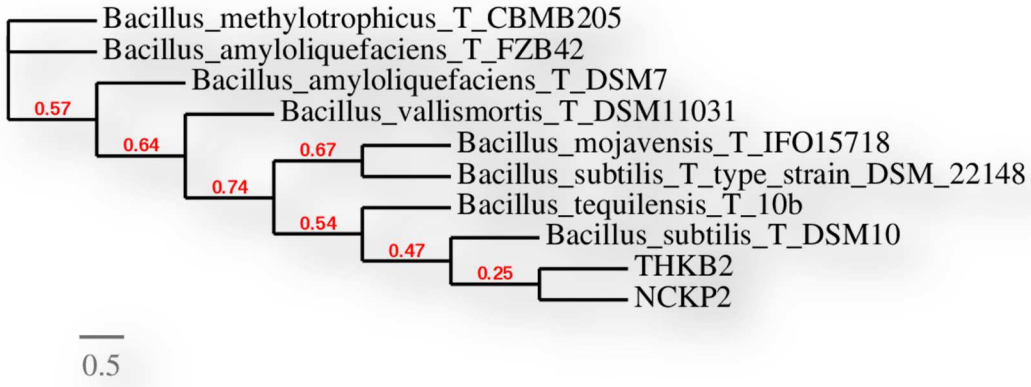


Figure 6b: Phylogenetic tree constructed using the top 8 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool with branch distance ignored for isolates THKB2 & NCKP2 from Great Smoky Mountains National Park. Branch support values are out of 1.0.

THKB1

RDP SeqMatch search results revealed isolate THKB1 16S rDNA sequence had a 99.3% identity to type strain *Bacillus licheniformis* (accession number CP000002). THKB1 and non-type strain *Bacillus* sp. HY13 2010 share a common ancestor according to phylogenetic analysis.

Table 9: List of type strain bacterial species with ≥ 90% 16S rDNA sequence match to isolate THKB1 and their percent identity to THKB1.

Isolate	Best SeqMatch Type Strain Results	Accession Number	% identity to Isolate
THKB1	*Bacillus licheniformis*	CP000002	99.3
	Bacillus aerius	AJ831843	97.2
	Bacillus sonorensis	AF302118	95.4

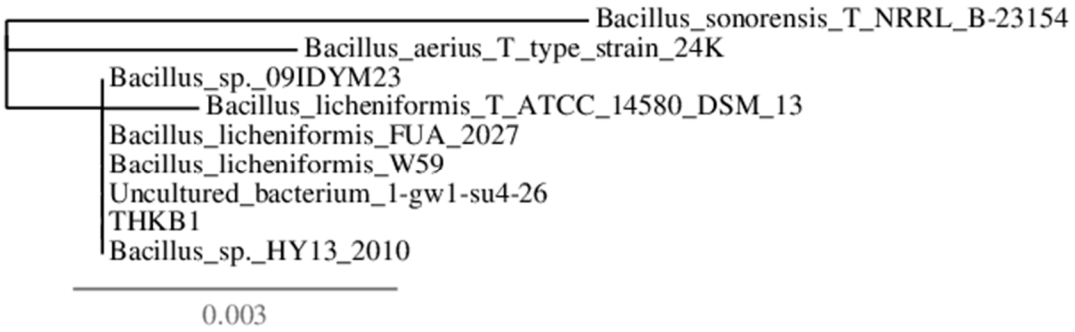


Figure 7a: Phylogenetic tree constructed using the top 8 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool for isolate THKB1 from Great Smoky Mountains National Park.

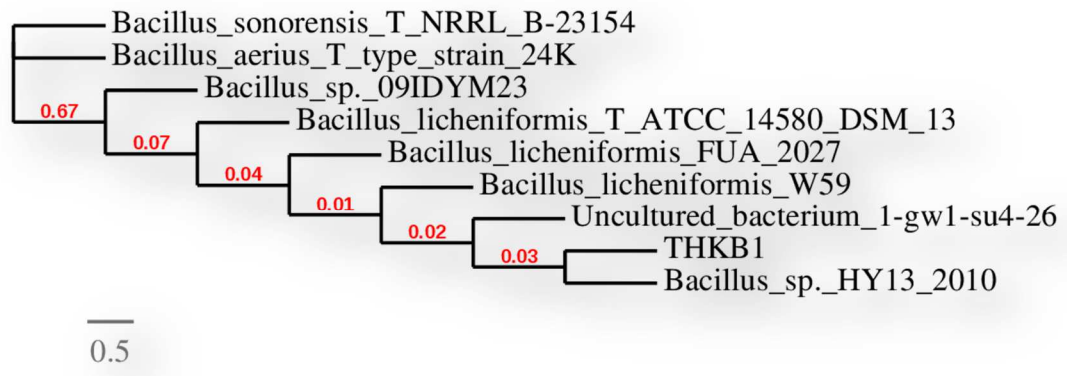


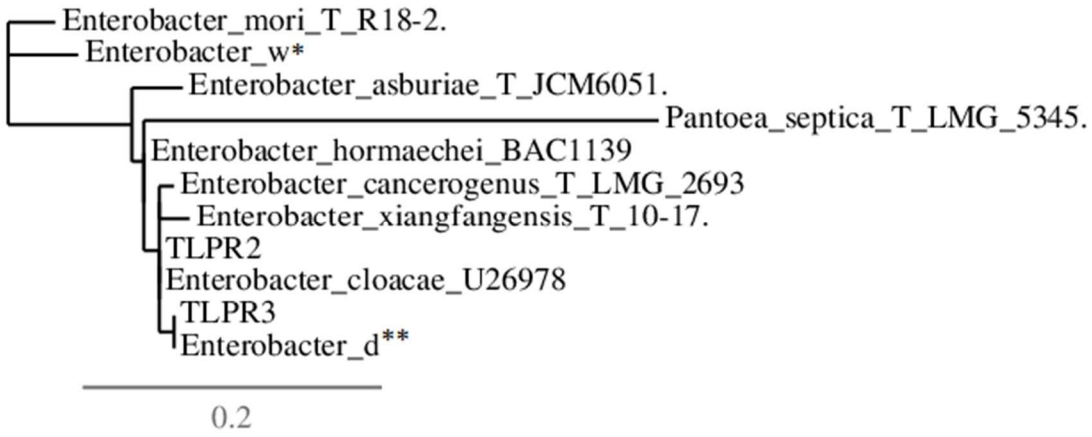
Figure 7b: Phylogenetic tree constructed using the top 8 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool with branch distance ignored for isolate THKB1 from Great Smoky Mountains National Park. Branch support values are out of 1.0.

TLPR2 &TLPR3

Isolates TLPR2 and TLPR3 had 16S rDNA sequences with 95.9 % and 95.6% identity, respectively, to *Enterobacter hormaechei* (accession number AJ508302). Phylogenetic analysis of both isolates showed TLPR2 shares a common ancestor with non-type strain *Enterobacter cloacae* (U26978) while TLPR3 shares a common ancestor with *Enterobacter ludwigii* (EN-119).

Table 10: List of type strain bacterial species with ≥ 90% 16S rDNA sequence match to isolates TLPR2 and TLPR3 and their percent identity to TLPR2 and TLPR3.

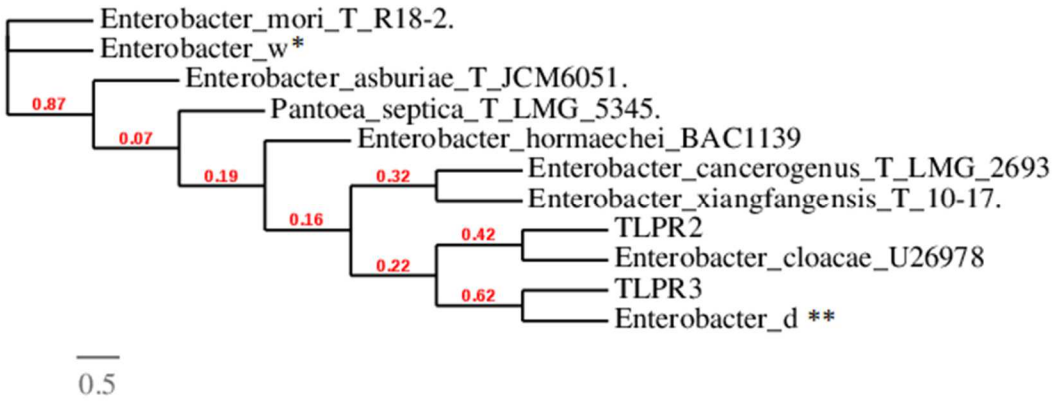
Isolate	Best SeqMatch Type Strain Results	Accession Number	% identity to Isolate
TLPR1	*Enterobacter hormaechei*	AJ508302	95.9
	Enterobacter ludwigii	AJ853891	94.5
	Enterobacter cancerogenus	LMG 2693	93.9
	Enterobacter asburiae	AB004744	93.8
TLPR3	*Enterobacter hormaechei*	AJ508302	95.6
	Enterobacter ludwigii	AJ853891	93.1
	Enterobacter cancerogenus	LMG 2693	92.6
	Enterobacter asburiae	AB004744	92.4
	Pantoea septica	EU216734	92.1



* *Enterobacter ludwigii* KS81

** *Enterobacter ludwigii* (T); type strain; EN-119

Figure 8a: Phylogenetic tree constructed using the top 8 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool for isolates TLPR2 & TLPR3 from Great Smoky Mountains National Park.



* *Enterobacter ludwigii* KS81

** *Enterobacter ludwigii* (T); type strain; EN-119

Figure 8b: Phylogenetic tree constructed using the top 8 sequence matches from the Ribosomal Database Project (type & non-type strains), and analyzed with MUSCLE multiple sequence alignment tool, Gblocks selective tool, PhyML phylogeny tool and rendered via TreeDyn tool with branch distance ignored for isolates TLPR2 & TLPR3 from Great Smoky Mountains National Park. Branch support values are out of 1.0.

DISCUSSION

Isolate Selection and Classification

The results suggest that initial enrichment and screening of bacteria via phenylethyl alcohol and crystal violet was successful. Isolates that were obtained through the Gram negative pipeline were related at the genus and species level to known Gram negative bacteria. Isolates obtained using the Gram positive pipeline were related to Gram positive bacterial species with the exception of THKB1, which was related to Gram positive bacteria but grew on *Pseudomonas* isolation agar. This was unexpected but recent literature on triclosan (selective agent in PIA media) revealed that many bacterial species including *Staphylococcus aureus* (a Gram positive species) and *Salmonella enterica* have developed resistance to triclosan because it is used as an antiseptic agent in hand soap and even in toothpaste (Carey & McNamara 2014). NCOR1, THKB2 and NCKP2 were related to Gram positive bacterial species while NCOR2, NCOR3, NCKP1, TLPR1, TLPR2 and TLPR3 were related to Gram negative bacterial species. Even though a limited number of isolates were selected and sequenced; sequencing results suggest a relatively high amount of species diversity considering the selective media that were employed.

According to the RDP Classifier tool results, all the Gram negative isolates belonged to the phylum Proteobacteria and family Enterobacteriaceae while the Gram positive isolates were in the phylum Firmicutes but are members of two different families, Staphylococcaceae and Bacillaceae. THKB1, THKB2 and NCKP2 are members of family Bacillaceae while NCOR1 belonged to family Staphylococcaceae. This result was expected because all the selective media used in this research to select for the isolates are designed to culture known pathogenic bacteria belonging to these families.

NCKP1

Isolate NCKP1 was initially identified as *Salmonella typhimurium* because it was the only bacterium that grew on the STIA plates. It was a surprise when analysis of the NCKP1 16S rDNA sequence using both a phlyogenetic tree and RDP tools suggested NCKP1 is a member of the genus *Serratia* (Table 3 & Figure 2a & b).

Serratia spp. are rod shaped Gram negative bacteria and some members of this genus produce pigments named prodigiosins (Hearn et al. 1970; Gerber 1975). Certain species of *Serratia* such as *S. marcescens* have clinical importance and this species is known to cause nosocomial infections (Grimont & Grimont 1992; Brenner 1984). *Serratia marcescens* are opportunistic pathogens that are able to form biofilms and cause respiratory and urinary tract infections (Yu 1979). SeqMatch analysis of NCKP1 matched the isolate with *S. marcescens* KRED; a type strain subspecies that was isolated from a wastewater treatment tank in Saku, Japan (Ajithkumar et al. 2003).

Even though isolate NCKP1 had a 98.4% identity to *S. marcescens* KRED, it did not exhibit the same red-colored pigment (prodigiosin) described in 2003 by Ajithkumar et al. about *S. marcescens* KRED cultured on trypticase soy agar (TSA) plates at room temperature. The reason isolate NCKP1 was cultured at 25°C was because *S. marcescens* strains are known to lose their pigmentation at higher temperatures (Rjazantseva et al. 1994). Phylogenetic analysis showed isolate NCKP1 to be a branched descendent of *S. marcescens* KRED and non-type strain *Serratia* sp. NT3. The isolate is also related to the ancestor of three *S. marcescens* subspecies (Figure 2b). Molecular genetic analysis of isolate NCKP1 revealed the isolated was a member of the genus *Serratia*, and was related to *S. marcescens*. Metabolic testing and environmental

parameter susceptibility testing is required to confirm how closely NCKP1 is related to *S. marcescens*.

TLPR1 & NCOR2

Both isolates TLPR1 and NCOR2 were isolated on Hektoen enteric agar plates, and 16S rDNA sequences matched these two isolates to the same genus and type strain species. Yet maximum likelihood phylogenetic analysis placed both isolates at different positions on the tree and TLPR1 was revealed to be the ancestor of NCOR2 (Figure 3a & b). Isolate TLPR1 and NCOR2 belong to the genus *Proteus*; and are related to *P. penneri* and *P. mirabilis,* respectively, according to phylogenetic analysis.

RDP SeqMatch on the other hand matched both isolates to type strain *Proteus mirabilis* (NCTC 11938). *Proteus mirabilis* and *P. penneri* are Gram negative facultative anaerobic bacilli that are known to cause nosocomial urinary infections in immunocompromised patients (O'Hara et al. 2000). These *Proteus* species are closely related and are sometimes misidentified (Kishore 2012). *Proteus mirabilis* strains are much more prevalent in soils and freshwater bodies than *P. penneri* and cause more infections (Cantón et al. 2006).

NCOR1

The genus *Staphylococcus* has more than 40 species and subspecies, with a majority of the species being opportunistic pathogens (Marsou et al. 1999). Staphylococci are widespread in nature and are known to cause infections in both humans and animals (Kloos & Schleifer 1975; Kloos & Bannerman 1995). Perhaps the most notable species of this genus is *S. aureus* which can cause methicillin-resistant infections commonly known as MRSA (Kloos & Bannerman 1995).

Staphylococcus sciuri was initially classified as an animal pathogen that caused infections in livestock and rodents as seen in the subspecies *S. sciuri* subsp. *rodentium* (Kloos & Schleifer 1975). In recent years, *S. sciuri* has crossed over to humans and is now observed to cause infections in humans (Kalawole & Shittu 1997). Based on the research findings in 1999 by Marsou et al., *S. sciuri* is a natural reservoir of the *mecA* gene. This gene is responsible for the methicillin resistance observed in *S. aureus* (Wu et al. 1996; Vandenesch et al. 2003). Results suggest isolate NCOR1 is a member of genus *Staphylococcus* and has a 99% identity to *S. sciuri*. This means NCOR1 could potentially be pathogenic to humans.

NCOR3

This isolate was obtained via the Gram negative bacteria selection pipeline and was isolated on *Pseudomonas* isolation agar (PIA). Unlike isolate NCKP1, it was not a surprise when 16S rDNA sequencing analysis revealed the isolate belonged to a different genus – *Providencia*, instead of *Pseudomonas* – because the isolate did not exhibit the distinct green to yellow-green or sometimes pink coloration of colonies that is observed in true *Pseudomonas* spp. when cultured on PIA plates (S. O'Connell, personal communication).

Providencia rettgeri is a ubiquitous opportunistic pathogen that causes disease in a variety of animals including humans (Ladds et al. 1996 and Traub et al. 1971). In humans however, *P. rettgeri* has mainly been associated with causing nosocomial urinary tract infections (UTIs), but can also cause diarrhea and ocular infections such as keratitis, although this is rare (Koreishi et al. 2006). Most *P. rettgeri* strains are susceptible to antibiotic and antimicrobial agents, but there have been cases of resistance observed in the last decade (Chander et al. 2006). On the phylogenetic tree constructed, NCOR3 shares a direct common ancestor with a non-type

strain of *P. rettgeri* (Figure 5a & b). NCOR3 could be a strain or subspecies of *P. rettgeri* given that it has a 98.3% identity to the species and is closely related to one of its strains.

THKB2, NCKP2 & THKB1

Three isolates were found to be related to the genus *Bacillus* in this study. *Bacilli* are ubiquitous and can be found abundantly in soil and water bodies. THKB2 and NCKP2 were isolated on MSA plates while isolate THKB1 was cultured on a PIA plate. THKB2 and NCKP2 had rapid observable growth rates compared to the other isolates and formed lawns on the MSA plates. Neither isolate fermented mannitol as the plates remained red and no color change was observed. On PIA plates isolate THKB1 presented a creamy yellow pigmentation. 16S rDNA analysis matched THKB2 and NCKP2 to *Bacillus subtilis* (>99% similarity) while THKB1 showed the highest similarity to *Bacillus licheniformis* (>99%). On phylogenetic trees (Figure 6a & 6b), THKB2 and NCKP2 share a common ancestor; which implies these two isolates are closely related. Dissimilar to THKB2 and NCKP2, THKB1 was further away from its highest (*Bacillus licheniformis*) match on the phylogenetic tree. It shared a common ancestor with *Bacillus* sp. HY13 2010 (>99% similarity).

B. subtilis are benign bacteria that form part of the natural microflora of the gastrointestinal tract in humans. In Italy, *B. subtilis* is commonly utilized as a probiotic to treat and/or prevent intestinal disorders, and can be purchased at local pharmacies (Mazza et al. 1992 and Oggioni et al. 1998). It is not recognized as a pathogen, but there was a rare case presented in 1998 by Oggioni et al. about a 73 year old man with chronic lymphocytic leukemia who developed a recurring case of *B. subtilis* induced septicemia. *B. licheniformis* on the other hand is an opportunistic pathogen that has been associated with food poisoning in humans and bovine toxemia and abortions (Salkinoja-Salonen et al. 1999 and Logan 2012).

TLPR2 &TLPR3

As previously mentioned in the results section, the RDP Classifier tool was unable to determine the genus that isolates TLPR2 and TLPR3 belonged to with 100% confidence. The isolates were classified as *Enterobacter* at 83% and 81% confidence, respectively. Many species that belong to the family Enterobacteriaceae are difficult identify via 16S rDNA alone because of the massive diversity within the family and diverse phenotypic characteristics (S. O'Connell, personal communication). Similar to *Staphylococcus*, most members of the genus *Enterobacter* are thought to be potentially pathogenic and cause opportunistic community and nosocomial infections (Sanders and Sanders 1997).

Although both isolates showed high similarity to *Enterobacter hormaechei* (>95%), phylogenetic analysis of isolates TLPR2 and TLPR3 revealed these two isolates were related to different species. TLPR2 shared a common ancestor with a non-type strain of *Enterobacter cloacae* while TLPR3 shared a common ancestor with *Enterobacter ludwigii. Enterobacter hormaechei* and *E. cloacae* are members of the *Enterobacter cloacae* complex (a group of six genetically and phenotypically similar *Enterobacter* species); and in 2005, Hoffmann et al. proposed a novel species, *E. ludwigii,* to join the complex (Hoffmann and Roggenkamp 2003). All three species of *Enterobacter* are opportunistic pathogens that cause nosocomial infections in immunocompromised patients (Mayhall et al. 1979 and Hoffmann et al. 2005).

CONCLUSION

A positive confirmation that an isolate was a potential pathogen was defined by the following: An isolate that had a 100% SeqMatch identity with a well-characterized pathogen, based on 16S rDNA sequence data, and shared an immediate ancestor with that pathogen on a phylogenetic tree would be considered as a potential pathogen as it was highly related to that species. The results revealed that none of the isolates sequenced had a 100% match to an obligate or opportunistic pathogen but a few isolates including NCOR3, TLPR2 and TLPR3 shared an immediate common ancestor with the opportunistic pathogens *P. rettgeri, E. cloacae and E. ludwigii,* respectively based on phylogeny. Most of the isolates with the exception of THKB2 and NCKP2 were most closely related to opportunistic pathogens.

Although none of the isolates had a 100% match to known pathogenic species, many of them matched (>97%) to opportunistic pathogens, which generally supports the hypothesis that animal pathogens might be found in this study. By general consensus, isolates with high sequencing match (≥99%) to a well documented bacterial species are considered to be closely related or sometimes identified as that bacterial species while a 97 to 99% identity in 16S rDNA is a criterion used to identify an organism at genus level and sometimes species level (Drancourt et al. 2004). Since many of the isolates fell between 97 to 99% identity, further tests are needed to confirm the true identity and pathogenicity of each species

Some tests that could be used to determine bacterial pathogenicity and infectivity include virulence gene cloning (Wu et al. 1996), running a susceptibility test to both natural and synthetic antibiotics and antimicrobials (as described in Edberg et al. 1996), or whole genome sequencing and probing for virulence genes. Infectivity could be tested simply by exposing

healthy mice to the isolate either via ingestion or injection of the isolate into the blood stream of the mice. Virulence factors such as exotoxins could also be tested via immunoassays such as enzyme-linked immunosorbent assay (ELISA) to for specific exotoxins (Kato et al. 1998, Zhu et al. 2014).

This study supports the idea that pathogenic bacteria exist in Great Smoky Mountains National Park. There were many isolates obtained during the selection process but due to time and resource constraints only 10 isolates were selected, thus there are pathogens that could have been missed in this work. Without these constraints, more isolates could have been cultured from the initial screening media (TSB + PEA and TSB + crystal violet). Isolates obtained would have undergone both 16S sequencing and phylogenetic analysis. And once the genus for each isolate was identified, the isolates would undergo a series of metabolic and environmental susceptibility test to confirm their identity. Pathogenicity would be tested using one or more of the methods mentioned above. Conversely a fully molecular approach could be used to identify bacterial pathogens that could be in the streams. This approach would rely on polymerase chain reaction (PCR) and microarray based assay as described in Järvinen et al. 2009.

This study used methods associated with culturing human pathogens. For this reason, many animal pathogens might have been selected against. This would be especially so for non-mammalian and avian species. Different selection procedures and variables could be employed such as the use of different selective media types and incubation temperature could be changed to find pathogens that were selected against using the current procedures.

It is possible that the four sites sampled for pathogens may not be reservoirs for pathogens or are not conducive for pathogen proliferation. Other sites such as camping grounds and stagnant water bodies in the park could be better sites for isolating human pathogens. This is

because the continuous flow of water could prevent the pathogens from settling in and proliferating while stagnated water-bodies are nutrient rich and conducive for growth; camping grounds are also more likely to harbor pathogens since there is a higher probability that people travel to those sites may unknowingly be carrying pathogens on their skin, clothes and could also introduce microbes from their food sources and untreated wastes.

REFERENCES

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J.. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25:3389-3402.

Altschul, S.F. (1997). Evaluating the statistical significance of multiple distinct local alignments. In Theoretical and Computational Methods in Genome Research. (S. Suhai, ed.), pp. 1-14, Plenum, New York.

Atlas, R. M. (2010).Handbook of Microbiological Media, Fourth Edition. CRC Press

Brenner, D.J. R.N. Krieg, T.J. Staley, and M.G. Garrity. (2005). Bergey's Manual of Systematic Bacteriology. New York: Springer Vol. 1-3.

Authors? (2012). Standard methods for the examination of water and wastewater: Including bottom sediments and sludges; 22nd ed. New York: American Public Health Association.

Brewster, J.D., and Paoli, G.C. (2013). DNA extraction protocol for rapid PCR detection of pathogenic bacteria. Analytical Biochemistry Vol. 442 (1):107-109.

Brisson, N., Delrieu, A.P., Samain, D., and Courvalin, P. (1988). Inactivation of lincosamide antibiotics in *Staphylococcus*. Identification of lincosamide o-nucleotidyltransferases and comparison of the corresponding resistance genes. The Journal of Biological Chemistry 263:15880–15887.

Cantón, R, Sánchez-Moreno, M.P., and Morosini Reilly, M.I. (2006). *Proteus penneri*. Enfermedades Infecciosas y Microbiología Clínica Vol. 24 (Supplement 1):8-13.

Carey, D.E., and McNamara, P.J. (2014). The impact of triclosan on the spread of antibiotic resistance in the environment. Frontiers in Microbiology. 5(780):1-11.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17(4):540-52.

Chander, Y., Goyal, S.M., and Gupta, S.C. (2006). Antimicrobial resistance of *Providencia* spp. isolated from animal manure. Veterinary Journal 172: 188-191.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M, Marsh, T., Garrity, G.M., and. Tiedje, J.M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research 37:141-145.

Dereeper, A., Guignon, V. Blanc, G, Audic, Buffet, S.,. Chevenet, F., Dufayard, J.-F, Guindon, S., Lefort, V, Lescot, M., Claverie, J.-M., and Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research 36:W465-469 (Web Server Issue).

Devriese, L.A., Poutrel, B., Killper-Bälz, R., and Schleifer, K.H. (1983). *Staphylococcus gallinarum* and *Staphylococcus caprae*, two new species from animals. International Journal of Systematic and Evolutionary Microbiology 33: 480–486.

Dey, R., Bodennec, J., Mameri. O.M. and Pernin, P. (2009). Free-living freshwater amoebae differ in their susceptibility to the pathogenic bacterium *Legionella pneumophila*. FEMS Microbiology Letters Vol.290(1): 10-17.

Drancourt, M., Berger, P., and Raoult, D. (2004). Systematic 16S rRNA gene sequencing of atypical clinical isolates identified 27 new bacterial species associated with humans. Journal of Clinical Microbiology 42:2197–2202.

Edberg, S.C., Gallo, P. and Kontnick, C. (1996). Analysis of the virulence characteristics of bacteria isolated from bottled, water cooler, and tap water. Microbial Ecology in Health and Disease. Vol. 9: 67-77.

Edgar ,R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32(5):1792-7.

Emerson, D., Agulto, L., Liu, H., and Liu, L. 2008. Identifying and characterizing bacteria in an era of genomics and proteomics. BioScience 58:925-936.

Findlay, S. (2010). Stream microbial ecology. Journal of the North American Benthological Society 29:170-181.

Geldreich, E. (1996). Pathogenic agents in freshwater resources. Hydrological Processes 10:315-333.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59(3):307-21.

Hoffmann, H., and Roggenkamp, A. (2003). Population genetics of the nomenspecies *Enterobacter cloacae*. Applied and Environmental Microbiology 69:5306-5318.

Hoffmann, H., Stindl, S., Stumpf, A., Mehlen, A., Monget, D. Heesemann, J., Schleifer, K.H., and Roggenkamp, A. (2005). Description of *Enterobacter ludwigii* sp. nov., a novel Enterobacter species of clinical relevance. Systematic and Applied Microbiology 28(3): 206–212.

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

Hullar, M.A.J., Kaplan, L.A., and Stahl, D.A. (2006). Recurring seasonal dynamics of microbial communities in stream habitats. Applied and Environmental Microbiology 72:713-722.

Järvinen, A.-K., Laakso, S., Piiparinen, P., Aittakorpi, A., Lindfors, M., Huopaniemi, L.,Piiparinen,H., Mäki, M. (2009). Rapid identification of bacterial pathogens using a PCR- and microarray-based assay.BioMed Central Microbiology 9:161-176.

Jenkins, M. (2007). Vegetation communities of Great Smoky Mountains National Park. Southeastern Naturalist 6:35-56.

Cabral, J.P.S. (2010). Water Microbiology. Bacterial pathogens and water. International Journal of Environmental Research and Public Health 7(10): 3657–3703.

Kalawole D.O. and Shittu A.O. (1997). Unusual recovery of animal staphylococci from septic wounds of hospital patients in lle-lfe, Nigeria, Letters in Applied Microbiology 24: 87–90.

Kloos ,W.E. and Bannerman, T.L. (1995). Staphylococci and Micrococci, in: Murray, P.R., Baron, E.J.O., Pfaller, M.A., Tenover, F.C., and Yolken, R.H. (Eds.), Manual of Clinical Microbiology, American Society for Microbiology Press, Washington DC, pp. 282–298.

Kloos, W.E. and Schleifer, K.H. (1975). Simplified scheme for routine identification of human *Staphylococcus* species, Journal of Clinical Microbiology 1: 82–88.

Kloos, W.E. and Schleifer, K.H. (1976). Characterization of *Staphylococcus sciuri* sp. nov. and its subspecies. International Journal of Systematic and Evolutionary Microbiology 26: 22–37.

Koreishi, AF, Schechter, BA, and Karp, CL. (2006). Ocular infections caused by *Providencia rettgeri*. Ophthalmology 113:1463–6.

Ladds, P.W., Bradley, J., and Hirst, R.G. (1996). *Providencia rettgeri* meningitis in hatchling saltwater crocodiles (*Crocodylus porosus*). Australian Veterinary Journal 74: 397-398.

Lamanna, C. and Mallette, F. (1953). Basic Bacteriology: Its Biological and Chemical Background. Williams and Wilkins Company, Baltimore, MD.
Leboffe, M.J. and Pierce, B.E. 2010. Microbiology: Laboratory Theory and Application, 3rd edition. Morton Publishing Company.
Logan, NA. (2012). *Bacillus* and relatives in foodborne illness. Journal of Applied Microbiology 112:417–429.
Mazza, P., Zani, F., and Martelli, P. (1992) Studies on the antibiotic resistance of *Bacillus subtilis* strains used in oral bacteriotherapy. Bollettino Chimico Farmaceutico 131:401–408.
Mayhall, C.G., Lamb, V.A., Gayle, W.E., and Haynes, B.W. (1979). *Enterobacter cloacae* septicemia in a burn center: Epidemiology and control of an outbreak. The Journal of Infectious Diseases 139(2):166-71.
Moore, V. (1912). Principles of Microbiology. Carpenter and Company, Ithaca, NY.
O'Connell, S.P., York, E.A., Collins, M.B., Rosbach, D.R,. Black, K.R, and Haney, W.B. (2007). An initial inventory of bacteria found within the soils and waters of Great Smoky Mountains National Park. Southeastern Naturalist Vol. 6:57-72.
O'Hara, C.M., Brenner, F.W., and Miller, J.M. (2000). Classification, identification, and clinical significance of *Proteus*, *Providencia*, and *Morganella*. Clinical Microbiology Reviews. Vol. 13 : 534-46.
Oggioni, M., Pozzi, G., Valensin, P., Galieni, P., and Bigazzi, C. (1998). Recurrent septicemia in an immunocompromised patient due to probiotic strains of *Bacillus subtilis.* Journal of Clinical Microbiology 36:325–326.
Pandey P.K., Kass, P.H., Soupir, M.L., Biswas, and S., Singh, V.P. (2014). Contamination of water resources by pathogenic bacteria. AMB Express Vol. 4:51.
Pelczar, M.J., Jr. (1957). Manual of Microbiological Methods. McGraw-Hill Book Co., New York, NY.
Reed, H. (1914). A Manual of Bacteriology: For Agricultural and General Science Students. Ginn and Company, Boston, MA.
Rjazantseva, IN, Andreeva, IN, and Ogorodnikova, TI. (1994). Effect of various growth conditions on pigmentation of *Serratia marcescens*. Microbios 79:155-161.
Salkinoja-Salonen, MS, Vuorio, R, Andersson, MA, Kämpfer, P, Andersson, MC, Honkanen-Buzalski, T, and Scoging, AC. (1999). Toxigenic strains of *Bacillus licheniformis* related to food poisoning. Applied and Environmental Microbiology 65:4637–4645.
Sanders, J.W.E. and Sanders, C.C. (1997). *Enterobacter* spp.: pathogens poised to flourish at the turn of the century. Clinical Microbiology Reviews 10:220–241.
Savage, D.C. (1977). Microbial ecology of the gastrointestinal tract. Annual Review of Microbiology 31:107–133.
Silver S. and Wendt L. (1967). Mechanism of Action of Phenylethyl alcohol:
Breakdown of the cellular permeability barrier. Journal of Bacteriology 93 (2):560-566.
Smibert, R.M. and Krieg, N.R. (1994). Phenotypic characterization. In P. Gerhardt, R. Murray, W. Wood, and N. Krieg (eds.), Methods for General and Molecular Bacteriology. ASM Press, Washington, DC.
Traub, W.H., Craddock, M.E., Raymond, E.A., Fox, M., and McCall, C.E. (1971). Characterization of an unusual strain of *Proteus rettgeri* associated with an outbreak of nosocomial urinary tract infection. Applied Microbiology 22:278–283.

Vandenesch, F., Naimi, T., Enright M.C., Lina, G., Nimmo, G.R., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M.E., and Etienne, J. (2003). Community-acquired methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes: worldwide emergence. Emerging Infectious Diseases 9:978–984.

Wang, Q, Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied Environmental Microbiology 73(16):5261-5267.

Wu, S., Piscitelli, C., de Lencastre, H., and Tomasz, A. (1996). Tracking the evolutionary origin of the methicillin resistant gene: cloning and sequencing of a homologue of mecA from a methicillin susceptible strain of *Staphylococcus sciuri*. Microbial Drug Resistance 2: 435–441.

Yu VL. (1979). *Serratia marcescens*: historical perspective and clinical review. The New England Journal of Medicine 300(16):887-893.

Zhang X, Zheng, Y, Fried, LE, et al. – list all authors (2011) Disruption of the mitochondrial thioredoxin system as a cell death mechanism of cationic triphenylmethanes. Free Radical Biology & Medicine 50:811–820.

Zhu, K., Dietrich, R., Didier, A., Doyscher, D., and Märtlbauer, E. (2014). Recent developments in antibody-based assays for the detection of bacterial toxins. Toxins 6(4):1325–1348.

APPENDIX

Water Pump Description:

The water pump used to this study was a hand drill powered pump (Flotec FPDMP21SA-P2 self priming drill pump) with about 5 ft of plastic hose attached to one end and a Pall Gelman Sciences capsule filter clamped to the other end. The free end of the hose had a steel mesh screwed to it. The hose and steel mesh were attached to two large plastic colanders of equal size (~ 8 inches in diameter) clamped together to form a sphere. A different hose was attached to the outlet end of the capsule filter. This hose emptied into a 6-gallon jerry can for measuring the volume of water sampled.

16S rDNA Sequences from ten bacterial isolates obtained from streams in Great Smoky Mountains National Park (key for isolates: NC = North Carolina, T = Tennessee; KP = Kephart Prong, OR = Oconaluftee River, HKB = Hickory King Branch, LPR = Little Pigeon River; 1, 2, 3 = sequential isolates obtained from each site).

NCOR1

1 CTATACATGC AGTCGAGCGA ACAGATGAGA AGCTTGCTTC TCTGATGTTA GCGGCGGACG
61 GGTGAGTAAC ACGTGGGTAA CCTACCTATA AGACTGGGAT AACTCCGGGA AACCGGGGCT
121 AATACCGGAT AATATTTTGA ACCGCATGGT TCAATAGTGA AAGACGGTTT CGGCTGTCAC
181 TTATAGATGG ACCCGCGCCG TATTAGCTAG TTGGTAAGGT AACGGCTTAC CAAGGCGACG
241 ATACGTAGCC GACCTGAGAG GGTGATCGGC CACACTGGAA CTGAGACACG GTCCAGACTC
301 CTACGGGAGG CAGCAGTAGG GAATCTTCCG CAATGGGCGA AAGCCTGACG GAGCAACGCC
361 GCGTGAGTGA TGAAGGTCTT CGGATCGTAA AACTCTGTTG TTAGGGAAGA ACAAATTTGT
421 TAGTAACTGA ACAAGTCTTG ACGGTACCTA ACCAGAAAGC CACGGCTAAC TACGTGCCAG
481 CAGCCGCGGT AATACGTAGG TGGCAAGCGT TATCCGGGAA TTATTGGGCG TAAAGCGCGC
541 GTAGGCGGTT TCTTAAGTCT GATGTGAAAG CCCACGGCTC AACCGTGGAG GGTCATTGGA
601 AACTGGGAAA CTTGAGTGCA GAAGAGGAGA GTGGAATTCC ATGTGTAGCG GTGAAATGCG
661 CAGAGATATG GAGGAACACC AGTGGCGAAG GCGGCTCTCT GGTCTGTAAC TGACGCTGAT
721 GTGCGAAAGC GTGGGGATCA AACAGGATTA GATACCCTGG TAGTCCACGC CGTAAACGAT
781 GAGTGCTAAG TGTTAGGGGG TTTCCGCCCC TTAGTGCTGC AGCTAACGCA TTAAGCACTC
841 CGCCTGGGGA GTACGACCGC AAGGTTGAAA CTCAAAGGAA TTGACGGGGA CCCGCACAAG
901 CGGTGGAGCA TGTGGTTTAA TTCGAAGCAA CGCGAAGAAC CTTACCAAAT CTTGACATCC
961 TTTGACCGCT CTAGAGATAG AGTCTTCCCC TTCGGGGGAC AAAGTGACAG GTGGTGCATG
1021 GTTGTCGTCA GCTCGTGTCG TGAGATGTTG GGTTAAGTCC CGCAACGAGC GCAACCCTTA
1081 AGCTTAGTTG CCATCATTAA GTTGGGCACT CTAGGTTGAC TGCCGGTGAC AAACCGGAGG
1141 AAGGTGGGGA TGACGTCAAA TCATCATGCC CCTTATGATT TGGGCTACAC ACGTGCTACA
1201 ATGGATAATA CAAAGGGCAG CGAATCCGCG AGGCCAAGCA AATCCCATAA AATTATTCTC
1261 AGTTCGGATT GTAGTCTGCA ACTCGACTAC ATGAAGCTGG AATCGCTAGT AATCGTAGAT
1321 CAGCATGCTA CGGTGAATAC GTTCCCGGGT CTTGTACACA CCGCCCGTCA CACCACGAGA
1381 GTTTGTAACA CCCGAAGCCG GTGGAGTAAC CTTTAGGAGC TAGC

NCOR2

1 TGCAGTCGAG CGGTAACAGG AGAAAGCTTG CTTTCTTGCT GACGAGCGGC GGACGGGTGA

61 GTAATGTATG GGGATCTGCC CGATAGAGGG GGATAACTAC TGGAAACGGT GGCTAATACC

121 GCATAATGTC TACGGACCAA AGCAGGGGCT CTTCGGACCT TGCACTATCG GATGAACCCA

181 TATGGGATTA GCTAGTAGGT GGGGTAAAGG CTCACCTAGG CGACGATCTC TAGCTGGTCT

241 GAGAGGATGA TCAGCCACAC TGGGACTGAG ACACGGCCCA GACTCCTACG GGAGGCAGCA

301 GTGGGGAATA TTGCACAATG GGCGCAAGCC TGATGCAGCC ATGCCGCGTG TATGAAGAAG

361 GCCTTAGGGT TGTAAAGTAC TTTCAGCGGG GAGGAAGGTG ATAAGGTTAA TACCCTTATC

421 AATTGACGTT ACCCGCAGAA GAAGCACCGG CTAACTCCGT GCCAGCAGCC GCGGTAATAC

481 GGAGGGTGCA AGCGTTAATC GGAATTACTG GGCGTAAAGC GCACGCAGGC GGTCAATTAA

541 GTCAGATGTG AAAGCCCCGA GCTTAACTTG GGAATTGCAT CTGAAACTGG TTGGCTAGAG

601 TCTTGTAGAG GGGGGTAGAA TTCCATGTGT AGCGGTGAAA TGCGTAGAGA TGTGGAGGAA

661 TACCGGTGGC GAAGGCGGCC CCCTGGACAA AGACTGACGC TCAGGTGCGA AAGCGTGGGG

721 AGCAAACAGG ATTAGATACC CTGGTAGTCC ACGCTGTAAA CGATGTCGAT TTAGAGGTTG

781 TGGTCTTGAA CCGTGGCTTC TGGAGCTAAC GCGTTAAATC GACCGCCTGG GGAGTACGGC

841 CGCAAGGTTA AAACTCAAAT GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCATGTGGTT

901 TAATTCGATG CAACGCGAAG AACCTTACCT ACTCTTGACA TCCAGCGAAT CCTTTAGAGA

961 TAGAGGAGTG CCTTCGGGAA CGCTGAGACA GGTGCTGCAT GGCTGTCGTC AGCTCGTGTT

1021 GTGAAATGTT GGGTTAAGTC CCGCAACGAG CGCAACCCTT ATCCTTTGTT GCCAGCACGT

1081 AATGGTGGGA ACTCAAAGGA GACTGCCGGT GATAAACCGG AGGAAGGTGG GGATGACGTC

1141 AAGTCATCAT GGCCCTTACG AGTAGGGCTA CACACGTGCT ACAATGGCAG ATACAAAGAG

1201 AAGCGACCTC GCGAGAGCAA GCGGAACTCA TAAAGTCTGT CGTAGTCCGG ATTGGAGTCT

1261 GCAACTCGAC TCCATGAAGT CGGAATCGCT AGTAATCGTA GATCAGAATG CTACGGTGAA

1321 TACGTTCCCG GGCCTTGTAC ACACCGCCCG TCACACCATG GGAGTGGGTT GCAAAAGATA

1381 GTAGGTAGCT TAACCTTCGG GAG

NCOR3

1 GCAGTCGAGC GGTAACAGGG GAAGCTTGCT TCCTCGCTGA CGAGCGGCGG ACGGGTGAGT
61 AATGTATGGG GATCTGCCCG ATAGAGGGGG ATAACTACTG GAAACGGTAG CTAATACCGC
121 ATAATCTCTC AGGAGCAAAG CAGGGGAACT TCGGTCCTTG CGCTATCGGA TGAACCCATA
181 TGGGATTAGC TAGTAGGTGA GGTAATGGCT CACCTAGGCG ACGATCCCTA GCTGGTCTGA
241 GAGGATGATC AGCCACACTG GGACTGAGAC ACGGCCCAGA CTCCTACGGG AGGCAGCAGT
301 GGGGAATATT GCACAATGGG CGCAAGCCTG ATGCAGCCAT GCCGCGTGTA TGAAGAAGGC
361 CCTAGGGTTG TAAAGTACTT TCAGTCGGGA GGAAGGCGTT GATGCTAATA TCATCAACGA
421 TTGACGTTAC CGACAGAAGA AGCACCGGCT AACTCCGTGC CAGCAGCCGC GGTAATACGG
481 AGGGTGCAAG CGTTAATCGG AATTACTGGG CGTAAAGCGC ACGCAGGCGG TTGATTAAGT
541 TAGATGTGAA ATCCCCGGGC TTAACCTGGG AATGGCATCT AAGACTGGTC AGCTAGAGTC
601 TTGTAGAGGG GGGTAGAATT CCATGTGTAG CGGTGAAATG CGTAGAGATG TGGAGGAATA
661 CCGGTGGCGA AGGCGGCCCC CTGGACAAAG ACTGACGCTC AGGTGCGAAA GCGTGGGGAG
721 CAAACAGGAT TAGATACCCT GGTAGTCCAC GCTGTAAACG ATGTCGATTT GAAGGTTGTT
781 CCCTTGAGGA GTGGCTTTCG GAGCTAACGC GTTAAATCGA CCGCCTGGGG AGTACGGCCG
841 CAAGGTTAAA ACTCAAATGA ATTGACGGGG GCCCGCACAA GCGGTGGAGC ATGTGGTTTA
901 ATTCGATGCA ACGCGAAGAA CCTTACCTAC TCTTGACATC CAGAGAACTT AGCAGAGATG
961 CTTTGGTGCC TTCGGGAACT CTGAGACAGG TGCTGCATGG CTGTCGTCAG CTCGTGTTGT
1021 GAAATGTTGG GTTAAGTCCC GCAACGAGCG CAACCCTTAT CCTTTGTTGC CAGCGATTCG
1081 GTCGGGAACT CAAAGGAGAC TGCCGGTGAT AAACCGGAGG AAGGTGGGGA TGACGTCAAG
1141 TCATCATGGC CCTTACGAGT AGGGCTACAC ACGTGCTACA ATGGCGTATA CAAAGAGAAG
1201 CGACCTCGCG AGAGCAAGCG GAACTCATAA AGTACGTCGT AGTCCGGATT GGAGTCTGCA
1261 ACTCGACTCC ATGAAGTCGG AATCGCTAGT AATCGTAGAT CAGAATGCTA CGGTGAATAC
1321 GTTCCCGGGC CTTGTACACA CCGCCCGTCA CACCATGGGA GTGGGTTGCA AAAGAAGTAG
1381 GTAGCTTAAC CTTCGGGAGG GCGCTA

THKB1

1 GCAGTCGAGC GGACCGACGG GAGCTTGCTC CCTTAGGTCA GCGGCGGACG GGTGAGTAAC
61 ACGTGGGTAA CCTGCCTGTA AGACTGGGAT AACTCCGGGA AACCGGGGCT AATACCGGAT
121 GCTTGATTGA ACCGCATGGT TCAATCATAA AAGGTGGCTT TTAGCTACCA CTTGCAGATG
181 GACCCGCGGC GCATTAGCTA GTTGGTGAGG TAACGGCTCA CCAAGGCGAC GATGCGTAGC
241 CGACCTGAGA GGGTGATCGG CCACACTGGG ACTGAGACAC GGCCCAGACT CCTACGGGAG
301 GCAGCAGTAG GGAATCTTCC GCAATGGACG AAAGTCTGAC GGAGCAACGC CGCGTGAGTG
361 ATGAAGGTTT TCGGATCGTA AAACTCTGTT GTTAGGGAAG AACAAGTACC GTTCGAATAG
421 GGCGGTACCT TGACGGTACC TAACCAGAAA GCCACGGCTA ACTACGTGCC AGCAGCCGCG
481 GTAATACGTA GGTGGCAAGC GTTGTCCGGA ATTATTGGGC GTAAAGCGCG CGCAGGCGGT
541 TTCTTAAGTC TGATGTGAAA GCCCCCGGCT CAACCGGGGA GGGTCATTGG AAACTGGGGA
601 ACTTGAGTGC AGAAGAGGAG AGTGGAATTC CACGTGTAGC GGTGAAATGC GTAGAGATGT
661 GGAGGAACAC CAGTGGCGAA GGCGACTCTC TGGTCTGTAA CTGACGCTGA GGCGCGAAAG
721 CGTGGGGAGC GAACAGGATT AGATACCCTG GTAGTCCACG CCGTAAACGA TGAGTGCTAA
781 GTGTTAGAGG GTTTCCGCCC TTTAGTGCTG CAGCAAACGC ATTAAGCACT CCGCCTGGGG
841 AGTACGGTCG CAAGACTGAA ACTCAAAGGA ATTGACGGGG GCCCGCACAA GCGGTGGAGC
901 ATGTGGTTTA ATTCGAAGCA ACGCGAAGAA CCTTACCAGG TCTTGACATC CTCTGACAAC
961 CCTAGAGATA GGGCTTCCCC TTCGGGGGCA GAGTGACAGG TGGTGCATGG TTGTCGTCAG
1021 CTCGTGTCGT GAGATGTTGG GTTAAGTCCC GCAACGAGCG CAACCCTTGA TCTTAGTTGC
1081 CAGCATTCAG TTGGGCACTC TAAGGTGACT GCCGGTGACA AACCGGAGGA AGGTGGGGAT
1141 GACGTCAAAT CATCATGCCC CTTATGACCT GGGCTACACA CGTGCTACAA TGGGCAGAAC
1201 AAAGGGCAGC GAAGCCGCGA GGCTAAGCCA ATCCCACAAA TCTGTTCTCA GTTCGGATCG
1261 CAGTCTGCAA CTCGACTGCG TGAAGCTGGA ATCGCTAGTA ATCGCGGATC AGCATGCCGC
1321 GGTGAATACG TTCCCGGGCC TTGTACACAC CGCCCGTCAC ACCACGAGAG TTTGTAACAC
1381 CCGAAGTCGG TGAGGTAACC TTT

THKB2

1 TGCAGTCGAG CGGACAGATG GGAGCTTGCT CCCTGATGTT AGCGGCGGAC GGGTGAGTAA
61 CACGTGGGTA ACCTGCCTGT AAGACTGGGA TAACTCCGGG AAACCGGGGC TAATACCGGA
121 TGGTTGTTTG AACCGCATGG TTCAAACATA AAAGGTGGCT TCGGCTACCA CTTACAGATG
181 GACCCGCGGC GCATTAGCTA GTTGGTGAGG TAACGGCTCA CCAAGGCAAC GATGCGTAGC
241 CGACCTGAGA GGGTGATCGG CCACACTGGG ACTGAGACAC GGCCCAGACT CCTACGGGAG
301 GCAGCAGTAG GGAATCTTCC GCAATGGACG AAAGTCTGAC GGAGCAACGC CGCGTGAGTG
361 ATGAAGGTTT TCGGATCGTA AAGCTCTGTT GTTAGGGAAG AACAAGTACC GTTCGAATAG
421 GGCGGTACCT TGACGGTACC TAACCAGAAA GCCACGGCTA ACTACGTGCC AGCAGCCGCG
481 GTAATACGTA GGTGGCAAGC GTTGTCCGGA ATTATTGGGC GTAAAGGGCT CGCAGGCGGT
541 TTCTTAAGTC TGATGTGAAA GCCCCCGGCT CAACCGGGGA GGGTCATTGG AAACTGGGGA
601 ACTTGAGTGC AGAAGAGGAG AGTGGAATTC CACGTGTAGC GGTGAAATGC GTAGAGATGT
661 GGAGGAACAC CAGTGGCGAA GGCGACTCTC TGGTCTGTAA CTGACGCTGA GGAGCGAAAG
721 CGTGGGGAGC GAACAGGATT AGATACCCTG GTAGTCCACG CCGTAAACGA TGAGTGCTAA
781 GTGTTAGGGG GTTTCCGCCC CTTAGTGCTG CAGCTAACGC ATTAAGCACT CCGCCTGGGG
841 AGTACGGTCG CAAGACTGAA ACTCAAAGGA ATTGACGGGG GCCTGACGGG GGCCCGCACA
901 AGCGGTGGAG CATGTGGTTT AATTCGAAGC AACGCGAAGA ACCTTACCAG GTCTTGACAT
961 CCTCTGACAA TCCTAGAGAT AGGACGTCCC CTTCGGGGGC AGAGTGACAG GTGGTGCATG
1021 GTTGTCGTCA GCTCGTGTCG TGAGATGTTG GGTTAAGTCC CGCAACGAGC GCAACCCTTG
1081 ATCTTAGTTG CCAGCATTCA GTTGGGCACT CTAAGGTGAC TGCCGGTGAC AAACCGGAGG
1141 AAGGTGGGGA TGACGTCAAA TCATCATGCC CCTTATGACC TGGGCTACAC ACGTGCTACA
1201 ATGGACAGAA CAAAGGGCAG CGAAACCGCG AGGTTAAGCC AATCCCACAA ATCTGTTCTC
1261 AGTTCGGATC GCAGTCTGCA ACTCGACTGC GTGAAGCTGG AATCGCTAGT AATCGCGGAT
1321 CAGCATGCCG CGGTGAATAC GTTCCCGGGC CTTGTACACA CCGCCCGTCA CACCACGAGA
1381 GTTTGTAACA CCCGAAGTCG GTGAGGTAAC

NCKP1

1 TGCAGTCGAG CGGTAGCACA GGGGAGCTTG CTCCCYGGGT GACGAGCGGC GGACGGGTGA

61 GTAATGTCTG GGAAACTGCC TGATGGAGGG GGATAACTAC TGGAAACGGT AGCTAATACC

121 GCATAACGTC GCAAGACCAA AGAGGGGGAC CTTCGGGCCT CTTGCCATCA GATGTGCCCA

181 GATGGGATTA GCTAGTAGGT GGGGTAATGG CTCACCTAGG CGACGATCCC TAGCTGGTCT

241 GAGAGGATGA CCAGCCACAC TGGAACTGAG ACACGGTCCA GACTCCTACG GGAGGCAGCA

301 GTGGGGAATA TTGCACAATG GGCGCAAGCC TGATGCAGCC ATGCCGCGTG TGTGAAGAAG

361 GCCTTCGGGT TGTAAAGCAC TTTCAGCGAG GAGGAAGGTG GTGARCTTAA TACGTTCATC

421 AATTGACGTT ACTCGCAGAA GAAGCACCGG CTAACTCCGT GCCAGCAGCC GCGGTAATAC

481 GGAGGGTGCA AGCGTTAATC GGAATTACTG GGCGTAAAGC GCACGCAGGC GGTTTGTTAA

541 GTCAGATGTG AAATCCCCGG GCTCAACCTG GGAACTGCAT TTGAAACTGG CAAGCTAGAG

601 TCTCGTAGAG GGGGGTAGAA TTCCAGGTGT AGCGGTGAAA TGCGTAGAGA TCTGGAGGAA

661 TACCGGTGGC GAAGGCGGCC CCCTGGACGA AGACTGACGC TCAGGTGCGA AAGCGTGGGG

721 AGCAAACAGG ATTAGATACC CTGGTAGTCC ACGCTGTAAA CGATGTCGAT TTGGAGGTTG

781 TGCCCTTGAG GCGTGGCTTC CGGAGCTAAC GCGTTAAATC GACCGCCTGG GGAGTACGGC

841 CGCAAGGTTA AAACTCAAAT GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCATGTGGTT

901 TAATTCGATG CAACGCGAGA ACCTTACCTA CTCTTGACAT CCAGAGAACT TAGCAGAGAT

961 GNTTTGGTGC CTTCGGGAAC TCTGAGACAG GTGCTGCATG GCTGTCGTCA GCTCGTGTTG

1021 TGAAATGTTG GGTTAAGTCC CGCAACGAGC GCAACCCTTA TCCTTTGTTG CCAGCGGTTC

1081 GGCCGGGAAC TCAAAGGAGA CTGCCAGTGA TAAACTGGAG GAAGGTGGGG ATGACGTCAA

1141 GTCATCATGG CCCTTACGAG TAGGGCTACA CACGTGCTAC AATGGCGTAT ACAAAGAGAA

1201 GCGACCTCGC GAGAGCAAGC GGACCTCATA AAGTACGTCG TAGTCCGGAT TGGAGTCTGC

1261 AACTCGACTC CATGAAGTCG GAATCGCTAG TAATCGTAGA TCAGAATGCT ACGGTGAATA

1321 CGTTCCCGGG CCTTGTACAC ACCGCCCGTC ACACCATGGG AGTGGGTTGC AAAAGAAGTA

1381 GGTAGCTTAA CCTTCGGGAG GGCGCTAC

NCKP2

1 TGCAGTCGAG CGGACAGATG GGAGCTTGCT CCCTGATGTT AGCGGCGGAC GGGTGAGTAA

61 CACGTGGGTA ACCTGCCTGT AAGACTGGGA TAACTCCGGG AAACCGGGGC TAATACCGGA

121 TGGTTGTTTG AACCGCATGG TTCAAACATA AAAGGTGGCT TCGGCTACCA CTTACAGATG

181 GACCCGCGGC GCATTAGCTA GTTGGTGAGG TAACGGCTCA CCAAGGCAAC GATGCGTAGC

241 CGACCTGAGA GGGTGATCGG CCACACTGGG ACTGAGACAC GGCCCAGACT CCTACGGGAG

301 GCAGCAGTAG GGAATCTTCC GCAATGGACG AAAGTCTGAC GGAGCAACGC CGCGTGAGTG

361 ATGAAGGTTT TCGGATCGTA AAGCTCTGTT GTTAGGGAAG AACAAGTACC GTTCGAATAG

421 GGCGGTACCT TGACGGTACC TAACCAGAAA GCCACGGCTA ACTACGTGCC AGCAGCCGCG

481 GTAATACGTA GGTGGCAAGC GTTGTCCGGA ATTATTGGGC GTAAAGGGCT CGCAGGCGGT

541 TTCTTAAGTC TGATGTGAAA GCCCCCGGCT CAACCGGGGA GGGTCATTGG AAACTGGGGA

601 ACTTGAGTGC AGAAGAGGAG AGTGGAATTC CACGTGTAGC GGTGAAATGC GTAGAGATGT

661 GGAGGAACAC CAGTGGCGAA GGCGACTCTC TGGTCTGTAA CTGACGCTGA GGAGCGAAAG

721 CGTGGGGAGC GAACAGGATT AGATACCCTG GTAGTCCACG CCGTAAACGA TGAGTGCTAA

781 GTGTTAGGGG GTTTCCGCCC CTTAGTGCTG CAGCTAACGC ATTAAGCACT CCGCCTGGGG

841 AGTACGGTCG CAAGACTGAA ACTCAAAGGA ATTGACGGGG GCCCGCACAA GCGGTGGAGC

901 ATGTGGTTTA ATTCGAAGCA ACGCGAAGAA CCTTACCAGG TCTTGACATC CTCTGACAAT

961 CCTAGAGATA GGACGTCCCC TTCGGGGGCA GAGTGACAGG TGGTGCATGG TTGTCGTCAG

1021 CTCGTGTCGT GAGATGTTGG GTTAAGTCCC GCAACGAGCG CAACCCTTGA TCTTAGTTGC

1081 CAGCATTCAG TTGGGCACTC TAAGGTGACT GCCGGTGACA AACCGGAGGA AGGTGGGGAT

1141 GACGTCAAAT CATCATGCCC CTTATGACCT GGGCTACACA CGTGCTACAA TGGACAGAAC

1201 AAAGGGCAGC GAAACCGCGA GGTTAAGCCA ATCCCACAAA TCTGTTCTCA GTTCGGATCG

1261 CAGTCTGCAA CTCGACTGCG TGAAGCTGGA ATCGCTAGTA ATCGCGGATC AGCATGCCGC

1321 GGTGAATACG TTCCCGGGCC TTGTACACAC CGCCCGTCAC ACCACGAGAG TTTGTAACAC

1381 CCGAAGTCGG TGAGGTAACC TTTAGGAGCC A

TLPR1

1 GCAGTCGAGC GGTAACAGGA GAAAGCTTGC TTTCTTGCTG ACGAGCGGCG GACGGGTGAG

61 TAATGTATGG GGATCTGCCC GATAGAGGGG GATAACTACT GGAAACGGTG GCTAATACCG

121 CATAATGTCT ACGGACCAAA GCAGGGGCTC TTCGGACCTT GCACTATCGG ATGAACCCAT

181 ATGGGATTAG CTAGTAGGTG GGGTAAAGGC TCACCTAGGC GACGATCTCT AGCTGGTCTG

241 AGAGGATGAT CAGCCACACT GGGACTGAGA CACGGCCCAG ACTCCTACGG GAGGCAGCAG

301 TGGGGAATAT TGCACAATGG GCGCAAGCCT GATGCAGCCA TGCCGCGTGT ATGAAGAAGG

361 CCTTAGGGTT GTAAAGTACT TTCAGCGGGG AGGAAGGTGA TAAGGTTAAT ACCCTTATCA

421 ATTGACGTTA CCCGCAGAAG AAGCACCGGC TAACTCCGTG CCAGCAGCCG CGGTAATACG

481 GAGGGTGCAA GCGTTAATCG GAATTACTGG GCGTAAAGCG CACGCAGGCG GTCAATTAAG

541 TCAGATGTGA AAGCCCCGAG CTTAACTTGG GAATTGCATC TGAAACTGGT TGGCTAGAGT

601 CTTGTAGAGG GGGGTAGAAT TCCATGTGTA GCGGTGAAAT GCGTAGAGAT GTGGAGGAAT

661 ACCGGTGGCG AAGGCGGCCC CCTGGACAAA GACTGACGCT CAGGTGCGAA AGCGTGGGGA

721 GCAAACAGGA TTAGATACCC TGGTAGTCCA CGCTGTAAAC GATGTCGATT TAGAGGTTGT

781 GGTCTTGAAC CGTGGCTTCT GGAGCTAACG CGTTAAATCG ACCGCCTGGG GAGTACGGCC

841 GCAAGGTTAA AACTCAAATG AATTGACGGG GGCCCGCACA AGCGGTGGAG CATGTGGTTT

901 AATTCGATGC AACGCGAAGA ACCTTACCTA CTCTTGACAT CCAGCGAATC CTTTAGAGAT

961 AGAGGAGTGC CTTCGGGAAC GCTGAGACAG GTGCTGCATG GCTGTCGTCA GCTCGTGTTG

1021 TGAAATGTTG GGTTAAGTCC CGCAACGAGC GCAACCCTTA TCCTTTGTTG CCAGCACGTA

1081 ATGGTGGGAA CTCAAAGGAG ACTGCCGGTG ATAAACCGGA GGAAGGTGGG GATGACGTCA

1141 AGTCATCATG GCCCTTACGA GTAGGGCTAC ACACGTGCTA CAATGGCAGA TACAAAGAGA

1201 AGCGACCTCG CGAGAGCAAG CGGAACTCAT AAAGTCTGTC GTAGTCCGGA TTGGAGTCTG

1261 CAACTCGACT CCATGAAGTC GGAATCGCTA GTAATCGTAG ATCAGAATGC TACGGTGAAT

1321 ACGTTCCCGG GCCTTGTACA CACCGCCCGT CACACCATGG GAGTGGGTTG CAAAAGAAGT

1381 AGGTAGCTTA ACCTTCGGGA

TLPR2

1 GCAGTCGAAC GGTAACAGGA AGCAGCTTGC TGCTTCGCTG ACGAGTGGCG GACGGGTGAG

61 TAATGTCTGG GAAACTGCCT GATGGAGGGG GATAACTACT GGAAACGGTA GCTAATACCG

121 CATAACGTCG CAAGACCAAA GAGGGGGACC TTCGGGCCTC TTGCCATCGG ATGTGCCCAG

181 ATGGGATTAG CTAGTAGGTG GGGTAACGGC TCACCTAGGC GACGATCCCT AGCTGGTCTG

241 AGAGGATGAC CAGCCACACT GGAACTGAGA CACGGTCCAG ACTCCTACGG GAGGCAGCAG

301 TGGGGAATAT TGCACAATGG GCGCAAGCCT GATGCAGCCA TGCCGCGTGT ATGAAGAAGG

361 CCTTCGGGTT GTAAAGTACT TTCAGCGGGG AGGAAGGCGA TAAGGTTAAT AACCTTGTCG

421 ATTGACGTTA CCCGCAGAAG AAGCACCGGC TAACTCCGTG CCAGCAGCCG CGGTAATACG

481 GAGGGTGCAA GCGTTAATCG GAATTACTGG GCGTAAAGCG CACGCAGGCG GTCTGTCAAG

541 TCGGATGTGA AATCCCCGGG CTCAACCTGG GAACTGCATT CGAAACTGGC AGGCTAGAGT

601 CTTGTAGAGG GGGGTAGAAT TCCAGGTGTA GCGGTGAAAT GCGTAGAGAT CTGGAGGAAT

661 ACCGGTGGCG AAGGCGGCCC CCTGGACAAA GACTGACGCT CAGGTGCGAA AGCGTGGGGA

721 GCAAACAGGA TTAGATACCC TGGTAGTCCA CGCCGTAAAC GATGTCGACT TGGAGGTTGT

781 GCCCTTGAGG CGTGGCTTCC GGAGCTAACG CGTTAAGTCG ACCGCCTGGG GAGTACGGCC

841 GCAAGGTTAA AACTCAAATG AATTGACGGG GGCCCGCACA AGCGGTGGAG CATGTGGTTT

901 AATTCGATGC AACGCGAAGA ACCTTACCTA CTCTTGACAT CCAGAGAACT TAGCAGAGAT

961 GCTTTGGTGC CTTCGGGAAC TCTGAGACAG GTGCTGCATG GCTGTCGTCA GCTCGTGTTG

1021 TGAAATGTTG GGTTAAGTCC CGCAACGAGC GCAACCCTTA TCCTTTGTTG CCAGCGGTCC

1081 GGCCGGGAAC TCAAAGGAGA CTGCCAGTGA TAAACTGGAG GAAGGTGGGG ATGACGTCAA

1141 GTCATCATGG CCCTTACGAG TAGGGCTACA CACGTGCTAC AATGGCGCAT ACAAAGAGAA

1201 GCGACCTCGC GAGAGCAAGC GGACCTCATA AAGTGCGTCG TAGTCCGGAT TGGAGTCTGC

1261 AACTCGACTC CATGAAGTCG GAATCGCTAG TAATCGTGGA TCAGAATGCC ACGGTGAATA

1321 CGTTCCCGGG CCTTGTACAC ACCGCCCGTC ACACCATGGG AGTGGGTTGC AAAAGAAGTA

1381 GGTAGCTTAA CCTTCGGGAG GGCGCT

TLPR3

1 TGCAGTCGAC GGTAACAGGA AGCAGCTTGC TGCTTCGCTG ACGAGTGGCG GACGGGTGAG

61 TAATGTCTGG GAAACTGCCT GATGGAGGGG GATAACTACT GGAAACGGTA GCTAATACCG

121 CATAACGTCG CAAGACCAAA GAGGGGGACC TTCGGGCCTC TTGCCATCGG ATGTGCCCAG

181 ATGGGATTAG CTAGTAGGTG GGGTAACGGC TCACCTAGGC GACGATCCCT AGCTGGTCTG

241 AGAGGATGAC CAGCCACACT GGAACTGAGA CACGGTCCAG ACTCCTACGG GAGGCAGCAG

301 TGGGGAATAT TGCACAATGG GCGCAAGCCT GATGCAGCCA TGCCGCGTGT ATGAAGAAGG

361 CCTTCGGGTT GTAAAGTACT TTCAGCGGGG AGGAAGGCGA TGAGGTTAAT AACCTTGTCG

421 ATTGACGTTA CCCGCAGAAG AAGCACCGGC TAACTCCGTG CCAGCAGCCG CGGTAATACG

481 GAGGGTGCAA GCGTTAATCG GAATTACTGG GCGTAAAGCG CACGCAGGCG GTCTGTCAAG

541 TCGGATGTGA AATCCCCGGG CTCAACCTGG GAACTGCATT CGAAACTGGC AGGCTAGAGT

601 CTTGTAGAGG GGGGTAGAAT TCCAGGTGTA GCGGTGAAAT GCGTAGAGAT CTGGAGGAAT

661 ACCGGTGGCG AAGGCGGCCC CCTGGACAAA GACTGACGCT CAGGTGCGAA AGCGTGGGGA

721 GCAAACAGGA TTAGATACCC TGGTAGTCCA CGCCGTAAAC GATGTCGACT TGGAGGTTGT

781 GCCCTTGAGG CGTGGCTTCC GGAGCTAACG CGTTAAGTCG ACCGCCTGGG GAGTACGGCC

841 GCAAGGTTAA AACTCAAATG AATTGACGGG GGCCCGCACA AGCGGTGGAG CATGTGGTTT

901 AATTCGATGC AACGCGAAGA ACCTTACCTA CTCTTGACAT CCAGAGAACT TAGCAGAGAT

961 GCTTTGGTGC CTTCGGGAAC TCTGAGACAG GTGCTGCATG GCTGTCGTCA GCTCGTGTTG

1021 TGAAATGTTG GGTTAAGTCC CGCAACGAGC GCAACCCTTA TCCTTTGTTG CCAGCGGTCC

1081 GGCCGGGAAC TCAAAGGAGA CTGCCAGTGA TAAACTGGAG GAAGGTGGGG ATGACGTCAA

1141 GTCATCATGG CCCTTACGAG TAGGGCTACA CACGTGCTAC AATGGCGCAT ACAAAGAGAA

1201 GCGACCTCGC GAGAGCAAGC GGACCTCATA AAGTGCGTCG TAGTCCGGAT TGGAGTCTGC

1261 AACTCGACTC CATGAAGTCG GAATCGCTAG TAATCGTGGA TCAGAATGCC ACGGTGAATA

1321 CGTTCCCGGG CCTTGTACAC ACCGCCCGTC ACACCATGGG AGTGGGTTGC AAAAGAAAGT

1381 AGGTAGCTTA ACGCTTCGGG AGGGCGCT