A NEW GENUS AND SPECIES OF *ENTEROBACTERIACEAE* ISOLATED FROM KEPHART PRONG, 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 in Biology

By

Kacie Fraser

Director: Dr. Sean O'Connell
Associate Professor of Biology
Biology Department

Committee Members: Dr. Malcolm Powell, Department of Biology
Paul Super, National Park Service

June 2017

ACKNOWLEDGEMENTS

I would like to thank my committee members, Dr. Malcolm Powell and Paul Super and my reader Dr. Robert Youker for their assistance and encouragement and in particular, my director Dr. Sean O'Connell for continual support, reassurance, and guidance. I also extend sincere thanks to the following people, Dr. Kathy Mathews, Rob McKinnon, and Tori Carlson without whom this thesis would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATBI: All Taxa Biodiversity Inventory
bp: Base pairs
BLAST: Basic Local Alignment Search Tool
CI: Consistency index
CSIs: Conserved Signature Insertion/Deletions
GTR: General Time Reversible
GSMNP: Great Smoky Mountains National Park
HI: Homoplasy index
MALDI-TOF: Matrix Assisted Laser Desorption/Ionization-Time of Flight
PCR: Polymerase chain reaction
R2A: Reasoner's 2 Agar
RC: Rescaled consistency index
RI: Retention index
RDP: Ribosomal Database Project

ABSTRACT

A NEW GENUS AND SPECIES OF *ENTEROBACTERIACEAE* ISOLATED FROM KEPHART PRONG, GREAT SMOKY MOUNTAINS NATIONAL PARK

Kacie Fraser

Western Carolina University (June 2017)

Director: Dr. Sean O'Connell

A bacterium was isolated from a water sample collected from the Kephart Prong in Great Smoky Mountains National Park. This isolate did not have a conclusive match to any known bacteria in the Ribosomal Database Project (RDP); therefore, it was tested to verify it as a novel species. The full 16S rRNA gene of the newly discovered bacterium was sequenced and results showed there to be a 100% confidence that it belonged in the *Enterobacteriaceae* family but only an 81% confidence that is belonged in the genus *Serratia* suggesting that this bacterium could represent a new genus in the family. The bacterium was characterized by its preferences and tolerance to temperature, pH, and salinity as well as for metabolic capabilities such as oxidase activity, ability to hydrolyze gelatin, ability to convert tryptophan into indole, and the ability to hydrolyze casein. Antibiotic resistance was also tested against penicillin, tetracycline, chloramphenicol, clindamycin, nalidixic acid, nitrofurantoin, and colistin. Results from these phenotypic tests suggested that the isolate is not closely related to any known enterobacter species. Based on the limitations of the above methods, the full genome was sequenced to better place it in the family *Enterobacteriaceae*. Results of whole genome comparisons to its most closely related relatives indicate that this species is quite novel. Phylogenetics were performed and demonstrate that a new genus and species of bacteria has been cultured. This work will aid in understanding the biodiversity found in Great Smoky Mountains National Park as well as the possible contributions

this bacterium provide to the ecosystem. Bacteria are also beneficial to humans and act as important resources in many different fields such as medicine, agriculture, biotechnology, and ecology. Without the knowledge and resources gained from bacteria these fields would not be as advanced as they are today. Further, conservation and restoration of ecosystems cannot occur without first understanding the biodiversity that makes up that these environments. The discovery of this isolate helps contribute to the understanding of the biodiversity in GSMNP as well as the understanding of the *Enterobacteriaceae* family.

CHAPTER ONE: INTRODUCTION

Bacterial life on earth began 3.8 billion years ago and the diversity of microbes has been growing ever since. Today we have an estimated 10^{30} bacterial cells on Earth and about 10^{9} bacterial species (S. O'Connell, personal communication). This dwarfs the diversity of *Eukarya* potentially by many orders of magnitude. Bacterial metabolic pathways, habitats, optimum temperature, pH tolerances, etc. are extremely diverse. Prokaryotes have been found to survive at temperatures above 100˚C, pH below 0, and in many environments without oxygen. They also have a great impact on the environment around them. The chemical composition of Earth's atmosphere has been mediated by microbial life (DeLong, 2000). It is thought that the Earth's atmosphere began as a mixture of hydrogen, methane, carbon dioxide, and nitrogen. As is the case now, this environment was not conducive for many living creatures to survive, but many bacteria can grow well under such conditions. Oxygen was not introduced into the atmosphere until 3.0 billion years ago when oxygenic photosynthesis evolved in cyanobacteria (Madigan et. al., 2015). Microbial life is responsible for creating today's atmosphere and allowing multicellular organisms to evolve and thrive.

Microbial life is also very important in maintaining soil as well as water nutrient levels. Autotrophs use the Calvin Cycle to produce organic carbon from carbon dioxide. This process is the main way that carbon is made available to life on Earth, and includes plant-based use of chloroplasts (Madigan et al., 2015). Nitrogen fixation is also a common process performed by prokaryotes. Nucleic acids and proteins all require significant amounts of nitrogen. Atmospheric nitrogen cannot be processed by most living organisms. However, nitrogen-fixing *Bacteria* and *Archaea* can form ammonia from atmospheric nitrogen and many of these nitrogen fixers are symbiotic with plants or other organisms (Madigan et al., 2015). Bacteria are also important for

human uses. Agricultural businesses would decline significantly if not for microbes. Microbes such as bacteria and fungi are essential in decomposing organic matter and recycling old plant material. Microbes make up the majority of all life on Earth and play a large role in all environments present on Earth today. Scientists have been able to identify and characterize over 10,000 archaeal and bacterial species but there are millions or billions more yet to be discovered (Madigan et al., 2015).

Great Smoky Mountains National Park (GSMNP) is home to hundreds of thousands of species of animals, plants, and microorganisms, however, the microbial diversity is by far the greatest and least understood. The All Taxa Biodiversity Inventory (ATBI) is attempting to inventory all of these organisms present in GSMNP. The ATBI began cataloguing organisms in 1998 and to date have inventoried over 18,000 species, with almost 1,000 of those being new to science (ATBI, 2015). Human activity has greatly reduced global biodiversity due to habitat loss, pollution, global warming, etc. (Nichols & Langdon, 2007). In order to restore or conserve the biodiversity that is left on the planet we must first understand it. Therefore, the ATBI is aiming for more than just a list of species that are present in the park but to have information on the distribution, abundance, and ecological role of each species they find. The scientists involved in the ATBI have learned a lot about the species' ranges and habitats and their interactions with each other (Nichols & Langdon, 2007). This information is valuable to conservation efforts because it indicates which sites in the park need extra protection, which species and locations are most effectively monitored as vital sign indicators of the health of the ecosystem, and when and where pest species need to be controlled. It also facilitates early detection of invasive species and determines which ecological relationships are under stress across the park (Nichols & Langdon, 2007). Understanding the biodiversity and species interactions in GSMNP is an important step

in conserving that biodiversity and restoring what damage human activity has wreaked on this ecosystem.

Of the over 18,000 species inventoried in GSMNP only about 500 of those are bacterial species. Bacterial diversity exceeds *Eukarya* diversity by a sizable amount but only a small portion of the species discovered in the park are bacterial. This discrepancy in the numbers suggests a problem or difficulty in inventorying microbes. In order to concretely name a bacterial species and find its place in the Tree of Life a pure culture is desired. However, it has been observed that only a very small fraction of microbes are recoverable as pure cultures in microbial growth media (Youssef et al., 2015). Large numbers of the bacterial species on the planet have yet to be identified and this is largely due to the fact that they are unculturable on traditional growth media. It is believed that when using growth media of any kind that 99-99.9% of all cells within an environmental sample are not being recovered. There are measures that can be taken to increase the likelihood of culturing a specific bacterial species, genus, or family such as selective media type, incubation time or conditions, or using more sensitive growth detection methods. These methods can be used when searching for less common bacteria; however, the vast majority of organisms are still missed and remain uncultured. For this reason it is extremely important to continue inventorying bacterial species in conjunction with researching new methods in culturing the unculturable.

Even though culturing is the preferred method of working with bacteria there are gene-based approaches that can place an organism systematically without having a pure culture. Microbiologists Carl Woese and Norman Pace were both pioneers in the use of the 16S rRNA gene as a phylogenetic marker (Youssef et al., 2015). The 16S ribosomal RNA gene sequence is used because it is found only in prokaryotic organisms, but can also be compared to the 18S

rDNA found in eukaryotes, it is a relatively short sequence, and has only changed a small amount over evolutionary time. Not only is it fast and relatively inexpensive to sequence but it does not require a pure culture in order to be analyzed. One can directly isolate DNA from an environmental sample and use it for study (Youssef et al., 2015). Sequencing 16S rRNA has allowed uncultured bacteria to be discovered and greatly increases our understanding of the magnitude of the bacteria and archaea domains. However, in order to legitimately name a novel species, current convention requires a culture be obtained.

Only a few species from *Enterobacteriaceae* have been discovered in GSMNP (ATBI, 2015, S. O'Connell, personal communication). According to the authoritative reference for keying bacterial species, Bergey's Manual of Systematic Bacteriology (The Williams and Wilkins Co.), *Enterobacteriaceae* have great heterogeneity in ecology, host range, and pathogenic potential for humans, animals, insects and plants (Holt, 2012). They are distributed worldwide in a varied range of habitats such as soil, water, fruits, meats, eggs, vegetables, grains, flowering plants and trees, and in animals from insects to humans. There are 44 genera in this family with an estimated 60 genera in total. There is also expected to be a large number of unnamed species in this family. In general, most genera are Gram negative rods about 0.3-1.5 microns in diameter. All but one genus, *Tatumella,* is motile by use of a flagellum. *Enterobacteriaceae* are chemoorgantrophs and facultative anaerobes. These bacteria optimally grow at temperatures between 22-35°C. Many species in the *Enterobacteriaceae* are known for their pathogenicity to humans and animals, e.g., *Salmonella* spp. and *Escherichia coli* belong to this family. It is important to study all members of this family because of their potential pathogenicity, if not for other natural history reasons such as their roles in nutrient cycling, commensal relationships with hosts, and as prey for other organisms.

The purpose of this study was to classify a novel isolate from a GSMNP stream that belonged in the family *Enterobacteriaceae.* This isolate did not have a conclusive match to any known bacteria in the Ribosomal Database Project; therefore, it was tested to verify it as a novel species and possibly a novel genus. The newly discovered bacterium was characterized by its growth requirements, its 16S rRNA gene, and its full genome sequence. There was 100% confidence that it belonged to the *Enterobacteriaceae* family but only an 81% confidence that is belonged to the genus *Serratia*. These initial findings suggest that this bacterium could be a new genus in the *Enterobacteriaceae* family and this study was undertaken to test this hypothesis.

CHAPTER TWO: METHODS AND MATERIALS

2.1: Sample Collection

Three 50-mL water samples were collected from Kephart Prong Trail (LatLong: 35º 35' 13", -83º 21' 25") in GSMNP (O'Connell et al., 2017). The samples were collected aseptically, and once in the lab a serial dilution was performed. One milliliter of sample was diluted in a 1:10 series from 10^{-1}-10^{-5}, and 100µl of the dilutions was plated on R2A media and incubated for 7 days at 25˚C. Seven colonies were chosen from each sample, streak-plated on R2A media plates, and incubated at 25˚C. Several weeks of re-streaking (every two to five days) was performed to isolate 21 different colonies of bacteria. Over time six species stopped growing so a total of 15 isolated species continued to the DNA extraction, 16S rDNA PCR, and DNA sequencing steps.

2.2: Partial 16S rDNA Sequencing

DNA was extracted from the isolates using the Ultra Clean Microbial DNA Isolation Kit (Mo Bio Inc., Solana Beach, CA). Once the DNA extraction was complete agarose gel electrophoresis was completed on all samples to ensure that the extraction was successful. Partial gene fragments for the 16S rRNA were targeted using PCR for sequencing and identification purposes. The total reaction consisted of 50µL with the following ingredients: 1X Master Mix Chemistry (Promega Corporation, Madison, WI), 0.25µM each primer (bacterial-specific primers 341F (5′-CCTACGGGAGGCAGCAG-3′) and 907R (5′-CCGTCAATTCCTTTRAGTTT-3′), and 1µL DNA template (Ishii et al., 2001).

A touchdown PCR approach was employed, consisting of 5 minutes of initial denaturation at 94˚C, then 30 cycles of denaturation at 94˚C for 1 minute, annealing for 1 minute, and extension for 3 minutes at 72˚C (Ishii et al., 2001). The annealing temperature in the first two rounds was at 65˚C, followed by one round each at 1˚C lower than the previous round, and finally eighteen rounds at 55˚C. A final extension for 7 minutes at 72˚C was employed and

then amplicons were stored at 4˚C. The PCR products from all 15 samples were then confirmed by agarose gel electrophoresis. The PCR products were cleaned using Amicon centrifugal filters (Millipore) and then sequencing PCR was performed using BigDye v3.1 chemistry (Applied Biosystems). Next the samples were cleaned using the BigDye XTerminator Cleaning Kit. The cleaned DNA was then sequenced on an Applied Biosystems 3130 capillary sequencer. Finch TV (Geospiza, 2012) and the Ribosomal Database Project (RDP) website (Altschul et al. 1997) were used to analyze and identify the sequences of DNA. RDP Classifer and RDP SeqMatch were used to determine each isolate's closest match. Search parameters for RDP SeqMatch were type strains, isolates, good quality sequences, and sequence sizes of both above and below 1,200 nucleotides. An isolate was chosen based on the quality of the sequence, low genus match in RDP, and efficient growth rate. An organism with a 100% match to the family *Enterobacteriaceae* but low match to an established genus in RDP was chosen for further observations.

2.3: Morphological and Growth Characteristics

Once the environmental isolate KF13 was chosen it was characterized using a variety of standard bacteriological tests. Negative staining with nigrosin was used to determine the shape and size of the cells and the culture was also Gram stained. Growth characteristics across temperature, pH, and salinity ranges were determined using R2A media plates. KF13 was plated on R2A media and incubated at 4˚C, 30˚C, 37˚C, and 50˚C to determine temperature preferences and tolerance. The isolate was plated on R2A nutrient agar with pH levels of 3, 5, and 9 to determine pH tolerance and was also plated on R2A nutrient agar with salinity levels of 1%, 3%, 5%, 7%, 10%, and 15% to determine salt preferences and tolerance. A control culture on R2A media with pH 7 and 0% salinity was incubated at 25˚C. All plates were grown for 48 hours.

The isolate was also tested for its oxygen requirement. It was plated on R2A and placed into a GasPak™ bag (Becton Dickinson) with an anaerobic indicator to determine its ability to grow without oxygen. The culture was observed after 72 hours. KF13 was also stabbed into motility medium with triphenyltetrazolium chloride to test for the presumptive presence of flagella. This was allowed to grow for 72 hours. A positive result was indicated by a red and turbid area extending away from the line of inoculation. Flagella staining was also conducted on cultures less than 24 hours old. *E. coli* was used as a positive control.

2.4: Metabolic Capabilities

Metabolic capabilities of the isolate were tested. All tests performed were done in triplicate. Catalase activity was determined by exposing the culture to hydrogen peroxide and oxidase activity was determined using BBL DrySlides (Becton Dickinson) which test for the presence of cytochrome C oxidase. A positive catalase test is indicated by gas bubble formation and a positive oxidase test is indicated by a blue/purple color change within 20 seconds. KF13's ability to metabolize different substances was determined using BBL EnteroPluri Tubes (Becton Dickinson). Results were observed after 48 hours at 25˚C. Lipase activity was determined by plating the isolate on spirit blue agar that contained emulsified safflower oil. *Staphylococcus aureus* was used as a positive control and inoculated plates were incubated at 30°C. Growth and lipolytic halos were looked for after 72 hours. The ability to hydrolyze gelatin was determined using a nutrient gelatin stab. *S. aureus* was used as a positive control and cultures were incubated at 25°C. Liquefaction of the medium, if present, would occur after a 7-day incubation period. KF13's ability to ferment mannitol and use it as a carbon source was also tested. The isolate was grown in a phenol red with mannitol medium at 30°C. *E. coli* was used as a positive control and any color change was observed after 48 hours. The ability of the isolate to convert tryptophan

into indole was determined by using an indole with nitrate-test medium. The inoculated tube was incubated at 30°C for 48 hours. After 48 hours Kovacs' Reagent was added to the tube and any color change was recorded. *E. coli* was used as a positive control. The isolate was also tested for the presence of the enzyme caseinase by using a casein agar. Plates were incubated at 30°C and casein hydrolytic halos were observed after 48 hours. *Pseudomonas aeruginosa* was used as a positive control. The ability of the isolate to hydrolyze starch was determined using a starch agar. Plates were incubated at 30°C for 24 hours. After incubation, the plates were flooded with an iodine reagent and clear halos would indicate a positive result. *Staphylococcus epidermidis* was used as a positive control. DNase activity was determined by using a DNase test agar. Plates were incubated at 30°C for 24 hours. After incubation, the plates were flooded with 1N hydrochloric acid. The acid was allowed to absorb into the media for two minutes. After diffusion, the presence of any clear halos around the microbial growth was recorded. *S. aureus* was used as a positive control. Lastly, the isolate's ability to reduce nitrate was determined by using nitrate broth containing a Durham tube. The inoculated tubes were stored at 30°C for 24 hours. After 24 hours any presence of bubbles in the Durham tube were recorded. If bubbles were not present, nitrate reagent A (alpha-naphthylamine dissolved in acetic acid), nitrate reagent B (sulfanilic acid dissolved in acetic acid), and zinc powder were added and any color change was recorded. *E. coli* was used as a positive control. All metabolic tests and media were employed as described in from Leboffe & Pierce (2010).

2.5: Antibiotic Resistance

Antibiotic resistance to seven different antibiotics was tested. Penicillin (10units), tetracycline (30µg), chloramphenicol (30µg), clindamycin (2µg), nalidixic acid (30µg), nitrofurantoin (300µg), and colistin (10µg) saturated discs were used (Becton Dickinson). A

lawn of KF13 was created on an R2A agar plate and the antibiotic disc was placed on the plate immediately after the lawn was spread. All antibiotics were tested in triplicate on the sample plate. Plates were incubated at 30°C for 24 hours and any inhibition of growth was observed.

2.6: Additional Sequencing

The culture was also sent to GeneWiz Labs (South Plainfield, New Jersey) in order obtain the sequence for the entire 16S rRNA gene. Finch TV (Geospiza, 2012) and the RDP website (Altschul et al. 1997) were used to analyze and identify the sequence DNA.

The isolate was sent to MR DNA labs (Shallowater, Texas) to obtain the full genome sequence. They prepared the library using the Nextera DNA Sample preparation kit and protocol (Illumina; MR DNA labs, personal communication). Fifty nanograms of DNA was used to prepare the library. DNA concentrations were determined using Qubit® dsDNA HS Assay Kit (Life Technologies). A limited five cycle PCR was performed. The final DNA concentrations for the isolate and library (184.40 ng/µl, 12.70 ng/µl) were measured using Qubit® dsDNA HS Assay Kit (Life Technologies) and the average library size (669 bp) was determined using Agilent 2100 Bioanalyzer (Agilent Technologies). The libraries were pooled in equimolar ratios of 2nM, and 10pM of the library pool was clustered using the cBot (Illumina) and sequenced paired end for 500 cycles using the HiSeq 2500 system (Illumina). MR DNA lab annotated all of the genes for their potential function using BaseSpace analysis tools from Illumina. . MR DNA lab also provided amino acid sequences for all of the predicted proteins and KF13 proteomic data was compared to data from a study conducted by Adeolu et al., 2016. Clustal Omega (Sievers et al., 2011) was used to align KF13 protein sequences to reference protein sequences to determine if conserved signature insertion/deletions were present in KF13. The program OneCodex (https://www.onecodex.com) was used to compare the whole genome data of the isolate to the

whole genome data of its closest relatives. The program Geneious R10 (http://www.geneious.com, Kearse et al., 2012) was used to align all the contigs together to determine if the whole genome had been sequenced.

To confirm the isolate's identity and taxonomic placement the isolate was also sent to Midi Labs (Newark, Delaware) to perform three additional tests: 16S rDNA sequencing, MALDI-TOF mass spectrophotometry, and fatty acid analysis. MALDI-TOF is an ionization technique that creates ions from large protein molecules, that can be separated by mass which can be used to identify or taxonomically place an organism (MIDI Labs, midilabs.com). Fatty acid analysis compared the fatty acid composition of KF13 to the fatty acid compositions of known organisms and 16S rDNA sequencing was used as a control that the correct culture was being analyzed.

2.7: Phylogenetics

2.7.1: Preliminary Phylogenetic Tree

A preliminary tree was constructed to determine which species and genera should be included in the final phylogeny. The 20 most closely related species to KF13 based on the full 16S rDNA were used to construct the preliminary tree (Table 7). *E. coli* was included as the tree outgroup. The top 20 matches, *E. coli*, and KF13 were aligned using Clustal Omega (Sievers et al., 2011). Preliminary phylogenetic trees were constructed by parsimony methods using the PAUP* 4.0 program (Swafford, 2002). A hundred heuristic bootstrap replications were used to assess the branch stability of the parsimony tree.

2.7.2: Expanded Phylogenetic Tree

The tree created in the preliminary phylogenetic analysis was used to determine what other species needed to be included in the final tree. Taxa that formed a polytomy with the

isolate in Figure 5 were used to find additional species. RDP SeqMatch was used to find closely related species to the taxa that formed a polytomy with the isolate. The additional species can be seen in Table 12. These additional species along with the original 20 and *E. coli* as an outgroup were used to create the final expanded phylogenetic tree. Clustal Omega was used to align KF13, all 36 related species, and *E. coli* (Sievers et al., 2011). Phylogenetic trees were constructed by parsimony methods using the PAUP* 4.0 program (Swafford, 2002). A hundred heuristic bootstrap replications were used to assess the branch stability of the parsimony tree.

CHAPTER THREE: RESULTS

3.1: Partial 16S rDNA Sequencing

RDP (Cole et al., 2014) searches of the 15 isolates using SeqMatch can be seen in Table 1. KF13 was chosen to continue with morphological testing, growth requirements, and metabolic capabilities testing because the sequence obtained was unambiguous in FinchTV, and the highest RDP species match was below 95%. RDP and BLAST (Altschul et al., 1997) searches both determined that the sequence for KF13 was not represented in their databases. RDP classifier revealed that KF13 was in the *Bacteria* domain, Proteobacteria phylum, Gammaproteobacteria class, Enterobacteriales order, and *Enterobacteriaceae* family with 100% confidence; however, it determined that KF13 was related to the genus *Serratia* with only 47% confidence. RDP SeqMatch revealed the highest species match was *Serratia fonticola* at 85.2% (Table 2).

3.2: Morphological and Growth Characteristics

KF13 was white in color and translucent with CFUs less than 1 mm in diameter (Figure 1). The cells were <1μm-1μm in length. The negative stain of KF13 showed coccobacilli cells and it was Gram negative (Figure 2). The growth characteristics yielded the following results: normal growth on the 4˚C, 25˚C, and 30˚C plates. No growth occurred on the 50˚C plate and only minimal growth occurred on the 37˚C plate. Growth occurred on pH 5, pH 7, and pH 9 plates. No growth occurred on the pH 3 plate. Salinity plates resulted in growth at 1%, 3%, and 5%, with variable growth on the 7% plate. No growth occurred on the 10% and 15% salinity plates. GasPak™ testing showed growth in the absence of oxygen. Growth data can be seen in Table 3. Motility testing was positive and flagella staining showed multiple flagella on each cell (Figure 3).

3.3: Metabolic Capabilities

Metabolic testing yielded the following results: catalase was positive and oxidase activity was negative and the EnteroPluri tubes yielded positive results for glucose fermentation and gas production, citrate utilization, acetoin production (Voges-Proskauer test), and slight positives in lactose and arabinose fermentation (n=3). All other tests in the EnteroPluri tubes were negative (Table 4). Gelatin liquefaction was negative while variably positive responses were observed for mannitol fermentation, lipase activity, and indole formation. Casein hydrolysis, DNase activity, and starch hydrolysis were not detected and nitrate but not nitrite was reduced. A summary of the metabolic test results can be seen in Table 5. These metabolic capabilities results were compared with 10 of the closest relatives to KF13 based on full 16S rRNA gene data (Table 6).

3.4: Antibiotic Resistance

Tetracycline, chloramphenicol, and nalidixic acid produced zones of inhibition for KF13. Colistin produced small zones of inhibition, approximately 1mm away from the disc. The isolate showed resistance to penicillin, clindamycin, and nitrofurantoin.

3.5: Additional Sequencing

RDP (Cole et al., 2014) and BLAST (Altschul et al., 1997) searches for the full 16S rDNA both determined that the sequence for KF13 was not represented in these databases. RDP Classifier revealed that KF13 was in the *Enterobacteriaceae* family with 100% confidence; however, it determined that KF13 was related to the genus *Hafnia* with only 81% confidence. RDP SeqMatch revealed the isolate's closest relatives to be species in the genera *Hafnia, Rahnella, Ewingella, and Serratia* (Table 7).

OneCodex comparisons of the isolate's whole genome to its closest relatives revealed the isolate to be closely related to *Rahnella* and *Hafnia* species (Figure 4). Whole genome sequencing resulted in a genome size of 5,289,934 bp with a 52.5% G-C content. There were 13

contigs ranging in size from approximately 3,835 nucleotides to 1,621,288 nucleotides. The annotated genes of the whole genome only show 18 out of the 20 tRNA molecules were present, suggesting that the genome sequence is incomplete. There were approximately 4,869 genes in the genome and 16% of those genes did not have a predicted function. Five 16S rRNA genes, 49 genes of viral origin, and 76 genes with putative pathogenic functions were also detected (Table 8). The protein sequence comparisons between KF13 and reference sequences based on a study conducted by Abeolu et al., 2016 can be seen in Figures 7 and 8.

The 16S rRNA gene sequence obtained by Midi Labs (Newark, Delaware) was aligned to the full 16S rRNA gene sequence of KF13 using Clustal Omega (Sievers et al., 2011) which resulted in a 100% alignment. Midi Labs determined KF13's closest matches to be *Klebsiella pneumoniae, Raoultella planticola,* and *Kluyvera intermedia* (Table 9). The MALDI-TOF data shows low matches to the genera *Serratia* and *Ewingella* (Table 10). The fatty acid analysis showed matches to the genera *Serratia, Rahnella, Pantoea,* and *Hafnia* (Table 11).

3.6: Phylogenetics

3.6.1: Preliminary Phylogenetic Tree

The initial phylogenetic analysis resulted in a strict consensus tree of 61 of the most parsimonious trees with a consistency index (CI) excluding uninformative characters of 0.4399 and a retention index (RI) of 0.7213. A preliminary parsimony tree can be seen in Figure 5. Based on these results, additional sequences (Table 12) were added for further analyses.

3.6.2: Expanded Phylogenetic Tree

The full phylogenetic analysis resulted in a strict consensus tree of 688 of the most parsimonious trees with a consistency index (CI) excluding uninformative characters of 0.3605 and a retention index (RI) of 0.7230. An expanded final parsimony tree can be seen in Figure 6.

CHAPTER FOUR: DISCUSSION

The results of metabolic, phenotypic, and genetic testing suggest that KF13 is a new genus of bacteria in the *Enterobacteriaceae* family. The isolate has many characteristics that corroborate it is in *Enterobacteriaceae.* Most members of this family are Gram-negative, bacillus in shape, facultative anaerobes, motile, can grow from 4˚C to 30˚C, are catalase positive, and oxidase negative (Brenner & Farmer, 2001). Genetic data from the 16S rRNA gene and the whole genome confirm the isolate is in the *Enterobacteriaceae* family. However, the isolate could not be placed into a specific genus.

The RDP Classifier results, based on the full 16S rRNA sequence, placed the isolate into the genus *Hafnia* with only 81% confidence. When metabolic and phenotypic data of the isolate was compared with that of two species of *Hafnia, H. paralevi* and *H. psychrotolerans*, only 66% and 77% of the results matched, respectively (Table 6). This phenotypic information suggested the isolate could possibly be a new species in the genus *Hafnia*, but most likely is a new genus due also to the very low genotypic matches. The other high matches from RDP SeqMatch, based on the full 16S rRNA gene sequence, were in the genera *Rahnella, Serratia, Yersinia,* and *Ewingella* (Table 7). The metabolic and other phenotypic comparisons between these genera and the isolate resulted in less than 66% of the traits matching. Based on all of the metabolic, phenotypic, and 16S rDNA results the isolate is most closely related to *Hafnia*, but it is not likely in this genus. The 16S rRNA gene sequence, MALDI-TOF, and fatty acid data generated by Midi Labs confirms that KF13 is most likely a new genus. All three of these analyses could not confidently place KF13 in a known genus. The best conclusion Midi Labs could propose is that KF13 is an enteric species. The phylogenetic data also shows the isolate to most likely be a new

genus of *Enterobacteriaceae*. The isolate formed a polytomy 100% of the time with *Hafnia psychrotolerans.* This suggests that KF13 is closely related to the genus *Hafnia*.

The whole genome data did not provide as much clear evidence as to which genus the isolate is most closely related to but it did show more evidence that the isolate is a new genus. Figure 4 shows a OneCodex comparison of the whole genome of the isolate to whole genomes of closely related genera: *Serratia, Rahnella, Hafnia*, and *E. coli* as an outgroup. The reference genomes appear as a single color each in the graph indicating each was placed in one genus each with 100% confidence. However, KF13 has two different colors in the graphical representation in Figure 4 indicating that the OneCodex program could not confidently place the genome in a single extant genus. The two colors that the isolate is being represented by belong to *Rahnella* and *Serratia.* The program found a large portion of similarities in the genomes of both *Rahnella* and *Serratia* to the genome of the isolate.

From the genome data, the isolate's G-C content was able to be estimated as 52.5%. The *Enterobacteriaceae* family has a G-C content between 38% and 60% (Brenner and Farmer, 2001). The isolate falls within the range typical for the *Enterobacteriaceae* family providing more evidence that the isolate does belong to this family. Species in the genus *Hafnia* have a G-C content of 48%-49%, the genus *Rahnella* have a G-C content of 51%-56%, and the genus *Serratia* have a G-C content of 52%-60%. Based on the G-C content the isolate does not fall in the genus *Hafnia* but does fit into the genera *Rahnella* or *Serratia.* The OneCodex and G-C content data suggest the isolate's closest relatives are *Rahnella* and *Serratia.*

The whole genome data also provided evidence as to some putative interactions the isolate may have with its environment. There are 76 genes that have putative pathogenic properties and suggest that the isolate could be at least an opportunistic pathogen. There were

several genes for invasin proteins which are required to initiate the internalization of a bacterium into a host (Dong et al., 2013). A recent study performed by Dong et al. (2013) investigated the role invasins play in the virulence of *Edwardsiella tarda.* They discovered that the invasin found in *E. tarda* plays an important role in hemolytic activity, biofilm formation, and virulence. The invasins may not be the only contributor to invasion of a host cell but they do play a large role in promoting virulence during infection (Dong et al., 2013). The presence of the invasin genes as well as the other potential genes involved in pathogenicity such as type IV secretion system proteins, adherence and invasion outer membrane proteins, and cell division inhibitor proteins gives evidence that the isolate is a pathogen. The isolate did not grow well at 37˚C so it is unlikely that it could infect most mammals, but fish and most insects regulate their body temperature based on their environment. Therefore, the isolate could be a fish or insect pathogen with a body temperature similar to that of the water or ambient temperatures. There are a wide range of *Enterobacteriaceae* species that are pathogenic to fish and insects. *Edwardsiella* species are well known to cause disease in eel, salmon, trout, and catfish species commonly found in the Pacific Northwest (Brenner and Farmer, 2001). *Yersinia ruckerii* is a fish pathogen that causes red mouth disease in salmonid fish species (Kumar et al., 2015). Many of these fish species are found in the streams and rivers of GSMNP. There are also several *Enterobacteriaceae* genera that cause disease in insects. For example, all *Photorhabdus* species are insect pathogens that form a mutualistic relationship with nematodes in which the bacteria kill the insect so the nematode can use the insect body for nutrients and the bacteria is able to colonize the nematode intestines (Mulley et al., 2015).

The whole genome data also resulted in 42 genes for fimbriae, 19 genes for pili, 52 genes for drug resistance, and 27 to counter toxic metals (Table 8). Fimbriae are appendages found on

bacterial cells that are thinner than flagella (Madigan et al., 2015). They are used to adhere to surfaces such as other bacterial cells, animal and plant cells, or inanimate objects. Pili are also thin hair like appendages found on bacterial cells. They are used to transfer genetic material between other bacterial cells or to attach to surfaces. Fimbriae and pili are not strictly associated with pathogenicity but they are often used to attach to or infect cells, so their presence gives more evidence that the isolate could be a pathogen.

The presence of drug resistance genes also gives more evidence that the isolate could be a pathogen as well as how the bacterium is interacting with its environment. The drug resistance proteins have been found to play a role in host-pathogen interactions and in bacterial pathogenesis (Sun et al., 2014). The drug efflux proteins have the ability to extrude many defensive compounds the host produces. Deletion of any of these drug protein domains has been found to reduce virulence. The drug resistance proteins also help bacteria invade the host cells as well as create biofilms. Inhibition of the efflux pumps have been shown to reduce the bacterium's ability to invade cells and to create biofilms. Drug resistance genes can suggest pathogenicity or they can explain how the isolate is interacting with other bacteria. Bacterial species share closely related niches and must compete for space and resources. They have developed antimicrobial agents to kill or repel each other. For example, *Pseudomonas aeruginosa* can produce two different molecules that stimulate the dispersal of other species from established biofilms (Hibbing et al., 2010). The presence of drug resistance genes could indicate that the isolate has the ability to ward off competing bacteria and perhaps fungi.

The whole genome data also showed 10 genes associated with chitin or lignin breakdown (Table 8). Chitin is found in the exoskeletons of insects and cell walls of fungi while lignin is found in cell walls of plants. The presence of these genes suggests that KF13 may play a role in

the breakdown of these components in its environment. KF13 may help nutrient cycling by breaking down chitin as a carbon and nitrogen source and breaking down lignin as a carbon source for growth. Chitinase has also been found to be a fungicide. A study performed by Duzhak et al. (2012) discovered that chitinases in *Serratia marcescens* had antifungal activity that cleaved chitin molecules in the fungal cell wall which suppressed the fungus's vital functions. This study shows that the presence of chitinases could indicate KF13 is a fungal pathogen, or more likely can produce an antimicrobial agent.

KF13 is likely a new genus of *Enterobacteriaceae,* closely related to *Hafnia.* There are only 3 named *Hafnia* species and they can be found in a broad range of environments. They have been found in humans and animals, including birds, and in natural environments such as soil, sewage, and water (Brenner and Farmer, 2001). *Hafnia* species are categorized as opportunistic pathogens because they have been known to cause illness but they do not always require a host. *Hafnia* have been found in caged and wild birds but none of them appeared sick (Brenner and Farmer, 2001). However, there are many illnesses that have been caused or have been worsened by the presence of *Hafnia*. Respiratory tract infection, septicemia, meningitis, abscesses, urinary tract infections, and wound infections have all be associated with *Hafnia* (Brenner and Farmer, 2001). These species are opportunistic pathogens that will take advantage if the patient has an underlying illness or predisposing factors.

The *Enterobacteriaceae* family is very large and diverse with 44 different genera. Finding a place for KF13 in this family has proven difficult because of how vast and diverse it is. The family is so large it has been suggested that it could be split into seven different families (Adeolu et al., 2016). The 16S rDNA method of classification might not be sufficient for this family so a recent study has reconstructed this family into seven new families using 1548 core

proteins, 53 ribosomal proteins, and four multilocus sequence analysis proteins based on the whole genome sequencing (Adeolu et al., 2016). They also discovered that there are conserved signature insertions/deletions (CSIs) that are distinct for each newly proposed clade. CSIs were hypothetically developed by a common ancestor; those insertions or deletions were then carried out through its descendants. The presence of the same or very similar CSIs suggests there is a group of related organisms. These protein sequences are unique enough to create seven distinct clades which have been proposed as the families *Enterobacteriaceae, Erwiniaceae, Pectobacteriaceae, Yersiniaceae, Hafniaceae, Morganellaceae,* and *Budviciaceae* (Adeolu et al., 2016). Based on the growth, metabolic, and genetic data it is predicted that KF13 would be in the new proposed *Hafniaceae* family. However, the proteomic data shown in Figures 7 and 8 suggest KF13 does not fit into this family. Adeolu et al. (2016) discovered four unique CSIs for the *Hafniaceae* and three unique CSIs for the *Yersiniaceae* family. KF13 only has one of the CSIs for *Hafniaceae* but has all the CSIs for *Yersiniaceae.* The newly proposed *Yersiniaceae* family is comprised of the genera *Yersinia, Serratia, Rahnella, Chania, Ewingella*, and *Rouxiella* (Adeolu et al., 2016). Figure 8 shows that KF13 has identical insertions to *Rahnella* species. For example, the TetR transcriptional regulator protein has an insertion between amino acids 43-89 and between amino acids 82-123. *Serratia* and *Yersinia* species have an F and a Y insertion, respectively. *Rahnella* species and KF13 have an L and a D insertion, respectively. The presence of these CSIs suggest that KF13 is most closely related to the *Yersiniaceae* family, more specifically the genus *Rahnella*. It cannot be ruled out that KF13 could even represent a new family in the order Enterobacterales. More work will have to be completed to determine if this is the cases

Bacteria are an important resource for humans in the form of biotechnology, food production, antibiotic sources, and advancing knowledge of ecology and evolution. Without knowledge of bacteria many fields of science and agriculture would not be as advanced as they are today. Bacteria also play a very large role in their ecosystems. They act as nutrient cyclers and nitrogen fixers, have positive and negative relationships with other organism, and they created and help sustain the atmosphere needed for most life. Understanding the diversity of bacteria is important for humans and the ecosystem. The ATBI is attempting to understanding the diversity of all life in Great Smoky Mountains National Park to assist with conservation. The discovery of KF13 as a novel genus contributes to the ATBI as well as the basic understanding of the *Enterobacteriaceae* family.

Over 1700 bacterial samples from GSMNP have been sequenced and identified and only 1.69% of them are in the *Enterobacteriaceae* family. Finding organisms from this family in the park is not very common. About half of the *Enterobacteriaceae* came from water samples taken from Kephart Prong. Eight of the *Enterobacteriaceae* species discovered were in the genus *Serratia* but none have previously been found from *Hafnia* or *Rahnella*. The discovery of KF13 suggests these microbes are either rare in GSMNP or they are difficult to culture or clone. The majority of the *Enterobacteriaceae* being found in water suggests they might play an important role in streams or rivers of the Park and more work is needed to understand their roles better.

CHAPTER FIVE: CONCLUSIONS

The isolate KF13 is most likely a new genus of bacteria in the *Enterobacteriaceae* family. The low matches to known species in RDP Classifer and RDP SeqMatch as well as the OneCodex genome analysis data give strong evidence that the isolate is a novel genus. The phenotypic and phylogenetic data suggest that the isolate is most closely related to the genus *Hafnia.* The OneCodex and proteomic data suggest it is most closely related to the genus *Rahnella.* More testing such as sequencing the remainder of the genome, multi-locus sequence alignments (MLSA), and full genome phylogenetics and perhaps proteomics are needed to discover a final taxonomic placement for KF13.

REFERENCES

Adeolu, M., Alnajar, S., Naushad, S., and Gupta, S.R. 2016. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, *66*(12):5575-5599. doi:10.1099/ijsem.0.001485.

Altschul, S.F., Madden, T.L., Schaffer, 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.

Brenner, D. J., and Farmer J.J. 2001. Order XIII. "*Enterobacteriales*." In Brenner, D.J., Krieg, N.R., Staley, J.T., Garrity, G. M. (Ed.), Bergey's Manual of Systematic Bacteriology. Volume 2, Part B: The *Gammaproteobacteria*, pp. 587-850. Williams & Wilkins. New York, NY.

Cole, J. R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research,* 42(Database issue): D633-D642. doi: 10.1093/nar/gkt1244 [PMID: 24288368].

Discover Life in America. All Taxa Biodiversity Inventory (ATBI). 2012. http://www.dlia.org/atbi/. Accessed December 8, 2013.

Dong, X., Fan, X., Wang, B., Shi, X., and Zhang, X. 2013. Invasin of *Edwardsiella tarda* is essential for its haemolytic activity, biofilm formation and virulence towards fish. *Journal of Applied Microbiology*, *115*(1):12-19.

Duzhak, A., Panfilova, Z., Duzhak, T., Vasyunina, E., and Shternshis, M. 2012. Role of prodigiosin and chitinases in antagonistic activity of the bacterium *Serratia marcescens* against the fungus *Didymella applanata*. *Biochemistry (00062979)*, *77*(8):910-916. doi:10.1134/S0006297912080123.

Gu, Z., Liu, Y., Shen, L., Liu, X., Xiao, N., Jiao, N., Liu, H., Zhou, Y., and Zhang, S. 2015. *Hafnia psychrotolerans* sp. nov., isolated from lake water. *International Journal of Systematic and Evolutionary Microbiology*, *65*(Pt 3):971-974. doi:10.1099/ijs.0.000049.

Hibbing, M.E., Fuqua, C., Parsek, M.P., and Peterson, S.S. 2010. Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, *8*(1):15-25.

Huys, G., Cnockaert, M., Abbott, S.L., Janda, J.M., and Vandamme, P. 2010. *Hafnia paralvei* sp. nov., formerly known as *Hafnia alvei* hybridization group 2. *International Journal of Systematic and Evolutionary Microbiology*, *60*(Pt 8):1725-1728. doi:10.1099/ijs.0.018606-0

Ishii, K., Fukui, M. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. 2001. *Applied and Environmental Microbiology*, 67(8):3753-3755.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., and Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12):1647-1649.

Kumar, G., Menanteau-Ledouble, S., Saleh, M., and El-Matbouli, M. 2015. *Yersinia ruckeri*, the causative agent of enteric redmouth disease in fish. *Veterinary Research*, *46:*103. doi:10.1186/s13567-015-0238-4.

Leboffe, M.J., & Pierce, B.E. 2010. *Microbiology: Laboratory Theory and Application.* (3rd ed.). Morton Publication Company. Englewood, CO.

Le Flèche-Matéos, A., Levast, M., Lomprez, F., Arnoux, Y., Andonian, C., Perraud, M., Vincent, V., Ar Gouilh, M., Thilberge, J.M., Vandenbogaert, M., Diancourt, L., Caro, V., Burguiére, A.M., and Manuguerra, J. 2015. *Rouxiella chamberiensis* gen. nov., sp. nov., a member of the family *Enterobacteriaceae* isolated from parenteral nutrition bags. *International Journal of Systematic and Evolutionary Microbiology*, *65*(Pt 6):1812-1818. doi:10.1099/ijs.0.000179.

Madigan, M., Martinko, J., Bender, K., Buckley, D., Stahl, D. 2015. Brock Biology of Microorganisms: 14th Edition. Pearson Prentice Hall. Upper Saddle River, NJ.

Mulley, G., Beeton, M.L., Wilkinson, P., Vlisidou, I., Ockendon-Powell, N., Hapeshi, A., Tobias, N.J., Nollmann, F.I., Bode, H.B., van den Elsen, J., French-Constant, R.H., and Waterfield, N.R. 2015. From Insect to Man: *Photorhabdus* Sheds Light on the Emergence of Human Pathogenicity. *Plos One*, *10*(12), e0144937. doi:10.1371/journal.pone.

Nichols, B.J. and Langdon, K.R. 2007. The Smokies All Taxa Biodiversity Inventory: History and progress. *Southeastern Naturalist,* 6(1):27-34.

O'Connell, S., York, E.A., Collins, M.B., Rosbach, D.T., Black, K.R., and Haney, W.B. 2007. An initial inventory of bacteria found within the soils and waters of Great Smoky Mountain National Park. *Southeastern Naturalist*, 6(1):57-72.

Posada, D., and Crandall, K.A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics (Oxford, England)*, *14*(9):817-818.

Sievers, F., Wilm, A., Dineen, D.G., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., and Higgins, D. 2011. Fast, scalable

generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology Article*, 7(1):539. *doi:10.1038/msb.2011.75*.

Sun, J., Deng, Z., and Yan, A. 2014. Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. *Biochemical and Biophysical Research Communications*, *453*(2):254-267. doi:10.1016/j.bbrc.2014.05.

Swofford, D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

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 and Environmental Microbiology,* 73(16):5261-5267. doi: 10.1128/AEM.00062-07 [PMID: 17586664].

Youssef, N., Couger, M., McCully, A., Criado, A., and Elshahed, M. 2015. Assessing the global phylum level diversity within the bacterial domain: A review. *Journal of Advanced Research*, 6(3):269-282.

TABLES AND FIGURES

Table 1: DNA sequence matches of the partial 16S rRNA gene for the 15 isolates collected from Kephart Prong, Great Smoky Mountains National Park, using the Ribosomal Database Project (RDP) software program SeqMatch including percentages that indicate the extent of the match to the known isolate represented in the RDP database (accession number shown beside each record). The isolate KF13 was chosen for further analysis.

	RDP: Type Strains/ Isolates	**% Match to Sample**
KF3	*Mucilaginibacter dorajii;* GU139697	85.2%
KF4	*Pseudomonas psychrophila*; AB041885	95.9%
KF5	*Mucilaginibacter angelicae*; HM627214 *Mucilaginibacter polysacchareus*; HM748604	90.1%
KF6	*Rhodoferax saidenbachensis*; FJ755906	85.1%
KF7	*Herminiimonas arsenicoxydans*; AY728038	93.6%
KF8	*Mucilaginibacter flavus*; HQ449707	94.5%
KF9	*Flavobacterium succinicans*; DSM 4002;	90.3%
KF10	*Paucibacter toxinivorans*; AY515390	92.3%
KF11	*Undibacterium parvum*; DSM 23061	98.0%
KF12	*Pseudomonas poae*; DSM 14936 *Pseudomonas trivialis*; DSM 14937 *Pseudomonas extremorientalis*; AF405328 *Pseudomonas simiae*; AJ936933	92.5%
KF13	***Serratia fonticola;* FJ790328**	**85.2%**
KF14	*Flavobacterium fluvii;* EU109724	92.8%
KF15	*Janthinobacterium agaricidamnosum*; Y08845	97.1%
KF16	*Flavobacterium aquidurense*; AM177392	89.5%
KF17	*Herminiimonas arsenicoxydans*; AY728038	*27.9%
KF18	*Pseudomonas congelans*; DSM 14939	94.2%

* very poor sequence read

Table 2: DNA sequence matches for the partial 16S rRNA gene for the bacterial isolate (KF13) obtained from Kephart Prong, Great Smoky Mountains National Park, North Carolina, using the Ribosomal Database Project (RDP) software program SeqMatch including percentages that indicate the extent of match of the KF13 isolate to each isolate or clone represented in the RDP database (accession numbers included beside each record).

RDP: Type Strains/ Isolates	**% Match to KF13**
Serratia fonticola; FJ790328	85.2%
Serratia fonticola; DSM 4476	85.0%
Hafnia psychrotolerans; DJC1-1	84.4%
Yersinia entomophaga; DQ400782	84.2%
Yersinia nurmii; FJ717338	84.2%
Rahnella aquatilis; DSM 4594	83.2%
Plesiomonas shigelloides; NCIMB9242	82.6%
Serratia myotis; KJ739884	82.4%
Serratia grimesii; DSM 30063	82.4%
Serratia quinivorans; DSM 4597	82.4%
Serratia liquefaciens; AJ306725	82.4%
Serratia proteamaculans; DSM 4543	82.4%
Yersinia enterocolitica; ATCC 9610	82.2%
Yersinia enterocolitica; FR729477	82.2%
Yersinia enterocolitica; FR729477	82.2%
Yersinia plymuthica; DSM 4540	82.2%
Yersinia intermedia; ATCC 29909	82.0%
Yersinia mollaretii; ATCC 43969	82.0%
Yersinia ruckeri; ATCC 29473	82.0%
Yersinia bercovieri; ATCC 43970	82.0%

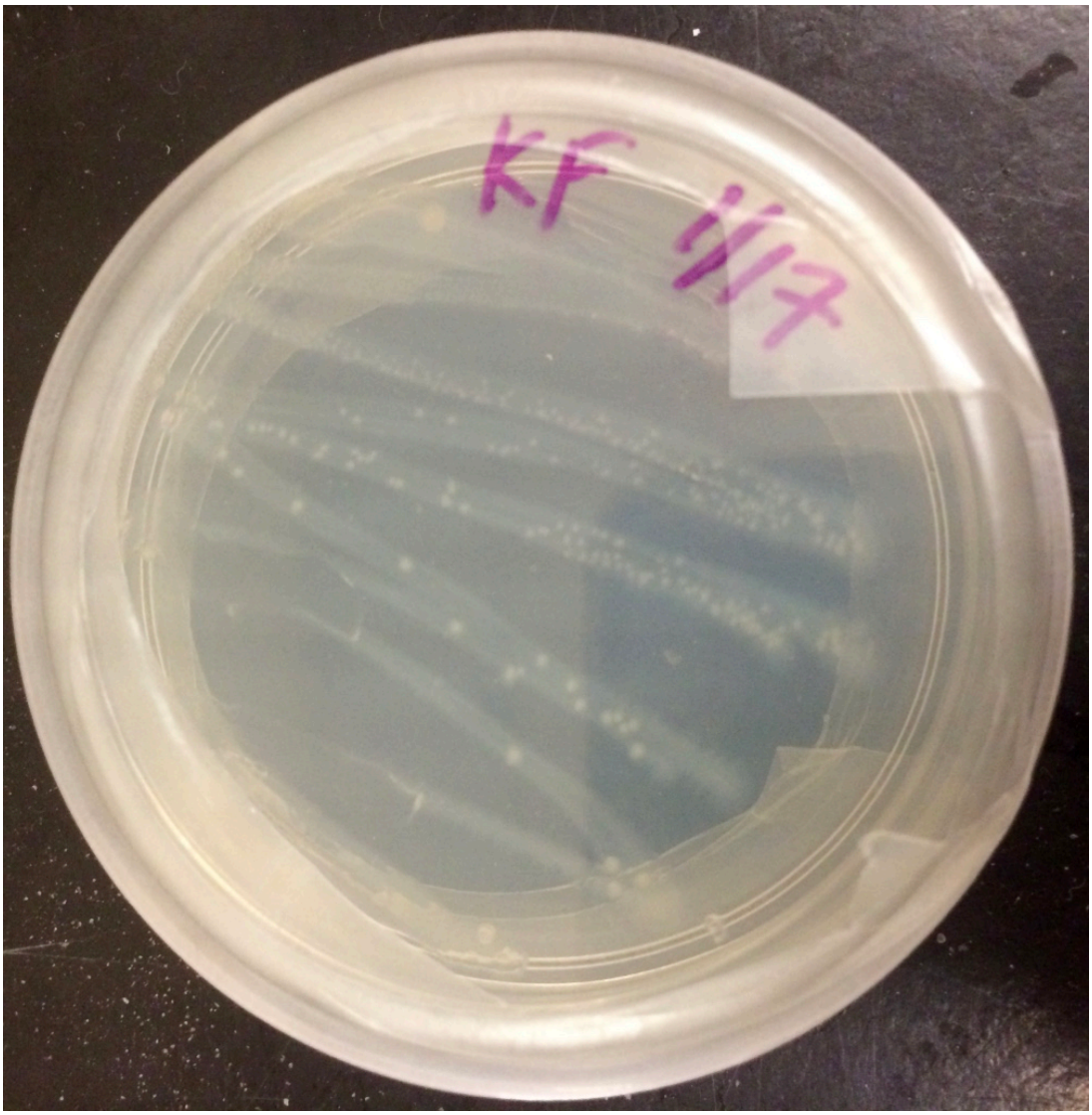

Figure 1: Isolate KF13 colony morphology. Creamy white to translucent with colonies 1-2mm in diameter.

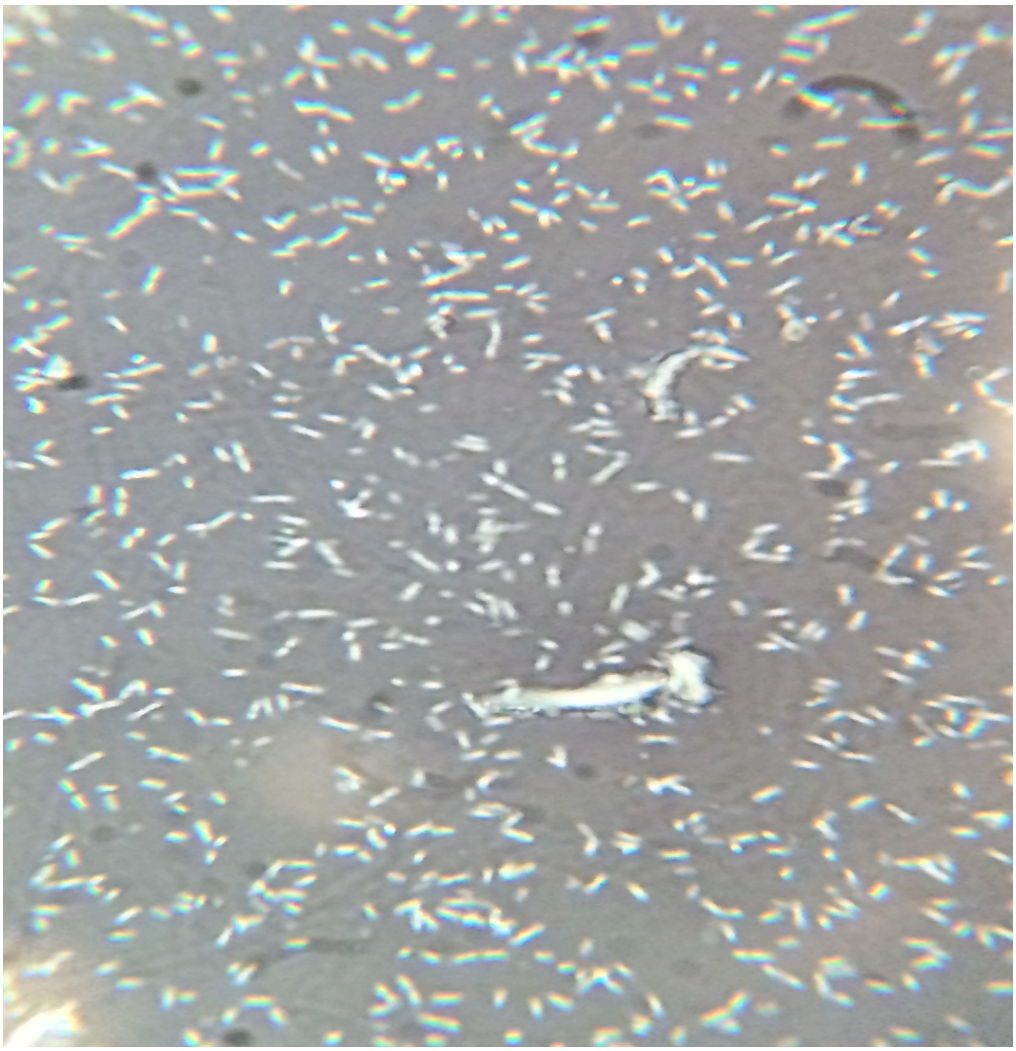

Figure 2: Isolate KF13 cell morphology. Negative staining of the isolate that shows coccobacillus cells less than 1µm in length using a light microscope 1000x oil immersion lens.

Table 3: Growth parameters for KF13 obtained from Kephart Prong, Great Smoky Mountains National Park, North Carolina, including ranges that permitted growth on R2A adjusted to various levels of NaCl, pH, and incubated at varying temperatures as well as requirement for O_2 (n=3).

Temperature	**Growth**
4ºC	+
25ºC	+
30ºC	+
37ºC	+*
50ºC	-
pH	**Growth**
3	-
5	+
7	+
9	+
NaCl	**Growth**
0%	+
1%	+
3%	+
5%	+
7%	+*
10%	-
15%	-
Oxygen	**Growth**
0%	+
20%	+

* Growth was not evident in all cases

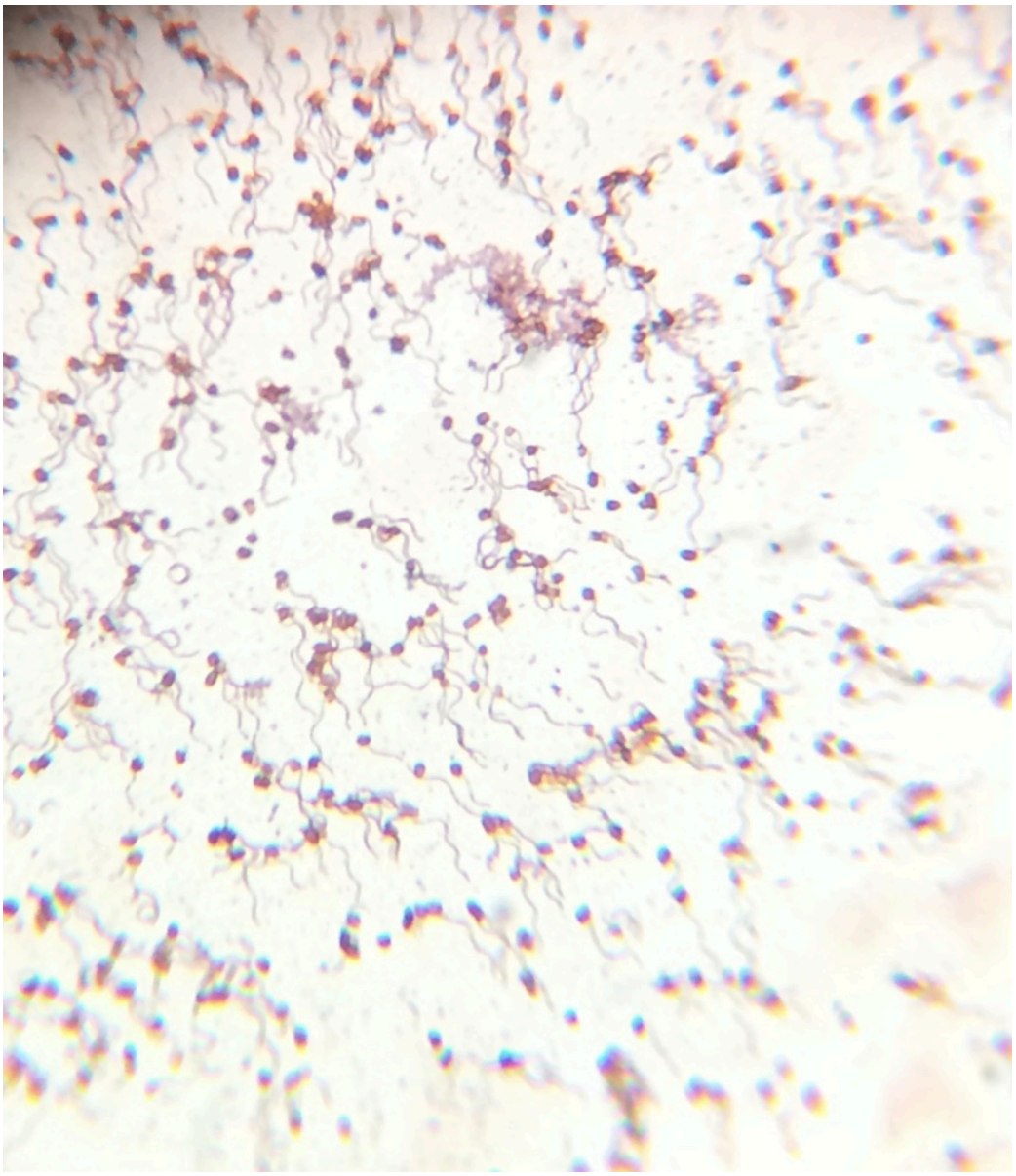

Figure 3: Flagella staining of KF13 using a light microscope at 1000x oil immersion lens.

Table 4: EnteroPluri Tube results for KF13, an isolate obtained from Kephart Prong, Great Smoky Mountains National Park, North Carolina indicating carbon and nitrogen sources utilized by the organism.

Test	**Biochemical Reactions**	**Result**
Glucose/ Gas	Glucose fermentation	+
	Gas production	+
Lysine	Lysine decarboxylation	-
Ornithine	Ornithine decarboxylation	-
H2S/Indole	Hydrogen sulfide production	-
	Indole production	-
Adonitol	Adonitol fermentation	-
Lactose	Lactose fermentation	Slight positive
Arabinose	Arabinose fermentation	Slight positive
Sorbitol	Sorbitol fermentation	-
Vogues-Proskauer	Acetoin production	+
Dulcitol/Phenylalanine	Dulcitol fermentation	-
	Phenylalanine deamination	-
Urea	Urea hydrolysis	-
Citrate	Citrate utilization	+

Table 5: A summary of some metabolic capabilities for KF13, an isolate obtained from Kephart Prong, Great Smoky Mountains National Park, North Carolina (v= variable result).

Test	**Result**
Catalase	+
Oxidase	-
Casein Hydrolysis	-
Starch Hydrolysis	-
Lipase Activity	v
Indole Production	Partial positive
Mannitol Fermentation	Partial positive
Nitrate Reduction	+
DNase Activity	-
Gelatin Hydrolysis	-

Table 6: Comparison of KF13's growth and metabolic capabilities to related species. Related species were determined by the closest matches to the 16S rRNA gene sequence using the Ribosomal Database Project (RDP) SeqMatch tool (d= differential result, nd= no data, v= variable result). *R.a.= Rahnella aquatilis; H.p.= Hafnia paralvei; H.ps. = Hafnia psychrotolerans; S.p.= Serratia plymuthica; S.g.= Serratia grimesii; Y.e.= Yersinia enterocolitica; E.a.= Ewingella americana; E.t.= Edwardsiella tarda*; C.b.= Coccobacillus; B.= Bacillus.

Test	KF13	*R.a.*	*H.p.*	*H.ps.*	*S.p.*	*S.g.*	*Y.e.*	*E.a.*	*E.t.*
Gram Stain	-	-	-	-	-	-	-	-	-
Cell Shape	C.b.	B.	B.	B.	B.	B.	C.b.	B.	B.
Growth at 4°C	+	+	+	+	+	+	+	+	+
Facultative anaerobe	+	+	+	+	+	+	+	+	+
Motility	+	d	+	+	+	+	+	+	+
Catalase	+	+	+	+	+	+	+	+	+
Oxidase	-	-	-	-	-	-	-	-	-
Casein Hydrolysis	-	nd	-	-	+	+	-	nd	nd
Starch Hydrolysis	-	nd	-	-	-	-	-	nd	nd
Lipase Activity	+	-	-	-	+	+	+	-	-
Indole Production	+	-	-	-	-	-	v	-	+
Mannitol Fermentation	-	nd	+	+	+	+	+	+	-
Nitrate Reduction	+	+	+	+	+	+	+	+	+
DNase Activity	-	-	-	-	+	+	-	-	-
Gelatin Hydrolysis	-	-	-	-	+	+	-	-	-
Glucose Fermentation/ Gas Production	+/+	+/+	+/+	+/+	+/+	+/+	+/-	+/-	+/+
Lysine Decarboxylation	-	-	+	-	-	+	-	nd	+
Ornithine Decarboxylation	-	-	+	-	-	+	+	-	+
H2S/Indole Production	-/-	-/-	-/-	-/-	-/-	-/-	-/v	-/-	+/+
Adonitol Fermentation	-		-	-	-	-	-	-	-
Lactose Fermentation	+	+	-	-	+	+	-	nd	-
Arabinose Fermentation	+	+	v	-	+	+	+	-	-
Sorbitol Fermentation	-	+	-	-	d	+	+	-	-
Vogues-Proskauer (Acetoin Production)	+	+	+	+	+	-	v	+	-
Dulcitol Fermentation/Phenylalanine Deamination	-/-	+/+	-/-	-/-	-/-	-/-	-/-	-/-	-
Urea Hydrolysis	-	-	-	-	-	-	+	-	-
Citrate Utilization	+	+	-	-	+	+	-	+	-

Table 7: DNA sequence matches for the full 16S rRNA gene for the bacterial isolate KF13 obtained from Kephart Prong, Great Smoky Mountains National Park, North Carolina, using the Ribosomal Database Project (RDP) software program SeqMatch including percentages that indicate the extent of the match of the isolate to each isolate or clone represented in the RDP database; accession numbers included. These sequences were used to construct the preliminary phylogenetic tree.

RDP: Type Strains/ Isolates	**% Match to KF13**
Hafnia psychrotolerans; DJC1-1	91.9%
Rahnella aquatilis; DSM 4594	91.8%
Ewingella Americana; JN175329	91.7%
Serratia plymuthica; DSM 454*0*	90.3%
Hafnia paralvei; ATCC 29927	89.9%
Yersinia nurmii; FJ717338	89.5%
Yersinia entomorphaga; DQ400782	89.4%
Yersinia enterocolitica; FR729477	89.4%
Yersinia enterocolitica; FR729477	89.4%
Yersinia enterocolitica; FR729477	89.4%
Yersinia enterocolitica; FR729477	89.4%
Yersinia enterocolitica; HF558392	89.4%
Yersinia enterocolitica; FR729477	89.4%
Yersinia enterocolitica; FR729477	89.4%
Yersinia enterocolitica; FR729477	89.4%
Yersinia enterocolitica; FR729477	89.4%
Yersinia enterocolitica; ATCC 9610	89.2%
Serratia grimesii; DSM 30063	89.1%
Yersinia intermedia; ATCC 29909	89.0%
Serratia fonticola; FJ790328	88.8%

Table 8: General genome features of KF13 isolated from Kephart Prong, Great Smoky Mountains National Park, North Carolina.

Genome Information	
Estimated Genome size (bp)	5,289,934
Estimated G+C content (%)	52.5%
Estimated Total number of genes	4,869
Total tRNA genes	97
16S rRNA genes	5 full, 2 partial
Genes with function prediction (%)	84%
Genes without function prediction (%)	16%
Genes of phage origin	49
Genes with pathogenicity implications	76
Genes for fimbriae	42
Genes for pili	19
Genes with drug resistance function	52
Genes to counter toxic metals	27
Genes for chitin and lignin breakdown	10

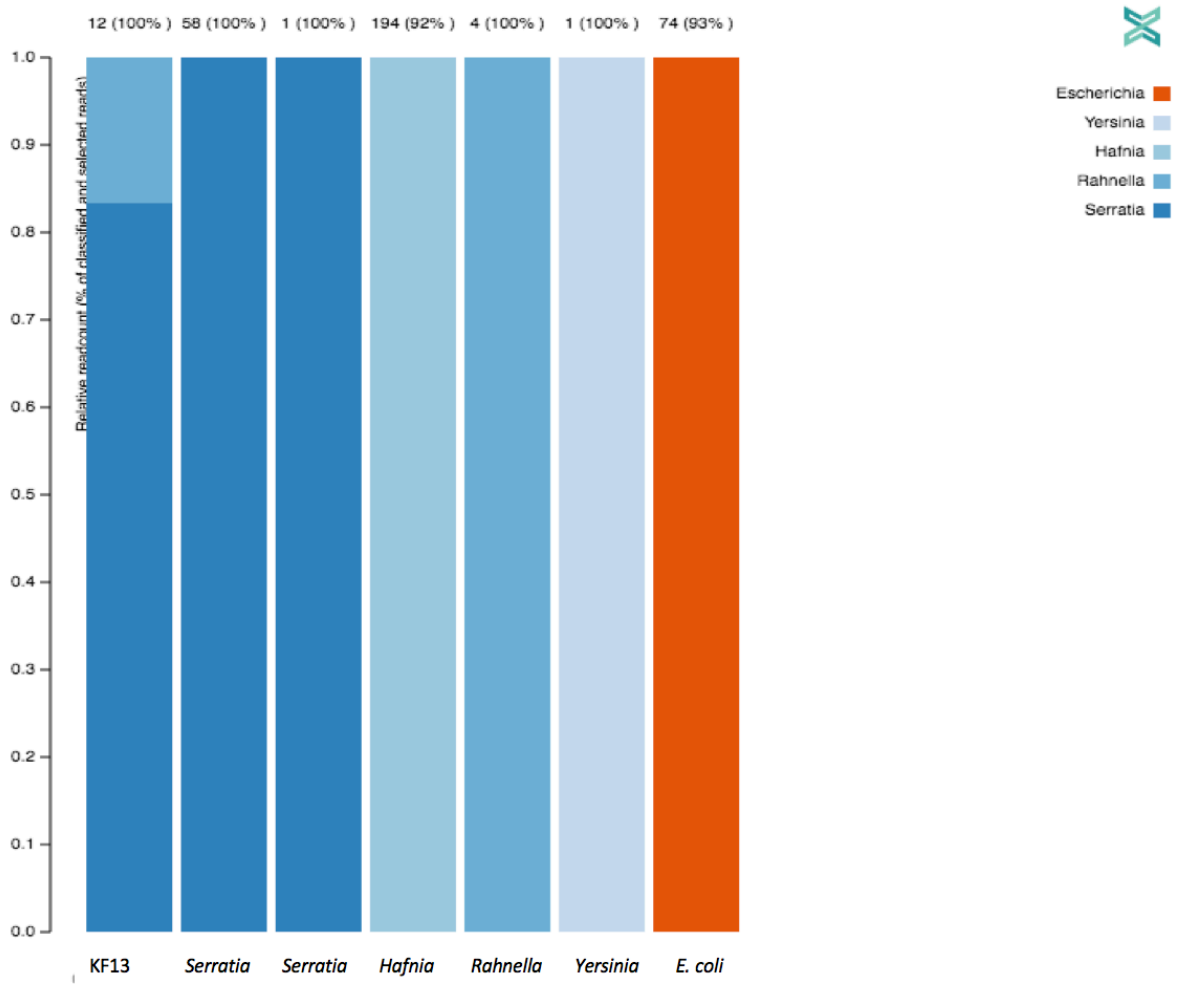

Figure 4: Comparison of the whole genomes of KF13 at the genus level to *Serratia, Hafnia, Rahnella, Yersinia,* and *E. coli* using the OneCodex software program showing that KF13 is most likely a new genus closely related to *Serratia* and *Rahnella.*

Table 9: Midi Labs 16S rRNA gene sequencing using the Sherlock® DNA database. Genetic relationships are expressed in percent genetic difference calcuated as the percentage of positions that differ when two sequences are aligned. Less than 3% difference indicates a species level match.

Midi Library Entry	**% Difference**
Klebsiella pneumoniae	3.60
Raoultella planticola	3.60
Kluyvera intermedia	3.60
Raoultella ornithinolytica	3.79
Obesumbacterium proteus	3.79
Klebsiella pneumoniae	3.79
Enterobacter aerogenes	3.88
Lelliottia amnigena	3.88
Hafnia alvei	3.88
Klebsiella pnumoniae	3.88

Table 10: MALDI-TOF mass spectorphotometer data from Midi Labs. A score value of 2.000-3.000 indicates a confidence level to the species, a score value of 1.700-1.9999 indicates a confidence level to the genus, and a score value below 1.699 indicates no match.

Organism	**Score**	**Source**
Serratia plymuthica	1.790	DSM 22907 DSM
Ewingella americana	1.741	19_QK 122 MHH
Serratia liquefaciens	1.715	M7545
Ewingella americana	1.701	RV412_A1_2010_01 LBK
Serratia plymuthica	1.668	DSM 30127 DSM
Serratia plymuthica	1.643	DSM 49 DSM
Serratia plymuthica	1.621	DSM 8572 DSM
Serratia liquefaciens	1.571	DSM 30065 DSM
Pseudomonas fragi	1.546	DSM 3456T HAM
Serratia liquefaciens	1.533	DSM 30125 DSM

Table 11: Fatty acid anaylsis from Midi Labs. A similarity index above 0.600 indicates a good species match, 0.400-0.600 may be a species match, lower than 0.400 or having several choices with similar values suggests the sample is not in the database.

Library	**Sim Index**	**Organism**
RTSBA6 6.21	0.817	*Serratia plymuthica*
	0.796	*Rahnella aquatilis*
	0.693	*Pantoea agglomerans*-GC subgroup C
	0.632	*Pantaea agglomerans*-GC subgroup B
	0.619	*Hafnia alvei*

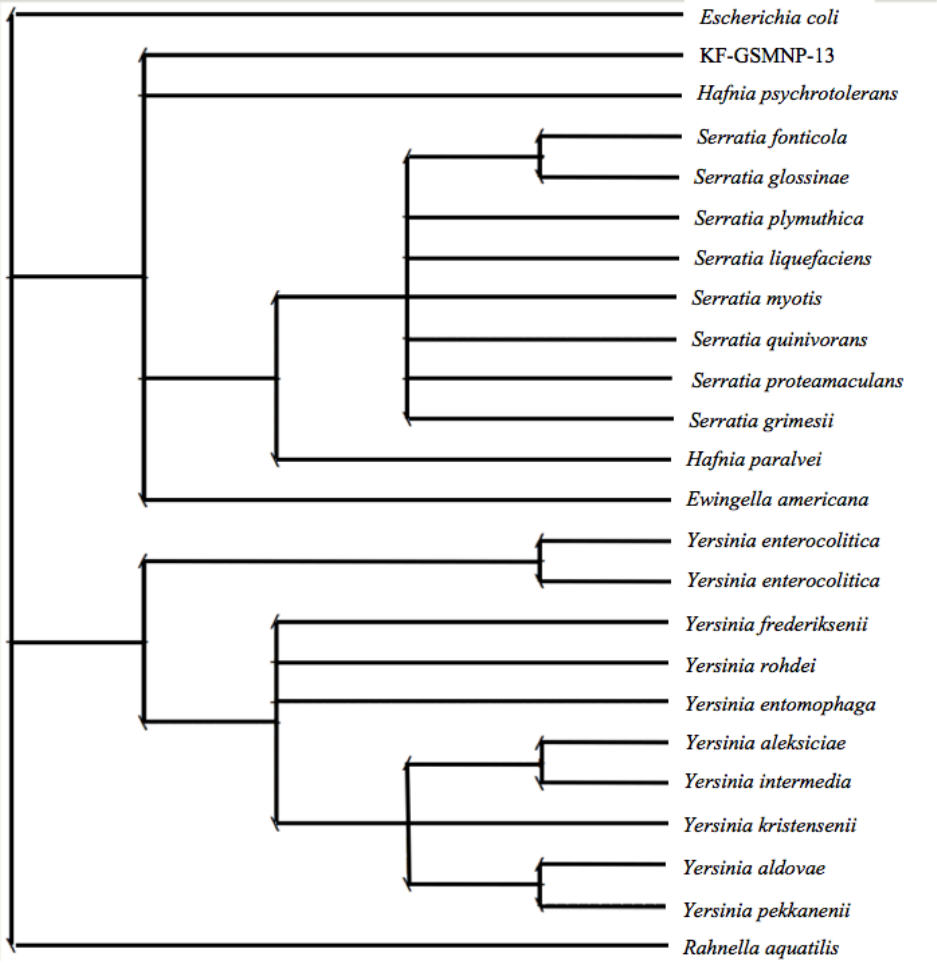

Figure 5: Strict consensus tree of 61 most parsiomonious trees using the top 20 sequence matches in RDP to KF13 based on full 16S rRNA gene sequence data (tree was constucted using the PAUP* 4.0 program and 100 heuristic bootstrap replications were used to assess the branch stability of the parsimony tree).

Table 12: Full 16S rRNA gene sequences of species related to the taxa that formed a polytomy with KF13 in the preliminary phylogenetic tree. Sequences were obtained from RDP and these were added to the original 20 to make the final phylogenetic tree.

RDP (Type Strains/ Isolates)	**Accession Numbers**
Hafnia alvei	ATCC13337
Serratia ficaria	DSM4569
Serratia entomophila	DSM12358
Serratia marcescens	DSM30121
Serratia odorifera	DSM4582
Serratia rubidaea	JCM1240
Serratia nematodiphila	DZ0503SBS1
Serratia symbiotica	GU394001
Serratia vespertilionis	KJ739885
Serratia aquatilis	KT387999
Serratia ureilytica	AJ854062
Edwardsiella tarda	ATCC15947
Edwardsiella ictaluri	JCM1680
Edwardsiella hoshinae	JCM1679

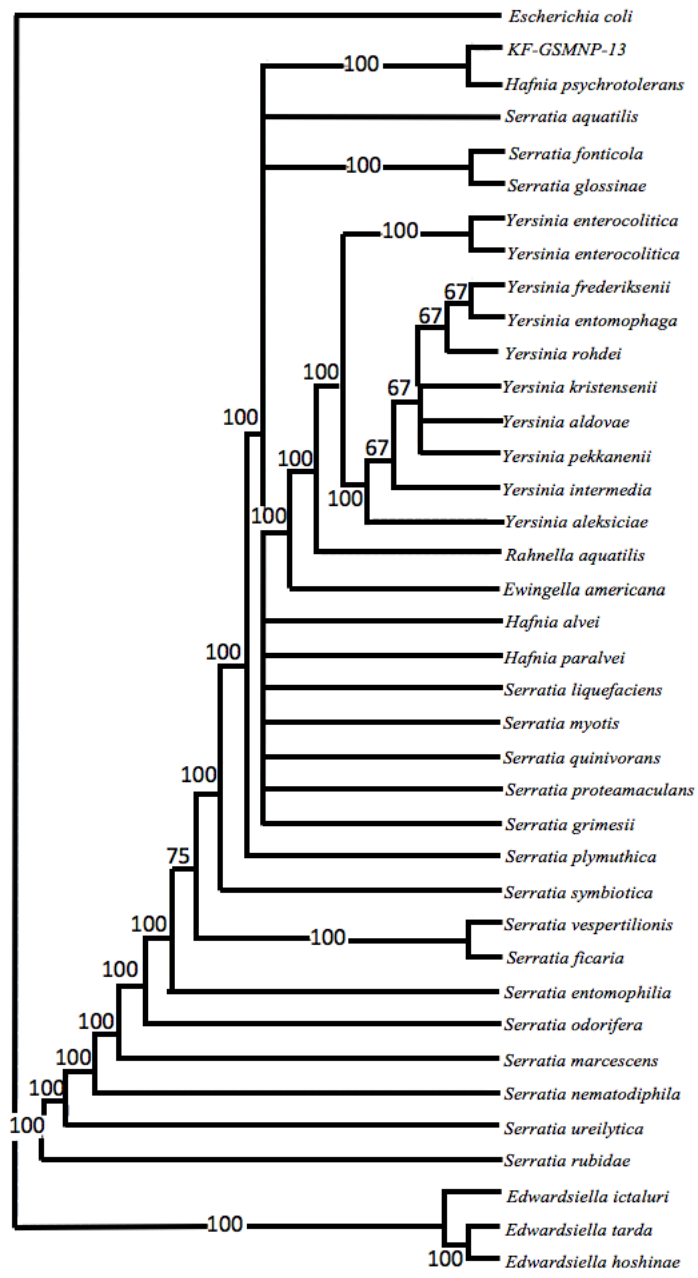

Figure 6: Strict consensus tree of 688 most parsiomonious trees using the top 20 sequence matches to KF13 based on full 16S rRNA gene sequence data as in Figure 2 but with the additional sequences seen in Table 10 (tree was constucted using the PAUP* 4.0 program and 100 heuristic bootstrap replications were used to assess the branch stability of the parsimony tree).

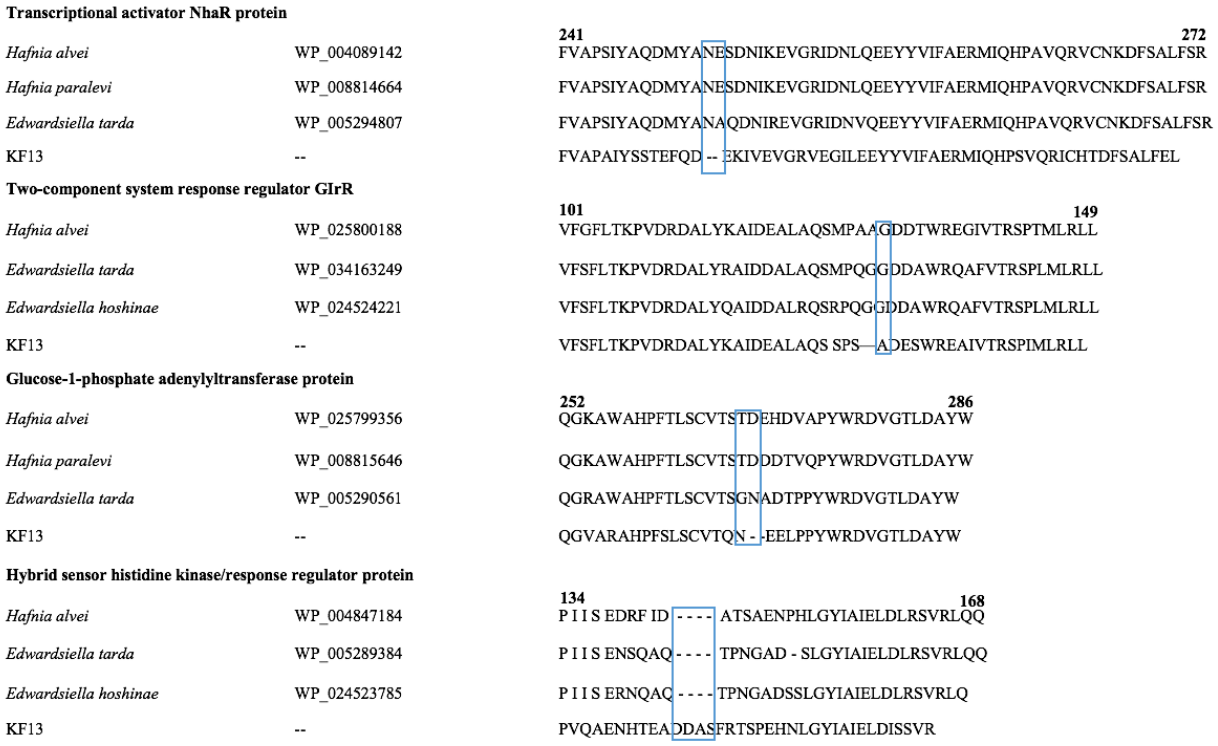

Figure 7: Four conserved signature insertions/deletions that distinguish the newly proposed *Hafniaceae* family. The CSIs found in three members (including GenBank accession number) of the newly proposed family are compared to KF13. The blue box indicates the position of an insertion or deletion.

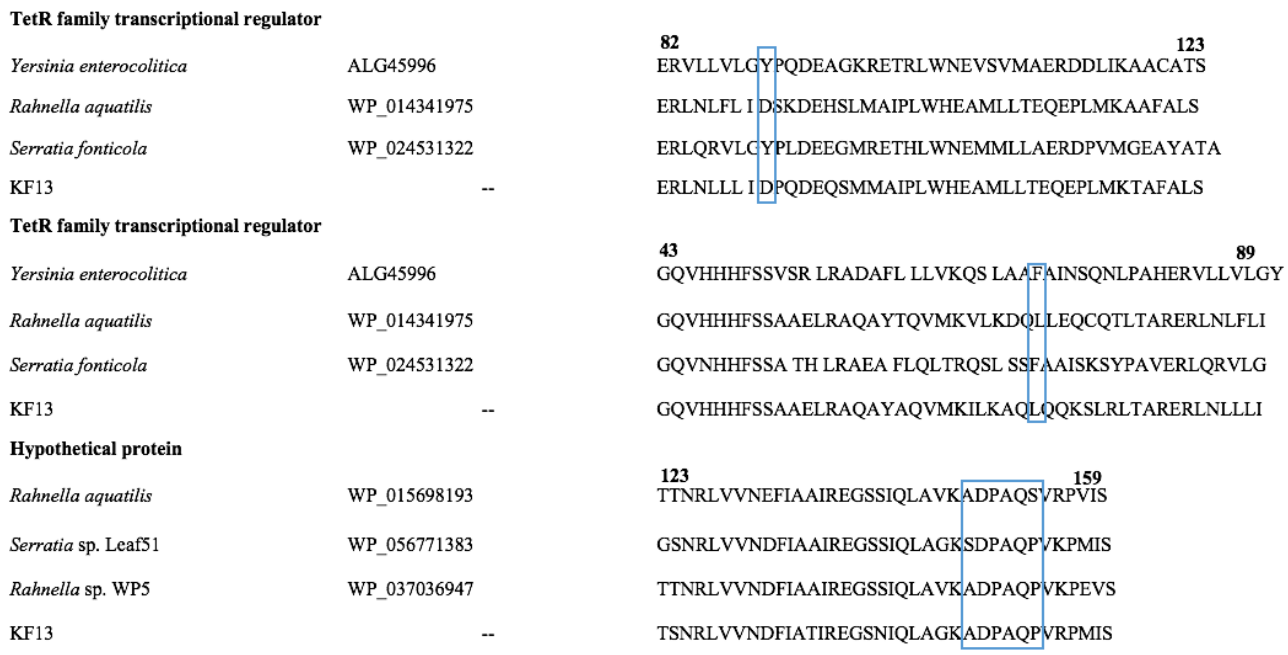

Figure 8: Three conserved signature insertions/deletions that distinguish the newly proposed *Yersiniaceae* family. The CSIs found in three members (including GenBank accession number) of the newly proposed family are compared to KF13. The blue box indicates the position of an insertion or deletion.