3-28-2006

The role of amplicon length heterogeneity-polymerase chain reaction in microbial community profiling and presumptive testing of bioagents

Melissa S. Doud
Florida International University

DOI: 10.25148/etd.FI15101220

To: Interim Dean Mark D. Szuchman
College of Arts and Sciences

This thesis, written by Melissa S. Doud, and entitled The Role of Amplicon Length Heterogeneity – Polymerase Chain Reaction in Microbial Community Profiling and Presumptive Testing of Bioagents, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

Juan Martinez

Kenneth Furton

Kalai Mathee, Major Professor

Date of Defense: March 28, 2006

The thesis of Melissa S. Doud is approved.

Interim Dean Mark D. Szuchman
College of Arts and Sciences

Interim Dean Stephan L. Mintz
University Graduate School

Florida International University, 2006

DEDICATION

This thesis is dedicated to the many cystic fibrosis patients that I have met during my research. May the quality of your life be improved by the information presented in this thesis and my future work.

ACKNOWLEDGMENTS

I would like to thank my mentor Kalai Mathee for letting me be a member of her incredible lab. To my fellow labmates and forensic peers, that put up with my mood swings and outbreaks of anger, thanks. You have been my Miami family and without you I may have long ago run away from this city. I would like to acknowledge the Department of Chemistry and Biochemistry for funding my education at FIU. Lastly, I would like to thank my family for being supportive of my journey down here. To my father, thanks for raising me so independent and passing on the gene for stubbornness and pride.

ABSTRACT OF THE THESIS

THE ROLE OF AMPLICON LENGTH HETEROGENEITY – POLYMERASE CHAIN REACTION IN MICROBIAL COMMUNITY PROFILING AND PRESUMPTIVE TESTING OF BIOAGENTS

by

Melissa S. Doud

Florida International University, 2006

Miami, Florida

Professor Kalai Mathee, Major Professor

Due to the threat of bioterrorist acts, there is a need to develop techniques that rapidly detect possible bioagents. Amplicon length heterogeneity-polymerase chain reaction (ALH-PCR) presumptively identifies eubacteria in samples by detecting differences between the lengths of the hypervariable regions of the 16S rRNA gene. To study the efficiency, reproducibility, and reliability of the technique, sputum from cystic fibrosis (CF) patients has been chosen as the model system. There is an abundance of microorganisms in the sputum of the CF lung. Using ALH-PCR, the complex microbial diversity and vast community composition in the lungs of the CF patients were studied. Twenty-four out of twenty-six CF samples were presumptively identified to contain *Pseudomonas aeurginosa*, a known CF pathogen. Sputum profiles were also compared over time and ALH was able to demonstrate that the CF lung flora is a dynamic community and may be affected by antibiotics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The discovery of DNA's structure in 1953 dramatically changed the scientific approach in the study of life but more importantly it has played a critical role in improving the United States (US) criminal justice system. DNA is now used not only as evidence that can convict but also to exonerate an individual for an offense. All organisms, human, animal, plant, viral, and even bacteria, contain DNA and thus have the potential to become the missing piece of information that is needed to solve a crime. DNA can also be manipulated to cause crime such as bioterrorism. Fortunately, with the use of molecular techniques, information can be extracted from the DNA that will lead to the apprehension of a criminal and the prevention of a bioterrorist act. The following introduction will discuss the role of DNA in the criminal justice system, the history of the bioterrorism, and the techniques that are used to study DNA.

I. Forensic science

Forensic science is a field that takes information and techniques from many academic areas including computer science, biology, chemistry, photography, accounting, and physics and applies it to the criminal justice system. Within this field are many specialties such as forensic biology. This field analyzes blood, semen, hair and other biological substances for characteristics that will establish the identity of the person that left the sample at a crime scene (81). In the US during the 1930s, human red cell antigens with enzyme and serum protein polymorphisms (ABO blood groups) could be

identified in humans and used to tell a suspect a part from a victim (55). This method was unable to distinguish individuals from one another due to the fact that there are only four blood types and with the Rh factor, eight discriminatory groups. For example, 38 % of people in the US have O positive blood (3). The development of DNA analysis drastically changed the role of biology in the justice system. In the past 30 years, forensic DNA analysis has made it possible to differentiate between and within human beings, plants, animal and even microbes. DNA can be used to identify paternity and link suspects to crimes; it can take as little as one hair follicle or a couple drops of blood to make a DNA profile. The courts have accepted human, plant, animal, and virus DNA as evidence and most recently microbial DNA has found a place in the criminal justice system.

II. History of human DNA analysis

Human DNA is highly conserved between individuals. Only one tenth of a percent of DNA varies from person to person. This amounts to approximately 3,000 million bases that differ (22). Yet, these differences can be used to separate one person from another and are the basis of DNA analysis (22). Alec Jeffreys of the University of Leicester, England coined the term DNA fingerprinting in 1985 (9). Jeffreys found that the human genome contains many minisatellite regions. These regions consist of tandemly repeated nucleotide sequences, all of which have a common core sequence. The length of each repeat varies in each individual (57). These variations can be detected using restriction fragment length polymorphism (RFLP). This technique uses restriction

endonuclease to digest DNA into specific fragments, which are then separated according to size by gel electrophoresis. Jeffreys produced two probes referred to as Multi-Locus Probes (MLP) to detect the fragments found in the minisatellite regions. The radioactively-labeled probes hybridized with the minisatellite fragments and then using Southern blot analysis, a series of bands similar to a barcode could be seen. The bands between 4 and 20 kilobases were unique to each individual. The forensic science community in the United Kingdom expected the MLP technique to be used in all of their major laboratories, but scientists ran into three problems. The method required large amounts of high molecular weight DNA to be extracted from crime scene samples, which wasn't always possible. Secondly, the DNA fragments from the control and the unknown need to be run on the same gel in order to accurately compare. Lastly, once the results were obtained, there was no easy way to convert the data into a form, which could be stored in a database (55). Even with these drawbacks, the MLP was already being used for identification purposes.

Jeffreys first used this technique to establish relationship between a boy and a mother (49). A little Ghanaian boy was trying to see his mother and sisters in the United Kingdom. Jeffreys was asked to confirm the child's identity so that he could enter the country. He used one restriction enzyme to cut the DNA. The products were then hybridized with two probes. He compared the boy's bands or DNA fingerprint to the print of his mother, brother, two sisters, and one unrelated individual. Bands that were found on the siblings' profiles but not found on the mother's profile were also on the boy's profile. Thus, the DNA profile indicated a parent-child relationship (49). In the US, Jeffreys' DNA fingerprinting technique was first implemented in a murder/rape trial

in Florida in the late 1980s (57). Randall Jones shot a man and woman; then dragged their bodies into the woods. Jones took their truck and pulled his car out of the mud, which was what had prompted him to commit the crime. Then, he went back into the woods and raped the woman, leaving behind a sample of his semen. An analysis of Jones' DNA and the semen at the crime scene was a match, which resulted in his conviction (57).

Other probes have been used with RFLP to determine genetic differences between people. These probes detect a specific variable number tandem repeat (VNTR) sequence. VNTR loci can have many alleles and the alleles vary in the length of the repeat sequence (89). During RFLP analysis, a restriction enzyme cuts the DNA at different bases depending on the enzyme. Each fragment produced will be a different length depending on the cut and the number of repeats (17). The VNTR probes produce simple banding patterns and can be highly specific when multiple probes are used (57).

In the early 1990s, forensic scientists became interested in polymerase chain reaction (PCR), which was developed by scientists at Cetus in 1985 (24). The technique amplifies template material from minimal amounts of extracted DNA (65). This new technology heralded the rapid development of STRs and SNPs that are currently used to profile human DNA (55).

STRs, short tandem repeats, are microsatellites that consist of small repeating units, usually 3-5 bases, with a maximum length of around 300 bases (55). The number of repeats at one STR locus or allele can vary from person to person but are not unique to one individual. When multiple STR loci are analyzed, a high degree of discrimination among individuals can be determined due to the variety of alleles present in a population

(10). In 2002, 13 STR core loci were chosen to add discriminatory power to DNA analysis (12). All 13 are highly polymorphic and 11 of them are located on different chromosomes (12). These loci are outside of the coding regions and thus have no functional significance (12). Two other markers, Amelogenin, though not STRs are a part of the kit. Amelogenin on the X and Y chromosomes are used to determine gender (15). STRs are also small and thus extremely stable even in old and degraded samples. This adds to their usefulness in forensic DNA analysis due to the fact biological evidence is not always preserved before reaching the laboratory.

Single nucleotide polymorphisms or SNPs can also differentiate people but only some of the SNP-based technology has been accepted by the criminal justice system. SNPs are the most abundant class of human polymorphisms (87). They are useful in forensic science due to their low mutation rates, abundance in the human genome, and short amplicon lengths (87). Currently, SNPs are used to study the Y chromosome and thus are indicators of ancestral lineage (47). As that SNPs are shorter than STRs, SNPs can be identified to a higher degree than STRs in degraded samples (20, 87). Researchers from the Forensic Science Service have developed a SNP kit similar to a STR multiplex kit, which can be used to process degraded samples for genetic information. The kit has been validated in the United Kingdom but this research has yet to be applied in the US justice system (20).

In the US, SNPs have been used to develop DNAWitness^{TM} 2.5 (21). This kit is used to determine genetic heritage from a human DNA sample of unknown origin. The sample is analyzed for numerous SNPs and then compared to a SNP database. The test will determine the percentage of genetic makeup amongst four population groups, Sub-

Saharan African, Native American, East Asian, and European (21). The ancestry of a person will to some extent dictate what a person looks like. The kit was used to help identify the Louisiana serial killer from 2003 (21). An eye witness told the police that the rapist was Caucasian. The SNP kit analyzed DNA from the related crime scene and discovered the owner of the DNA was 85 % Sub-Saharan African and 15 % Native American. This lead the police to believe that the rapist was not Caucasian and they altered their investigation. This information led to the arrest of Derrick Todd Lee. He was convicted of first-degree murder and aggravated rape (21).

III. Nonhuman DNA in forensic science

There are cases where human DNA evidence is absent, but crime scenes can be rich in nonhuman DNA. The techniques used to analyze human DNA and other molecular methods can be utilized to study plant, animal, and even bacterial DNA. Nonhuman samples can be used as evidence in crimes against humans, such as murder or kidnapping, and even bioterrorism.

Plant DNA analysis in forensic science. State v. Bogan in 1992 at Arizona was the first case to use plant DNA as trace evidence (84). A woman was killed and her body dumped in the Arizona desert. Near the body laid a beeper, which could be traced to Mr. Bogan. He claimed that he picked up the victim in his truck and she stole his wallet and beeper. The detectives needed to establish that the suspect had been to the crime scene and plant DNA provided that missing link. Police officers found a damaged Palo Verde

tree, *Cercidium spp.* at the crime scene and a seedpod in the back of Bogan's truck (84). Molecular geneticist Tim Helentjaris of the University of Arizona, Tucson was asked to use DNA profiling to match the seedpod to the damaged tree. At first, success was doubted but Helentjaris found a high degree of genetic variation in the tree, which was necessary to actually say that there is individuality in Palo Verde trees. Using randomly amplified polymorphic DNA or RAPD analysis, Helentjaris matched the seedpod to the tree at the crime scene. RAPD uses generic DNA primers that contain as few as ten bases and bind to multiple sites in the genome. Each primer can produce a reproducible profile of the PCR amplified fragments (105).

Sometimes, possession of certain plants is considered a crime. *Cannabis sativa* or marijuana is illegal in most of the US. Prior to the mid-1970s, most of the marijuana was grown in Mexico. Since the US and Mexico began eradication programs, a domestic growing industry has sprung up and is currently thriving in the US. Most plants originate from a few breeding stocks from the west coast. Thus, most marijuana plants are cloned. DNA profiling could determine the extent of genetic variation in marijuana plants and then the profile can link a cloned plant to its parent. Determining all of the parents and creating a database will be the difficult part. Typing *Cannabis sativa* can be done using RAPD, amplified fragment length polymorphism (AFLP), or STR. RAPD is inexpensive and simple to perform but difficult to reproduce due to the difference in PCR primers binding to target DNA sequences. Also, scientists can score faint bands on the agarose gel differently. AFLP can distinguish between individuals using PCR amplification of restriction fragments to which adaptor oligomer sequences have been attached. This method seems to be the best for identification purposes. STR markers in *Cannabis sativa*

have recently been discovered and more research needs to be performed to determine the usefulness of this marker in the identification of the plant (16).

Animal DNA for forensic analysis. Animal DNA was not used in trials until 1997. In Beamish v. Her Majesty The Queen, a man was convicted of murder by the evidence his cat left. On Prince Edward Island, a woman disappeared from her home. Weeks later, a man's jacket with blood on it was discovered in the woods wrapped up in a plastic bag. The blood matched the missing woman, whom was later found in a shallow grave. White cat hair was found inside the lining of the coat. Using DNA analysis with 10 STRs, the family cat was matched to the cat hair found in the jacket. The victim's common-law husband was arrested for the crime, as he owned the white cat (84).

Often the possession of an animal, alive or dead, is a crime. Wildlife forensic science uses DNA analysis to link an animal to a poacher. One wildlife case involved a poacher whom captured and the knifed a protected wild bore sow in a national park in Italy (52). The carcass was hidden in the woods until nightfall. Unfortunately for the poacher, the park staff found the carcass before he could return to pick it up. Detectives managed to find a potential suspect and in his home they found a blood stained knife. Scientists were able to say that the blood on the knife belonged to the same wild species of boar that was killed. Using this blood evidence, the man was convicted of poaching and cruelty to animals (52). Animal DNA is used in many of these wildlife forensic cases that involve the poaching of endangered species or the selling of parts from these animals (50, 98).

Insects in forensics. Insects have been used in forensic entomology to tell the time since death of a human being (54). This has been accomplished by identifying the insect species and the stage of growth it is found at. Larvae and puparium species (*Calliphora sp.*) can be difficult to speciate in their immature stages and therefore molecular methods have been used to determine identity. Mitochondrial DNA can distinguish between species and sub-species due to its sequence variation caused by a high mutation rate. Mitochondrial DNA is composed of two hypervarible regions (I and II) and various genes. One gene of interest codes for two subunits of the enzyme cytochome oxidase (97). The sequence of these subunits have been determined for *Calliphora sp.*, a forensically important insect found on corpses (54). Insect larvae can be sequenced and then compared to this bank to determine if the larva belongs to this species. With the correct identification of the insect, forensic entomologist can accurately determine time since death.

Viral DNA. Since HIV and AIDS has become an epidemic, it is not surprising to know that HIV has found its way into the criminal justice system. People have been convicted of knowingly infected others with human immunodeficiency virus (HIV) from transmission. HIV can also been used as a weapon as described in the case of Dr. Richard Schmidt and his nurse Janice Trahan (19, 61). They had been having an extramarital affair but Nurse Trahan decided to end it when she realized Dr. Schmidt would never leave his wife. They stayed friends and also continued to work together. Dr. Schmidt was even kind enough to still administer vitamin B-12 shots to Nurse Trahan. But one needle was contaminated with HIV from one of his patients. Nurse

Trahan did contract HIV and tuberculosis. Researchers at Bayor College of Medicine in Houston, Texas profiled the HIV strain in Trahan's blood and then matched it to a strain that belonged to one of the doctor's other patients. Schmidt was convicted for spreading HIV to his nurse and sentenced to 50 years of hard labor (19, 61).

Brian Stewart also injected HIV into someone he wanted out of his life (1). On February 1992, Brian went to the hospital to visit his illegitimate 11-month-old son. Brian appeared to be a loving, caring father but a few months later he separated from the boy's mother and refused to accept paternity. During these months and through the next four years, the child was constantly ill and finally on May 17, 1996 was diagnosed with an advance stage of AIDS. Through process of elimination, health care workers determined that the boy had not accidentally been affected with HIV and detectives began to develop a case against Brian Stewart. Detectives and the prosecutors proved that Stewart, a phlebotomist, used a contaminated needle on his son more than six years ago when the child was first hospitalized. On December 5, 1998, Stewart was found guilty of first degree assault and sentenced to life in prison (1).

IV. Microbial forensics

Human, animal, plant, and virus DNA has been used to convict the guilty. Microorganisms also have a role in forensic science. Microbial forensics is an emerging field that "can be defined as a scientific discipline dedicated to analyzing evidence from a bioterrorism act, biocrime, or inadvertent microorganism/toxin release for attribution purposes"(8).

Bioterrorism. Microbial forensics began to receive more attention from the general public after the US anthrax letter attacks in 2001 (8). From October 2nd to November 16th, 11 civilians were diagnosed with inhalation anthrax and another 11 contracted cutaneous anthrax, five of the inhalation cases were not curable. Of the 22 cases, 20 had either handled mail contaminated with anthrax spores or were exposed to contaminated mail processing worksites (39). Anthrax is caused by the gram-positive bacterium, *Bacillus anthracis*, that has two forms, the metabolically active bacterium form and the hard to kill endospore (Figure 1) (66). This bacterium has been identified as a biological warfare agent by the United Nations in 1969, World Health Organization in 1970, United Nations Office of Disarmament Affairs in 1992, Australia Group in 1992, North Atlantic Treaty Organization in 1996, Center for Disease Control in 2000, and again by the United Nations Office in 2001 (29). *B. anthracis* possesses many characteristics that make it ideal to be used as a lethal weapon. It can be found in the environment and is easy to cultivate in makeshift laboratories (29). The bacteria can also cause illness in humans and animals (66). In humans, anthrax presents flu-like symptoms such as fever, chills, weakness, nausea, etc. Cutaneous anthrax is the most common while inhalation anthrax, which is rare, has a higher rate of mortality (66). Once diagnosed, antibiotics such as penicillin and ciprofloxacin can treat the infection (66). Human vaccines at this time have limited usefulness (66).

Today, the general public thinks of bioterrorism as a war that uses biological materials that can be spread vastly by air, water, or weapons. And in the end, these organisms or bioproducts will cause wide-spread chaos, destruction, and death. The political or science definition of a biological warfare is "the military employment of

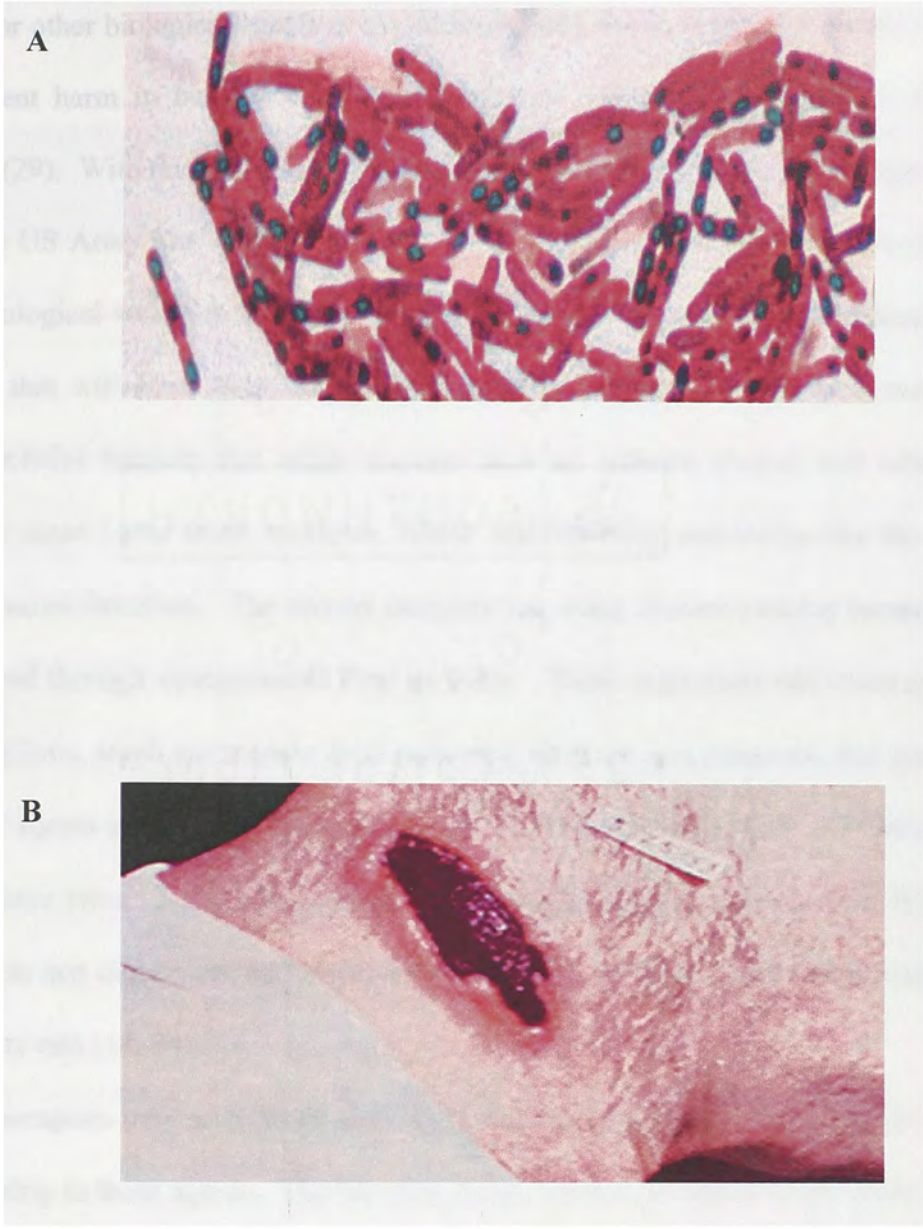


Figure 1. *Bacillus anthracis.* (A) vegative cells or sporangia stained red and mature spores stained green, (http://www.textbookofbacteriology.net/Bacillus.hml), (B) cutaneous anthrax lesion 15 days after transmission (14).

microbial or other biological agents or toxins to produce death, temporary incapacitation, or permanent harm in humans or to kill or damage animals or plants for a military objective" (29). With this definition, substances can be deemed as a biological agent.

The US Army has created a Critical Agent List that contains three categories of possible biological weapons (Table 1) (14, 34). Category A agents are considered to be substances that will most likely cause mass causalities if used as biological weapons. This list includes bacteria that inflict diseases such as, anthrax, plague, and tularemia; viruses that cause Lassa fever, smallpox, Ebola, and Marburg; and toxins like the neural toxin that causes botulism. The second category has many disease causing bacteria that can be spread through contaminated food or water. These organisms can cause typhoid fever, brucellosis, staph enterotoxin food poisoning, Q fever, gas gangrene, and glanders. Category C agents are all involved in emerging infectious diseases, such as hemorrhagic fever and hanta virus. These agents make good biological weapons, because of their ease of production and dispersion, and availability. Also, if used they retain a high morbidity and mortality rate (14, 34).

Biological weapons may seem to be a strong arsenal to have but there is a negative side to participating in these agents. The bacteria, fungi, viruses, or toxins could easily infect the people trying to use them as a weapon. The bioagents also infect the environment being conquered thus making it of no use to the new occupants. Regardless of these few limitations, world leaders, government officials, terrorists, and even civilians have set out to develop the art of biological warfare and exploit their capabilities of causing mass destruction (34).

Table 1. US Army's critical agent list (34).

Category A	Category B	Category C
Anthrax	Typhoid fever	Yellow fever
Plague	Brucellosis	Tuberculosis
Tularemia	Q fever	Hemorrhagic fever
Lassa fever	Gas gangrene	Hanta virus
Smallpox	Glanders	
Ebola virus		
Marburg Virus		
Botulism		

Historical perspective. Since the US anthrax attacks, bioterrorism has become a household name. Yet, the idea and use of biological warfare has been around for centuries. Historians state one of the earliest uses of a biology agent in a dispute was during the ninth to seventh centuries BC. The Assyrians poisoned the wells of their enemies with ergot (34). This toxic fungus causes nausea, headaches, vomiting, seizing, coma, and death. Since then, various rulers employed biological agents to poison or kill their enemies. Sixth century BC, an Athenian poisoned wells with a deadly herb (34).

Then, the Spanish tainted wine with the blood of leprosy patients in 1495. Another example occurs two hundred years later, when Poles fired spheres containing rabid dog saliva at their enemies (34).

World War II ushered a new era of warfare. Americans made the atom bomb and the Japanese delved into the world of biological warfare (34). Lieutenant General Ishii Shiro was a medical doctor and a commander of Unit 731 of the Imperial Japanese Army (30, 34). He has been credited with creating a massive biological warfare laboratory located in Japanese invaded Manchuria. Shiro began working with biological weapons prior to World War II but it was during this time that he made great progress in understanding how anthrax, plague, and glanders affected humans. The research of Unit 731 included developing a porcelain bomb containing anthrax spores, determining the effect of rice contaminated with plague-infected fleas would have on a civilian population, and exposing POWs and criminals to anthrax, plague, brucellosis, botulism, gas gangrene, and cholera. During a Soviet military tribunal in December 1949, it was determined that the biological weapons developed by Shiro and his men were used "against the Soviet Union in 1939, against Chinese cities during 1940-1944, and against Chinese troops in 1942" (34). In total, the Japanese biological warfare program killed at least 8,000 people in biowarfare experiments and countless others from related epidemics. Twelve Japanese defendants were found guilty of war crimes but Lieutenant General Ishii was never indicted. His freedom was guaranteed by the American government in exchange for the data resulting from the experiments (34).

The US was interested in biological weapons prior to the Soviet military tribunal. Camp Detrick now Fort Detrick in Maryland was the first American biological weapons

facility to open in 1943 (2, 34). After the Americans received the added information from the Japanese, three other facilities opened up, Horn Island, Mississippi; Granite , Utah; and Vigo, Indiana (2, 34). The first experiments involved brucellosis and Venezuelan equine encephalitis. In the 1950s, the US constructed a four and an eight story windowless building to study the dispersal rate of anthrax spores using various weapons. Throughout the era surrounding the Korean War, tularemia, Q fever, and anthrax were produced, weaponized, and agent dispersion methods, rates, and patterns were studied. Throughout the Vietnam War, the US continued to stockpile bioagents and began to produce botulinum toxin, staph enterotoxin, smallpox, rice blast, rye stem rust, and wheat stem rust. In the mid-1960s, civilians began to learn of the world's biological warfare programs. Protesters picketed Fort Detrick and countries around the world began to rethink the use of biological agents in warfare. The US and 102 other nations signed the International Biological Weapons Convention in 1972 stating they would no longer develop, possess, or stockpile agents in "quantities that have no justification for prophylactic, protective or other peaceful purposes" (34). Currently the US Army Medical Research Institute of Infectious Diseases (USAMRIID) located at Fort Detrick, is the only remaining biological weapons program in the US (34).

Throughout the second half of the twentieth century, the former Soviet Union, Germany, Iraq, and Great Britain also developed their own biological warfare programs (2, 18, 34). Currently, it is believed that there are other countries that possess or can make these weapons. These countries include Libya, North and South Korea, Israel, Syria, China, Egypt, Cuba, Vietnam, India, South Africa, Russia, and Iraq (18, 34).

Despite that fact that many countries have the knowledge and the resources to employ biological weapons against their enemies, few have used them (34).

Biopreparedness. With the threat of biowarfare, the US government has taken action to prevent, diagnose, treat, and trace biological agents. The Center of Disease Control (CDC) is the leading agency in "biopreparedness." The organization has developed a plan, which has five components: preparedness and prevention, detection and surveillance, laboratory response, communication, and training and research (13, 34). Preparedness and prevention is rooted in quick identification of a bioagent. Plans are being made to distribute antibiotics or vaccines and other medical equipment to the area of need in highly time-efficient manner (13, 34). Detection and surveillance relies on smaller government agencies such as local police and fire department to be able to quickly detect the presence of an agent. The CDC will provide these agencies with funds to be used in bioagent detection (13, 34).

In 1999, Congress gave $121 million to the CDC to improve their laboratory capability and response system (34). With these funds, the Laboratory Response Network (LRN) was developed (Figure 2). LRN is comprised of the Association of Public Health Laboratories, FBI, CDC labs, USAMRIID, and clinical labs throughout the states. The network works on a four-tiered concept beginning with the local hospitals being the first ones to find suspicion samples. These samples are sent to county or state labs for retesting. If the sample is confirmed to be related to a bioagent it is sent to the third-tiered state lab for further testing. Eventually, the suspicious sample is sent to the CDC lab or USAMRIID (13, 34, 74).

The communication aspect of the CDC's plan is led by the Health Alert Network (HAN). This system ensures that all aspects of the US government including the White House, Federal Bureau of Investigation, Central Intelligence Agency, state and local governments share information regarding possible biological warfare acts (13, 34).

The last part of the CDC's mission is to supply information to individuals regarding the possible threats that may arise from the use of bioagents. The organization also wants to make technologies available to outside groups that can be used in the detection and treatment of dangerous biological agents. In regards to research, the CDC itself is working on new applications to prevent and treat bioagents (18, 34).

The CDC is not the only organization that has acknowledged the need for the development of a plan and the tools to prepare and fight against biological warfare. In June 2002, the American Academy of Microbiology (AAM) held a colloquium to discuss bioterrorism, specifically the role of microbial forensics (42). As stated earlier microbial forensics is defined as "a scientific discipline dedicated to analyzing evidence from a bioterrorism act, biocrime, or inadvertent microorganism/toxin release for attribution purposes"(8). The Academy decided that it was necessary to find techniques that would identify microorganisms that were used in bioattacks and then be able to trace their origins. During the meeting, participants identified areas in which more research and information was needed in order to ensure that the aims of microbial forensics would be accomplished. Their recommendations included: educating the local law enforcement agencies and medical first responders, developing new techniques that would identify biocrime organisms, find ways in which the source of these organisms could be traced,

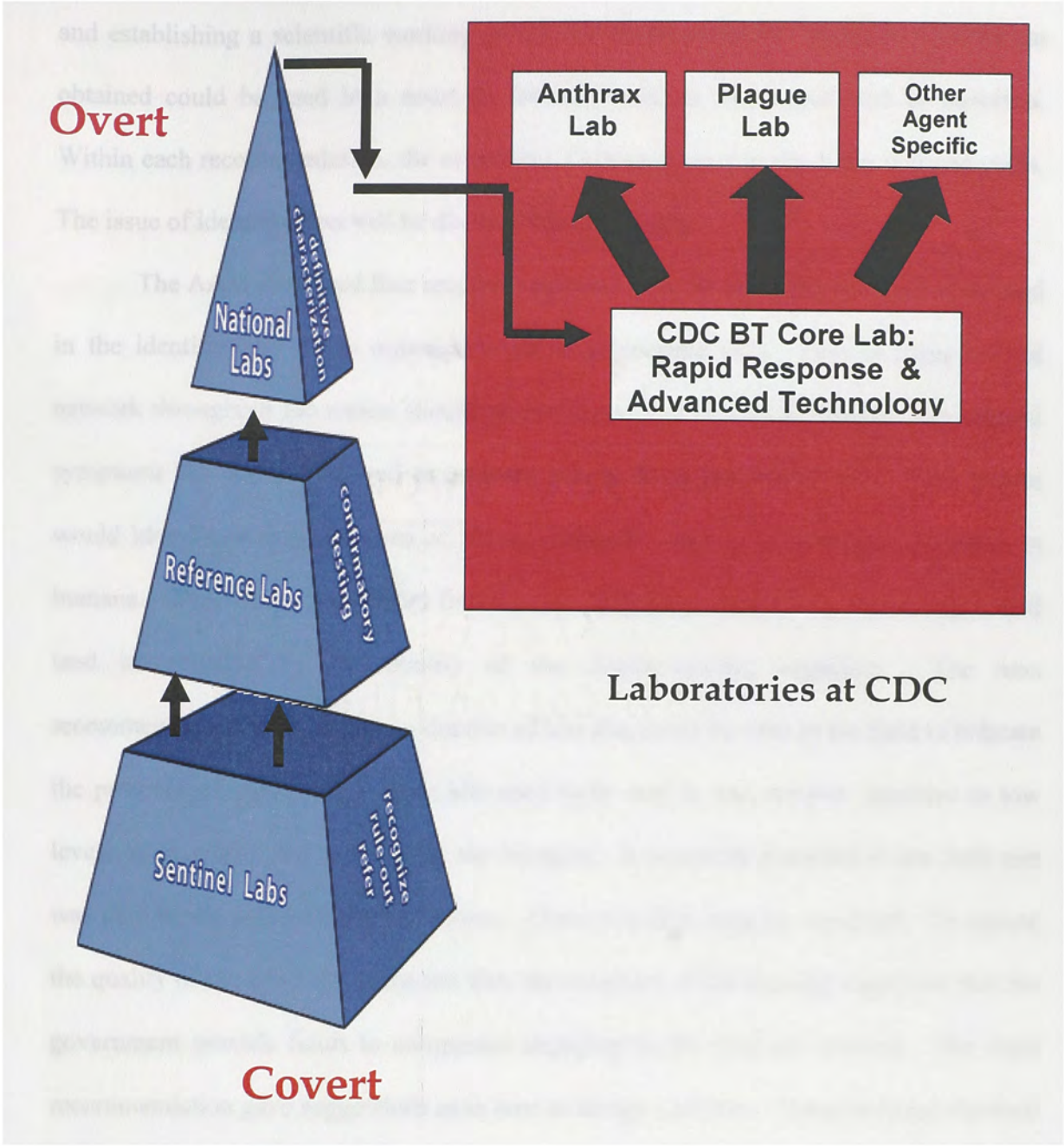


Figure 2. Laboratory Response Network structure for agent testing and sample flow. Local hospitals detect suspicious samples and send to sentinel labs for further testing. Possible pathogens are further characterized using reference and national labs, such as the CDC (13).

and establishing a scientific working group that could assure the scientific information obtained could be used in a court of law to prosecute the perpetrators or terrorists. Within each recommendation, the committee discussed ways to reach the outlined goals. The issue of identification will be discussed in detail (42).

The AAM discussed four recommendations as to the best plan of action to be used in the identification of an organism used in a biocrime (42). First, a computerized network throughout the nation should be developed that will track "disease and unusual symptoms for humans, as well as animals, plants, food, and water" (42). This system would identify unusual patterns of illness outbreaks, specifically flu-like symptoms in humans. The pattern can predict future cases, help pinpoint the original exposure, and lead investigators to the identity of the disease-causing organism. The next recommendation involved the production of kits that could be used in the field to indicate the presence of a bioagent. These kits need to be easy to use, reliable, sensitive to low levels of exposure, and specific for the bioagent. It would be preferred if one field test was able to identify multiple organisms. These kits then must be validated. To ensure the quality of the biocrime agent test kits, the members of the meeting suggested that the government provide funds to companies engaging in the field kit research. The third recommendation gave suggestions as to how to design such kits. These included classical phenotypic assays, sequencing, single nucleotide polymorphisms, and pathogenicity arrays using 16S rRNA probes. It was noted that most techniques have disadvantages that make positive, rapid identification difficult. Therefore, the AAM states that it would be more accurate to use multiple techniques in the test kits. The colloquium further suggests that genomic sequence and other molecular databases be implemented into

research behind the bioagent identification kits. Another source to tap would be from the fields of environmental microbiology and microbial ecology. These disciplines have developed techniques to detect and identify microorganisms in natural communities. Their approach may lead to be useful in biocrime detection. Overall, the AAM has recognized the need for the development of assays that will quickly identify the presence of a bioagent. These kits should be based on newly discovered technologies and modified classical microorganism identification assays (42). Once protocols and kits are established, quality assurance and quality control standards must be set to demonstrate the reliability of the techniques. It is extremely important that results obtained from novel techniques will stand up in a court of law and pass either the Frye Test or the Daubert Standard, which are the rules of scientific admissibility (81). These well-established protocols then can be used not to only identify the pathogen involved in the attack but also the source of the microorganism (42).

The US government also plans to institute a national microbial forensics system that will help in the fight against bioterrorism. The FBI has already formed the Scientific Working Group on Microbial Genetics and Forensics (SWGMGF) (60). Similar working groups have been used in other areas of forensics, such as Scientific Working Group on DNA Analysis methods. The SWGMGF is focusing on quality assurance and guidelines for laboratories, establishing criteria for development and validation of methods to characterize microbes, and establishing design criteria for information databases (60). To further show the role microbial forensics plays in bioterrorism, The National Institute of Justice in November 2002 solicited scientists to develop new techniques to be used in the field of forensic DNA, including the exploration of bacterial DNA (68). In the past,

traditional culturing methods have been used to identify microbes. Only a small subset of bacteria can actually be cultured in a laboratory, thus many organisms including novel bacteria cannot be identified using these techniques (104). With the advent of molecular biology techniques, bacteria no longer need to be isolated and communities of bacteria can be studied (35).

V. Molecular techniques used to identify bacteria

Before the advent of polymerase chain reaction (65), eubacteria were identified based on morphology, physiology, and other whole cell properties. These properties are said to be too simple or volatile to correctly indicate phylogenetic relationships (72). Currently, organism identification and its phylogeny can be better studied using molecular markers and molecular techniques (72).

Molecular markers. Markers or molecular chronometers have DNA sequences that change randomly in time but overall have a constant rate of evolution. This clocklike behavior can be seen in genes that have a high degree of functionally constancy (102). Molecular chronometers should also have a large phylogenetic range and therefore be present in most to all organisms (72). Lastly, the molecular marker needs to have a large size with regions that are evolutionarily independent (102). One family of genes fills all of these requirements to become the "ultimate molecular chronometers", the rRNA genes (72). Ribosomal ribonucleotide acids (rRNAs) have a constant function, are found in all organisms, and the sequences change at different positions and at different rates (72).

Ribosomal RNA. Ribosomal RNA (rRNA) represents 80%-90% of the total mass of cellular RNA in both eukaryotes and prokaryotes. Ribosomal RNA is directly associated with ribosomes (ribonucleoproteins) that play a role in protein synthesis (48, 70, 71). The basic form of the ribosome is conserved throughout organisms although changes in overall proportions and size of RNA and protein can be seen. All ribosomes in a cell are identical and in prokaryotes they consist of a small (30S) subunit that is composed of 16S rRNA and 21 proteins and the large (50S) subunit (Figure 3). The larger unit contains 23S rRNA, 5S rRNA, and 31 proteins (48, 70, 71).

16S rRNA. The 16S ribosomal subunit is most commonly used in studies of phylogeny (44). This gene is composed of alternating evolutionary conserved and variable regions (Figure 4) (95). The presence of conserved regions can be attributed to the fact that ribosomes have a function and sequence mutations in the conserved regions **PCR.** Polymerase chain reaction (PCR) is a technique used to detect specific sequences of DNA from organisms by molecular amplification (65). The technique amplifies template material from minimal amounts of extracted DNA by annealing and extending two oligonucleotide primers that flank the target region in duplex DNA (Figure 5). First, the DNA is denatured and then each primer hybridizes to one of the two separated strands. The annealed primers are then extended on the template strand with a DNA polymerase.

These three steps, which occur at unique temperatures, are repeated in a cycle and DNA is amplified (24, 65). There have been multiple techniques that implement PCR technology to study microbial communities including the three discussed below.

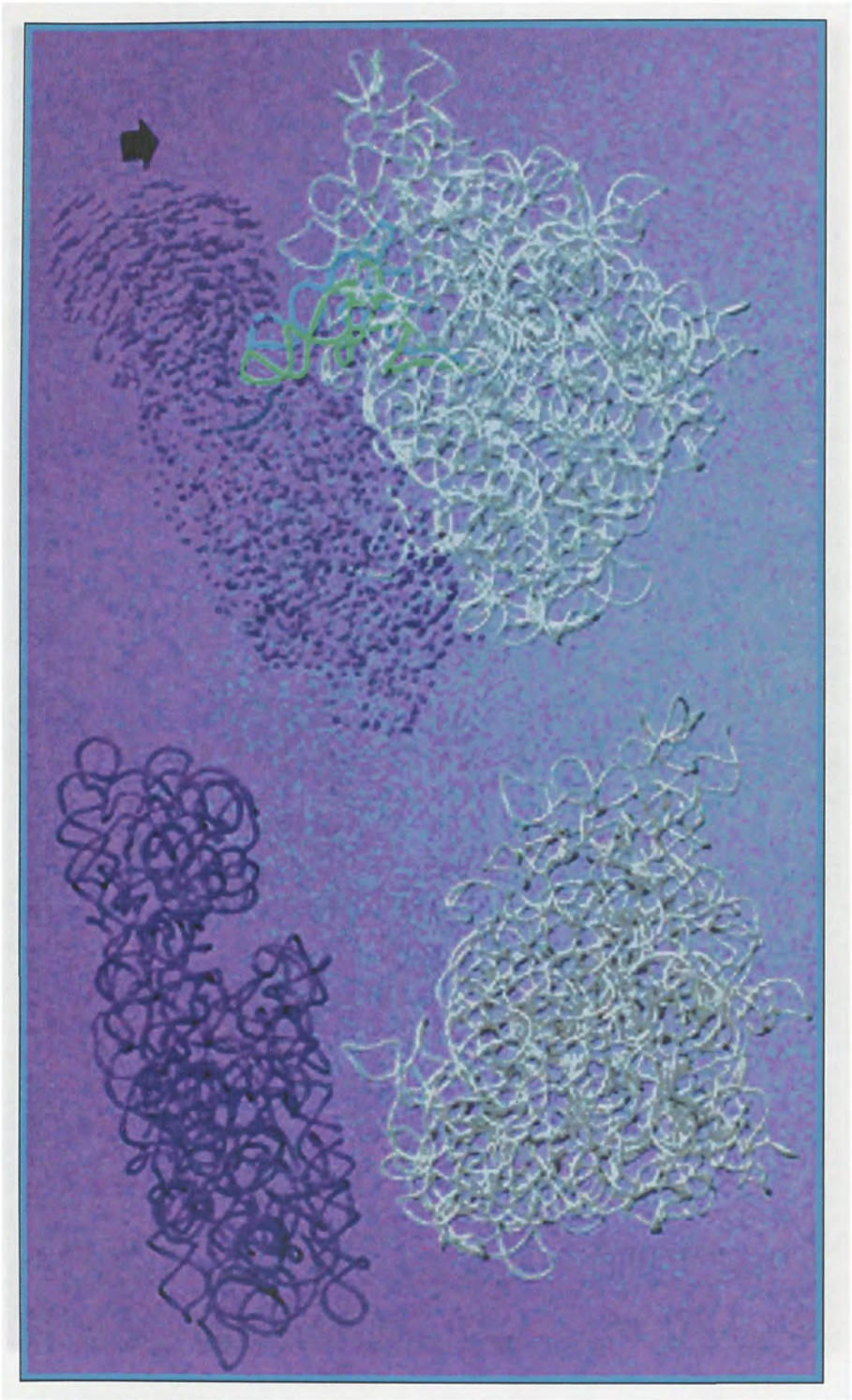


Figure 3. Ribosome subunits. The purple 30S subunit joins the gray 50S subunit to form the 70S ribosome (103).

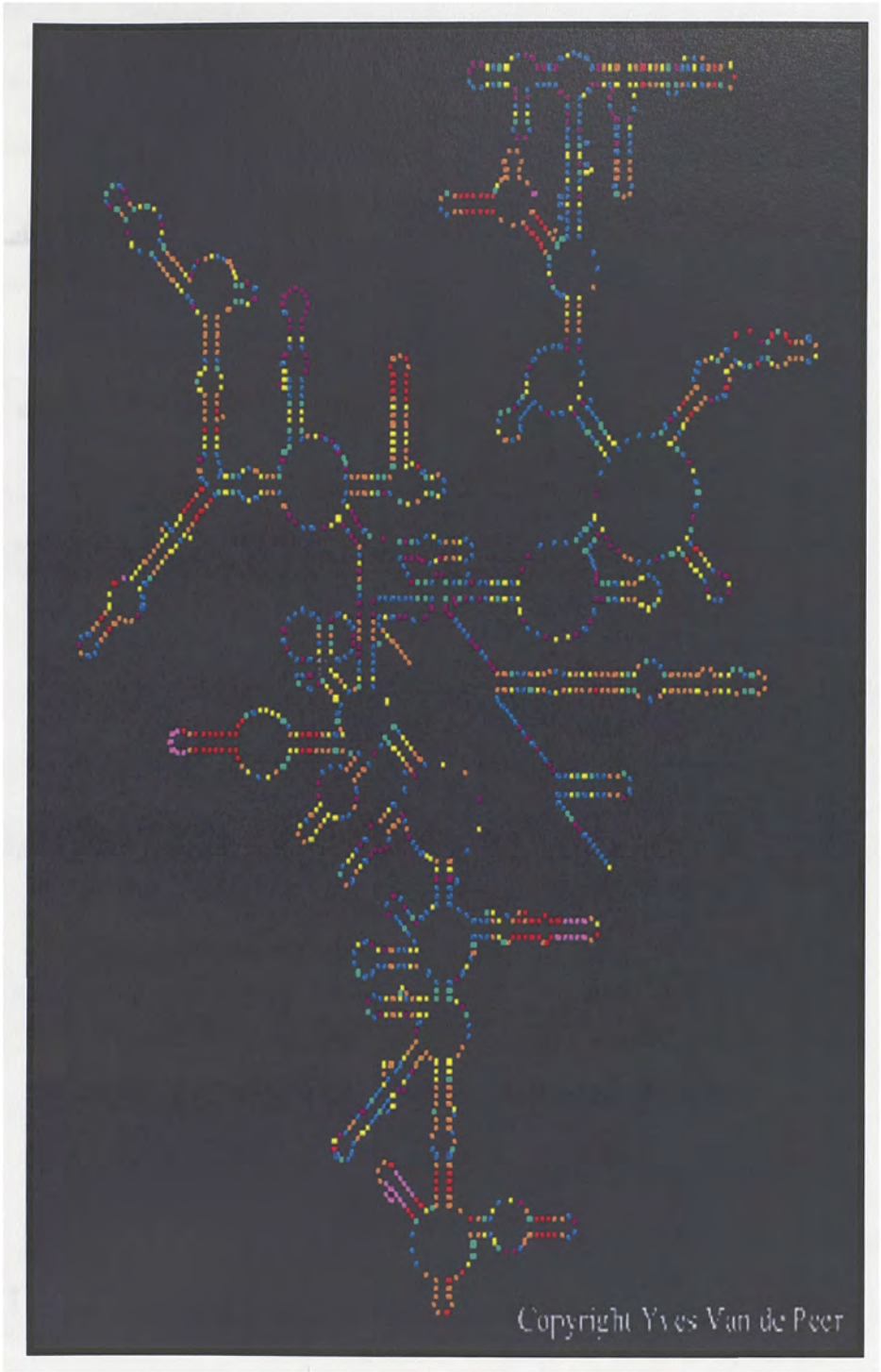


Figure 4. 16S rRNA secondary structure map. Purple indicates absolutely conserved regions, blue represents regions with a high amount of conservation. Red shows areas that are most variable (95).

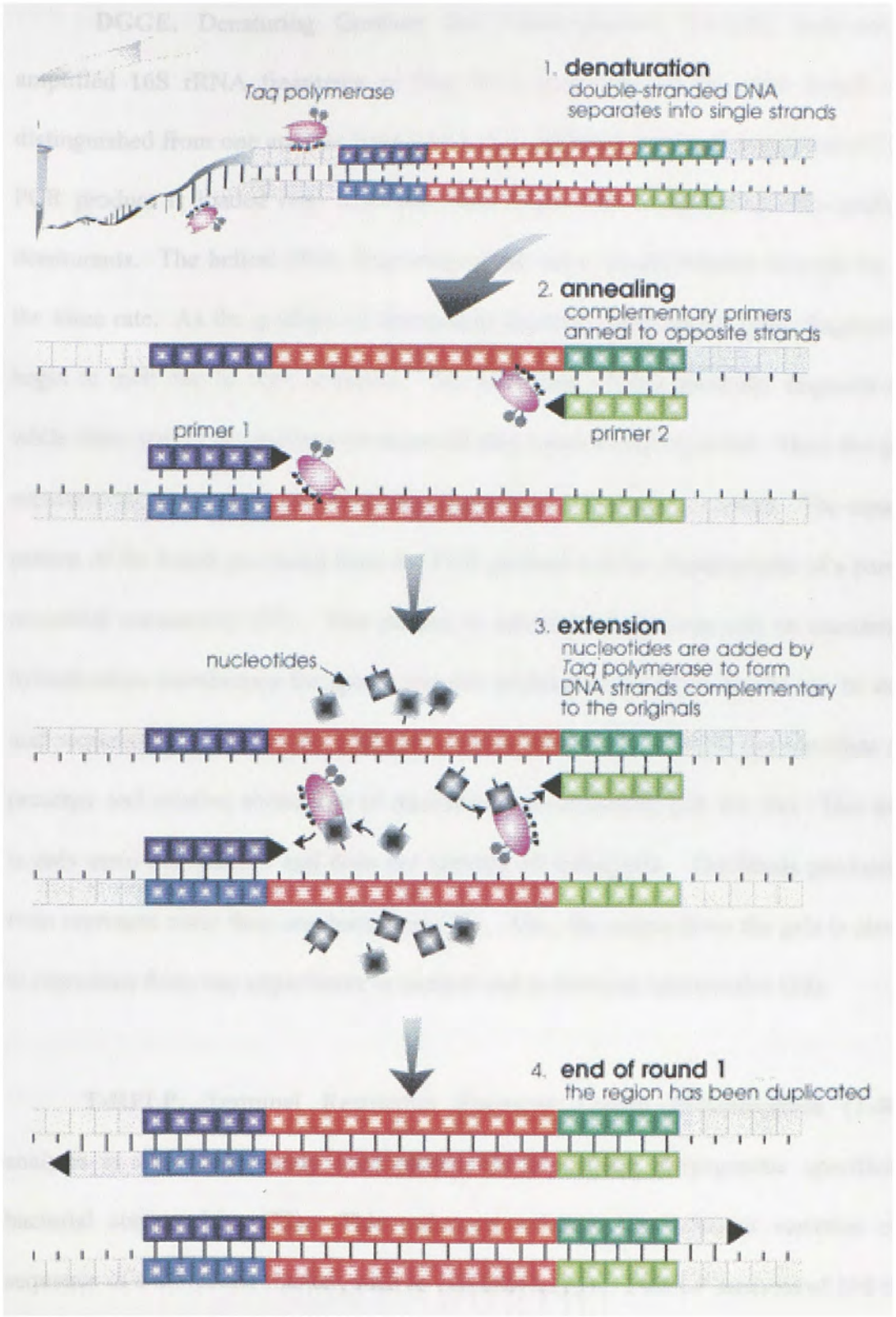


Figure 5. PCR amplification (37).

DGGE. Denaturing Gradient Gel Electrophoresis (DGGE) separates PCR amplified 16S rRNA fragments so that DNA fragments of the same length can be distinguished from one another based upon their different base-pair sequences (67). The PCR product is loaded onto a polyacrylamide gel that is composed of a gradient of denaturants. The helical DNA fragments of the same length migrate through the gel at the same rate. As the gradient of denaturants increase, some of the DNA fragments will begin to melt due to their sequence. The migration of this particular fragment ceases while other fragments continue to move till they reach a melting point. Thus, the gel has separated the same sized 16S rRNA fragments by their sequence content. The separation pattern of the bands produced from the PCR product will be characteristic of a particular microbial community (67). This pattern or microbial fingerprint can be transferred to hybridization membranes for specie specific probing or the DNA bands can be excised and sequenced. Through the addition of these techniques, DGGE can elucidate to the presence and relative abundance of microbes in a community (28, 67, 94). This method is only semi-quantitative and does not identify all eubacteria. The bands produced can even represent more than one bacterium (83). Also, the output from the gels is also hard to reproduce from one experiment to another and in between laboratories (28).

T-RFLP. Terminal Restriction Fragment Length Polymorphism (T-RFLP) analysis is the most widely used method in identifying phlyogenetic specificity in bacterial communities (32). This technique relies on the inherent variation of the sequence of a molecular marker, such as 16S rRNA (23). T-RFLP analysis of 16S rRNA genes includes PCR amplification of those genes with terminal labeling, digestions of the

PCR product with restriction enzymes, and detection of terminal restriction fragment by an automated DNA sequencer or capillary electrophoresis (32). The resulting output consists of a microbial profile where each detected is a specific fragment length produced from the cut PCR product. Each represents one or more bacteria that have a similar sequence and fragment length of the amplified region of the 16S rRNA gene. T-RFLP profiles can be used for differentiation of communities, identifying specific organisms in populations, and for comparison of the relative phylotype richness and structure of communities (23). This method has been successful in differentiation of bacterial communities that are in marine samples, digestive tracts of fish, activated sludge samples, soil samples, and sputum samples of cystic fibrosis patients (23, 32, 63, 79). This method is fast and data can be easily replicated for statistical analysis. The major downfall of T-RFLP is that many bacteria produce similar fragment sizes and thus not all profiles are species specific. Some fragments might even represent more than one genus (23). There are also inherent problems using restriction enzymes, such as incomplete digestion, which can produce DNA fragments that do not correlate to the correct bacterium (63).

ALH. Amplicon Length Heterogeneity (ALH) method is able to profile a diverse microbial community (5, 63, 78, 79, 90, 91). ALH-PCR relies on specific regions of DNA that show variation in sequence length (63). For example, the 16S rRNA gene has conserved and hypervariable regions (Figure 4) (95). These hypervariable regions vary in length and sequence. Through the use of ALH-PCR, the lengths of these regions are analyzed to distinguish between groups of organisms (63). In this technique, forward and reverse primers bind to the conserved regions of the 16S rRNA marker (Figure 6) and

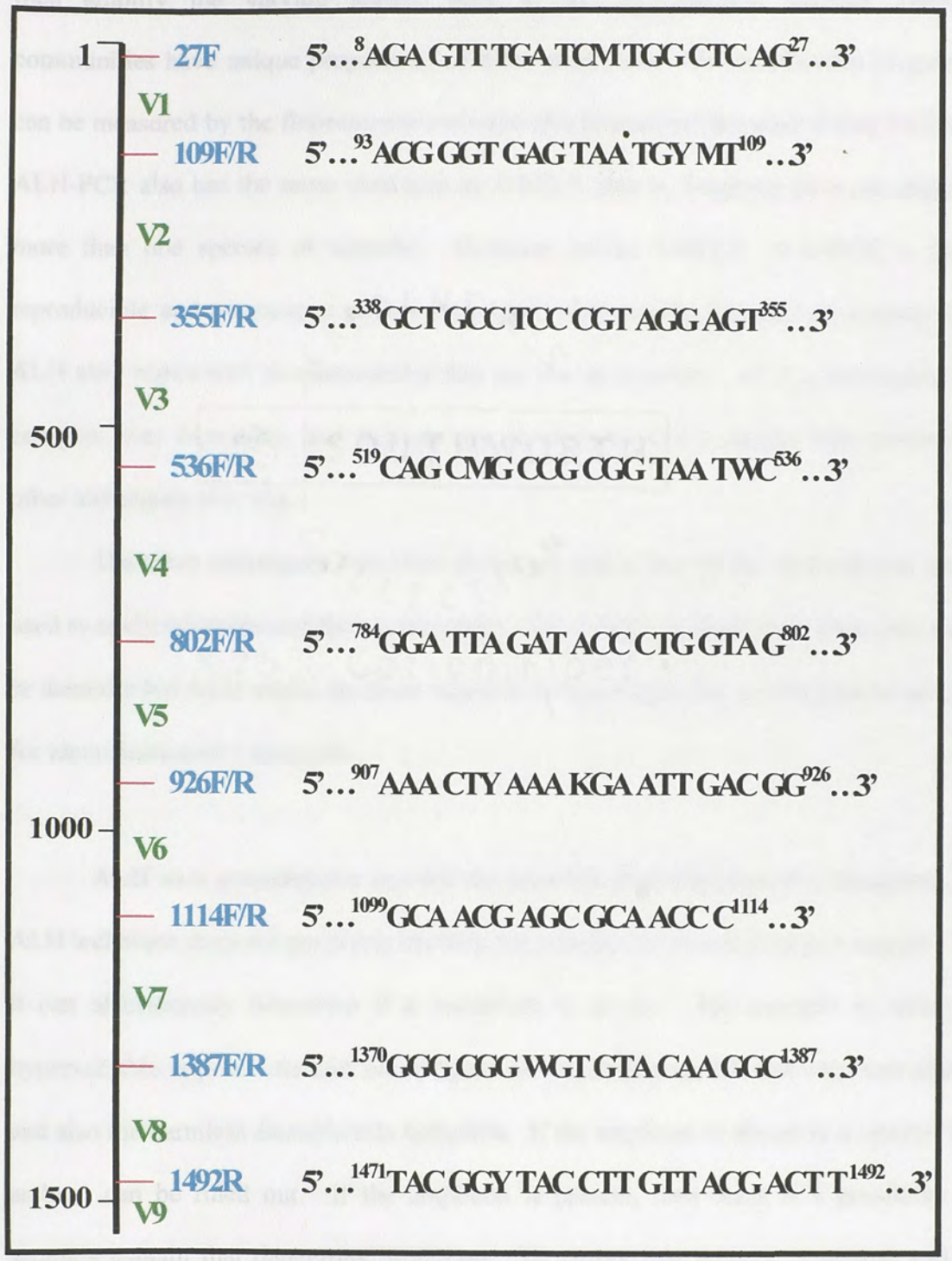


Figure 6. Linear view of 16S rRNA gene. Primer sequences are displayed at each conserved region.

then amplify the varying lengths seen in the hypervariable regions. Different communities have unique proportions of these amplicons. The relationship proportions can be measured by the fluorescence emission of a labeled primer used during PCR (90). ALH-PCR also has the same weakness as T-RFLP, that is, fragment sizes can represent more than one species of microbe. However unlike T-RFLP, ALH-PCR is highly reproducible and produces a greater resolution of the whole bacterial community (93). ALH also works well in communities that are low in diversity. ALH is technically less complex, less expensive, and analysis can be performed in a shorter time period than other techniques (63, 93).

The three techniques described above are just a few of the methods that can be used to study microbes and their community. All of these methods have their own merits or demerits but ALH seems the most suited to be developed into a presumptive test used for identification of a bioagent.

ALH as a presumptive test for the possible identification of a bioagent. The ALH technique does not positively identify the presence of a bacterium in a sample. Yet, it can affirmatively determine if a bacterium is absent. For example in using the hypervariable regions one and two (Figure 6), the amplicon 356 can represent anthrax and also the harmless *Desulfocella halophil*a. If the amplicon is absent in a sample, then anthrax can be ruled out. If the amplicon is present, then there is a possibility the samples contain that dangerous bacterium. To ensure that anthrax is present and not some harmless bacterium, another test such as sequencing would be performed. Therefore, ALH could be used as a presumptive test for a bioagent. If an amplicon that

represents a harmful bacterium is found, then a further confirmatory test, like sequencing would be performed.

Due to the dangers in using bioagents such as *B. anthracis*, *Yersinius pestitus*, and *Vibro cholerae*, a model system was chosen to prove that amplicon length heterogeneity could be used as a bioagent presumptive test.

VI. Cystic Fibrosis as the model system

ALH-PCR can profile bacterial communities found in soil and other environmental samples but it too can be used with biological samples, such as sputum from cystic fibrosis (CF) patients (62, 79). CF is an autorecessive disease, which is seen in one and 2500 Caucasian live births in the US (73). CF is due to a mutation in the gene located at chromosome 7q31, which encodes for the protein called cystic fibrosis transmembrane conductance regulator (CFTR) (46, 77, 96). A defect in the CFTR protein leads to a malfunctioning cyclic-AMP-activated chloride channel in secretory epithelia (46). This defect in the lung leads to the inability to secrete chloride and to the excess re-absorption of sodium (46). Thus, there is a decreased fluid secretion and the mucus becomes immobilized and adheres to the epithelial cells. Overproduction of mucus in the airway results in congestion of the respiratory tract and increases susceptibility to bronchopulmonary infection. The CF patients often suffer from infection with *Pseudomonas aeruginosa*, *Burkholderia cepacia*, *Haemophilus influenza*, and *Staphylococcus aureus* (26, 41). A person's susceptibility to a certain pathogen can vary but it does tend to be age related. *S. aureus* is the predominant pathogen in children,

followed by *H. influenzae* in early childhood, and *P. aeruginosa* during adolescence (4). Eighty percent of adults are chronically infected with *P. aeruginosa* (4). These chronic infections cause inflammation and eventually lung damage; therefore, lung disease is the major cause of morbidity and mortality among these patients (36).

Bacterial infections are identified using traditional culturing techniques which may fail to accurately identify the bacteria present in the lung. Misidentification has been shown to be a problem when studying CF infections (88, 101). This can impair the successfulness of a patient's treatment (43, 51, 56, 86). Studies have shown that patients have been diagnosed with *B. cepacia* using culturing techniques but with more stringent molecular techniques, the infection was shown to be *B. gladioli* or *S. maltophilia* (58, 59). *P. aeruginosa*, which is found in the CF lung, can be difficult to identify. CF-derived isolates of this bacterium may have phenotypic diversity due to loss of pigment production, mucoidy, and the presence of a rough lipopolysaccharide, hence hampering identification (88). Occasionally, *P.* aeruginosa is identified as another bacterium. One studied showed that 770 CF isolates were initially identified as *B. cepacia.* Rescreening these samples with polyphasic analysis that included selective media and biochemical tests, demonstrated that 88 samples were misidentified of which 11 of them were *P. aeruginosa* (86). Misidentification problems can be reduced or completely eliminated by using genotype-based molecular identification methods (88).

It has also been estimated that less than 1 % of eubacteria in the environment can be cultured (80, 99, 100). Thus, these identification methods would fail to detect all pathogens that may be causing the lung infection. Fortunately, with the advent of molecular microbiology techniques, culturing for identification purposes can be

circumvented. ALH is the chosen molecular method to study eubacterial infections of the CF lung.

VII. Hypothesis and aims

ALH-PCR is a highly versatile technique that has been used in ecology based fields to detect the presence of bacteria in soil, sputum, and water (62, 63, 90, 93). This technique has proven to be robust, inexpensive, and a valuable tool in molecular identification (63, 93). ALH-PCR could potentially be used as a presumptive test in the fight against bioterrorism by determining the absence of an amplicon length that may represent a human pathogen. It also can be used to study the bacterial community in the CF lung with little to no modification as shown previously (62). This project proposes to validate past CF studies and increase the information attained from ALH by studying more hypervariable regions of the 16S rRNA genes. Also, the ALH-PCR technique will be implemented to study the bacterial dynamics of the CF lung. Specifically, I will perform four aims:

Aim I: Analyze the metagenomic DNA from the CF sputa samples using V1 + V2 region of the 16S rRNA gene to validate a previous study.

Aim II: Design a multiplex consisting of the V1 and V3 regions of the 16S rRNA gene.

Aim III: Analyze CF sputa samples using the newly designed multiplex.

Aim IV: Determine if ALH can identify differences in patient sputum profile over time.

MATERIALS AND METHODS

Sample collection. Seven sputum samples were collected from CF patients (identity withheld) attending the Joe Dimaggio Children's Hospital (Hollywood, FL). Samples are referred to by JD followed by a number. Nineteen sputum samples obtained from CF patients attending the University of Miami, School of Medicine (Miami, FL) under FIU IRB approval # 033004-02. Samples are referred to by UM followed by a letter. All 26 expectorated sputum samples were frozen at -20 °C prior to DNA extraction.

Strains, plasmids and primers. The strains and primers used in this study are listed in Table 2. All bacteria were cultured in Luria-Bertani (LB) broth or on LB agar at 37 °C.

Genomic DNA extraction. *Escherichia coli* DH5α and *Pseudomonas aeruginosa* PAO1 (Table 2) were streaked out and grown at 37 °C on a LB agar plate. An isolated colony was inoculated into LB broth. The culture was allowed to grow overnight at 37 °C. Genomic DNA was extracted using DNAzol (Molecular Research Center, Cincinnati, Ohio) as per manufacturer's protocol. One milliliter of the growth culture was placed in each eppendorf tube. The bacteria cells were pelleted by centrifugation at 10,000 g for 5 minutes. The supernatant was discarded and the remaining pellet was resuspended in 1 ml of DNAzol. Tubes were inverted five times in a 3 minute time period. Five hundred µl of 100 % ethanol was added to each tube.

Table 2. Bacterial strains and primers used in this study.

Strain/Primer	Genotype	Reference
Escherichia coli		
DH5α	F⁻ ϕ80d*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *deoR recA1 endA1 hsdR17*(r_k^-, m_k^+) *phoA supE44 λ- thi-1 gyrA96 relA1*	New England Biolabs
Pseudomonas aeruginosa PAO1	Prototype	(33)
Primers*		
355R	5'-GCTGCCTCCCGTAGGAGT-3'	
P2	5'-TTCCCCACGCGTTACTCACC-3'	
338F	5'-ACTCCTACGGGAGGCAGCAG-3'	
518R	5'-ATTACCGCGGCTGCTGG-3'	
27F	5'-AGAGTTTGATCMTGG-3'	
P1	5 –TTCCCCACGCGTTACTCACC- 3'	

*M is a choice between A or C base.

Tubes were vortexed to homogenize mixture completely. The precipitated DNA was then centrifuged at 5,000 X g for 3 minutes. The supernatant was discarded and the DNA pellet was washed twice by the addition of 500 µl of 75 % ethanol. The pellet was dried and then eluted with 30 µl of diethylpyrocarbonate (DEPC) water. The genomic DNA was stored at 4 °C till further use.

Sputa DNA extraction. Metagenomic DNA was extracted from the samples using the method of Reischl *et al* (76). Each individual sputum sample was split into three equal aliquots and placed in an eppendorf tube. The triplicate extraction was used to determine reproducibility of extraction technique and to increase the sample size for statistical purposes. The mass of the sputum was calculated by subtracting the weight of the tube from the weight of the tube containing the sputum. Equal volumes of 2 % NaOH (Sigma-Aldrich, St. Louis, MO) and 1 % N-acteyl-L-cysteine (Sigma-Aldrich, St. Louis, MO) were added to each tube. The tubes were then vortexed. The sputa were pelleted by centrifugation at 10,000 X g for 10 minutes. The supernatant was discarded and the resulting pellet was suspended in 100 µl of extraction buffer (1 % Triton X-100, 0.5 % Tween-20, 10 mM Tris-HCl, 1 mM EDTA). The samples were placed in liquid nitrogen for 3 minutes followed by boiling water for 1 minute. This was repeated four more times to ensure complete lysis of the bacteria. The cell debris was pelleted by a second centrifugation at 10,000 X g for 5 minutes. The supernatant containing the metagenomic DNA was then transferred to a new tube and cleaned using the GeneClean Spin Kit (Qbiogene, Irvine, CA) as per manufacture's protocol. Four hundred microliters of

GeneClean Spin Glassmilk (propriety silica matrix and chaotropic binding salt with a Tris Boric EDTA Modifier) was added to each GeneClean Spin Filter. An amount of 100 to 300 µl of extracted metagenomic DNA was added to each filter. The tubes were inverted every minute for 5 minutes to ensure that the DNA solution binds to the Glassmilk. The tubes were centrifuged at 15,500 X g for 5 minutes. The filtrate was removed from the catch tubes. Then, 500 µl of GeneClean Spin New Wash (NaCl, Tris, and EDTA) was added to the filters. The tubes are centrifuged twice at 15,500 X g for 30 seconds. The catch tubes were emptied and then spun dry for 2 minutes at 15,500 X g. The spin filters were transferred to new microcentrifuge tubes. Fifty microliters of GeneClean Spin Elution Solution (RNase, DNase, pyrogen-free water) was added to the filters. The tubes were centrifuged at 15,500 X g for 30 seconds and then another 30 µl of elution solution was added. The DNA again was eluted for 30 seconds at 15,500 X g. The eluted DNA was refrigerated at 4 °C till further use.

Agarose gel electrophoresis. A one percent (weight/volume) agarose electrophoresis gel (BioWhittaker Molecular Applications, Rockland, ME) was run to determine the successfulness of the genomic DNA extractions. A 1.5 % (weight/volume) agarose electrophoresis gel was used to determine the successfulness of a PCR reaction. Electrophoresis was carried out in 1 X Tris-Borate-EDTA (TBE) buffer (0.89 M Tris-Borate, 0.02 M EDTA, pH 8.0). A 1-kb and 100 base pair (bp) molecular ladders were loaded for size reference for the genomic DNA and the PCR product gels, respectively (New England Biolabs, Beverly, MA). Gels were stained with ethidium bromide and viewed with ultraviolet light.

DNA quantification and dilution. Extracted DNA was quantified using a Versafluor Fluorometer (BioRad, Hercules, CA) with the 360 nm excitation and 460 nm emission filters. Hoechst dye at a 1 mg/ml concentration was used to determine DNA concentrations expected to range between 10 to 500 ng/ml. The Hoeschst dye was prepared by adding 100 µl of 10 mg/ml Hoeschst dye (Bio Rad, Hercules, CA) to 900 µl of sterile water. This dye was stored in the dark at 4 °C. A 10 mg/ml of DNA standard was also prepared by mixing 10 µl of 1 mg/ml calf thymus DNA with 100 µl of 10 X TEN Assay Buffer (100 mM Tris, 2 M NaCl, 10 mM EDTA, pH 7.4) (Bio Rad, Hercules, CA) and 890 µl of sterile water. This reagent was stored at 4 °C. The assay solution was prepared by mixing 2.5 µl of 1 mg/ml Hoeschst dye with 2.5 ml of 10 X TEN assay buffer and 22.5 ml sterile water in a 50 ml conical tube. A standard curve was made to ensure accuracies of quantifications (Table 3). To quantify the unknown samples, 2 ml of the unknown sample was added to 2 ml of the 0.10 mg/ml Hoeschst assay solution. The relative fluorescent units were recorded. The DNA samples were then diluted to a final concentration of 10 ng using DEPC water.

Samples containing less than 1 nanogram of DNA were quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Carlsban California). DNA was quantified as per manufacture's protocol. The Quant-iT™ PicoGreen® reagent was diluted 200 fold with 1X Tris EDTA buffer (TE). For the high-range standard curve, a 2 µg/ml stock solution of calf thymus DNA is prepared in Tris EDTA buffer. The highest amount quantified was 1 µg/ml, which was measured from a cuvette containing 1,000 µl of the diluted Quant-iT™ PicoGreen® reagent. The second point on the standard curve was 100 ng/ml prepared from 100 µl of picogreen and 900 µl of TE. The third point was

Table 3. DNA quantification using Hoescht's dye for standard curve

Total DNA (ng)	DNA Stock Solution (µg/ml)	DNA Volume (2 µl)	0.1 µg/ml Hoeschst Dye Solution (ml)
Blank	---	---	2
20	10	2	2
50	10	5	2
100	10	10	2

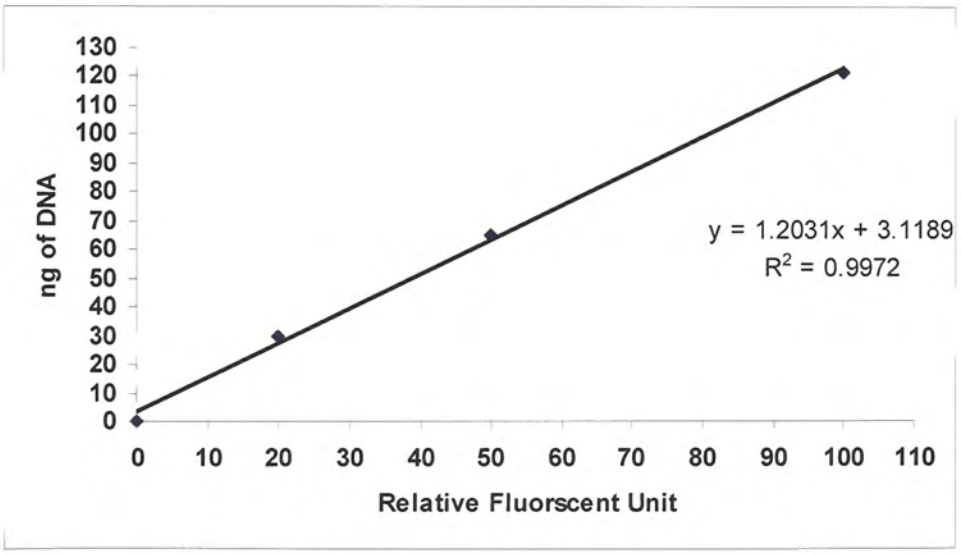


Figure 7. DNA quantification standard curve. Hoescht's dye was used to determine the relative fluorescent units produced by a standard amount of DNA.

at the concentration of 10 ng/ml prepared from 10 µl of picogreen and 990 µl of TE. To measure the unknown samples, 999 µl of water was added to 1000 µl of picogreen. The unknown DNA at the amount of 1 µl was added to each solution and the relative fluorescence was measured. The concentration of the unknown DNA was interpreted from the standard curve.

PCR amplification of 16S rRNA genes for metagenome profiling. The metagenomic DNA derived from the extracted sputum samples were amplified using eubacterial primers for the 16S rRNA gene. The hypervariable region 1 (V1) plus the hypervariable region 2 (V2) of the 16S rRNA gene were amplified using 27F-6FAM forward primer and the 355R reverse primer (Table 2). The first (V1) and third variable regions (V3) were multiplexed using fluorescent primers P1F-6FAM and 338F-6FAM and reverse primers P2R and 518R (Table 2). All forward primers are fluorescently labeled with 6-FAM™ (Integrated DNA Technologies, Shokie, IL). Final concentrations of the V1+V2 PCR reaction mixture was: 1 X PCR buffer, 2.5 mM $MgCl_2$, 0.25 mM dNTPs, 0.5 µM forward and reverse primers, 0.1 % BSA, 0.025 U of AmpliTaq Gold LD DNA polymerse™ (Applied Biosystems, Foster City, CA), and DEPC water to make up the final volume of 20 µl. One microliter of 10 ng DNA template was added per reaction. The final concentration of the multiplex reaction was: 1 X PCR buffer, 2.5 mM $MgCl_2$, 0.25 mM dNTPs, 1 µM forward and reverse primers, 0.1 % BSA, 0.025 U of AmpliTaq Gold LD DNA polymerse™ (Applied Biosystems, Foster City, CA), and DEPC water to make up the final volume of 20 µl. One microliter of 10 ng DNA template was added per reaction.

The following parameters set on a MJ Research Peltier Thermal Cycler 200 (Waltham, MA) were used to amplify the selected fragments: initial denaturation at 95 °C for 11 minutes, 25 cycles of denaturation at 95 °C, annealing at 55 °C and extension at 72 °C, each for 1 minute and a final elongation at 72 °C for 10 minutes. A negative control containing water was amplified for every PCR master mix to check for contamination. As described earlier, an agarose gel was run to confirm the success of amplification. A positive control was performed using either PAO1 or *E. coli* DNA.

ALH analysis. ALH analysis was performed by the Forensic DNA Profiling Facility (Florida International University, Miami, FL). The fragments of the PCR products were detected using the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA). A formamide-size standard mix was prepared using a 96:1 ratio of Hi-Di™ (highly deionized) formamide and GeneScan™ 500 ROX™ internal standard (Applied Biosystems, Foster City, CA). Each PCR product was denatured by adding 0.50 µl to 9.5 µl of the formamide-size standard mix for 2 minutes at 95 °C and snap cooling for 5 minutes on ice. Each sample was run for 28 minutes on the ABI Prism 310. Fragments were separated by capillary electrophoresis using polymer POP-4, matrix DS-3O_6FAM_HEX_NED_ROX and filter D (Applied Biosystems, Foster City, CA).

The output was collected and analyzed using GeneMapper® Software Version 3.7 (Applied Biosystems, Foster City, CA) and Microsoft Excel (Microsoft, Redmond, WA). The ABI Prism™ Genotyper software analysis parameters were set to the local Southern Size calling, no correction, and the minimum noise threshold was set at 70 fluorescent units (63). For V1 + V2, amplicons were called if between 300 and 400 base pairs.

Amplicons for V1 region were between 60 and 120 bases where as the V3 region contained s between 159 and 290 bp.

Three replicate profiles from separate PCR reactions for each triplicate extraction of each sample were compared to evaluate the reproducibility of the fragments produced by PCR and the reproducibility of the extractions. Hence, nine profiles were produced for each sample. Fragments were further analyzed if a) the fragment was present in two out of three of the ALH profiles from each aliquot and b) the relative height ratio was equal or greater than 1 % (63). Relative height ratios for the ALH fingerprint technique were calculated by dividing each individual height by the total height of each electropherogram (63).

Statistical analysis. Data was analyzed using MS Excel (Microsoft, Redmond, WA), and the PRIMER 5 statistical software (PRIMER E Ltd., Plymouth Marine Laboratory, United Kingdom).

Descriptive statistics were performed on the three PCR replicates produced from the three aliquots of every sample and the mean relative ratios from each aliquot were used in the following analyses. The three averaged aliquots were then summed together to produce an overall profile per sample. Macros were developed for MS Excel to expedite these calculations (Robert Sautter, Florida International University, Miami, FL). Analysis of similarity between patients' samples at one location, groups of samples between two locations, and patient profiles over time were based on Bray – Curtis similarity index. This index calculates relationships between samples using the following formula:

$$BC_{jk} = 100 \left\{ 1 - \left| \frac{\sum^{P} i=1 \; y_{ij} - y_{ik}}{\sum^{P} i=1 \; (y_{ij} + y_{ik})} \right| \right\}$$

Where BC_{jk} is the similarity between the *j*th and *k*th sites and y_{ij} represents the abundance for the *i*th species in the *j*th site (75). In addition, multi-dimensional (MDS) cluster analysis was used to compare sample data (6, 31). Shannon-Weaver description statistics were used to determine diversity.

Presumptive identity analysis. Using the size standard, sizes in base pairs were determined. The base pair lengths were then searched against a database of bacterial genera 16S rRNA hypervariable region amplicon lengths (Amplicon Length Heterogeneity Analysis System [ALHAS], George Mason University). One amplicon length may represent more than one bacterium and thus only a presumptive identity of that amplicon can be achieved. Virtual analysis was performed for bacteria of interest that were not found in ALHAS as described in Miller *et al.* Bacterial sequences from NCBI were used to identify expected amplicon lengths for each variable region being studied.

Picogreen is more effective for quantification of small amounts of metagenomic DNA. Extracted DNA was at first quantified using a Versafluor Fluorometer with Hoechst dye (BioRad, Hercules, CA). Hoescht dye with a DNA stock solution of 10 µg/ml measures DNA from 100 ng to as little as 1 ng. Many of the quantified samples contained DNA at amounts near or below the lower threshold of Hoechst dye thus resulting in negative RFU values. These samples were then quantified using Picogreen. The resulting RFU values were above 9999 RFU and thus could not be read by the instrument. Hence, the amount of DNA in the samples was at the higher threshold end of picogreen and at the lower threshold of Hoechst dye. Aliquots of these samples were diluted by a factor of ten and quantified with picogreen (Table 4).

Table 4. Comparison of DNA quantification methods. Samples giving a negative RFU using Hoechst's dye were diluted by a factor of ten and quantified using picogreen.

Sample	Hoescht's Dye RFU	Dilute Picogreen RFU
UMC2	-2	580
UMP1	-9	502
UMS1	-14	437
UMT2	-14	432
JD2b	-22	259

Reproducibility. DNA was extracted from 26 CF sputa samples using the method of Reischl *et al* (76). Each sample was extracted in triplicate to determine the reproducibility of the extraction technique. The extracted samples were amplified using V1 + V2 (27F and 355R) (Table 2) and V1 _V3 multiplex primers (P1 and P2, 338F and 518R, respectively) (Table 2). The electropherogram output from three replicates amplified during the same PCR reaction were analyzed using Bray-Curtis similarity index and non-metric multidimensional scaling analysis. Figure eight illustrates the results obtained using samples from patient UMA, UMB, JD1, and JD2. The amplicons obtained from the V1 + V2 PCR reaction varied from extraction to extraction and no clustering was observed (Figure 8, Panel A). The lack of clustering indicates that the extractions based on presence of amplicons and amplitude was more dissimilar than similar. This trend was similarly seen in the V1_V3 region (Figure 8, Panel B).

In addition, PCR reproducibility was tested. Each aliquot was amplified three times for V1 + V2 region and V1_V3 multiplex and analyzed via the ABI 310 Genetic Analyzer (Applied Biosystems, Foster City, CA). The resulting ALH profiles from JD1 and UMO were overlaid and compared (Figure 9). Blue, black, and green represent each of the three PCR reactions. The three colors overlap at each amplicon length, which is plotted on the X-axis. The amplicon heights varied per reaction. The PCR reproducibility was seen in all other sample aliquots analyzed in this study (data not shown).

PCR amplification of 16S rRNA genes for metagenome profiling. All CF sputum samples were analyzed for eubacteria using the 16S rRNA gene (Figure 6). The

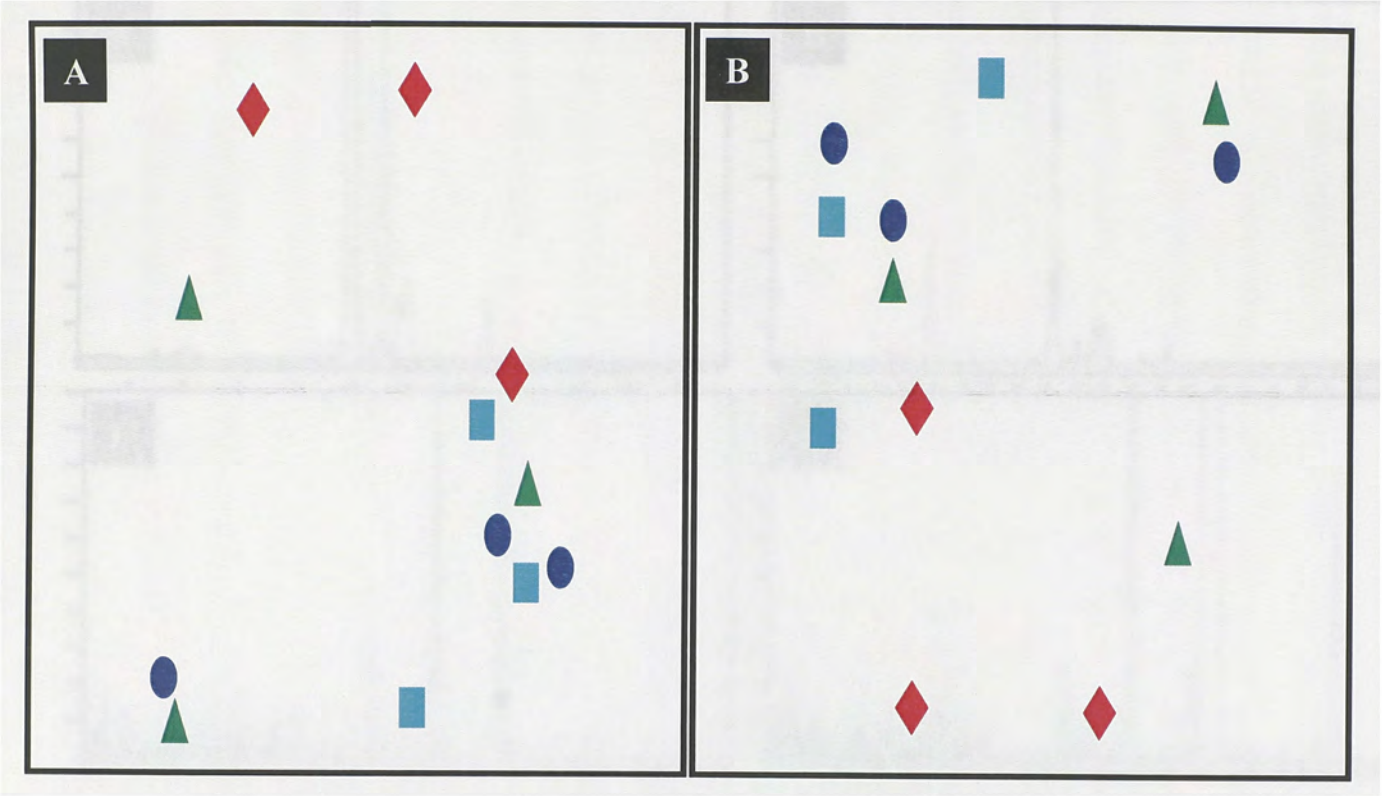


Figure 8. Reproducibility of extraction. Panels A and B represent analyses of V1+V2 and V1, V3 regions, respectively. The amplitude and heights of peaks present in an electropherogram were analyzed using the Bray – Curtis similarity coefficient and multidimensional scaling analysis. Symbols ▲, ■, ●, and ◆, represent Patient UMA, UMB, JD1 and JD2, respectively.

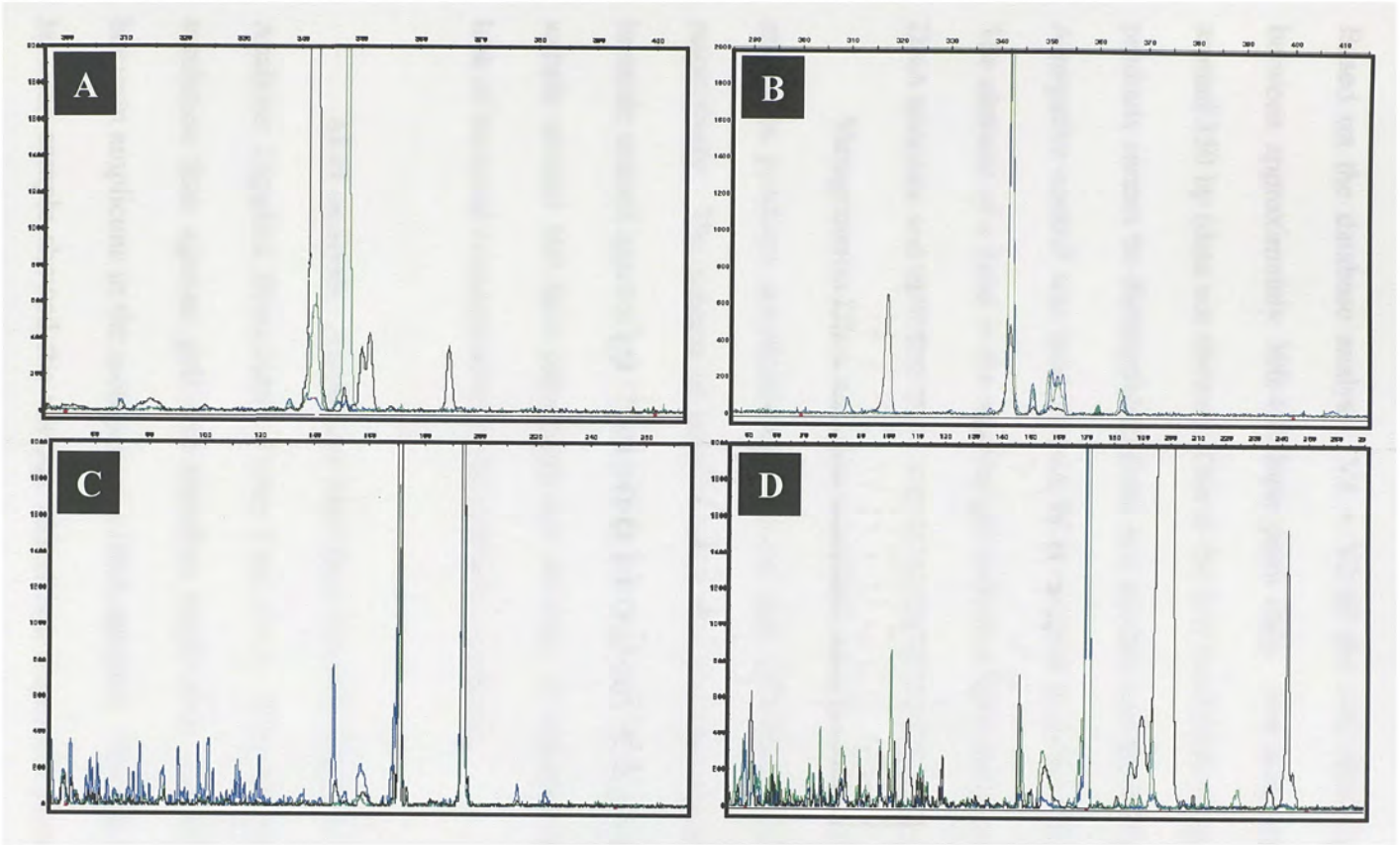


Figure 9. PCR Reproducibility. A single metagenome sample was amplified three times (indicated by blue, black, and green, respectively. and the resulting electropherograms were superimposed. Panels A and B represent V1+V2 and V1, V3 electropherograms, respectively from Patient JD1. Panels C and D represent V1+V2 and V1, V3 electropherograms, respectively from Patient UMO.

V1 + V2 region of the ribosomal gene were successfully amplified and the PCR products were detected on a 1.5 % agarose gel stained with ethidium bromide (data not shown). Based on the database analysis, V1 + V2 of the 16S rRNA gene are known to vary between approximately 300-400 base pairs (bp). The samples produced a wide band around 350 bp (data not shown). Due to the low resolution of agarose gels, multiple PCR products cannot be distinguished from one another and therefore appear as a wide band. A negative control was used in each PCR reaction to detect for bacterial contamination. The absence of a band in the agarose gel indicates PCR did not work due to the lack of a DNA template and therefore there was no bacterial contamination.

Metagenomic DNA was also amplified using a multiplex PCR reaction. V1, V3 multiplex produces amplicons between 60 and 120 bases and between 150 and 290, respectively. The success of amplification was also determined using 1.5 % ethidium bromide stained agarose gel. Two bands, one for each PCR product, can be seen for each sample around 100 base pairs (data not shown). A negative control demonstrated the lack of bacterial contamination in the multiplex reactions.

ALH analysis. Amplicons were then detected using the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA). This instrument has a much higher resolution than agarose gels and therefore single base pair differences were detected between amplicons in the metagenomic DNA samples. The electropherograms produced by each sample showed the various 16S rRNA fragments present in the metagenomic DNA. The eubacterial profiles from the CF patients were analyzed using ANOSIM, Shannon-Weaver Statistics, Bray-Curtis similarity index, and non-metric

multidimensional scaling analysis. Initially, the patient's sputum was divided into three aliquots and the metagenomic DNA was extracted. Each of these three extractions was then amplified three times for each region and analyzed via the ABI 310 Genetic Analyzer (Applied Biosystems, Foster City, CA). Amplicons that were present in two out of three electropherograms of the aliquots were added together to generate one profile per sample and then used for further analysis. The amplicon lengths and intensity of amplicons from V1 + V2 and V1_V3 multiplex were converted to ratio data and use for further analyses.

The abundance of amplicon lengths present in the CF sputa. Our analysis shows that the CF lung can contain one or more species of eubacteria. The complexity of the microbial community found in a CF patient's lung was assessed using Shannon-Weaver statistics. This information was then depicted in graphs which show the amplicon lengths and normalized relative fluorescent units for the averaged three aliquots.

The amplicon sizes from the 26 samples ranged from 335 to 374 bases for the V1 + V2 region (Figure 10). Patient JD6 had the most complex profile with 11 amplicons present for the V1 + V2 region (Figure 10, Panel R). Patients UMF, UMG, UMH, UMJ, UMK, UMP, UMT had the least complex ALH profiles with only one amplicon present at 342 bp (Figure 10, Panel T). The remaining 18 samples produced ALH profiles containing multiple amplicons (Figure 10). Amplicon 342 was the predominate amplicon observed in 24 out of 26 samples.

A
B
C
D
Relative Ratio (± SE)
1.2
1.0
0.8
0.6
0.4
0.2
0
Amplicon Length (BP)
374
367
365
360
359
358
357
356
353
352
351
350
349
348
347
345
342
341
340
339
335

E
F
G
H
Relative Ratio (± SE)
1.2
1.0
0.8
0.6
0.4
0.2
0
Amplicon Length (BP)
374
367
365
360
359
358
357
356
353
352
351
350
349
348
347
345
342
341
340
339
335

I
J
K
L
Relative Ratio (± SE)
1.0
0.8
0.6
0.4
0.2
0
Amplicon Length (BP)
374
367
365
360
359
358
357
356
353
352
351
350
349
348
347
345
342
341
340
339
335

M
N
O
P
Relative Ratio (± SE)
1.0
0.8
0.6
0.4
0.2
0
Amplicon Length (BP)
374
367
365
360
359
358
357
356
353
352
351
350
349
348
347
345
342
341
340
339
335

Q
R
S
T
Relative Ratio (± SE)
1.2
1.0
0.8
0.6
0.4
0.2
0
Amplicon Length (BP)
374
367
365
360
359
358
357
356
353
352
351
350
349
348
347
345
342
341
340
339
335

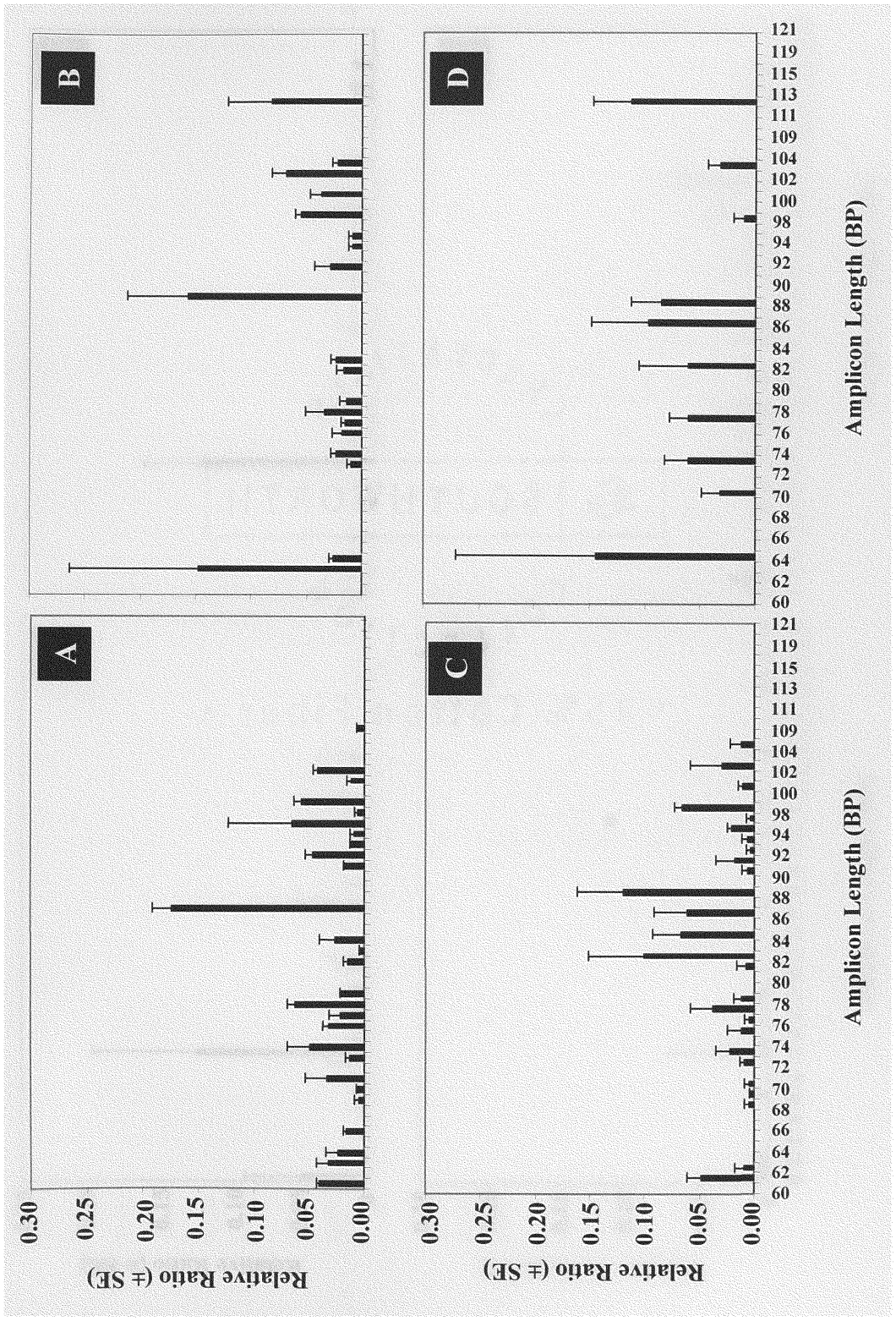
Figure 10. Hypervariable region V1 + V2 amplicon data graphs. Panels A – E represents patients UMA – UME. Panels F – I refers to patients UML, UMM, UMN, UMO, respectively. Panels J, K, and L represents patients UMQ, UMR, and UMS, respectively. Panels M, N, O, P, Q, R, S indicates profiles of JD1 – JD7. Panel T represents profiles from patients UMF, UMG, UMH, UMJ, UMK, and UMT.

V1 profiles ranged from having 60 to 120 bp present (Figure 11). All 26 samples contained multiple amplicons in the profile. The most complex profile contained 49 amplicons for patient JD4 (Figure 11, Panel X). The least diverse profile with ten amplicons belonged to patient UMD (Figure 11, Panel D). The predominant amplicon was found at 98 base pair length and present in 24 out of 26 profiles.

The V3 region contained amplicons from 157 to 225 (Figure 12). Profile from UMA was the least diverse with only one amplicon at 171 bp (Figure 12, Panel A). Twenty-five out of 26 profiles contained more than one amplicon with patient UMT producing the most with 13 amplicon (Figure 12, Panel S). Amplicon 171 was the predominant peak found in all 26 samples.

CF sputum samples produced similar ALH profiles. The eubacterial communities found in the CF sputum were compared to one another to establish the uniqueness of a patient's microbial flora. Samples were compared within the CF center and between centers, Joe DiMaggio Children's Hospital (JD) and University of Miami (UM).

To determine if ALH profiles were similar between patients that were treated at the same CF center, the profiles for each sample were compared to samples attained from the same location. The samples were compared using Bray – Curtis similarity index and non-metric multidimensional scaling analysis (MDS). Each region V1 + V2, and V1_V3 separated and combined was analyzed to determine the highest similarity seen between two patients.


A
B
C
D
Relative Ratio (± SE)
Amplicon Length (BP)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
60
62
64
66
68
70
72
74
76
78
80
82
84
86
88
90
92
94
98
100
102
104
109
111
113
115
119
121

E
F
G
H
0.25
0.20
0.15
0.10
0.05
0
Relative Ratio (± SE)
121
119
115
113
111
109
104
102
100
98
94
92
90
88
86
84
82
80
78
76
74
72
70
68
66
64
62
60
Amplicon Length (BP)

I
J
K
L
Relative Ratio (± SE)
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
Amplicon Length (BP)
121
119
115
113
111
109
104
102
100
98
94
92
90
88
86
84
82
80
78
76
74
72
70
68
66
64
62
60

M
N
O
P
Relative Ratio (± SE)
0.25
0.20
0.15
0.10
0.05
0
Amplicon Length (BP)
60
62
64
66
68
70
72
74
76
78
80
82
84
86
88
90
92
94
98
100
102
104
109
111
113
115
119
121

Q
R
S
T
Relative Ratio (± SE)
Amplicon Length (BP)
0.25
0.20
0.15
0.10
0.05
0
60
62
64
66
68
70
72
74
76
78
80
82
84
86
88
90
92
94
98
100
102
104
109
111
113
115
119
121

U
V
W
X
Relative Ratio (± SE)
Amplicon Length (BP)
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
60
62
64
66
68
70
72
74
76
78
80
82
84
86
88
90
92
94
98
100
102
104
109
111
113
115
119
121

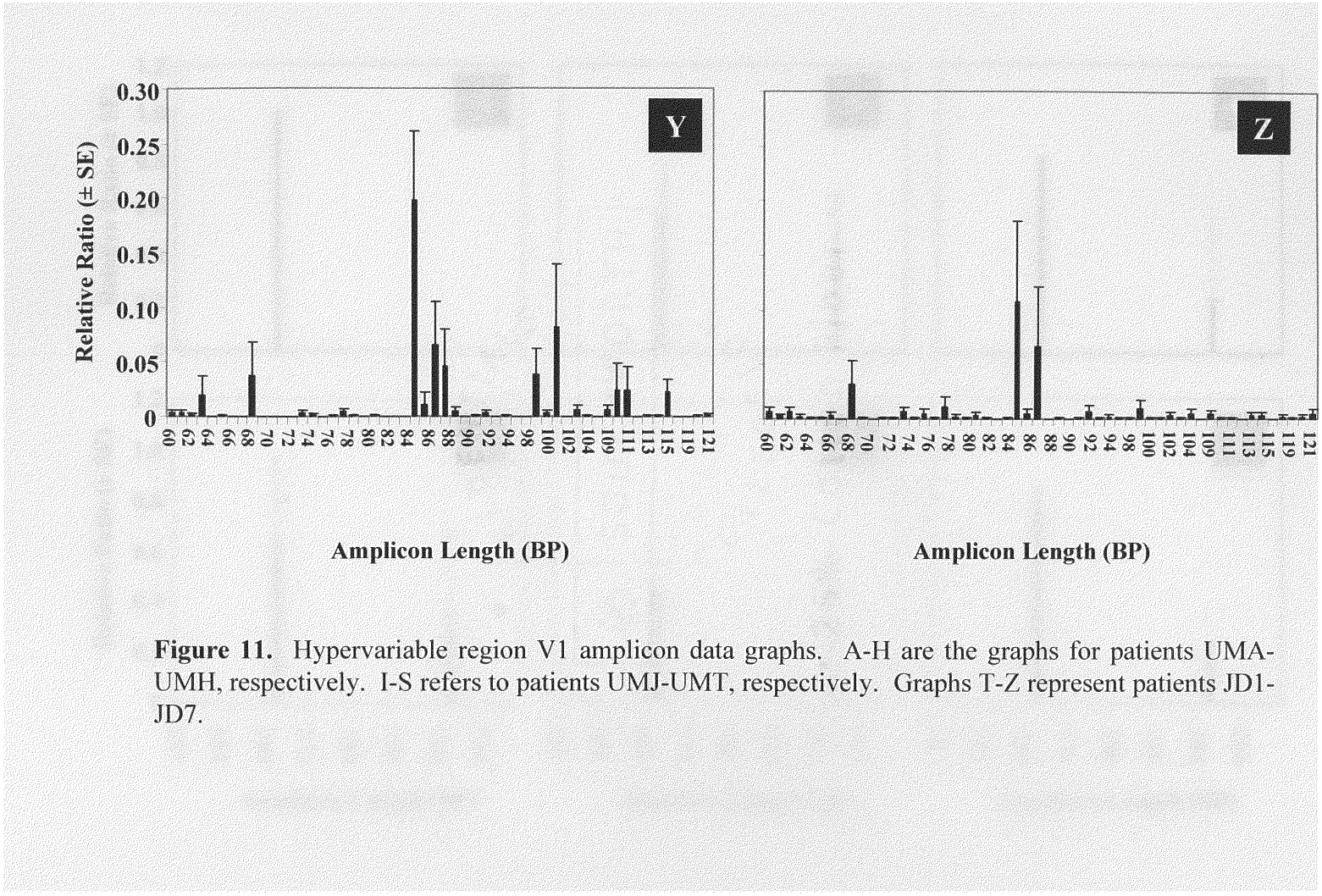


Figure 11. Hypervariable region V1 amplicon data graphs. A-H are the graphs for patients UMA-UMH, respectively. I-S refers to patients UMJ-UMT, respectively. Graphs T-Z represent patients JD1-JD7.

A
B
C
D
E
F
Relative Ratio (± SE)
1.2
1.0
0.8
0.6
0.4
0.2
0
Amplicon Length (BP)
214
194
190
180
174
169
162
157

G
H
I
J
K
L
Relative Ratio (± SE)
1
0.8
0.6
0.4
0.2
0
Amplicon Length (BP)
214
194
190
180
174
169
162
157

M
N
O
P
Q
R
1.0
0.8
0.6
0.4
0.2
0
Relative Ratio (± SE)
214
194
190
180
174
169
162
157
Amplicon Length (BP)

S
T
U
V
W
X
Relative Ratio (± SE)
1.2
1.0
0.8
0.6
0.4
0.2
0
Amplicon Length (BP)
214
194
190
180
174
169
162
157

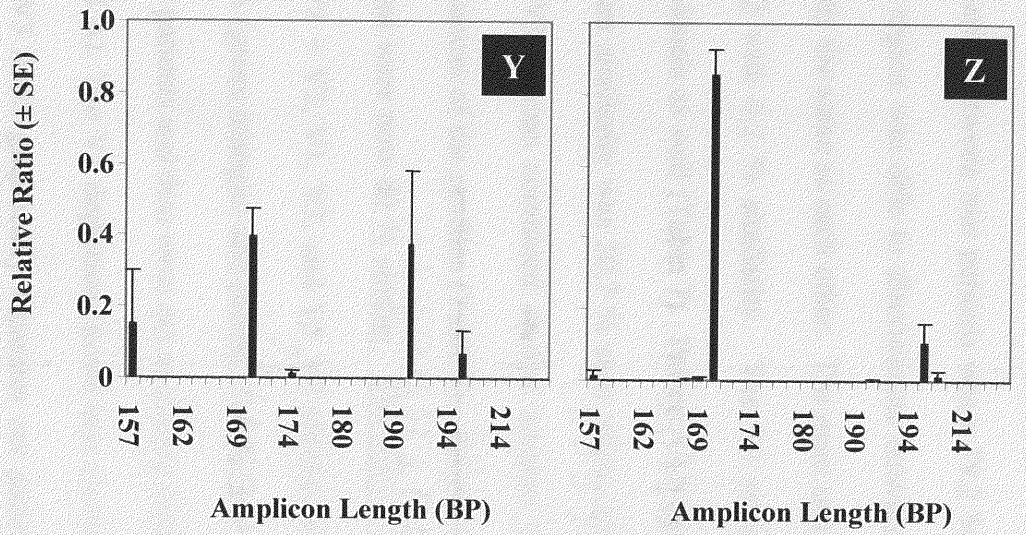


Figure 12. Hypervariable region V3 amplicon data graphs. A-H are the graphs for patients UMA-UMH, respectively. I-S refers to patients UMJ-UMT, respectively. Graphs T-Z represent patients JD1-JD7.

With the UM patients using the V1 + V2 region, patient UMF profile was the most similar to other patients (Table 5). That patient profile was exactly the same as profiles from patient UMG, UMH, UMJ, UMK, UMP, and UMT. Patient UMR had less than 36.2 % similarity with 16 patients but had a 91.0 % similarity with patient UML.

Patients were differentiated more using the V1 region (Table 6). The highest similarity between two patients was a 79.2 % match between patients UMF and UMG. This region was able to discriminate between all patients as that no two profiles were exactly the same as each other. The most unique profile was between patients UMD and UMQ with 6.2 % similarity. The V3 region was not able to differentiate between individuals as well (Table 7). Patient UMH and UMS were 95.5 % similar and the least amount similarity was 20.7 % seen when comparing patient UMB and UMK. The V1 and V3 regions combined caused samples to be more similar and thus lowered the uniqueness of the profiles (Table 8). Patient UMF and UMG when analyzed using both regions were now 80 % similar. The Bray – Curtis similarity was also plotted on MDS for V1 + V2, V1, V3, and V1_V3 regions combined. Visually with V1 + V2, patient UMA (green triangle) and patient UMQ (blue diamond) were less similar than any of the other patients and thus were not as closely clustered to the rest of the symbols (Figure 13, Panel A). The V1 depiction shows less clustering for all patients (Figure 13, Panel B). The V3 MDS graphically separates out patient UMA (green triangle) and UMB (blue triangle) (Figure 13, Panel C). The V1 and V3 region combined shows similar clustering as the V3 region alone (Figure 13, Panel D).

Patients from the JD center were compared in the same manner as above. For the V1 + V2 region, patient JD3 and JD7 had a 93.5 % similarity (Table 9, Panel A). This

Table 5. V1 + V2 region compared amongst UM patients.

	UMA	UMB	UMC	UMD	UME	UMF	UMG	UMH	UMJ	UMK	UML	UMM	UMN	UMO	UMP	UMQ	UMR	UMS	UMT
UMA	100.0																		
UMB	50.3	100.0																	
UMC	53.0	86.4	100.0																
UMD	50.0	94.5	85.0	100.0															
UME	50.0	92.5	79.0	94.0	100.0														
UMF	50.0	92.5	79.0	94.0	98.0	100.0													
UMG	50.0	92.5	79.0	94.0	98.0	100.0	100.0												
UMH	50.0	92.5	79.0	94.0	98.0	100.0	100.0	100.0											
UMJ	50.0	92.5	79.0	94.0	98.0	100.0	100.0	100.0	100.0										
UMK	50.0	92.5	79.0	94.0	98.0	100.0	100.0	100.0	100.0	100.0									
UML	43.0	42.2	43.0	42.0	40.0	40.0	40.0	40.0	40.0	40.0	100.0								
UMM	49.8	93.0	80.6	95.5	97.5	98.5	98.5	98.5	98.5	98.5	40.8	100.0							
UMN	50.0	93.5	81.0	93.0	92.0	92.0	92.0	92.0	92.0	92.0	40.0	91.5	100.0						
UMO	63.0	75.4	80.0	77.0	75.0	75.0	75.0	75.0	75.0	75.0	43.0	74.6	76.0	100.0					
UMP	50.0	92.5	79.0	94.0	98.0	100.0	100.0	100.0	100.0	100.0	40.0	98.5	92.0	75.0	100.0				
UMQ	54.0	43.2	44.0	41.0	41.0	39.0	39.0	39.0	39.0	39.0	44.0	39.8	39.0	52.0	39.0	100.0			
UMR	36.0	36.2	36.0	36.0	36.0	36.0	36.0	36.0	36.0	36.0	91.0	35.8	36.0	36.0	36.0	36.0	100.0		
UMS	50.0	89.4	80.0	88.0	88.0	88.0	88.0	88.0	88.0	88.0	40.0	87.6	95.0	75.0	88.0	39.0	36.0	100.0	
UMT	50.0	92.5	79.0	94.0	98.0	100.0	100.0	100.0	100.0	100.0	40.0	98.5	92.0	75.0	100.0	39.0	36.0	88.0	100.0

A match between two patient profiles is indicated by 100.0 %. The blue indicates that patient UMF had a profile that was the most similar to other patients. Patient UMR in green had the most unique profile.

Table 6. The V1 region compared amongst UM patients.

	UMA	UMB	UMC	UMD	UME	UMF	UMG	UMH	UMJ	UMK	UML	UMM	UMN	UMO	UMP	UMQ	UMR	UMS	UMT
UMA	100.0																		
UMB	36.5	100.0																	
UMC	40.5	53.4	100.0																
UMD	27.3	41.5	40.7	100.0															
UME	26.5	27.0	36.0	21.9	100.0														
UMF	24.6	25.0	28.9	18.9	50.9	100.0													
UMG	18.9	19.2	24.4	18.2	59.6	79.2	100.0												
UMH	33.3	25.4	33.3	15.0	39.3	54.8	48.1	100.0											
UMJ	26.2	35.0	34.7	17.1	54.0	43.8	42.9	42.9	100.0										
UMK	27.8	19.5	22.0	16.0	46.4	52.6	57.1	63.5	36.9	100.0									
UML	35.3	20.9	19.6	12.5	33.8	17.9	17.1	19.0	32.6	25.3	100.0								
UMM	34.2	24.3	34.4	28.6	31.0	47.5	43.1	64.6	38.8	56.7	12.3	100.0							
UMN	35.3	29.1	40.0	32.9	40.0	59.0	52.8	56.7	43.5	41.9	21.7	53.1	100.0						
UMO	18.2	22.2	23.3	20.0	43.1	53.8	40.9	41.4	36.7	37.7	21.6	40.0	56.1	100.0					
UMP	8.2	25.0	16.2	6.9	41.0	15.0	18.8	17.4	50.0	19.5	25.8	9.3	22.2	33.3	100.0				
UMQ	13.3	23.3	14.8	6.2	43.5	29.8	25.6	18.9	47.3	29.2	34.8	8.0	23.1	37.2	58.1	100.0			
UMR	34.7	23.5	18.6	9.9	29.0	22.2	18.2	20.3	42.3	18.8	56.5	9.1	26.5	40.7	42.6	44.4	100.0		
UMS	20.5	34.8	25.8	10.4	51.7	33.9	39.2	36.9	68.7	33.3	29.6	29.0	34.4	40.0	55.8	48.0	48.5	100.0	
UMT	18.0	31.7	16.3	7.3	34.9	28.1	21.4	28.6	55.6	21.5	30.2	17.9	29.0	36.7	50.0	61.8	42.3	65.7	100.0

A match between two patient profiles is indicated by 100.0 %. Green indicates that patient UMF had a profile that was the most similar to another patient. Blue shows that patient UMQ had the most unique profile.

Table 7. The V3 region compared amongst UM patients.

	UMA	UMB	UMC	UMD	UME	UMF	UMG	UMH	UMJ	UMK	UML	UMM	UMN	UMO	UMP	UMQ	UMR	UMS	UMT
UMA	100.0																		
UMB	58.1	100.0																	
UMC	56.0	60.4	100.0																
UMD	40.6	38.9	74.7	100.0															
UME	33.3	39.1	67.9	62.4	100.0														
UMF	33.3	39.1	67.9	88.0	67.1	100.0													
UMG	29.8	24.7	55.2	78.5	50.7	80.0	100.0												
UMH	35.0	41.0	70.6	74.4	88.2	77.9	60.3	100.0											
UMJ	35.0	41.0	70.6	67.8	92.6	72.1	54.8	93.9	100.0										
UMK	33.3	20.7	54.7	80.0	50.0	81.4	81.3	60.3	54.4	100.0									
UML	45.2	46.2	83.3	75.7	69.5	67.8	59.4	71.9	71.9	59.3	100.0								
UMM	33.3	39.1	67.9	84.8	80.0	87.1	68.0	91.2	85.3	68.6	67.8	100.0							
UMN	33.3	39.1	67.9	88.0	77.1	90.0	70.7	88.2	82.4	71.4	67.8	97.1	100.0						
UMO	31.8	28.6	60.0	83.7	54.2	84.7	93.5	64.3	58.6	84.7	63.9	72.2	75.0	100.0					
UMP	27.7	32.7	58.5	77.5	85.4	82.8	68.3	86.3	86.3	67.5	60.7	89.2	89.2	72.0	100.0				
UMQ	30.1	35.4	62.6	82.1	61.7	83.2	71.7	70.3	64.8	71.1	64.6	77.9	80.5	75.8	77.1	100.0			
UMR	36.4	35.0	68.7	79.7	60.2	70.7	65.7	71.3	65.1	64.7	84.7	70.7	70.7	70.1	65.3	71.8	100.0		
UMS	34.1	40.0	69.2	69.9	88.4	73.9	58.1	95.5	94.0	56.5	70.7	87.0	84.1	62.0	85.2	69.4	70.2	100.0	
UMT	36.4	42.5	72.7	72.9	81.2	76.7	60.1	89.9	86.8	58.6	73.9	85.7	85.7	65.7	78.7	73.2	74.6	91.6	100.0

A match between two patient profiles is indicated by 100.0 %. Green indicates that patient UMH had a profile that was the most similar to another patient. Blue shows that patient UMK had the most unique profile.

Table 8. The V1 and V3 regions combined to compare amongst UM patients.

	UMA	UMB	UMC	UMD	UME	UMF	UMG	UMH	UMJ	UMK	UML	UMM	UMN	UMO	UMP	UMQ	UMR	UMS	UMT
UMA	100.0																		
UMB	39.8	100.0																	
UMC	44.4	55.3	100.0																
UMD	31.8	40.6	56.3	100.0															
UME	29.4	32.3	53.3	47.5	100.0														
UMF	28.3	31.2	50.0	62.3	62.6	100.0													
UMG	24.0	21.9	42.4	58.7	52.8	79.8	100.0												
UMH	34.0	31.8	52.5	50.7	73.1	70.7	57.0	100.0											
UMJ	29.7	37.4	53.0	47.3	80.4	63.0	51.5	76.2	100.0										
UMK	30.2	20.0	39.6	56.0	49.0	73.1	75.4	61.3	48.8	100.0									
UML	38.4	29.1	46.9	45.2	55.4	48.0	44.4	49.5	55.0	45.7	100.0								
UMM	33.8	30.7	52.3	63.4	65.7	75.4	61.7	82.6	70.0	65.0	45.2	100.0							
UMN	34.5	33.3	54.7	66.7	66.0	80.6	66.0	77.8	69.3	62.4	48.8	83.3	100.0						
UMO	24.2	25.1	43.9	61.3	51.3	76.5	81.8	57.6	52.0	72.1	48.0	63.3	69.7	100.0					
UMP	18.1	29.0	42.6	57.0	76.5	69.0	60.3	70.4	77.6	57.6	49.7	72.0	74.3	65.0	100.0				
UMQ	21.2	29.1	42.9	57.3	57.4	70.4	62.6	56.6	60.0	60.9	54.1	60.3	65.7	67.3	74.1	100.0			
UMR	35.4	28.1	43.9	51.3	50.3	55.1	52.5	53.5	57.0	49.7	72.4	50.3	55.7	61.2	59.9	64.3	100.0		
UMS	26.1	37.0	48.7	47.0	77.6	61.9	53.3	76.4	85.6	49.5	53.8	69.0	68.3	55.8	78.8	64.0	62.9	100.0	
UMT	25.1	36.0	44.7	46.0	66.3	60.9	49.2	68.3	75.6	46.5	54.8	63.0	66.3	56.9	71.7	70.1	62.9	82.8	100.0

A match between two patient profiles is indicated by 100.0 %. Green indicates that patient UMF had a profile that was the most similar to other patients.

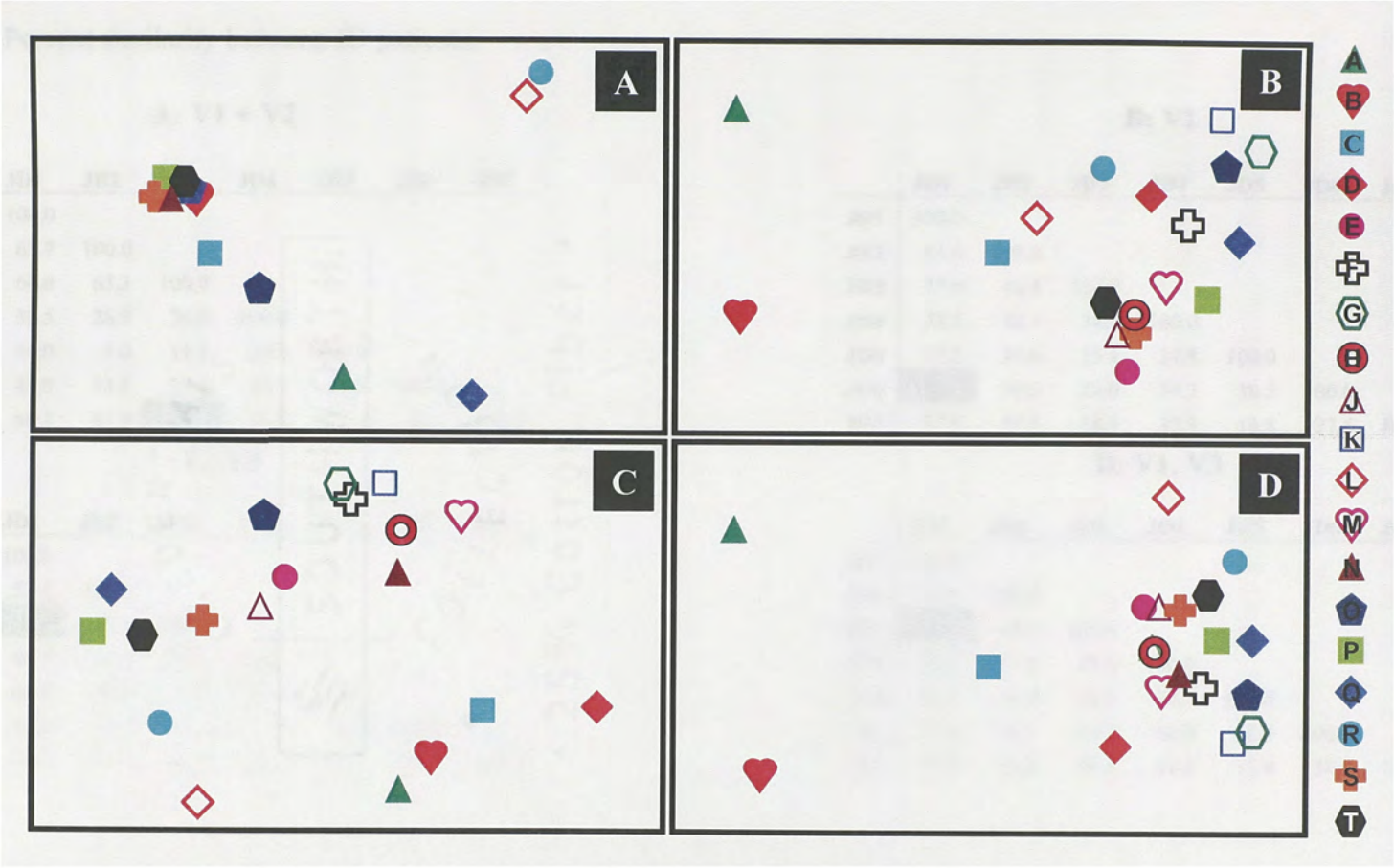


Figure 13. Similarity between UM patients. Multidimensional scaling analysis based on Bray – Curtis similarity coefficient. Panels A, B, and C represent (V1+V2), V1, and V3 regions, respectively. Panel D represents V1 and V3 regions combined.

Table 9. Percent similarity between JD patients.

A: V1 + V2

	JD1	JD2	JD3	JD4	JD5	JD6	JD7
JD1	100.0						
JD2	65.7	100.0					
JD3	64.0	63.3	100.0				
JD4	53.5	26.9	30.0	100.0			
JD5	14.0	1.0	11.1	20.0	100.0		
JD6	43.0	33.2	35.4	38.0	33.3	100.0	
JD7	60.7	63.0	93.5	26.9	11.1	29.1	100.0

B: V1

	JD1	JD2	JD3	JD4	JD5	JD6	JD7
JD1	100.0						
JD2	41.0	100.0					
JD3	37.9	46.4	100.0				
JD4	52.5	48.7	34.5	100.0			
JD5	37.2	39.6	15.4	24.8	100.0		
JD6	66.7	30.0	20.0	39.2	38.5	100.0	
JD7	37.9	57.1	38.9	37.9	19.8	27.5	100.0

C: V3

	JD1	JD2	JD3	JD4	JD5	JD6	JD7
JD1	100.0						
JD2	57.1	100.0					
JD3	78.5	49.3	100.0				
JD4	60.3	30.3	50.4	100.0			
JD5	62.8	58.4	51.4	46.5	100.0		
JD6	47.9	30.3	36.5	41.7	39.4	100.0	
JD7	67.1	58.7	60.4	47.1	49.1	31.7	100.0

D: V1, V3

	JD1	JD2	JD3	JD4	JD5	JD6	JD7
JD1	100.0						
JD2	50.8	100.0					
JD3	66.3	48.5	100.0				
JD4	57.1	37.6	45.6	100.0			
JD5	48.2	48.0	34.7	34.2	100.0		
JD6	57.6	30.2	29.7	40.4	38.8	100.0	
JD7	58.6	58.3	56.4	44.4	35.8	30.0	100.0

Panel A and B are representative of V1 + V2 and the V1 regions, respectively. Panel C refers to the V3 region. Panel D is the similarity seen when the V1 and V3 regions are combined. A match between two patient profiles is indicated by 100.0 %. Green indicates that patient had a profile that was the most similar to other patients.

was decreased to 38.9 % when analyzed using the V1 region (Table 9, Panel B). The highest similarity was 66.7 % between patients JD1 and JD 6 using V1. The V3 region again was less discriminatory given 78.5 % similarity between patients JD1 and JD3 (Table 9, Panel C). Combined, the V1 and V3 regions produced a 66.3 % match between JD1 and JD3 (Table 9, Panel D). JD3 and JD7 were now 56.4 % similar. The Bray – Curtis similarity was also plotted on MDS for V1 + V2, V1, V3, and V1_V3 regions combined (Figure 14). Very little clustering was seen for all regions except V1 + V2 and it would appear that the patients were different from one another.

The sample profiles attained from UM were compared against JD samples using MDS derived from the Bray-Curtis similarity index and ANOSIM (Figure 15). The V1 + V2 region appears to cluster some of the UM patients separately from the JD patients (Figure 15, Panel A). Three out of the seven JD patients were separated from the UM patients. The samples that clustered together indicate the profiles were similar based on the presence of amplicons and their corresponding heights. Using ANOSIM, a 58.4 % similarity ($p < 0.006$) was calculated between the two groups. The V1 region produced 52.6 % similarity ($p < 0.001$) between the groups of patients. Visually, the two groups separated into two clusters (Figure 15, Panel B). The V3 region showed a lack of clustering and there was an overall 73.4 % similarity (Figure 15, Panel C). The V1, V3 regions did graphically separate the patients by treatment location but two JD patients were clustered with UM patients (Figure 15, Panel B). Statistically, there was a 57.8 % similarity ($p < 0.001$) between the two sets of patients.

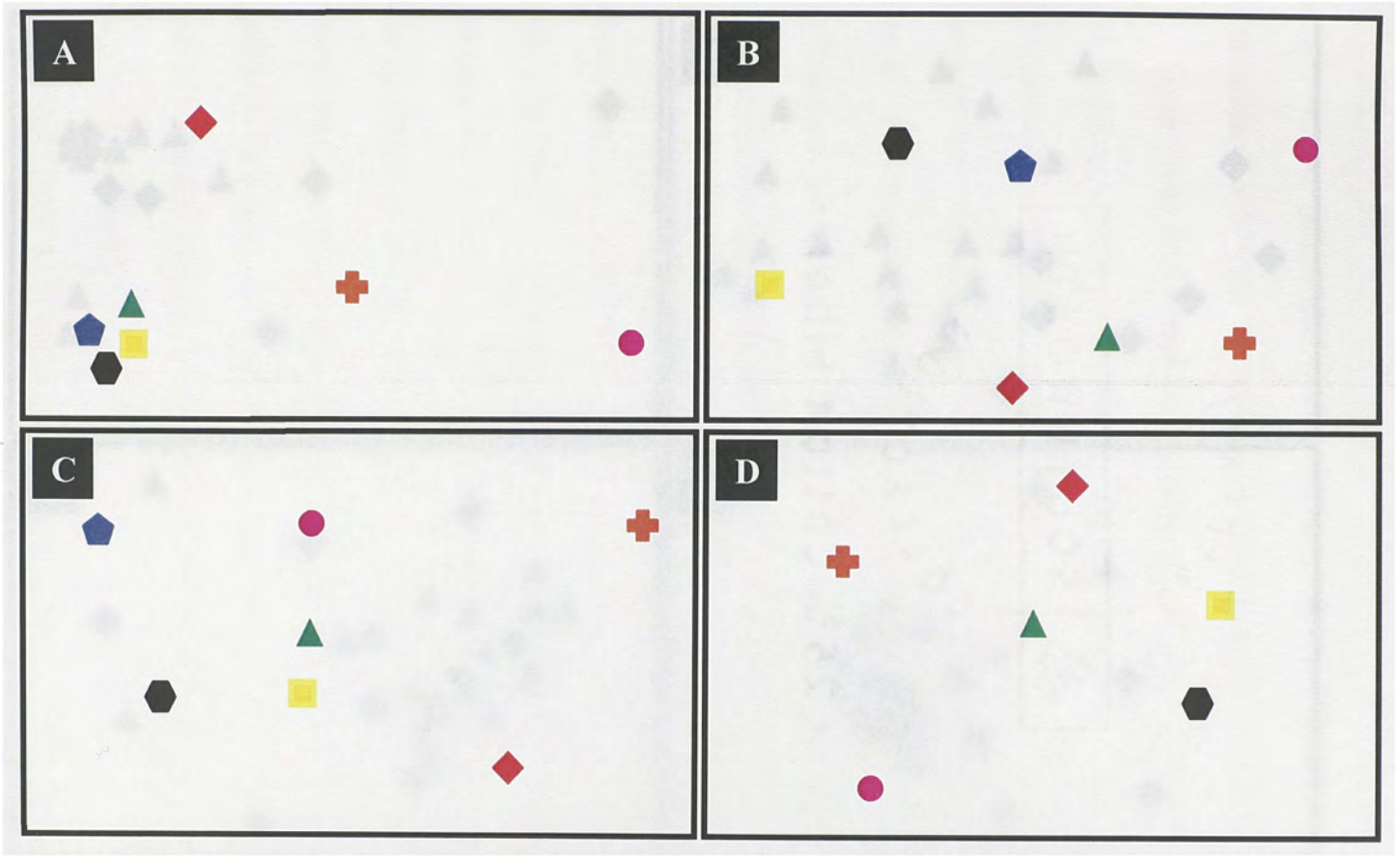


Figure 14. Similarity between JD patients. Multidimensional scaling analysis based on Bray – Curtis similarity coefficient. Green, blue, yellow, red, purple, orange, black symbols represent patients JD1, JD2, JD3, JD4, JD5, JD6, and JD7, respectively. Panels A, B, and C represent V1+V2, V1, and V3 regions, respectively. Panel D represents V1 and V3 regions combined.

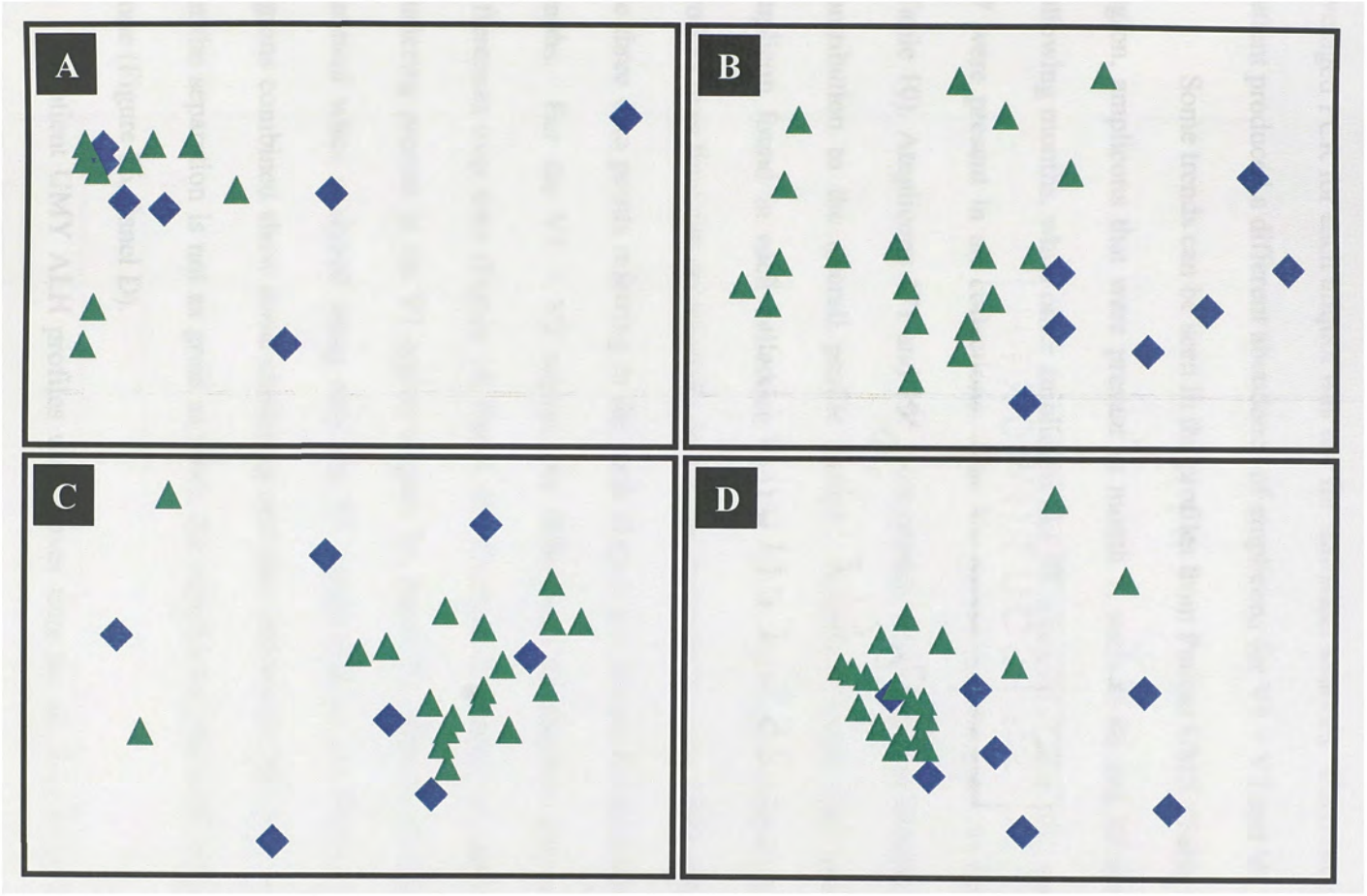


Figure 15. Similarity between hospitals. Multidimensional scaling analysis based on Bray – Curtis similarity coefficient. Blue and green triangles corresponds to JD and UM patients, respectively. Panels A, B, and C represent V1+V2, V1, and V3 regions, respectively. Panel D represents V1 and V3 regions combined.

Patient CF flora changes over time. Two patients referred to as Patient UMX and Patient UMY gave a sputum sample every other month for five months. The three averaged PCR for each aliquot was use for statistical analysis. Each time period for each patient produced a different abundance of amplicons for V1 + V2 and V1_V3 regions.

Some trends can be seen in the profiles from Patient UMX (Table 10). For the V1 region, amplicons that were present in month 1, such as 86 and 88 disappeared in the following months, while other amplicons like 98 appeared (Table 10). Amplicons 73 and 77 were present in all collections. The V3 region did not seem to fluctuate as much (Table 10). Amplicons 171 and 195 were present during all five months but the percent contribution to the overall profile varied. Amplicon length 342 was the dominant amplicon found in each collection for the V1 + V2 region (Table 10). Changes in profiles over time can occasionally be noticed more easily using MDS (Figure 16). There are three data points referring to the each aliquot per sample for each time period in the graphs. For the V1 + V2 region, the differences in aliquots vary as much as the differences over time (Figure 16, Panel A). Changes over time can easily be seen in the clustering present in the V1 region (Figure 16, Panel B). Nearly, all time points appear identical when analyzed using only the V3 region (Figure 16, Panel C). The V1_V3 regions combined show some clustering and thus differences between collection months but the separation is not as great as when the samples are analyzed using the V1 region alone (Figure 16, Panel D).

Patient UMY ALH profiles varied over time for all three hypervariable regions (Table 11). For the V1 region, amplicons that were present in month 1, such as 62 and 88 disappeared in the following months, while other amplicons for example, 85 appeared

Table 10. Amplicon contributions for the Patient UMX over a five-month time period. Each peak is present in the average contribution calculated from the overall profile.

16S Region	Amplicon Size (bp)	Month 1 Contribution (%)	Month 3 Contribution (%)	Month 5 Contribution (%)
V1	60	0	8.45	0
	61	0	8.45	0
	62	0	0	5.18
	73	11.52	6.34	2.29
	77	13.55	8.45	3.77
	85	0	0	6.36
	86	14.85	0	0
	88	17.58	0	0
	91	0	6.34	0
	98	0	8.45	5.77
	102	0	0	8.52
	103	0	23.92	0
	108	0	0	36.92
	112	24.21	4.23	0
	113	0	4.23	0
	115	0	4.23	0
	119	0	0	4.37
V3	171	40.28	45.19	72.06
	189	0	0	17.01
	194	0	32.20	0
	195	59.72	21.43	10.93
V1+V2	342	97.13	100	92.32

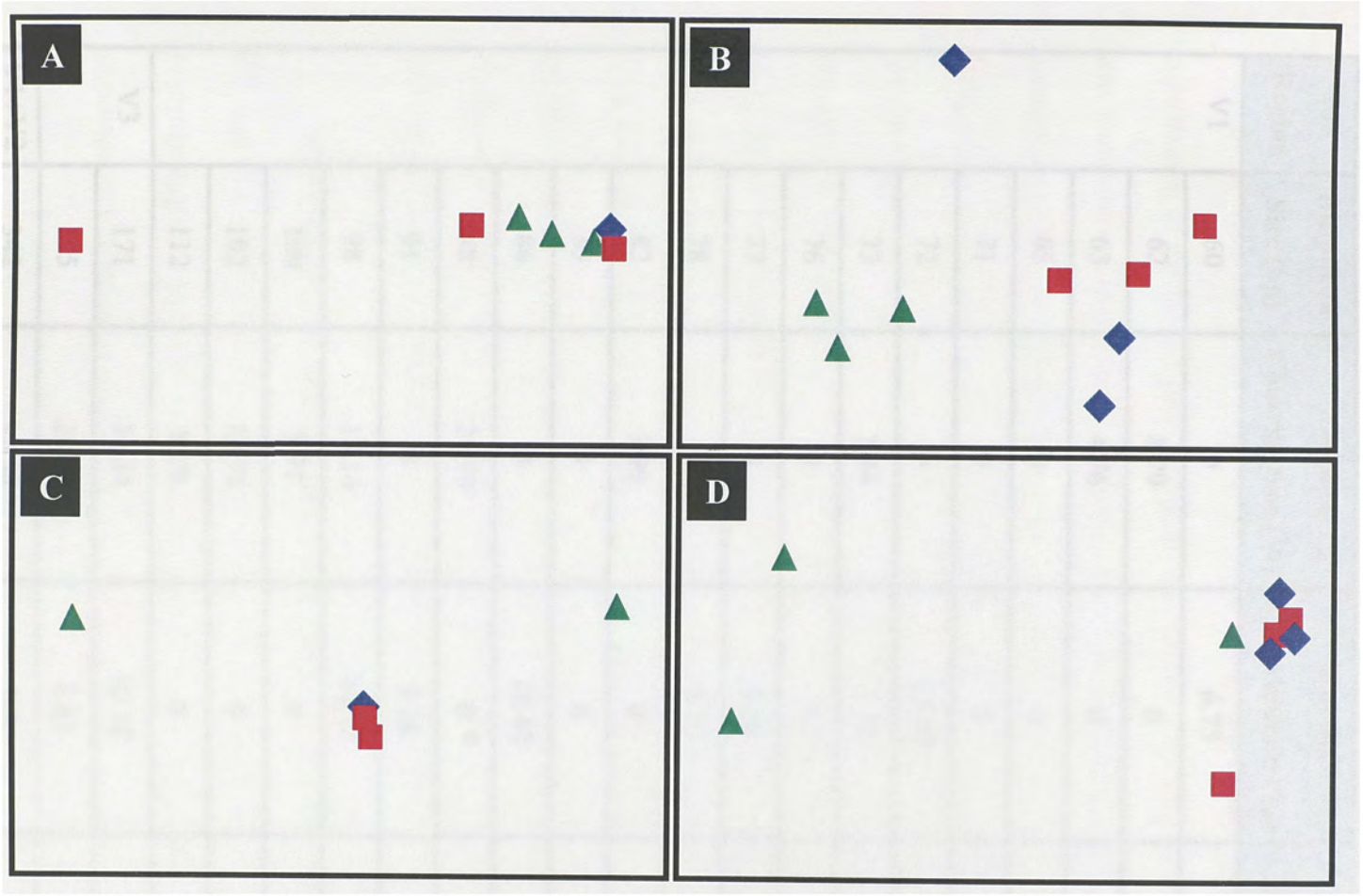


Figure 16. Patient UMX profiles over a period of time. Green triangles, blue diamonds and red squares represent month 1, 3 and 5, respectively. Each symbol is one aliquot of the sample collected at a certain time period. Panels A, B, and C represent V1+V2, V1, and V3 regions, respectively. Panel D represents V1 and V3 regions combined.

Table 11. Amplicon contributions for the Patient UMY over a five-month time period. Each peak is present in the average contribution calculated from the overall profile.

16S Region	Amplicon Size (bp)	Month 1 Contribution (%)	Month 3 Contribution (%)	Month 5 Contribution (%)
V1	60	0	6.75	4.79
	62	8.79	0	0
	63	4.76	0	0
	65	0	0	0
	71	0	0	4.76
	72	0	25.32	0
	73	3.44	7.72	0
	76	0	0	5.77
	77	0	3.28	0
	78	0	7.72	0
	82	3.99	0	0
	85	0	0	8.55
	86	0	18.42	0
	88	24.79	0	0
	91	0	3.28	0
	98	11.14	10.13	7.51
	100	5.97	0	0
	102	11.91	0	38.59
	112	8.79	0	5.47
V3	171	71.34	97.18	91.67
	195	28.66	2.82	8.33
V1+V2	342	91.58	100	38.50
	348	9.42	0	45.50

(Table 11). Amplicon 98 was present at all three time points. Amplicons 171 and 195 contributed the most to the V3 profiles. Over the five months, amplicon 195 decreased and then increased for month five. Amplicon 171 was the dominate amplicon seen in all collections. Amplicons 171 and 195 were present during all five months but the percent contribution to the overall profile varied. For the V1 + V2 region, amplicon length 342 was the dominant amplicon found in the first two time periods but then was drastically decreased for the last month (Table 11). The fluctuating profiles can be depicted using MDS (Figure 17). For the V1 + V2 region, the profiles generated from the fifth month clearly cluster separately from the other time periods (Figure 17, Panel A). Clustering can also been seen between the first and third month profiles. Changes over time can easily be seen in the clustering present in the V1 region and though the aliquots from the third month are spread apart they are still more closely clustered than points from the other time periods (Figure 17, Panel B). All symbols cluster very close to one another, except for one aliquot from month three, when analyzed using only the V3 region (Figure 17, Panel C). The V1_V3 regions combined show some clustering and thus differences between profiles at the different time points (Figure 17, Panel D).

Presumptive identity analysis. The amplicons present in the patient's V1 + V2, V1, and V3 profiles can be compared against the Amplicon Length Heterogeneity Analysis System (ALHAS) database to determine to possible identity (Table 12). One amplicon can represent a multitude of organisms. For example, amplicon 342 was present in 24 out of the 26 samples and it could represent *Pseudomonas* species, *Burkholderia* species, and *Ralstonia* species. Amplicon length 98 for the V1 region was

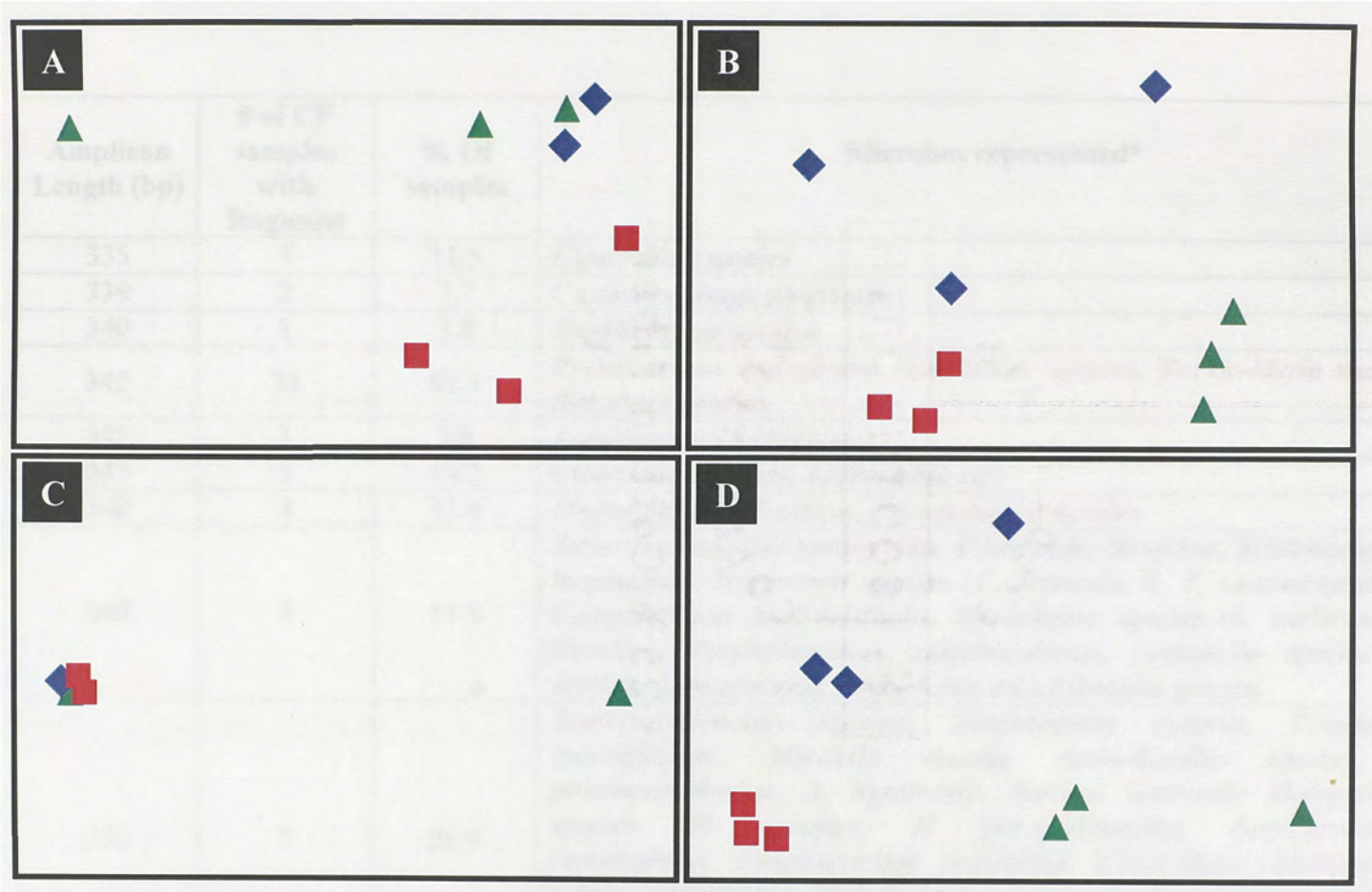


Figure 17. Patient UMY profiles over a period of time. Green triangles, blue diamonds and red squares represent month 1, 3 and 5, respectively. Each symbol is one aliquot of the sample collected at a certain time period. Panels A, B, and C represent V1+V2, V1, and V3 regions, respectively. Panel D represents V1 and V3 regions combined.

Table 12. Amplicon Length Heterogeneity Analysis System (ALHAS) database search for the most organisms represented by the various V1+V2 fragment lengths present in the metagenomic CF sputum samples.

Amplicon Length (bp)	# of CF samples with fragment	% Of samples	Microbes represented*
335	3	11.5	*Clostridium* species
339	2	7.7	*Capnocytophaga gingivalis,*
340	1	3.8	***Burkholderia* species**
342	24	92.3	***Pseudomonas aeurginosa* and other species, *Burkholderia* species, *Ralstonia* species**
345	1	3.8	***Pseudomonas aeruginosa***
347	5	19.2	*Clostridium* species, *Escherichia coli*
348	4	15.4	***Haemophilus influenzae, Pseudomonas* species**
349	3	11.5	***Stenotrophomonas maltophilia,*** *Bacteroides forsythus, Bifidobacterium inopinatum, Treponema* species (*T. denticola II, T. saccharophilum*), *Campylobacter hyointestinalis, Spirochaeta* species (*S. zuelzerae, S. litoralis*), *Porphyromonas asaccharolytica, Legionella* species (*L. erythra, L. bozemanii*), *Escherichia coli, Eikenella* species.
350	7	26.9	***Stenotrophomonas* species,** *Xanthomonas cynarae, Treponema succinifaciens, Moritella viscosa, Actinobacillus* species *(A. pleuropneumoniae, A. lignieresii), Sarcina ventriculi, Haemophilus* species *(H. parasuis, H. paragallinarum), Saccharococcus thermophilus, Fusobacterium prausnitzii, Clostridium subterminale, Vibrio aestuarianus, Photobacterium angustum, Pasteurella* species *(P. aerogenes, P. canis, P. volantium, P. dagmatis, P. avium, P. gallinarum, P. stomatis, P. volantium, P. multocida), Arcanobacterium pyogenes.*

351	4	15.4	*Brevibacterium* species, *Bacillus* species *(B. subtilis, B. fusiformis) Desulfotomaculum nigrificans, Clostridium* species *(C. estertheticum, C. botulinum), Arcobacter* species *(A. cryaerophilus, A. nitrofigilis, A. skirrowii, A. butzleri), Flavobacterium mizutaii, Haemophilus somnus.*
352	5	19.2	*Bacteroides Fragilis, Bacillus* species, *Clostridium* species, *Treponema species*
353	1	3.8	*Porphyromonas circumdentaria, Porphyromonas endodontalis Porphyromonas gingivalis*
357	1	3.8	*Bacillus* species, *Chlyamdia* species
360	2	7.7	*Abiotrophia defectiva*
365	10	38.4	*Lactobacillus mali, Vagococcus fluvialis, Tissierella praeacuta,* Unidentified Bacteria
367	2	7.7	Unidentified Bacteria
374	1	3.8	Unidentified Bacteria

* Organisms in bold and red are known pulmonary flora and CF pathogens.

present in 21 samples and the most likely hit in the database belonged to unidentified oral bacteria. The V3 region for 23 patients contained a length of 195 base pairs, which may represent *P. aeruginosa.* The ALHAS databases for the V1 and V3 regions do not contain many human pathogenic related bacteria and therefore many of the amplicon lengths seen in the profiles did not correspond to a bacterium in the database. Therefore, virtual analysis was performed to determine the possible lengths that may be present for the most common CF pathogens. Sequences from the 16S rRNA gene of these pathogens were analyzed to determine the amplicon length that would be amplified if the organism's DNA was amplified using the V1 and V3 primers (Table 12). Through the use of virtual analysis, it was determined that the most common CF pathogens would have a length of 197 for the V3 region, except for *Ralstonia picketti*, which falls at 187. No profiles contained the 197 or 187 amplicon. For the V1 region, amplicon length 86 was present in ten out of the 26 profiles. Nine profiles contained an amplicon at 88 bp. Twelve profiles produced an amplicon at 91, two profiles for 92, and zero for 96.

Table 13. Virtual analysis of common CF pathogens for V1 and V3 regions.

Pathogen	V1 Region (bp)	V3 Region (bp)	V1 + V2 Region (bp)
Alcaligenes xylosoxidans	86	197	348
Burkholderia gladioli	88	197	340, 342
Burkholderia cepacia	88	197	340, 342, 343
Haemophilus influenzae	92	197	348
Pseudomonas aeruginosa	86	195, 197	342
Ralstonia picketti	92	187	342
Staphylococcus aureus	91	197	355
Stenotrophomonas maltophilia	96	197	349

DISCUSSION

ALH analysis of 16S rRNA genes can provide rapid detection of both known and novel organisms that may be present in many complex microbial communities (63, 90). Prior work has determined that this known ecological application can be extended to the analysis of complex biological communities such as sputa from CF patients (62). This technique can determine the most abundant amplicon lengths found in a community and with the use of ALHAS the amplicon can be presumptively identified. Thus, the ALH technique could be used to attentively identify pathogenic organisms and thus be used as a tool against bioterrorism. CF sputa were chosen again to further study the properties of ALH. The flora present in the CF lung is a dynamic community that is altered by age of patient, illness, and antibiotics. It is hypothesized that these changes over time can be studied using ALH. Also, it is hypothesized that the use of multiple hypervariable regions of the 16S rRNA genes can produce more information that will discriminate between pathogens thus leading to quicker identity match. The ALH technique has been proven to be robust and reproducible when analyzing CF sputa, marine water, and soil (5, 63, 78, 79, 91). There is great potential for the use of ALH as a diagnostic tool with bacterial infection. Also, forensic science may benefit from the use of ALH as it can be used to determine the absence of a bioagent.

Modification of quantification protocol needed to ensure accurate data analysis. The quantity of metagenomic DNA extracted from the CF sputum varied from sample to sample. It was determined that the amount commonly found in these samples

were near one nanogram. This amount is difficult to quantify using both Hoescht dye and Picogreen due to the fact that the amount is at the lower threshold and the higher threshold for the dyes, respectively. To overcome, it is recommended that Picogreen and a smaller amount of DNA be used to quantify. Accurate quantification of metagenomic DNA is important to ensure that PCR success is comparable from sample to sample. As the PCR products are detected using the genetic analyzer, the electropherogram output depicts the amount of DNA present in the sample by the amplitude of the amplicon. For example, if a sample thought to contain one nanogram of DNA but actually contains less DNA is amplified and then detected, the resulting amplitude will not be reflective of the actual amount of that pathogen present in the sample. The sample may contain a higher amount of bacteria. It may be possible to achieve more accurate quantification by using real time PCR (45).

Extraction reproducibility altered the overall profile of a sputum sample. Each sample was split into three aliquots and amplified. Extraction reproducibility is often not seen in the samples (Figure 8). Not all amplicon lengths were present in each aliquot. Some sample profiles vary more between aliquots then between samples. The lack of reproducibility may be caused by two factors. First, sputum is very viscous and sticky, thus making it difficult to make a homogenized mixture. The extraction buffer might also cause the reproducibility error. It is possible that the buffer preferentially lyses some cells. To ensure that the patient profile is complete, it is recommended that the DNA extracted from the aliquots be combined and homogenized and all further analysis be produced from the homogenized DNA.

PCR amplification does not bias eubacterial profiles. Each aliquot was amplified three times for the V1 + V2 and the V1_V3 regions. The resulting profiles were overlaid and compared. As show in figure 9, amplicons were present in at least two out of the three replicates for two patients. Similar results were seen in all patients' replicates (data not shown). Amplitude height variances were seen in with some peaks. This may be caused from quantification errors and normal PCR biases but this did not affect the reproducibility of PCR. ALH-PCR reproducibility was also shown in previous studies. A similar study performed with CF sputa samples from Denmark showed that profiles produced from replicate PCR were consistent and that only amplitude height varied (62). ALH-PCR has also been used to study the microbial community in soil from Miami-Dade, Florida. In this study, each soil sample was analyzed nine times and data was consistent from each PCR and therefore again showing that ALH-PCR is reproducible (64). This same trend was seen in studies performed with petroleum-contaminated soils, water, and sputum (63, 79, 90).

The eubacterial community is similar between CF patients in South Florida. Sputum samples obtained from patients attending CF treatment centers, Joe Dimaggio Children's Hospital (JD) or University of Miami (UM) were analyzed using the V1 + V2, V1, V3 and the V1_V3 region combined. Profiles produced from the V1 + V2 region showed a 58.4 % similarity between patients attending the JD and UM centers. Many samples were more closely clustered to the other center than to the patient's actual treatment center (Figure 15, Panel A). In a previous study, V1 + V2 region was used to

discriminate between patients and it was determined that the highest amount of similarity seen between two patients was 72 % (62). Both studies demonstrate that the V1 + V2 region is unable to completely distinguish between profiles that are being compared. The V1 eubacterial profiles from the UM and the JD centers were also compared and one of the CF samples was 58.4 % similar to another sample. The V3 profiles from the clinics showed that the patient lung flora was the most similar (73.4 %) between treatment centers. When comparing the regions discriminatory power, it was determined that the V1 region alone distinguished the treatment centers from one another more than V1 + V2, V3, and V1_V3 regions combined. Not only did the V1 region produce the lowest percentage of similarity, JD and UM samples were more closely clustered within the appropriate treatment centers (Figure 14 and 15, Panel B). The V1 region still produced a significant amount of similarity between the two groups of patients. Therefore, many of the patients either from UM or JD had similar microbes in their lung flora. The fact that V1 showed the patients' profiles were 41.6 % different means that there are some unique pathogens found in UM patients versus JD patients. This uniqueness can lead us to conclude that there is a center-specific profile for each treatment center. It is not surprising to see differences in eubacterial profiles when comparing the two CF centers. Research has shown that CF infections can vary from one location to another. In one study, *Pseudomonas cepacia* infections were found in patients from the United Kingdom, Canada, and the eastern United States (40). Multilocus enzyme electrophoresis was used to determine that *B. cepacia* (formerly *P. cepacia*) isolates clearly differ according to origin. The study showed that the isolates within one clinic had the same enzyme type and therefore, it was deduced that the pathogen is readily transferred from patient to

patient in the treatment center (40). Geographical location also influenced the presence of a strain. Infections seen at clinics within one region, such as eastern United States were more related to each other than to another location like the United Kingdom (40). The influence location has on CF infections was also seen in a study that tracked *P. aeruginosa* infections in children (25). It was determined that children being treated at the urban CF center acquired *P. aeruginosa* at an earlier age than the patients at the more rural center (25). This study again demonstrates that pathogens can be spread from patient to patient in social settings. Both of these studies highlight the same trend seen in CF treatment centers in south Florida; that is, location can influence the overall eubacterial profile.

The center-specific profile that caused the patients to cluster within their respective treatment center was best seen with the analysis performed using the V1 profile. The discriminatory power of the V1 region may arise from two factors. First, the profiles in general have more amplicons per sample in that region than any other region (Figure 11). Therefore, there are a higher number of data points in each sample that can be used to compare samples and increase the statistical power of the Bray-Curtis coefficient. Also, it is known that the V1 region has more length heterogeneity than any other region of the 16S rRNA gene. Therefore, an amplicon may still represent more than one bacterium but that representative number will be smaller than the number of an amplicons belonging to the V1 + V2 region. Hence, the V1 region profiles have multiple amplicons which may belong to one amplicon length in the V1 + V2 or V3 regions. This increased number of V1 amplicons leads to the regions ability to discriminate between samples better than the other regions. Another study showed that the V1 has a high

power of discrimination when compared to other regions of the 16S rRNA gene. When combining the V1 region with the V3 region, eubacterial profiles produced from soil in Miami-Dade, Florida were more unique than profiles from the V1 + V2 region (64). The V1 region was able to provide more information as to the bacteria present in the soil than the V1 + V2 region (64). CF sputa profiles were less similar to one another using the V1 region as compared to the V1 + V2 region. Thus, this leads us to conclude the V1 region is a more informative region than the V1 + V2 region for the analysis of CF sputa. In this study, the V3 did not add information to the analysis.

To determine if patients within a CF center had unique microbial communities, the ALH profiles from 19 UM patients treated at the University of Miami (UM) were compared against each other. Seven patients from Joe DiMaggio Children's Hospital (JD) were also studied. Statistically, ALH using either V1 + V2, V1, or V3 regions of the 16S rRNA gene cannot discriminate between patients at either the UM or JD centers (Tables 5 – 9). The V1 + V2 profiles were identical for seven UM patients and therefore this region was unable to discriminate between patients. V1 again proved to have a higher discriminatory power compared to the other regions but two patients were still found to be 79 % similar (Table 6 and 9). The V3 region produced higher similarity (Table 7 and 9). Graphically, all regions show at least two patients overlapping each other and therefore they have similar eubacterial profiles (Figure 13). This lack of discrimination with patients at one center was also statistically seen with the JD samples. The V1 region was able to differentiate two patients up to 33 % but no profile was completely unique. Graphically, it appears that each region was able to differentiate one profile from another as that no samples are overlapping (Figure 14). The MDS analysis

might be biased due to the extremely small sample size, meaning that the slightest difference in amplitude height can drastically pull two data points away from one another. With the use of similarity index, it can be concluded that patients within one treatment center have similar eubacterial profiles and this is the same trend seen in the other studies mentioned above (25, 40). But the small differences within a profile are best detected using the V1 region. This region's high power of discrimination again agrees with past analyses as described earlier.

Overall, the 26 sputum samples obtained from South Florida patients produced very similar profiles with all regions used in this study. Many of the profiles shared amplicons, such as 342 in the V1 + V2 region. The presence of a dominant amplicon in a profile might obscure the rarer amplicons and bias the analysis to say that the patients' profiles are more similar to one another than they actually are. It is thought that there are only a few pathogens commonly found in the lungs of CF patients due to clinical diagnostic procedures (27). These organisms may represent the dominant amplicons in the profile. With the use of molecular technique, it has been discovered that many more organisms plague the CF lung (79). Perhaps, these bacteria represent the rarer amplicons and are responsible for some of the differences in disease manifestation seen between CF patients.

ALH can identify microbial community changes over time. CF sputum sample from Patient UMX and Patient UMY were analyzed every other month for five months. During this time, patients were taking a combination of antibiotics, such as azithromycin, bactrim, tobramycin, and levaquin. Azithromycin is a macrolide antibiotic

that inhibits protein synthesis. It is used to treat against most gram-positive aerobes including *Staphylococcus* species and *H. influenza*, which are found in the CF lung (106). It has also been recently found to reduce the production of mucoid alginate in biofilm-producing *P. aeruginosa*. Biofilm growth of the organism is directly related to its ability to be more resistant to antibiotics (69). Bactrim, another drug is a combination of sulfamethoxazole and trimethoprim. This is used to prevent and treat infections and in the case of the CF patient, to prevent *Pneumocystis carinii* pneumonia (85). Tobramycin, an aminoglycoside, is given to patients diagnosed with *P. aeruginosa* infections to treat that infection (92). Levaquin or levofloxacin, is a second generation fluoroquinolone (82). It can be used against gram-negative infections such as *Pseudomonas* and can also treat against some gram-positive organisms (38, 82). The last drug being administered was colistin, a polymyxin E, that is used to treat against antibiotic resistant strains of *P. aeruginosa* (11).

During a five month period, Patient UMX was taking azithromycin, bactrim and tobramycin. Shortly before the third collection the patient stopped taking tobramycin. Thus, antibiotic treatment varied little during these six months. As with the drug treatment, the profiles over time were quite similar. For V1 + V2, Patient UMX's profile was predominately amplicon 342 for all collections (Table 10). This amplicon had an abundance between 92 and 100 % and it may be presumptively identified as *P. aeruginosa* which concurs with the patient's clinical diagnoses. Amplicons in the V1 and V3 region slightly changed (Table 10). Visually, it can be seen that aliquots within one time period can vary as much as the samples at each month (Figure 16). As shown in previous analysis, V1 has the highest discriminatory power out of the regions analyzed.

For comparisons over time, V1 clusters the aliquots together for each time period more so than other regions. V1 demonstrates that Patient UMX's profile did change over the course of the five month period (Figure 16, Panel B).

Patient UMY's profile was more dynamic than Patient UMX. Patient UMY's sputum showed a constant level of abundance for amplicon 342 between the first two samplings (Table 11). Then, there was a drastic decrease of abundance for amplicon length 342 between the fifth and third month. This change in profile may correspond to the change in the patient's treatment. The patient during the first month had been taking levaquin but it was replaced with colistin between the second and third sampling. This may have caused the pathogen associated with that amplicon to disappear. As the abundance of amplicon 342 decreased, amplicon 348 emerged. It is possible that as one pathogen was cleared and it was replaced with another bacterium that fit that niche (7). The decreasing 342 amplicon may represent *P. aeruginosa*, which the patient was diagnosed with. Levaquin and colistin are both known for treating against this pathogen. There are studies that have found levaquin is not as effective as the first generation fluoroquinolones in treating *P. aeruginosa* and Colistin has been known to effective against antibiotic resistant strains of *P. aeruginosa* (11, 82). It could be hypothesized that Patient UMY was infected with an antibiotic resistant strain of *P. aeruginosa* and the infection was reduced when the stronger antibiotic was administered. The patient's lung infection was then dominated by a bacterium represented by amplicon 348, which may belong to *H. influenzae*. Changes were also seen in the V1 and V3 regions (Table 11). Amplicon 195 in the V3 region decreased over time and this was presumptively identified as *P. aeruginosa*.

The dynamic lung microbial community was also seen in the MDS graphs (Figure 17). The V1 and the combined V1_V3 regions combined distinguished samples between sampling time the best (Figure 17). The V3 region alone did not separate any samples from one another and therefore did not add discriminatory power to the combined V1_V3 analysis (Figure 17). So again, V1 has the highest discriminatory power. ALH again demonstrated that a patient eubacterial profile changed over time.

Combining the ALH profiles from the three regions depicts Patient UMY having a more variable microbial community than patient UMX. This could be due to various factors such as age and antibiotic treatment (4, 53). Patient UMX was a 23 year old Hispanic woman while Patient UMY was a 60 year old Caucasian. These differences could lend to the fact that both patients produced unique profiles and their respective lung flora adapted to the presence of antibiotics differently.

Clearly, ALH can be used to analyze changes in a microbial community. Armed with the patient antibiotic information and samples over a long enough period of time, it should be possible to determine the effectiveness of the drug against pathogens in the CF lung. This could eventually lead to more effective treatments.

Presumptive analysis reveals sputa contain common CF pathogens. One of the downfalls to the ALH technique is that the organism found in a sample cannot be conclusively identified. Using the ALHAS database, the amplicon length present in a profile can be presumptively identified. Amplicons of interest can be later cloned and sequenced to ascertain the name of the organism present. With the use of ALHAS, it was determined that common amplicons present in the CF sputa are potentially pathogens

associated with this disease. Of relevant possible pathogens, amplicons were present that may represent *H. influenzae*, *S. maltophilia*, *Actinobacillus pleuropneumonia*, *Bacteriodes fragilis*, *Abiotrophia defective*, and *Sarcina ventriculi*. These pathogens were also found in CF sputum samples in an earlier study (62, 79). Specifically, all samples produced an amplicon length of 342, which could be *Pseudomonas* species, *Burkholderia* species, or *Ralstonia* species. The UM patients were clinically diagnosed with *P. aeruginosa* and therefore it would be likely that amplicon 342 was indicative of *P. aeruginosa* not *Burkholderia*. As that no patient information was attained from JD samples, no assumption can be made. *Burkholderia cepacia* amplifies a 343 base pair fragment and this amplicon was not present in any sample. The UM patients were not diagnosed to have *B. cepacia* and therefore, it was not expected to be seen in the profiles.

Presumptive analysis for the amplicons present in the V1 region revealed little information as that ALHAS database did not contain many CF related pathogens. The V3 region furthered strengthened the hypothesis that the patients were infected with *Pseudomonas* as that amplicon 195 may represent *P. aeruginosa*. Virtual analysis was performed to identify if any of the amplicons present in the profiles represented a common CF pathogen. Profiles may have contained the pathogens *Alcaligenes xylosoxidans*, *B. cepacia*, *B. gladioli*, *P. aeruginosa*, *H. influenzae*, *R. picketti*, and *S. aureus*. No profiles contained an amplicon that corresponds to *Stenotrophomonas maltophilia*. The virtual analysis for the V3 region indicates that almost all of the common CF pathogens listed above fall at 197 base pair. Therefore, it is not surprising that the V3 region produced little to no power of discrimination as discussed earlier. Yet, no profiles contained an amplicon at 197. It is possible that a strain not virtually

analyzed may fall at an amplicon seen in the profiles. Therefore, virtual analysis of these organisms needs to be confirmed experimentally. These isolate pathogens need to be analyzed using ALH and the corresponding V1 and V3 primers to determine the amplicon size that would be seen on the electropherogram.

To strengthen the use of ALHAS analysis, the database should be augmented with information derived from clinically isolated strains of the most common CF pathogens. Also to conclusively identify an organism, the corresponding amplicons it produces should be found in all three regions. Once a possible organism is recognized, cloning and sequencing should be performed to achieve identity.

Conclusions

The use of ALH in the treatment of CF patients. The ALH technique is able to show the complexity of the flora present in the CF lung. It was also proved that it is possible to observe how the microbial community in the lung over time changes and may lead to deciphering what factors drive the change. ALH can be used to gather much information that is not known about the CF lung and can also be used to further strengthen the details that are understood. To identify more trends between treatment, patient information, and microbial profile a long term study with ALH would need to be undertaken.

To identify what factors drive a community, a larger patient sample size is needed. Also, patients need to donate a sample at specific intervals for a few years. Information regarding age, sex, ethnicity, *CFTR* mutation, and antibiotic treatment must

also be attained with the samples. To further increase the power of presumptive analysis, clinically identified isolates from CF patients need to be analyzed using ALH and the resulting information needs to supplement the static ALHAS database. Lastly, cloning and sequencing needs to be performed on the samples containing rare amplicons to determine if there are any novel CF pathogens present in the sputum sample. With a new pathogen identified, drug treatment can be altered to target that organism.

CF sputum may also be studying using ALH and other genes. In this study, ALH was used with the 16S rRNA gene. Due to the relative small microbial community present in the CF lung, it may be possible to gather more information by using the interspacer region of the ribosomal genes with ALH. In the future, ALH can also be used to study the fungal community that is associated with the CF lung.

ALH can be used as presumptive test for bioagents. ALH was used to presumptively identify organisms within the CF lung. The results produced from the model system can be extrapolated and used in many other microbial communities, such as soil, water, or food that may have been contaminated with a human pathogen. Though ALH does not identify a pathogen, it does prove the absence of the organism in the sample being analyzed. Therefore, the technique can be used to determine if there is a possibility of a bioagent, such as *B. anthracis*, being present in a sample. If the amplicon that corresponds to the bioagent is detected, then further more expensive testing can be performed. If ALH determines the pathogen's amplicon length for that specific region is absent, then no further testing needs to be done.

In order to implement this technique into forensics, ALH must be used to test samples that contain bioagents or their close relatives. If it proves to identify the absence of such agents, then validation studies must be performed. Once that is accomplished, the test needs to be commercially packaged. ALH kits can then be sold to law enforcement that will use them as a first response in a potential bioterrorism act. Therefore, ultimately ALH could be used in microbial forensics and in the fight against bioterrorism.

In conclusion, ALH is a molecular-based technique that can elucidate much information about the microbial community found in the CF lung. It could lead to new treatments, discovery of new pathogens, and ultimately improve the quality of a CF patient's life. The technique can also be utilized in forensic science. ALH can be used as a quick presumptive test to identify the absence of a bioagent.

LITERATURE CITED

1. Missouri v. Stewart. 18 S.W. 3d 75.

2. **Alibek, K., and S. Handleman.** 2000. Biohazard: The chilling true story of the largest covert biological weapons program in the world-told from inside by the man who ran it. Dell Publishing, New York.

3. **American Red Cross of Wood County.** 2000. Q&A: ABO Blood Groups.

4. **Beringer, P. M., and M. D. Appleman.** 2000. Unusual respiratory bacterial flora in cystic fibrosis: microbiologic and clinical features. *Current Opinion in Pulmonary Medicine* **6:**545-550.

5. **Bernhard, A. E., D. Colbert, J. McManus, and K. G. Field.** 2005. Microbial community dynamics based on 16S rRNA gene profiles in a Pacific Northwest estuary and its tributaries. *FEMS Microbiology Ecology* **52:**115-128.

6. **Blackwood, C. B., T. Marsh, S.-H. Kim, and E. A. Paul.** 2003. Terminal restriction fragment length polymorphism data analysis for quantitative comparison of microbial communities. *Applied Environmental Microbiology* **69:**926-932.

7. **Boni, M. F., and M. W. Feldman.** 2005. Evolution of antibiotic resistance by human and bacterial niche construction. *Evolution* **59:**477-491.

8. **Budowle, B., S. E. Schutzer, A. Einseln, L. C. Kelley, A. C. Walsh, J. A. L. Smith, B. L. Marrone, J. Robertson, and J. Campos.** 2003. Public health: building microbial forensics as a response to bioterrorism. *Science* **301:**1852-1853.

9. **Butler, J. M.** 2001. Forensic DNA typing: biology and technology behind STR markers. Academic Press, San Diego.

10. **Butler, J. M., and D.J. Reeder** 2003, Sept. 15 1997, posting date. Short tandem repeat DNA internet base. The National Institute of Justice. [Online.]

11. **Canton, R., N. Cobos, J. de Gracia, F. Baquero, J. Honorato, S. Gartner, A. Alvarez, A. Salcedo, A. Oliver, and E. Garcia-Quetglas.** 2005. Antimicrobial therapy for pulmonary pathogenic colonisation and infection by *Pseudomonas aeruginosa* in cystic fibrosis patients. *Clinical Microbiology and Infection* **11:**690-703.

12. **Carey, L., and L. Mitnik.** 2002. Trends in DNA forensic analysis. *Electrophoresis* **23:**1386-1397.

13. **Center for Disease Control and Prevention.** 2004. Bioterrorism and public health preparedness. http://www.cdc.gov/programs/bt.htm.

14. **Center for Disease Control and Prevention.** 2004. Emergency preparedness and response: bioterrorism agents. http://www.bt.cdc.gov/agent/agentlist-category.asp.

15. **Corach, D., L. Filgueira Risso, M. Marino, G. Penacino, and A. Sala.** 2001. Routine Y-STR typing in forensic casework. *Forensic Science International* **118:**131-135.

16. **Coyle, H. M., T. Palmbach, N. Juliano, C. Ladd, and H.C. Lee.** 2003. An overview of DNA methods for the identification and individualization of marijuana. *Croatian Medical Journal* **44:**315-321.

17. **Curran, T.** 1997, posting date. Forensic DNA analysis: technology and application. Library of Parliament-Parliamentary Research Branch. [Online.]

18. **Davis, C. J.** 1999. Nuclear blindness: An overview of the biological weapons programs of the former Soviet Union and Iraq. *Emerging Infectious Diseases* **5:**509-512.

19. **Day, A., American Prosecutors Research Institute.** 2001. Nonhuman DNA testing increases DNA's power to identify and convict criminals. *Silent Witness* **6.**

20. **Dixon, L. A., C. M. Murray, E. J. Archer, A. E. Dobbins, P. Koumi, and P. Gill.** 2005. Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. *Forensic Science International* **154:**62-77.

21. **DNA Print Genomics.** 2003. Product and services: forensics. http://www.dnaprint.com/welcome/productsandservices/forensics/.

22. **Drell, D.** 2003. DNA Forensics, vol. 2003. U.S. DOE human genome project. http://www.ornl.gov/TechResources/Human_Genome/elsi/forensics.html.

23. **Dunbar, J., L. O. Ticknor, and C. R. Kuske.** 2001. Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacterial communities. *Applied Environmental Microbiolgy* **67:**190-197.

24. **Erlich, H. A., D. Gelfand, and J. Sninsky.** 1991. Recent advances in the polymerase chain reaction. *Science* **252:**1643-1651.

25. **Farrell, P. M., G. Shen, M. Splaingard, C. E. Colby, A. Laxova, M. R. Kosorok, M. J. Rock, and E. H. Mischler.** 1997. Acquisition of *Pseudomonas aeruginosa* in children with cystic fibrosis. *Pediatrics* **100:**1-9.

26. **Govan, J. R. W., and G.S. Harris.** 1986. *Pseudomonas aeruginosa* and cystic fibrosis: unusual bacterial adaptation and pathogenesis. *Microbiological Sciences* **3:**302-308.

27. **Govan, J. R. W., J. W. Nelson.** 1993. Microbiology of cystic fibrosis lung infections: themes and issues. *Journal Of The Royal Society of Medicine* **86:**11-18.

28. **Grant, A., and L. A. Ogilvie.** 2004. Name that microbe: rapid identification of taxa responsible for individual fragments in fingerprints of microbial community structure. *Molecular Ecology Notes* **4:**133-136.

29. **Guillemin, J.** 2005. Biological weapons (columbia contemporary issues in national security policy). Columbia University Press, New York.

30. **Harris, S. H.** 1989. Factories of death: Japanese biological warfare, 1932-45 and the American cover-up. Routledge, New York.

31. **Hartmann, M., B. Frey, R. Kolliker, and F. Widmer.** 2005. Semi-automated genetic analyses of soil microbial communities: comparison of T-RFLP and RISA based on descriptive and discriminative statistical approaches. *Journal of Microbiological Methods* **61:**349-360.

32. **Hiraishi, A., M. Iwasaki, and H. Shinjo.** 2000. Terminal restriction pattern analysis of 16S rRNA genes for the characterization of bacterial communities of activated sludge. *Journal of Bioscience and Bioengineering* **90:**148-156.

33. **Holloway, B. W., and A. F. Morgan.** 1986. Genome organization in *Pseudomonas*. *Annual Review of Microbiology* **40:**79-105.

34. **Holmes, C.** 2003. Spores, plagues and history: The story of anthrax. Durban House, Dallas.

35. **Horswell, J., S. J. Cordiner, E. W. Maas, T. M. Martin, K. B. W. Sutherland, T. W. Speir, B. Nogales, and A. M. Osborn.** 2002. Forensic comparison of soils by bacterial community DNA profiling. *Journal of Forensic Science* **47:**350-353.

36. **Hull, J.** 2003. Basic science of cystic fibrosis. *Current Paediatrics* **13:**253-258.

37. **Inman, K., and N. Rudin.** 1997. Interpretation of DNA typing results. CRC Press, Washington, D.C.

38. **Ishida, H., Y. Ishida, Y. Kurosaka, T. Otani, K. Sato, and H. Kobayashi.** 1998. In vitro and in vivo activities of levofloxacin against biofilm-producing *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy* **42:**1641-1645.

39. **Jernigan, J. A., D. S. Stephens, D. A. Ashford, C. Omenaca, M. S. Topiel, M. Galbraith, M. Tapper, T. L. Fisk, S. Zaki, T. Popovic, R. F. Meyer, C. P. Quinn, S. A. Harper, S. K. Fridkin, J. J. Sejvar, C. W. Shepard, M. McConnell, J. Guarner, W. J. Shieh, J. M. Malecki, J. L. Gerberding, J. M. Hughes, and B. A. Perkins.** 2001. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. *Emerging Infectious Diseases* **7:**933-944.

40. **Johnson, W. M., S. D. Tyler, and K. R. Rozee.** 1994. Linkage analysis of geographic and clinical clusters in *Pseudomonas cepacia* infections by multilocus enzyme electrophoresis and ribotyping. *Journal of Clinical Microbiolgy* **32:**924-930.

41. **Karpati, F., and J. Jonasson.** 1996. Polymerase chain reaction for the detection of *Pseudomonas aeruginosa, Stenotrophomonas maltophilia* and *Burkholderia cepacia* sputum of patients with cystic fibrosis. *Molecular and Cellular Probes* **10:**397-403.

42. **Keim, P.** 2002. Presented at the microbial forensics: a scientific assessment, Burlington, Vermont, 2003.

43. **Kiska, D. L., A. Kerr, M. C. Jones, J. A. Caracciolo, B. Eskridge, M. Jordan, S. Miller, D. Hughes, N. King, and P. H. Gilligan.** 1996. Accuracy of four commercial systems for identification of *Burkholderia cepacia* and other gram-negative nonfermenting bacilli recovered from patients with cystic fibrosis. *Journal of Clinical Microbiology* **34:**886-891.

44. **Klappenbach, J. A., P. R. Saxman, J. R. Cole, and T. M. Schmidt.** 2001. rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Research* **29:**181-184.

45. **Klein, D.** 2002. Quantification using real-time PCR technology: applications and limitations. *Trends in Molecular Medicine* **8:**257-260.

46. **Kunzelmann, K.** 1999. The cystic fibrosis transmembrane conductance regulator and its function in epithelial transport. *Reviews of physiology, biochemistry, and pharmacology* **137:**1-70.

47. **Lessig, R., M. Zoledziewska, K. Fahr, J. Edelmann, M. Kostrzewa, T. Dobosz, and W. J. Kleemann.** 2005. Y-SNP-genotyping - a new approach in forensic analysis. *Forensic Science International* **154:**128-136.

48. **Lewin, B.** 2002. Protein Synthesis, p. 100,139-140, 159-164, Genes VII. University Press, Oxford.

49. **Lewin, R.** 1986. DNA fingerprints in health and disease. *Science* **233:**521-522.

50. **Linacre, A., H. M. Hsieh, L. H. Huang, L. C. Tsai, Y. C. Kuo, H. H. Meng, and J. C. I. Lee.** 2004. Identification of rhino from medicinal powders and sculptures. *International Congress Series* **1261**:595-595.

51. **Lipuma, J. J.** 1998. Molecular tools for epidemiologic study of infectious diseases. *Pediatric Infectious Disease Journal* **17**:667-675.

52. **Lorenzini, R.** 2005. DNA forensics and the poaching of wildlife in Italy: A case study. *Forensic Science International* **153**:218-221.

53. **Lyczak, J. B., C.L.Cannon, and G. B. Pier.** 2002. Lung infections associated with cystic fibrosis. *Clinical Microbioogyl Review* **15**:194-222.

54. **Malgorn, Y., and R. Coquoz.** 1999. DNA typing for identification of some species of *Calliphoridae*: An interest in forensic entomology. *Forensic Science International* **102**:111-119.

55. **Martin, P. D., H. Schmitter, and P. M. Schneider.** 2001. A brief history of the formation of DNA databases in forensic science within Europe. *Forensic Science International* **119**:225-231.

56. **Marty, N.** 1997. Epidemiological typing of *Stenotrophomonas maltophilia*. *Journal of Hospital Infection* **36**:261-266.

57. **Marx, J. L.** 1988. DNA fingerprinting takes the witness stand. *Science* **240**:1616-1618.

58. **McDowell, A., E. Mahenthiralingam, J. E. Moore, K. E. Dunbar, A. K. Webb, M. E. Dodd, S. L. Martin, B. C. Millar, C. J. Scott, M. Crowe, and J. S. Elborn.** 2001. PCR-based detection and identification of *Burkholderia cepacia* complex pathogens in sputum from cystic fibrosis patients. *Journal of Clinical Microbiolgy* **39**:4247-4255.

59. **McMenamin, J. D., T. M. Zaccone, T. Coenye, P. Vandamme, and J. J. LiPuma.** 2000. Misidentification of *Burkholderia cepacia* in US cystic fibrosis treatment centers: an analysis of 1,051 recent sputum isolates. *Chest* **117**:1661-1665.

60. **Members of the Scientific Working Group on Microbial Genetics and Forensics** October 2003, posting date. Quality assurance guidelines for laboratories performing microbial forensic work. Federal Bureau of Investigation. [Online.]

61. **Metzker, M. L., D. P. Mindell, X. M. Liu, R. G. Ptak, R. A. Gibbs, and D. M. Hillis.** 2002. Molecular evidence of HIV-1 transmission in a criminal case. *Proceedings of the National Academy of Sciences of the United States of America* **99**:14292-14297.

62. **Miller, S.** 2003. Molecular identification of microbes in sputum of cystic fibrosis patients. Master's of Forensic Science. Florida International University, Miami.

63. **Mills, D. K., K. Fitzgerald, C. D. Litchfield, and P. M. Gillevet.** 2003. A comparison of DNA profiling techniques for monitoring nutrient impact on microbial community composition during bioremediation of petroleum-contaminated soils. *Journal of Microbiological Methods* **54:**57-74.

64. **Moreno, L. I.** 2005. Molecular and chemical characterization of three Miami-Dade soil types for forensic comparison. Florida International University, Miami.

65. **Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich.** 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **51:**263-273.

66. **Murray, P. R., K. S. Rosenthal, G. S. Kobayashi, S. George, and M. A. Pfaller.** 2002. Medical microbiology, 4 ed. Mosby, Inc., St. Louis.

67. **Muyzer, G., E. C. de Waal, and A. G. Uitterlinden.** 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied Environmental Microbiolgy* **59:**695-700.

68. **National Institute of Justice** 2003, posting date. Solicitation for forensic DNA research and development for FY 2003. U.S. Department of Justice. [Online.]

69. **Nicolau, D. P., M. A. Banevicius, C. H. Nightingale, and R. Quintiliani.** 1999. Beneficial effect of adjunctive azithromycin in treatment of mucoid *Pseudomonas aeruginosa* pneumonia in the murine model. *Antimicrobial Agents and Chemotherapy* **43:**3033-3035.

70. **Nomura, M., and V. A. Erdmann.** 1970. Reconstitution of 50S ribosomal subunits from dissociated molecular components. *Nature* **228:**744-748.

71. **Nomura, M., E. A. Morgan, and S. R. Jaskunas.** 1977. Genetics of bacterial ribosomes. *Annual Review of Genetics* **11:**297-347.

72. **Olsen, G. J., and C. R. Woese.** 1993. Ribosomal RNA: a key to phylogeny. *FASEB Journal* **7:**113-123.

73. **Patient-registry.** 2002. Cystic fibrosis foundation patient registry annual data report. Cystic Fibrosis Foundation.

74. **Popovic, T., and M. Glass.** 2003. Laboratory aspects of bioterrorism-related anthrax - from identification to molecular subtyping to microbial forensics. *Croatian Medical Journal* **44:**336-341.

75. **Rees, G. N., D. S. Baldwin, G. O. Watson, S. Perryman, and D. L. Nielsen.** 2004. Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms: application of multivariate statistics. *Antonie van Leeuwenhoek* **86:**339-347.

76. **Reischl, U., M. Pulz, W. Ehret, and H. Wolf.** 1994. PCR based detection of *Mycobacterium* in sputum samples using a simple and reliable DNA extraction protocol. *Biotechniques* **17:**844-845.

77. **Riordan, J. R., J.M. Rommens, B. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plavsic, J.L. Chou, M.L. Drumm, M.C. Iannuzzi, F.S. Collins, and L. Tsui.** 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245:**1066-1073.

78. **Ritchie, N. J., M. E. Schutter, R. P. Dick, and D. D. Myrold.** 2000. Use of length heterogeneity PCR and fatty acid methyl ester profiles to characterize microbial communities in soil. *Applied Environmental Microbiolgy* **66:**1668-1675.

79. **Rogers, G. B., C. A. Hart, J. R. Mason, M. Hughes, M. J. Walshaw, and K. D. Bruce.** 2003. Bacterial diversity in cases of lung infection in cystic fibrosis patients: 16S ribosomal DNA (rDNA) length heterogeneity PCR and 16S rDNA terminal restriction fragment length polymorphism profiling. *Journal of Clinical Microbiology* **41:**3548-3558.

80. **Roszak, D. B., and R. R. Colwell.** 1987. Survival strategies of bacteria in the natural environment. *Microbiol Review* **51:**365-379.

81. **Saferstein, R. (ed.).** 2002. Forensic science handbook, 2nd ed, vol. 1. Prentice Hall, Upper Saddle River, New Jersey.

82. **Scheld, M. W.** 2003. Maintaining fluorquinolone class efficacy: review of infuencing factors. *Emerging Infectious Diseases* **9**.

83. **Sekiguchi, H., M. Watanabe, T. Nakahara, B. Xu, and H. Uchiyama.** 2002. Succession of bacterial community structure along the Changjiang River determined by denaturing gradient gel electrophoresis and clone library analysis. *Applied Environmental Microbiolgy* **68:**5142-5150.

84. **Sensabaugh, K.** 1998. Non-human DNA Evidence. *38 Jurimetrics Journal* **1-16**.

85. **Sharma, G.** June 22, 2005, posting date. Cystic fibrosis. [Online.]

86. **Shelly, D. B., T. Spilker, E. J. Gracely, T. Coenye, P. Vandamme, and J. J. LiPuma.** 2000. Utility of commercial systems for identification of *Burkholderia*

cepacia complex from cystic fibrosis sputum culture. *Journal of Clinical Microbiolgy* **38:**3112-3115.

87. **Sobrino, B., M. Brion, and A. Carracedo.** 2005. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International* **154:**181-194.

88. **Spilker, T., T. Coenye, P. Vandamme, and J. J. LiPuma.** 2004. PCR-based assay for differentiation of *Pseudomonas aeruginosa* from other *Pseudomonas* species recovered from cystic fibrosis patients. *Journal of Clinical Microbiolgy* **42:**2074-2079.

89. **Strobeck, C.** 2003, posting date. DNA Fingerprinting. [Online.]

90. **Suzuki, M., M. S. Rappe, and S. J. Giovannoni.** 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. *Applied Environmental Microbiolgy* **64:**4522-4529.

91. **Suzuki, M. T., M. S. Rappe, Z. W. Haimberger, H. Winfield, N. Adair, J. Strobel, and S. J. Giovannoni.** 1997. Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. *Applied Environmental Microbiology* **63:**983-989.

92. **Taylor, C.** 2002. Cystic fibrosis medicine, the use of nebulised tobramycin (TOBI) http://wwwcysticfibrosismedicinecom/.

93. **Tiirola, M. A., J. E. Suvilampi, M. S. Kulomaa, and J. A. Rintala.** 2003. Microbial diversity in a thermophilic aerobic biofilm process: analysis by length heterogeneity PCR (LH-PCR). *Water Research* **37:**2259-2268.

94. **Torsvik, V., F. L. Daae, R. A. Sandaa, and L. Ovreas.** 1998. Novel techniques for analyzing microbial diversity in natural and perturbed environments. *Journal of Biotechnology* **64:**53-62.

95. **Van de Peer, Y., S. Chapelle, and P. De Wachter.** 1996. A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Research* **24:**3381-3391.

96. **Wainwright, B. J., P. J. Scambler, J. Schmidtke, E. A. Watson, H. Y. Law, M. Farrall, H. J. Cooke, H. Eiberg, and R. Williamson.** 1985. Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature* **318:**384-385.

97. **Wallman, J. F., and S. C. Donnellan.** 2001. The utility of mitochondrial DNA sequences for the identification of forensically important blowflies (Diptera: *Calliphoridae*) in southeastern Australia. *Forensic Science International* **120:**60-67.

98. **Wan, Q., and S. Fang.** 2003. Application of species-specific polymerase chain reaction in the forensic identification of tiger species. *Forensic Science International* **131:**75-78.

99. **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345:**63-65.

100. **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA sequences reveal uncultured inhabitants of a well-studied thermal community. *FEMS Microbiology Review* **6:**105-115.

101. **Whitby, P. W., H. L. N. Dick, P. W. Campbell, III, D. E. Tullis, A. Matlow, and T. L. Stull.** 1998. Comparison of culture and PCR for detection of *Burkholderia cepacia* in sputum samples of patients with cystic fibrosis. *Journal of Clinical Microbiology* **36:**1642-1645.

102. **Woese, C. R.** 1987. Bacterial evolution. *Microbiology Review* **51:**221-271.

103. **Yarris, L.** 2003. X-ray crystal images shed more light on protein synthesis. http://www.lbl.gov/Science-Articles/Archive/ALS-ribosome-structure.html.

104. **Yeates, C.** 1996. Microbial diversity. *Microbiology Australia* **17:**229-231.

105. **Yoon, C. K.** 1993. Botanical witness for the prosecution. *Science* **260:**894-895.

106. **Zhanel G., M. Dueck, D. Hoban, L. Vercaigne, J. Embil, A. Gin, and J. Karlowsky.** 2001. Review of macrolides and ketolides: focus on respiratory tract infections. *Drugs* **61:**443-498.