3-9-2009

# A Metagenomic Analysis of Microbial Contamination in Aviation Fuels

Jerrod P. McComb


# A METAGENOMIC ANALYSIS OF MICROBIAL CONTAMINATION IN AVIATION FUELS

THESIS

Jerrod P. McComb, Captain, USAF

AFIT/GEM/ENV/09-M11

**DEPARTMENT OF THE AIR FORCE**
**AIR UNIVERSITY**

**AIR FORCE INSTITUTE OF TECHNOLOGY**

**Wright-Patterson Air Force Base, Ohio**

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

AFIT/GEM/ENV/09-M11

# A METAGENOMIC ANALYSIS OF MICROBIAL CONTAMINATION IN AVIATION FUELS

THESIS

Presented to the Faculty

Department of Systems and Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Engineering Management

Jerrod P. McComb, BS

Captain, USAF

March 2009

AFIT/GEM/ENV/09-M11

# A METAGENOMIC ANALYSIS OF MICROBIAL CONTAMINATION IN AVIATION FUELS

Jerrod P. McComb, BS
Captain, USAF

Approved:

__///SIGNED///_____ _____
Charles A. Bleckmann (Chairman) Date

__///SIGNED///_____ _____
Michael L. Shelley (Member) Date

__///SIGNED///_____ _____
Lisa M. Brown (Member) Date

## Abstract

Microbial contamination in aviation fuel arises due to the impracticality of keeping fuel tanks sterile and the inevitable presence of water from condensation. Microbial contaminants in aviation fuels are a concern because of their potential to degrade the fuel, accelerate corrosion within the fuel tank, and threaten flight safety. This research aids in mitigating those problems by comprehensively characterizing the microbial communities affecting aviation fuels. Advances in molecular biological techniques have allowed for the identification of microorganisms which were not identified by the traditional culture-based methodologies used in previous studies**.** This study employed a molecular method known as 16S rDNA gene analysis to describe the microbial communities in aviation fuel. The microbial communities in JP-8, Jet A, and biodiesel were evaluated at the phylum and genus levels of taxonomy. The JP-8 community was found to be much richer than both the Jet A and biodiesel community. The biodiesel community was found to be a subset of the JP-8 community. A small subset of microorganisms was found to exist across all three fuels while the majority of identified microorganisms were endemic to a single fuel type. Rarefaction analysis showed that further sampling is likely to reveal additional diversity.

*For my children*

**Acknowledgments**

I would like to express my sincerest appreciation to my thesis advisor, Dr. Charles Bleckmann, and my sponsors at the University of Dayton Research Institute for their support in this endeavor. Your guidance and expertise were crucial to my success and beyond all expectations. Thank you for laying the foundation that made my work possible.

I would like to thank my fellow AFIT/GEM students. My experience here at AFIT has been a pleasure and I am truly blessed to have shared it with such an extraordinary group of individuals. We were a great team and I look forward to serving with you throughout our careers.

I am also very grateful for my parents and grandparents. You taught me the importance of higher education, and instilled in me the desire to succeed in any endeavor. Your unconditional love and support have enabled me to pursue my dreams. Your examples as role models have made me who I am today. For that I am grateful.

Finally, special thanks go to my beautiful wife for her love, motivation, support and understanding throughout the thesis process. Thanks for listening and helping me keep things in perspective. Without your encouragement and patience this thesis would not have been possible. Therefore, this is as much a result of your dedication and commitment as it is mine.

Jerrod P. McComb

## Table of Contents

Page

Page

## List of Figures

Page

## List of Tables

Page

# A METAGENOMIC ANALYSIS OF MICROBIAL CONTAMINATION IN AVIATION FUELS

## Chapter I: Introduction

### Overview

This chapter discusses the general topic of microbial contamination in aviation fuels. An overview of the pertinent subject areas as well as the methodology used in this thesis effort is provided. Reasons why this research is needed, the motivation behind it, and the thesis objectives are presented. This chapter also provides an outline of the remaining chapters of the document. The chapter concludes with some definitions of important terminology and overarching principles used in this research effort.

### Background

Microorganisms populate every conceivable environment, both familiar and exotic, from the surface of human skin, to rainforest soils, to hydrothermal vents in the ocean floor, and new information is constantly being discovered concerning their existence, prevalence and mechanisms for survival (Harwood, 2008). One environment that has not been intensely studied, and therefore presents many unanswered questions, is aviation fuel.

In every feasible environment, microbes are exploiting locally available energy sources to survive and thrive; aviation fuel systems are no exception. Aviation fuel systems are an ideal environment for the proliferation of microorganisms, as all physiological requirements for their growth (oxygen, carbon, water, etc.) are normally

present (Swift, 1988). It has been known since 1895 that microorganisms are capable of utilizing hydrocarbons as a source of metabolic energy; however little is known about the exact microorganisms responsible (Chelgren, 2008; Rauch, 2008; Zobell, 1946). Information such as community composition, degradation pathways and microbial interactions have yet to be fully researched (Rauch, 2008). Some hydrocarbon environments have been studied more comprehensively than others. Literature on the topic of oil and fuel spills, the use of microorganisms for bioremediation, and microorganisms in soil have been explored to a much greater degree than microbial contamination in aviation fuels (Van Hamme, Singh, & Ward, 2003).

Furthermore, many of the articles that have dealt with aviation fuels have typically tested or characterized only select species using traditional culture-based methods (Hedrick, Carroll, Owen, & Pritchard, 1963). For example, a study conducted by Hedrick et al. examined nineteen species representative of those commonly found in aviation fuel and concluded that more species remained viable when inoculated in pure cultures than when inoculated in mixed (composite) cultures (Hedrick et al., 1963). However, the ability to culture a microorganism in a lab (*in vitro*) does not necessarily divulge its function in a community (*in situ*) (Amann, Ludwig, & Schleifer, 1995; Hedrick et al., 1963). Consequently, caution must be exercised when extrapolating results from *in vitro* studies, with relatively few species, to the complexity of natural microbial communities in an ecosystem, which are known to encompass extraordinary diversity (Whitman, Coleman, & Wiebe, 1998).

An additional reason for further study of microbial contamination in aviation fuel is that many of the previous studies were conducted prior to the advent of the

revolutionary DNA/RNA analysis tools available today (Handelsman, 2004). The vast majority of microorganisms cannot be cultured *in vitro,* and therefore cannot be directly studied or controlled in a laboratory setting. Due to these limitations, a mere 1% of microorganisms are estimated to have been isolated using traditional culture methods (Amann et al., 1995; Hugenholtz, Goebel, & Pace, 1998).

Today's new molecular methodologies allow us to examine the elusive 99% of the uncultured microorganisms by examining their DNA sequences (Pace, 1997). This thesis effort utilized a method known as 16S rDNA gene analysis to characterize the microbial communities in aviation fuel. Numerous studies of this nature have been conducted over the past decade; each with astonishing results and discoveries (Cloud et al., 2002; Drancourt et al., 2000; Hagstrom et al., 2002; Nogales et al., 2001; Vasanthakumar, Handelsman, Schloss, Bauer, & Raffa, 2008).

This thesis effort utilized bacterial sequence data from the University of Dayton Research Institute (UDRI) Energy and Environmental Engineering Division laboratory, consisting of 3126 16S rDNA sequences from aviation fuel samples collected from a wide array of airframes covering a diverse geographical range of operational Air Force bases and commercial airports. Following a trimming and editing procedure described in the methodology chapter of this report, 1186 sequences were used for diversity estimation and library comparison analysis. The software packages used to analyze the data from the 16S rDNA gene sequencing method were the Ribosomal Database Project (RDP) Release 10 Update 7 Aligner, Classifier and format download programs, Distance Based Operational Taxonomic Unit (OTUs) and Richness Determination (DOTUR) version 1.53 (Schloss & Handelsman, 2005), Library Shuffle (∫-LIBSHUFF) Version 1

(Schloss, Larget, & Handelsman, 2004; Singleton, Furlong, Rathbun, & Whitman, 2001), and Shared OTUs and Similarities (SONS) Version 1 (Schloss & Handelsman, 2006a). These software packages are described in detail in the literature review section and their usage explained in the methodology section. They allowed for a complete characterization of the microbial communities into phyla and genera, and produced parameters that described the diversity and statistical similarities of each community. Microbial communities in the various fuel types were compared and any effects on the microbial diversity or composition noted. This information was used to draw inferences about the nature of the microbial communities contaminating aviation fuel.

**Problem Statement**

Microbial growth in aviation fuel storage tanks and aircraft wing tanks cause fuel filter plugging, corrosion, fuel degradation and increased maintenance costs associated with these problems (Rauch et al., 2005). Although Air Force researchers have been aware of these problems since at least 1956, when the first operational Air Force problem attributed to microorganisms occurred, no solution has been found (Bakanauskas, 1958; Finefrock & London, 1966; Rauch et al., 2005). This may be attributed to several factors such as little public knowledge of the problem, no visible problems, no recent major issues, adequate treatment available for symptoms, and difficulty determining a cause-effect relationship between microorganisms and problems (Balster, Chelgren, Strobel, Vangsness, & Bowen, 2006). Microbial colonization and subsequent degradation of aviation fuel is not an immediate health problem facing the masses. For fear of repercussion or unwanted negative attention, the motivation to hide the problem is high

and there is currently little reason to share contamination incidences (Balster et al., 2006). Also, with the development of effective biocides such as di-ethylene glycol monomethyl ether (di-EGME), a commonly used anti-icing agent and known biocide, there is little motivation to resolve the problem as continuing to treat the symptoms seems to be sufficient (Meshako, Bleckmann, & Goltz, 1999).

**Purpose of Research**

The problem of microbial contamination continues today, and with the expected increase in usage of biodiesel and other alternative fuels, the problems associated with microbial contamination are expected to increase (Robbins & Levy, 2004). In order to understand and improve mitigation of these problems it is necessary to characterize to the greatest extent possible the microbial consortia affecting our aviation fuels and the systems that utilize those fuels. One potential way forward is to determine which microbes are present when a problem is noted, find out what genes are expressed in those microbes' DNA that result in deleterious effects on the fuel systems (i.e., storage tanks and aircraft), then determine a way to block the expression of those genes. As a prerequisite to this approach, it is vital to answer the basic questions of what microorganisms currently exist, their frequency, and which ones contribute most significantly to the formation of biofilms and other types of aviation fuel contamination. This thesis effort is a continuation of the first study to apply molecular tools to the characterization of microbial communities in aviation fuel (Denaro, 2005). The results provided here will enhance the current understanding of the microorganisms present in

aviation fuels, also known as microbial contamination. The aviation fuels studied in this effort were JP-8, Jet A, and biodiesel.

**Research Objectives**

The primary objectives of this research were to:

1. Characterize the bacterial communities in the various aviation fuels by exploring community membership.
2. Investigate the effects of fuel type on microbial diversity and community structure.

The results of this study provide a qualitative characterization of the microbial communities responsible for contamination of aviation fuel supplies as well as a thorough quantitative investigation of the relationship between fuel type and microbial diversity. This thesis effort provides researchers with a baseline from which to further study the molecular dynamics and behavior of the microbial contaminants commonly found in aviation fuel and brings researchers one step closer to finding a specifically targeted, permanent and reliable solution to a longstanding problem in the military and civilian aviation sectors—microbial contamination in aviation fuel.

**Thesis Organization**

Chapter 2 examines the breadth of literature currently available regarding microbial contamination in aviation fuels. The history of microbial contamination, conditions required for microbial growth, problems associated with microbial growth, routes of microbial infection, and microorganisms previously identified in aviation fuels

are discussed. Additionally, the various types of aviation fuels and fuel additives are introduced. Finally, the past and present methods of detection and analysis of microorganisms are presented and explained, including the 16S rDNA gene sequencing and comparative analysis methods used in this study.

Chapter 3 describes the methodology used in this thesis. This section examines the steps taken, from sample extraction to laboratory procedures to sequence analysis, which resulted in the outputs displayed and described in Chapter 4. The purpose of this chapter is to provide instructions so that results may be validated or the methodology applied to future DNA sequence libraries.

Chapter 4 explores the results produced by the analysis methodology used in this thesis. This chapter focuses upon reviewing the outputs from the various software packages and putting them in an appropriate format from which conclusions can be drawn. Charts and figures are provided to include pie charts for community composition, graphs comparing the microbial diversity and diagrams comparing the compositional makeup of each of the DNA sequence libraries.

Chapter 5 introduces the conclusions and recommendations of this thesis effort. These conclusions are based upon both the results of the methodology and the overall experiment itself in terms of lessons learned and what could have been done differently. Suggestions for future research are also included in this chapter.

**Definitions**

*Bacterium (pl. bacteria), n.* — A single cell microorganism characterized by the absence of defined intracellular membranes that define all higher life forms. Potential food sources range from single carbon molecules to complex polymers, including plastic (ASTM, 1999).

*Bioburden, n.* — The level of microbial contamination (biomass) in a system (ASTM, 1999).

*Biocide, n.* — A poisonous substance that can kill living organisms (ASTM, 1999).

*Biodeterioration, n.* — The loss of commercial value or performance characteristics, or both, of a product (fuel) or material (fuel system) through biological processes (ASTM, 1999).

*Biofilm, n.* — A film or layer of microorganisms, biopolymers, water, and entrained organic and inorganic debris that forms as a result of microbial growth at phase interfaces (liquid-liquid, liquid-solid, liquid-gas, and so forth) (ASTM, 1999).

*Biosurfactant, n.* — A surface-active substances produced by living cells. They have the properties of reducing surface tension, stabilizing emulsions, promoting foaming and are generally non-toxic and biodegradable. Biosurfactants enhance the emulsification of hydrocarbons, have the potential to solubilize hydrocarbon contaminants and increase their availability for microbial degradation (Rahman & Gakpe, 2008).

*Contamination, n.* — The process of making inferior or impure by admixture, as well as to making unfit for use by the introduction of unwholesome or undesirable elements (Merriam-Webster Online, 2002). In the case of aviation fuel contamination, the undesirable elements are free phase water, solid particulates, and microorganisms.

*Consortium (pl. consortia), n.* — A microbial community comprised of more than one species that exhibits properties not shown by individual community members. Consortia often mediate biodeterioration processes that individual taxa cannot (ASTM, 1999).

*Free Phase Water n.* — Visible layer of water separate from the fuel within the same container. Water has three adverse effects in fuel systems. It does not burn in the engine, it freezes at low temperatures encountered during high altitude flights, and it provides an environment in which microorganisms can grow (Hemighaus et al., 2006).

*Metagenomics, n.* — The study of genetic material recovered from environmental samples. Traditional microbiology and microbial genome sequencing rely upon cultivated clonal cultures. This relatively new field of genetic research enables studies of organisms that are not easily cultured in a laboratory as well as studies of organisms in their natural environment (Handelsman, 2004).

*Microbially Induced Corrosion (MIC), n.* — Corrosion that is enhanced by the action of microorganisms in the local environment (ASTM, 1999).

*Microorganism or Microbe, n.* — An organism that is microscopic (usually too small to be seen by the naked human eye). Microorganisms are very diverse and include bacteria,

fungi, archaea, among others. All references to microorganisms or microbes in this thesis refer to bacteria.

*Phylogenetics, n.* — The study of evolutionary relatedness among various groups of organisms (e.g., species, populations), which is discovered through molecular sequencing data. Experience shows that closely related organisms have similar DNA sequences; more distantly related organisms have more dissimilar sequences (Fitch & Margoliash, 1967; Woese & Fox, 1977).

*Sulfate Reducing Bacteria (SRB), pl., n.* — Any bacteria with the capability of reducing sulfate to sulfide. The term SRB applies to representatives from a variety of bacterial taxa that share the common feature of sulfate reduction. SRB are major contributors to MIC (ASTM, 1999).

*Taxa, pl., n.* — The units of classification of organisms based on their relative similarities. Each taxonomic unit (group of organisms with greatest number of similarities) is assigned, beginning with the most inclusive, to a phylum, division, class, order, family, genus, and species (ASTM, 1999).

## Chapter II: Literature Review

### Overview

This chapter reviews and summarizes the literature regarding microbial contamination in aviation fuels. It covers the history of microbial contamination, problems associated with microbial growth in aviation fuel systems, conditions required for microbial growth, as well as a summary of the microorganisms that have been identified in previous studies. Additionally, the various types of aviation fuels and fuel additives are presented. The 16S rDNA gene sequencing and comparative analysis method used in this study will be introduced and explained. Finally, the database used to characterize the microbial contamination, and the software packages used to calculate the various diversity parameters in the analysis will be introduced, and their capabilities and limitations discussed.

### Historical Background

Reports of microbial contamination in petroleum products have been well documented over the past century (Finefrock, Killian, & London, 1965; Robbins & Levy, 2004; Zobell, 1946). The first documented case of microbial colonization of petroleum products was in 1895. The fungi *Botrytis cinera* was reported to have penetrated a thin layer of paraffin wax, a substance that was previously considered to be biologically inert—numerous studies were to follow (Zobell, 1946). One of the earliest reports of microbial contamination in fuels was reported in the 1930's, when bacteria was recognized as being responsible for accelerated corrosion and increased sulfur content in aircraft fuel storage systems (Neihof, 1988). Further research proved that

microorganisms were able to utilize hydrocarbons as a sole carbon source (Bushnell & Haas, 1941).

It was not until the 1950's that the US Air Force began to take notice when reports of microbial contamination problems in aviation gasoline (early 1950's) and aviation kerosene (late 1950's) began to surface (Bakanauskas, 1958; Finefrock & London, 1966). In 1956, flight operations of the B-47 and KC-97 were impaired when malfunctions in the aircrafts' fuel control systems (B-47) and refueling equipment (KC-97) were noted. Investigation of the problem showed an accumulation of sludge in the aircraft's fuel tanks. The sludge accumulation was subsequently traced back to a brown sludge found in the water-bottoms of the underground fuel storage tanks from which the aircraft had been refueled. Closer inspection of the sludge material found that it contained large numbers of living bacteria and their associated metabolic by-products. These findings certified that the presence of microorganisms resulting in sludge accumulation was a common occurrence in fuel tanks used to store aviation fuels (Bakanauskas, 1958).

In 1958 a US Air Force B-52 crash was directly attributed to the clogging of fuel screens and filters (Finefrock & London, 1966). The clogging appeared to be due to the presence of some form of fuel contaminant and ice formation (Finefrock & London, 1966). Many more organizations, to include US Navy and Royal Austrian Navy, were also becoming aware of the existence of microbial contamination in their jet fuel storage areas (Finefrock & London, 1966). These widespread findings prompted an Air Force wide investigation consisting of 11 different contractual efforts to further investigate the subject of microbial contamination in aviation fuels (Finefrock et al., 1965).

One effort examined 72 samples from aircraft fuel systems and were able to characterize 43 microorganisms and classify them into nine genera of bacteria and three genera of fungi (Edmonds & Cooney, 1967). Another investigation studied 19 species of bacteria that were believed to be representative of the types that naturally occur in aircraft fuel tanks and storage systems (Hedrick et al., 1963). By inoculating the microorganisms in pure cultures of hydrocarbon fuel medium the species that remained active after five months of inoculation were selected as candidates for the study of contamination control techniques (Hedrick et al., 1963).

Microbial contamination of military and civilian aircraft remained a top priority into the 1960's (Neihof, 1988). Several factors were occurring at that time that may have led to the steady increase in occurrences including the conversion to jet engines, new wing tank configurations, and the conversion to kerosene type fuels (Maurice, Lander, Edwards, & Harrison, 2001; Neihof, 1988). However, by 1963, fewer problems due to microbial contamination in jet fuel were being reported (Rauch, 2008). The decline was attributed to the inclusion of a fuel system icing inhibitor, Ethylene Glycol Monomethyl Ether (EGME) which was introduced as a fuel additive for JP-4 in 1962, and was found to have biocidal properties (Finefrock & London, 1966; Meshako et al., 1999; Neihof & Bailey, 1978). Better housekeeping procedures (i.e. proactive maintenance via improved water bottom removal) are believed to have contributed to the minimization of microbial contamination in aviation fuel as well (Neihof, 1988).

**Types of Aviation Fuel**

Aviation fuel is a specialized type of petroleum-based fuel used to power aircraft. It is generally of a higher quality than fuels used in less critical applications such as heating or road transportation (Hemighaus et al., 2006). The primary function of aviation fuel is to provide propulsive energy to the aircraft. Therefore the composition of aviation fuels has been primarily determined by specifications based upon performance and operational requirements. These include energy content, combustion characteristics, lubricity, stability, fluidity, corrosion protection and volatility, among others (Hemighaus et al., 2006). Availability and cost also play a factor (Hemighaus et al., 2006). Besides providing a source of energy, fuel is also used as a hydraulic fluid in engine control systems and as a coolant for certain fuel system components (Hemighaus et al., 2006).

It was recognized soon after the first jet-powered aircraft flew that the current aviation fuel, avgas, was unacceptable for long-term use due to problems caused by its chemical properties (Maurice et al., 2001). Problems included engine malfunctions at certain altitudes due to volatility and lubricity issues. For example, the lead in early fuels caused erosion of the turbine blades (Maurice et al., 2001). These issues led to efforts to find a better fuel. It was found that if the kerosene fraction of crude oil was used instead of gasoline many of the problems would be alleviated and some additional benefits were made available (Maurice et al., 2001). For example, the range the planes could fly was increased, less soot was produced, and combustors had to be replaced less frequently (Maurice et al., 2001).

Many types of fuel can be manufactured from crude oil, each with its own specific use (Maurice et al. 2001). Aviation fuel is the kerosene cut from the distillation

of petroleum and is a mixture of thousands of hydrocarbons (Edwards, 2003). It consists primarily of long, single branched chains of carbon and hydrogen, or alkanes, ranging from 10 carbons in length to 20 carbons in length (Rauch, 2008). While the major component of jet fuel is alkanes, there are typically small amounts of aromatic hydrocarbons, sulfur species, nitrogen species and trace metals (Rauch, 2008). Hydrocarbon chain length and size (molecular weights or carbon numbers) is restricted by the operational requirements for the product, for example, freezing point or smoke point (Maurice et al., 2001).

Aviation fuels are sometimes classified as kerosene or naphtha-type. Kerosene-type fuels include Jet A, Jet A1, JP-5 and JP-8. Kerosene-type jet fuels have a carbon number distribution between about 8 and 16 carbon numbers (Gaylarde, Bento, & Kelley, 1999). Naphtha-type jet fuels, sometimes referred to as "wide-cut" jet fuel, include Jet B and JP-4 (Hemighaus et al., 2006). Naphtha-type jet fuels have a carbon number distribution between about 5 and 15 carbon numbers (Hemighaus et al., 2006).

Due to distinctive flight missions, the specifications vary for military and civilian aviation fuels. A fuel specification is simply a method for those involved (users and producers) to ascertain and manage the desired traits of each type of aviation fuel (Hemighaus et al., 2006). Military and civilian aviation went through several variations or specifications of fuel before finding one that worked for the customer and refiners (Maurice et al., 2001). Most current jet fuels are described using five main specifications. The three specifications in civil use are American Society for Testing and Materials (ASTM) D 1655, British Defense Standard (Def Stan) 91-91, and Gosudartsvennye Standarty (GOST) 10227. Def Stan 91-91 replaced Directorate of

Engine Research and Development (DERD); the first jet fuel specification published was DERD 2482 in England in 1947 (Edwards, 2003; Hemighaus et al., 2006). GOST 10227 are the Russian specifications. The Joint Check List has been established by international oil companies to ensure standardization of jet fuel deliveries around the world under Jet A-1/Def Stan 91-91 (Edwards, 2003). Military fuel currently uses two specification—Military Detail (MIL-DTL) specification 83133E for JP-8 and JP-8 +100 and MIL-DTL-5624 for JP-5 (MIL-DTL-83133E, 1 April 1999) (Edwards, 2003).

Military and commercial aviation primarily use five types of fuel: Jet A, Jet A-1, JP-5, JP-8, and JP-8+100. Jet A and Jet A-1 are used by commercial carriers in the US and overseas, respectively, while JP-5, JP-8, and JP-8+100 are used by the military. Further description of the specific fuels analyzed in this thesis effort will be provided in the sections to follow.

*Fuel Additives*

Aviation fuel often contains additives to reduce the risk of icing due to low temperatures at higher altitudes, explosion due to static buildup in transport and storage, amongst other undesirable effects (Hemighaus et al., 2006). Additives account for the principal differences between current commercial and military aviation fuels (specifically Jet A and JP-8). Military fuels, signified by the term JP (Jet Propulsion), contain three or more additives. Jet A, used commercially in the United States, usually contains no additives at all or perhaps only an antioxidant (Hemighaus et al., 2006). Fuel additives are fuel-soluble chemicals added in small amounts to enhance or maintain properties that are important to fuel performance or fuel handling (Hemighaus et al., 2006). Typically,

additives are derived from petroleum based raw materials and their function and chemistry are highly specialized. Only small amounts, in the part per million (ppm) range, are required to induce the desired effects (Hemighaus et al., 2006).

Additives are used in varying degrees in all petroleum derived fuels, but the situation with aviation fuels is unique in that only those additives specifically approved may be added to jet fuel (Hemighaus et al., 2006). All jet fuel specifications list approved additives along with allowed concentrations. Some additives are required to be added, some are optional, and others are approved for use only by an agreement between the buyer and seller. Table 1 lists some of the main additives approved for use in the various aviation fuels.

**Table 1. Additive types in aviation fuels**

| Additive Type | Jet A | Jet A-1 | JP-4 | JP-5 | JP-8 |
|---|---|---|---|---|---|
| Antioxidant | Allowed | Required | Required | Required | Required |
| Metal Deactivator | Allowed | Allowed | Agreement | Agreement | Agreement |
| Electrical Conductivity/ Static Dissipater | Allowed | Required | Required | Agreement | Required |
| Corrosion Inhibitor/ Lubricity Improver | Agreement | Allowed | Required | Required | Required |
| Fuel System Icing Inhibitor | Agreement | Agreement | Required | Required | Required |
| Biocide | Agreement | Agreement | Not Allowed | Not Allowed | Not Allowed |
| Thermal Stability | Not Allowed | Not Allowed | Not Allowed | Not Allowed | Agreement |

(Derived from Hemighaus, 2006)

The two additives of interest when dealing with microbial contamination are the biocide and fuel system icing inhibitor (FSII) due to their antimicrobial properties (Hemighaus et al., 2006). No other military or commercial additives in current use are known to have toxic effects on microorganisms (Chelgren, 2008); however several are under consideration (Meshako et al., 1999). The role of FSII is to mix with any water

that develops and reduce the freezing point of the resulting mixture to prevent the production of ice crystals (Neihof & Bailey, 1978). FSII is a required additive for the military and optional for commercial aviation, while biocides are not allowed for the military but still optional for commercial aircraft (Hemighaus et al., 2006; Meshako et al., 1999). An icing inhibitor is unnecessary for commercial aircraft because they have fuel filter heaters (Hemighaus et al., 2006; Meshako et al., 1999). An icing inhibitor is required in military aircraft because fuel filter heaters are not used as every available pound is used on hardware to mission critical performance parameters (Hemighaus et al., 2006).

*Military Aviation Fuels*

Two types of JP fuel are currently being used by the U.S. Military. The Navy and Marine Corps use JP-5 during carrier operations. The Air Force, Navy and Marine Corps use JP-8 during land-based operations. Both are kerosene-type fuels. The primary difference between JP-5 and JP-8 is the flash point. JP-5 has a higher minimum flash point, which provides an additional level of safety in handling jet fuel in the unforgiving environment of carrier aviation (Hemighaus et al., 2006). A brief history of military jet fuels is provided in Table 2.

**Table 2. History of military aviation fuel**

| Fuel | Year Introduced | Type | Freeze Point °C max | Flash Point °C min | Comments |
|---|---|---|---|---|---|
| JP-1 | 1944 | kerosene | -60 | 43 | obsolete |
| JP-2 | 1945 | wide-cut | -60 | | obsolete |
| JP-3 | 1947 | wide-cut | -60 | | obsolete |
| JP-4 | 1951 | wide-cut | -72 | | obsolete |
| JP-5 | 1952 | kerosene | -46 | 60 | US Navy / Marine Corps fuel |
| JP-6 | 1956 | kerosene | -54 | | XB-70 program, obsolete |
| JPTS | 1956 | kerosene | -53 | 43 | Higher thermal stability, lower freezing point |
| JP-7 | 1960 | kerosene | -43 | 60 | Lower volatility, higher thermal stability |
| JP-8 | 1979 | kerosene | -47 | 38 | US Department of Defense fuel |
| JP-8+100 | 1996 | kerosene | -47 | 38 | US Air Force fuel containing an additive that provides improved thermal stability |

(Derived from Hemighaus, 2006)

Combat experience in Vietnam demonstrated that jet aircraft damage (and losses) due to the use of JP-4 was clearly higher than damage encountered by the Navy using JP-5 which has a higher minimum flash point (Maurice et al., 2001). This difference in aircraft damage and losses was the motivation behind the development of JP-8. JP-8 is essentially a common civilian jet fuel, Jet A, with a military additive package. This package contains three components: FSII to prevent water in the fuel from freezing, corrosion inhibitors (CI) to prevent fuel pump failures, and Static Dissipater Additive (SDA) to prevent mishaps due to static discharge while refueling (Graef, 2003). The desire to move toward a single fuel, coupled with the JP-4 safety hazards, led the Air Force to begin the conversion of all its aircraft and fuel systems to JP-8 in 1993 (Maurice et al., 2001). Conversion was completed in 1995.

Unfortunately, the heavier JP-8 led to increased maintenance costs at Air Force bases worldwide (Maurice et al., 2001). Fuel degradation was found to have caused fouling/coking in engine fuel nozzles, fuel controls, and fuel manifolds costing millions

per year (Maurice et al., 2001). This led to a joint government/industry/academia program to develop an additive package for JP-8. The additive agreed upon contained a detergent/dispersant (fuel injector cleaner), in addition to the standard additives. JP-8 with the additive package, added at approximately 250 ppm (1 quart of additive to 1000 gallons of fuel), is referred to as JP-8 + 100 (Maurice et al., 2001). JP-8 + 100 was introduced in 1994. The "plus 100" additive allows the bulk fuel temperature to increase by 55°C (from 163°C to 218°C) without generating harmful fuel system deposits, thereby increasing the thermal stability of the fuel (Maurice et al., 2001). The Air Force is now converting all fighters, bombers, trainers, and many cargo aircraft to JP-8 + 100 (Maurice et al., 2001). JP-8 is projected to remain in use at least until 2025 while JP-8 + 100 is being integrated (Defense Energy Support Center, 1998).

*Civilian Aviation Fuels*

While the military had been utilizing jet fuel since the early 1940's, commercial aviation did not emerge until about the 1950s (Hemighaus et al., 2006; Maurice et al., 2001). By the early 1960s, the civilian sector began to play a significant role in aviation. The main difference between civilian aviation fuels and JP-8 is the additive package or lack thereof, JP-8 containing the additive package. Jet A is used in the United States while most of the rest of the world uses Jet A-1 (Hemighaus et al., 2006). The important difference between the two fuels is that Jet A-1 has a lower maximum freezing point than Jet A (Jet A: –40°C, Jet A-1: –47°C) (Hemighaus et al., 2006). The lower freezing point makes Jet A-1 more suitable for long international flights, especially on polar routes during the winter. The choice of Jet A for use in the United States is driven by concerns

about fuel price and availability. Many years of experience have shown that Jet A is suitable for use in the United States (Hemighaus et al., 2006).

The only other jet fuel that is commonly used in civilian turbine engine-powered aviation is called Jet B, a fuel in the naphtha-kerosene region that is used for its enhanced cold-weather performance. Jet B's lighter composition makes it both more dangerous to handle and more expensive, and it is thus restricted only to areas where its cold-weather characteristics are absolutely necessary (Hemighaus et al., 2006).

*Alternative Aviation Fuels*

Alternative aviation fuels hold the potential for significant economic, operational and environmental benefits and the introduction of biofuels into aviation fuel systems is currently underway. In early 2008, Virgin Atlantic flew a Boeing 747 with one engine operating on a 20% biofuel mix of babassu oil and coconut oil from London to Amsterdam (Bradley, 2008). Of importance to microbiologists is that these fuels are readily biodegradable and it is probable that they would be subject to increased microbial growth during storage (Robbins & Levy, 2004). These fuels are mixtures of fatty acid methyl esters, which can be burned straight or utilized in blends with diesel fuel (Robbins & Levy, 2004). Biodiesel fuels are prepared from vegetable oils (i.e., soybean oil) or animal fats and exhibit similar chemical and physical properties as petroleum prepared diesel fuels except that the biodiesel fuels contain no aromatics or sulfur (Robbins & Levy, 2004).

In order to be viable in the commercial aviation industry, biodiesel must overcome several technical hurdles. However, the task is not insurmountable and there is

no single issue making bio-fuel unfit for widespread use (Daggett, Hendricks, Walther, & Corporan, 2007). The primary concern with biodiesel is its low temperature properties. Biodiesel has a freezing point near 0°C causing it to gel much faster than petro diesel during cold weather use (Danigole, 2007). Additionally, the increased viscosity can cause fuel filter clogging, as well as increased cloud formation from burning the fuel (Danigole, 2007). It has been shown that a twenty percent blend of biodiesel with petrodiesel reduces the freezing point enough to enable the use of biodiesel under most conditions experienced by diesel-based transportation (Daggett et al., 2007). In this study, the microbial contaminants in JP-8, Jet A, and biodiesel will be examined and compared.

**Growth Requirements of Microorganisms**

Jet fuel is sterile when it is first produced due to the high temperatures of the refinery process (Hemighaus et al., 2006). However, it quickly becomes contaminated by microorganisms that are ever present in air, water, or fuel system into which the sterile fuel is being added (Chesneau, 1988). Aviation fuel provides the necessary food (hydrocarbons), water, and most of the basic nutrients required by microorganisms (Robbins & Levy, 2004). Microorganisms require free water, an organic nutrient source for energy, inorganic nutrients and proper temperature and pH for growth (Vaccari, Strom, & Alleman, 2006). Some microorganisms require oxygen for growth, while other microorganisms grow in the absence of oxygen. Figure 1 depicts a fuel storage tank demonstrating its capability to provide all of the growth requirements for microorganisms. Microbes may also be able to metabolize some fuel additives, such as

the surfactants, as nutrient sources although others have inhibitory behavior (Gaylarde et al., 1999; Rahman & Gakpe, 2008). Some bacterial cells and fungal spores can survive dormant in dry fuel for months to several years (*Hormoconis resinae*) (Robbins & Levy, 2004). Cells require water for growth and reproduction therefore the bioburden in fuel tanks exists primarily at the fuel/water interface (Figure 2) where all their growth requirements can be provided (ASTM, 1999).

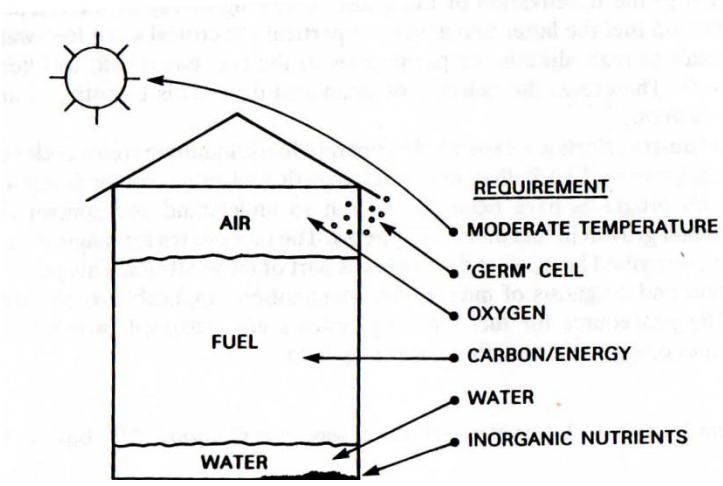



**Figure 1. Fuel tanks provide all requirements for microbial growth (Swift, 1988)**

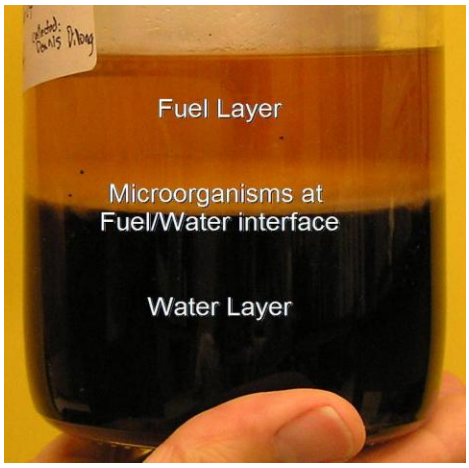



**Figure 2. Picture of a biofilm at the fuel/water interface**

*Water*

Free water is a fertile growing environment for microorganisms as it is the primary requirement for microbial growth (Gaylarde et al., 1999). When fuel is first delivered to the fuel tank, there may be little or no free water present. Free water becomes available from rainwater (especially in storage tanks with "floating roof" tops), ship ballast water, water leaking through faulty tank seals and vents in the system, residue from tank cleaning and in the fuel delivery (Chesneau, 1988; Robbins & Levy, 2004). Water also exists due to the inevitable presence of condensation. As the fuel cools, water will condense and free water droplets will form on the sides and bottom of the tank. Water is heavier than fuel, so it generally falls to the bottom of the tank. As microbes start to grow, cellular metabolism produces more free water (water is an end product of hydrocarbon degradation). *Hormoconis resinae* can produce 0.94g water per liter of fuel after four weeks (Robbins & Levy, 2004).

Dissolved water is also present in the fuel. The amount of water solubility in fuel is related to the hydrocarbon chain length, the presence of an aromatic structure, and

temperature (Robbins & Levy, 2004). Shorter chain paraffin dissolve more water than the longer chain paraffin (Robbins & Levy, 2004). An aromatic hydrocarbon can dissolve five times more water than straight chain hydrocarbons (Robbins & Levy, 2004). There is 1 part per million of dissolved water in aviation kerosene fuel for every degree Celsius (C) above zero (Gaylarde et al., 1999). Based on these facts kerosene fuels are more susceptible to microbial attack than other hydrocarbon fuels because they have a greater capacity to absorb dissolved water (Robbins & Levy, 2004).

*Organic nutrients – hydrocarbons*

There are an abundance of nutrient sources available for microorganisms in the fuel storage tank. Hydrocarbons (80 to 89% carbon) serve as a carbon source for a wide variety of microorganisms (Atlas, 1981; Rauch et al., 2005; Zobell, 1946). Microorganisms can metabolize straight chain aliphatic hydrocarbons and the lower molecular weight cyclic and aromatic molecules found in petroleum fuel for their energy production (Robbins & Levy, 2004). Microorganisms start to degrade these fuel hydrocarbons at the same time, but at different rates of activity. Straight chain alkanes are degraded the most rapidly (Atlas, 1981). The branched alkanes, cycloalkanes and aromatics are more slowly degraded (Atlas, 1981).

*Oxygen*

Oxygen is used by aerobic microorganisms to generate energy for growth. Obligate aerobic microorganisms require oxygen for respiration and biosynthesis (Vaccari et al., 2006). Facultative aerobic microorganisms, such as *Escherichia coli*, may grow aerobically in the presence of oxygen or fermentatively in the absence of oxygen

(Robbins & Levy, 2004). Microorganisms such as *Pseudomonas* utilize oxygen for aerobic respiration, but may use nitrate for anaerobic respiration (Robbins & Levy, 2004). Kerosene fuel may contain > 300 ppm of dissolved oxygen (Robbins & Levy, 2004).

Anaerobic microorganisms, such as sulfate reducing bacteria (SRB), are microorganisms that grow in the absence of oxygen. They are unable to generate energy by using oxygen as a terminal electron acceptor (Vaccari et al., 2006). SRB have been isolated from contaminated fuel tanks that were generally heavily fouled with microorganisms (Robbins & Levy, 2004). Heavy contamination of aerobic microorganisms in the water bottoms can produce biomass formation with anaerobic conditions underneath. Also, oxygen can be depleted by aerobic microbial respiration creating anaerobic conditions in areas of the water bottom (Robbins & Levy, 2004).

*Inorganic nutrients*

The major inorganic nutrients needed for microbial growth and metabolism include nitrogen, sulfur, phosphorus, potassium, magnesium, calcium and iron (Vaccari et al., 2006). Trace elements of cobalt, copper, manganese, molybdenum, selenium and zinc are also required by most microorganisms (Vaccari et al., 2006). Sodium chloride, tungsten and nickel may be needed by some microorganisms (Robbins & Levy, 2004). These inorganic nutrients are available in tank sediment, water and dust. Phosphorus is considered to be one of the major growth limiting factors in fuel since it is present at less than 1 ppm (Gaylarde et al., 1999). Reportedly, fuel additives can provide these

nutrients, such as nitrogen and phosphorus both from organic amines and nitrogen and sulfur from gum inhibitors (Robbins & Levy, 2004).

*Temperature*

Each microorganism has a range of minimum, optimal and maximum temperature that affects its growth and survival. As the temperature increases within this range, the metabolism of the microorganism increases (Vaccari et al., 2006). Above the maximum temperature, cellular metabolism ceases to function and the microorganism dies (Vaccari et al., 2006). The optimal temperature for the growth of most fuel microorganisms is 25°C to 30°C (Robbins & Levy, 2004). The average moderate temperature in the fuel tank is 20°C to 30°C (Robbins & Levy, 2004). However, microbial growth has been reported in fuel with temperatures ranging from -2°C to 55°C (Robbins & Levy, 2004).

*pH*

Microbial growth has been discovered at extreme pH levels of < 1.0 for acidophiles to 13.0 for alkalophiles (Vaccari et al., 2006). In general, the majority of bacteria prefer a neutral pH (Vaccari et al., 2006). Fungi prefer slightly acidic conditions (pH 4-6) for growth and SRB grow best at pH 7.5 (range of growth is pH 5 to pH 9 (Robbins & Levy, 2004). The pH of a fuel storage tank water bottom is generally between 6 and 9, so pH should not limit the ability of most microorganisms to grow in this environment (ASTM, 1999). Seawater, used as ballast in marine vessels, has a pH of approximately 8 (Neihof, 1988). Hydrocarbon-utilizing microorganisms can lower the water bottom pH by producing organic acids (Gaylarde et al., 1999). SRB can raise the

water bottom pH by removing the organic acids that are produced by the hydrocarbon-utilizing microorganisms (Robbins & Levy, 2004).

**Microorganisms Commonly Found in Aviation Fuel**

Microorganisms found in aviation fuels include bacteria and fungi (yeasts and molds). In 1946, ZoBell noted that almost one hundred species of bacteria, yeasts, and molds covering thirty genera had been described which can attack at least one type of hydrocarbon (Zobell, 1946). This number has grown as detection techniques have evolved (Denaro, 2005; Rauch et al., 2005). Although there is consistency among studies from the 1950's to the late 1990's which show that although many types of microorganisms have been discovered in fuel systems only a few have the ability to survive and multiply in tank bottoms and other water associated with aviation fuel (Bakanauskas, 1958; Crum, Reynolds, & Hedrick, 1967; Edmonds & Cooney, 1967; Gaylarde et al., 1999; Hedrick et al., 1963). Organisms of concern appear to be a part of the normal environmental population (Van Hamme et al., 2003; Zobell, 1946). Although some organisms appear most commonly in fuel systems, they do not seem to be particularly specialized for the hydrocarbon environment and appear to have other occupations in the natural environment (Van Hamme et al., 2003; Zobell, 1946) . Table 3 summarizes the results of several microbial contamination studies.

## Table 3. Microbial contaminants isolated from aviation fuels (1958-2005)

| **Bacteria** | JP-4 1958-1966 | Jet A 1988-1997 | Jet A-1 1998-1999 | JP-8 2002 | JP-8 2003 | Jet A 2005 |
|---|---|---|---|---|---|---|
| *Acidovorax* | | | | | | x |
| *Acinetobacter* | | x | x | | | |
| *Arthrobacter* | x | | x | | x | x |
| *Aerobacter* | | x | x | | | |
| *Aeromonas* | | x | x | | | |
| *Alcaligenes* | | x | x | | x | x |
| *Aquabacterium* | | | | | | x |
| *Aquasprillum* | | | | | | x |
| *Bacillus* | x | x | x | x | x | x |
| *Bradyrhizobium* | | | | | | x |
| *Brevibacterium* | x | | x | | | |
| *Burkholderia* | | | | | | x |
| *Caulobacter* | | | | | | x |
| *Clostridium* | | | | | | x |
| *Curtobacterium* | | | | | | x |
| *Desulfovibrio (SRB)* | x | x | x | | | |
| *Diaphorobacter* | | | | | | x |
| *Dietzia* | | | | | x | |
| *Escherichia* | x | | | | | x |
| *Enterobacter* | | | x | | | |
| *Ewingella* | | | | | | x |
| *Flavobacterium* | x | x | x | | | x |
| *Granulicatella* | | | | | | x |
| *Haemophilus* | | | | | | x |
| *Herbaspirillum* | | | | | | |
| *Kocuria* | | | | | x | |
| *Lactococcus* | | | | | | x |
| *Leucobacter* | | | | | x | |
| *Methylobacterium* | | | | | | x |
| *Microbacterium* | | | | | | x |
| *Micrococcus* | x | x | x | | x | |
| *Mycobacterium* | | | | | | x |
| *Padoraea* | | | | | | x |
| *Pantoea* | | | | | x | |
| *Photorhabdus* | | | | | | x |
| *Phyllobacterium* | | | | | | x |
| *Propionibacterium* | | | | | | x |
| *Pseudomonas* | x | x | x | | | x |
| *Rahnella* | | | | | | x |
| *Ralstonia* | | | | | | x |
| *Rhizobium* | | | | | | x |
| *Rhodococcus* | | | | | | x |
| *Rothia* | | | | | | x |
| *Serratia* | | | x | | | |
| *Sphingomonas* | | | | | x | x |
| *Staphylococcus* | | | | | x | x |
| *Streptococcus* | | | | | | x |
| *Streptomyces* | | | x | | | |
| *Wolinella* | | | | | | x |

(Derived from Denaro, 2005; Rauch, 2005)

These microbes appear to be widely and abundantly distributed in nature where they may be of considerable importance in the carbon cycle and to various industries (Van Hamme et al., 2003; Zobell, 1946). For example, the microbial oxidation of hydrocarbons may help to account for the rapid disappearance of petroleum which pollutes fields and waterways, for the deterioration of certain rubber products both natural and synthetic, for the spoilage of cooling oils, for the depreciation of oiled or asphalt-surfaced highways and for the biodegradation of petroleum or its products stored in the presence of water (Zobell, 1946).

A majority of microorganisms readily degrade the alkane constituents of hydrocarbon fuels (Watkinson & Morgan, 1990). Alkanes, with the exception of $C_4$ and below are very water insoluble or hydrophobic (Rauch, 2008). Therefore, microorganisms must utilize adaptations to access the straight-chain molecules (Watkinson & Morgan, 1990). Most microbes utilize secreted biosurfactants to solubilize the alkanes prior to metabolizing them (Rauch, 2008). Unfortunately, the biosurfactants have deleterious effects on fuel systems (Rahman & Gakpe, 2008). Once the microorganisms sequester the alkane molecules there are two main routes of metabolism. The first route is through sub-terminal oxidation or the addition of a carbonyl group on a non-terminal carbon (Rauch, 2008). This carbon is then oxidized further to form acetate which then enters into the citric acid cycle to produce energy through respiration (Rauch, 2008). The other major route used for aerobic metabolism of an alkane is conversion of the alkane to an alcohol which then proceeds through the same pathway as for fatty acid metabolism, called β-oxidation (Rauch, 2008). Regardless of the pathway used,

microorganisms are indeed capable of aerobically degrading the hydrocarbon in fuel and using it as an energy source.

## Problems Associated with Microbial Contamination in Aviation Fuels

While the metabolism of hydrocarbons is obviously beneficial to the microorganisms there are several detrimental consequences from the vantage point of the fuel when uncontrolled microbial growth is allowed to develop. Microbial growth in aviation fuel systems cause fuel filter plugging, corrosion of the fuel tank, fuel degradation and increased maintenance costs and safety concerns associated with these problems and others (Rauch et al., 2005). Table 4 highlights many of the problems that have been shown to result from microbial contamination in aviation fuels.

**Table 4. Problems associated with microbial contamination (Graef, 2003)**

| Problem |
|---|
| Sludge formation |
| Aluminum corrosion and deterioration of structural properties of aluminum alloys |
| Injector fouling |
| Degradation of fuel quality |
| Decreased life of engine parts due to breakdown of hydrocarbons |
| Interference with engine performance (flameouts) |
| Corrosion of fuel storage tanks and distribution equipment |
| Malfunction of fuel gauges |
| Increased water content of fuel |
| Increased sulfur content of fuel |
| Clogged fuel lines |
| Oxygen and hydrogen scavenging |
| Sulfate reduction |
| Biosurfactant production/Biofilm formation |
| Additive and fuel molecule metabolism |
| Damage to organic coatings |
| Failure of water separators |

Two of the most commonly recognized symptoms of microbial contamination are microbially induced corrosion (MIC) and plugged fuel filters caused by biofilms. The following sections provide a thorough description of these unfavorable symptoms.

*Biofilms*

A major problem associated with microbial contamination is the formation of biofilms. Biofilms are structured and organized accumulations of microbes in matrices of extracellular polymeric substances (EPS), proteins, nucleic acids, and other components (Chelgren, 2008; Costerton, Lewandowski, Caldwell, Korber, & Lappin-Scott, 1995; Davey & O'Toole, 2000; Zhang, Choi, Dionysiou, Sorial, & Oerther, 2006). Biofilms are essential for the transfer of metabolic products and for allowing nutrients, including oxygen, to flow through the system (Costerton et al., 1995; Davey & O'Toole, 2000). Despite this flow of oxygen it is still possible to have anaerobic pockets and places where denitrification can occur within the biofilms (Chelgren, 2008; Costerton et al., 1995; Davey & O'Toole, 2000).

Observation has shown that microorganisms normally exist as a member of an ordered biofilm ecosystem and are not free floating (Davey & O'Toole, 2000). Biofilms may be somewhat advantageous for microbes because they provide some measure of shelter, protection, and homeostasis; multispecies biofilms may also allow substrate exchange, dispersal and/or removal of metabolites, or the formation of syntrophic relationships (Davey & O'Toole, 2000). Syntrophic relationships are a subset of symbiotic relationships where two metabolically diverse microbes are reliant on the other

to use specific substrates, normally for energy manufacturing (Davey & O'Toole, 2000) (Vaccari et al., 2006).

The formation of biofilms can be influenced by many different factors including what microbes are present, flow conditions, nutrient availability, and local environmental parameters (Davey & O'Toole, 2000). These biofilms may be composed of a single species or a consortium of species (Davey & O'Toole, 2000). The microbes that initially colonize the surface are believed to alter surface properties and thus permit the attachment of other microbes less able to colonize at the beginning; these are known as pioneer species (Zhang et al., 2006). The microbes present in a biofilm alter the pH, oxygen availability, and types and levels of ions at the metal-solution boundary and thus influence corrosion (Gaylarde et al., 1999).

Research has indicated that initial colonization may be the result of certain bacterial populations and not the total biomass. This suggests further research on controlling biofilms should concentrate on these specific bacterial populations (Zhang et al., 2006). One potential pioneer species that may allow further development of a biofilm is *Acinetobacter* (Zhang et al., 2006). A characteristic of *Acinetobacter* that is thought to play a role in its ability to be a pioneer colonizing species is its motile structure (flagella) that it uses to move about surfaces, and its ability to form branching filaments (Zhang et al., 2006)

Biofilms are directly responsible for fuel filter plugging. Two distinct mechanisms can cause this problem. When flocs of biomass are transported through the fuel system and are trapped in the filter medium, they can restrict flow. Direct observation of filters plugged by this mechanism reveal masses of slime on the filter

element's external surfaces (Bakanauskas, 1958). Alternatively, microbial contaminants may colonize filter media. The biofilms they produce within the filter medium's matrix eventually plug the filter (ASTM, 1999).

Biofilms play a key role in support of MIC (Chelgren, 2008). Biocides are used to prevent biofilm formation; however, due to the nature of the biofilm structure, a biocide may not be able to penetrate the inner parts of the biofilm (Hemighaus et al., 2006).

*Microbially Induced Corrosion*

One of the most widely recognized and most serious effects of microbial contamination is MIC or biocorrosion. Corrosion itself is an electrochemical process in which a charge difference develops in adjacent areas of the storage tank metal surface (Robbins & Levy, 2004). The water bottom in contact with the metal surface of the storage tank creates many micro areas acting as anodes and cathodes (Angell, 1999). Electrons will flow from the anode (area of lower potential) to the cathode where they are consumed by different reactions (water and oxygen, water and hydrogen ion, hydrogen and sulfate, etc.) depending on the nature of the environment (Robbins & Levy, 2004). At the anode, pitting corrosion is initiated by the loss of metal ions into solution. MIC is usually caused by the activity of a microbial consortium rather than a single species, similar to the formation of biofilms (Beech & Sunner, 2004). Figure 3 depicts a simplified scheme of microbially induced corrosion beneath a bacterial colony.

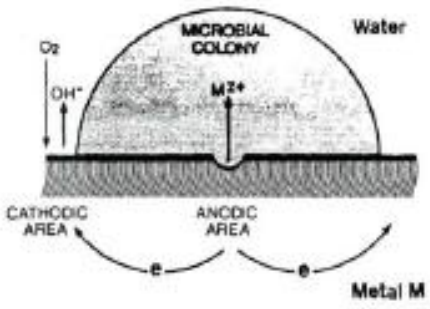



**Figure 3. Simplified scheme of MIC beneath a bacterial colony (Videla, 2001)**

The processes by which microorganisms cause accelerated corrosion include: microbial layers (sludge) on metal surfaces causing metal pitting or corrosion due to differing charge potentials between the covered and uncovered areas (the areas of lower potential will be attacked); fungi, such as *H. resinae*, producing organic acids causing the water bottom pH to drop; SRB (mainly *Desulfovibrio* and the more oxygen tolerant *Desulfotomaculum*) reducing sulfates in the water bottom to produce hydrogen sulfide; by utilizing the phosphate and nitrate components in corrosion inhibitors for growth, effectively removing the corrosion protection and indirectly aiding in the corrosion process; aerobic organisms using up the available oxygen, creating an oxygen deficient area in which SRB may thrive; and SRB producing the enzyme hydrogenase that can depolarize metal surfaces by removing hydrogen directly which makes the surface more porous and brittle (Angell, 1999; Beech & Sunner, 2004; Robbins & Levy, 2004; Zhang et al., 2006).

**Microbial Detection/Analysis Methods**

Work on hydrocarbon biodegradation by microorganisms started around 1906 (Bushnell & Haas, 1941). A key tool to study microorganisms that has evolved over the

recent decades is the emergence of enhanced DNA/RNA analysis methods. Until recently, microbial analysis relied almost solely on culture methods which do not recover all organisms present in a community. It is estimated that less than 10% of bacteria seen by direct count techniques can also be cultured (Head, Saunders, & Pickup, 1998; Hugenholtz et al., 1998). Some studies suggest that as few as 1% of microbes found in the environment have so far been cultivated and identified (Amann et al., 1995; Hugenholtz et al., 1998). Advancements in ribosomal DNA (rDNA) analysis have permitted the characterization of a wide spectrum of environmental contaminants without the requirement of cultivability (Amann et al., 1995; Clarridge, 2004; Handelsman, 2004; Head et al., 1998).

*Culture-based Methods*

Historically, microorganisms in fuels were detected not because of observation of the growth of the bacteria or fungi themselves but rather the results or symptoms of their growth (i.e. biofilms, MIC, foul odor, etc.). Nonetheless, when early researchers were attempting to determine the causes of these symptoms the first solution was to attempt to culture the microorganisms in the lab. While culturing microorganisms has several drawbacks, it is still the most widely used method to detect microbial contamination in the field (Rauch, 2008).

The most commonly used culture-based detection method is to test for "colony forming units" (CFUs). Microbial colonies are formed when several cycles of microbial cell reproduction occurs. Each colony forming unit is indicative of the presence of an individual, viable microbial cell that has reproduced. Each of the reproduced cells go on

and reproduce and so on until there are enough cells to make a small spot or "colony" of microbes visible to the naked eye (Rauch, 2008). Not only is the presence of the colony important since it indicates the presence of the original cell but, the physical appearance of the spot including color and morphology gives insight into the type of cell present (Edmonds, 1965). To use this test method samples of suspected fuel or water bottom are streaked onto an agar plate and incubated (Graef, 2003; Rauch et al., 2005). After a designated amount of time the colonies are counted. Several test kits are available commercially. The kits provide the appropriate agar media in a portable testing container as well as information on how to determine contamination level (low, medium, high) of the fuel tank by counting the number colonies that develop (Graef, 2003; Rauch et al., 2005).

The advantage to these cultivation methods is that cultures, or colonies, are physically available for further study; however, due to the challenge of growing microbes on agar plates, only a small percentage of the microbes will actually grow, resulting in a low estimate of bacterial diversity (Amann et al., 1995). This inability to culture most microorganisms is one of the biggest challenges in microbiology. It is now widely accepted that most cells that can be seen under a microscope are viable but not culturable (Amann et al., 1995). This inability to culture the vast majority of environmental microbes hampered early efforts to comprehensively analyze the issue of microbial contamination (Chelgren, 2008).

*Molecular-based Methods*

Significant progress has been realized with the arrival and expansion of molecular techniques and metagenomics analysis tools. Having side-stepped many of the limitations of cultivation-based studies, a dramatic rise in the number of recognized bacterial phylum has resulted; the decade from 1988-1998 saw a tripling in identifiable bacterial phylum (Brock, 1987; Hugenholtz et al., 1998). Efforts using molecular biology to identify environmental microbes first occurred over thirty years ago when it was realized that phylogenetic relationships among bacteria, as well as other life-forms, could be found via comparison of a stable region of the genetic code (Clarridge, 2004; Head et al., 1998; Woese & Fox, 1977). For the first time, researchers were able to classify and survey microbial communities in a relatively unbiased way and effectively explore microbial interactions *in situ*. Today's molecular methods allow us to examine the elusive 99% of the uncultured microorganisms by examining their DNA sequences (Pace, 1997). The molecular method used in this thesis effort is known as the 16S rDNA gene analysis method and is enhanced by the polymerase chain reaction.

**The Polymerase Chain Reaction**

The 1980's saw the inception of a revolutionary technique, the Polymerase Chain Reaction or PCR, which dramatically sped up the DNA analysis process by permitting amplification of only a select region or gene of interest (Amann et al., 1995; Mullis et al., 1986). As a result PCR is a technique that is widely used in molecular biology today (Appenzeller, 1990).

DNA is a nucleic acid that contains the genetic instructions for life. The main role of DNA molecules is to provide long-term storage of genetic information. DNA is organized into structures called chromosomes. Chromosomes are made up of many segments of genetic information and these segments are known as genes (Vaccari et al., 2006). Chemically, DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds. These two strands run in opposite directions to each other and are therefore complementary. Attached to each sugar is one of four types of molecules called bases (adenine, thymine, guanine and cytosine). It is the sequence of these four bases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of nucleotides within proteins. Gene function can often be inferred from the nucleotide sequence, either from protein structure or comparison to known genes (Kersey & Apweiler, 2006). The code is read by copying stretches of DNA into the related nucleic acid RNA, in a process called transcription (Vaccari et al., 2006). Transcribed stretches can then be duplicated in a process called DNA replication. This is the natural process that PCR is used to enhance.

PCR derives its name from one of its key components, a DNA polymerase, used to amplify a piece of DNA by *in vitro* replication. DNA polymerase is an enzyme that reads singled stranded DNA and synthesizes its complementary strands by using the original piece of DNA as a template. As PCR progresses, the DNA generated, as well as the original, is used as a template for replication. This sets in motion a chain reaction in which the DNA template is exponentially amplified. The result is a highly concentrated solution of only the gene or segment of a gene selected for analysis (Vaccari et al., 2006).

In order to point the DNA polymerase in the proper direction for replication a primer is used. A primer is a short strand of nucleotides (approximately 20 bp) that serves as a starting point for DNA replication. The choice of appropriate primers to amplify the beginning of a replicated gene is highly dependent on the project's research goals. In this project, the goal was to identify and differentiate between as many bacteria as possible from the aviation fuel samples. Therefore, primers were constructed from the conserved regions at the beginning of the gene (forward primer) and at the cutoff (reverse primer) (Figure 4) (Baker, Smith, & Cowan, 2003; Clarridge, 2004). These primers are often referred to as "universal" because they are built from the conserved regions that all bacteria have. However, no primer can be designed to completely anneal to all bacteria since there is variability between bacteria and other organisms (Baker et al., 2003). The "universal" primers used in this project introduced bias into the results, because they were designed to anneal to bacterial genes, but could anneal to genes from other organisms that are not within the bacterial domain (Baker et al., 2003). Furthermore, they may not anneal well to the genes of some bacteria (Baker et al., 2003).

**16S rDNA "Fuelbug" Forward Primer: TGG AGA GTT TGA TCC TGG CTC A**

**16S rDNA "Fuelbug" Reverse Primer: GCT GCT GGC ACG TAG TTA GC**

**Figure 4. Forward and reverse primers used for PCR**

### 16S rDNA Gene Analysis Method

Biologically defining organisms with molecular technology uses the concept of phylogeny. A molecular basis for this concept was reviewed by Olsen and Woese in

1993 (Olsen & Woese, 1993). This review stated that the majority of essential genes in a genome share a common heritage or evolutionary history. A gene mutates over time and, theoretically, this change can be measured, compared, and ultimately the relation between two DNA sequences can be established (Woese, 1987). This is referred to as an organism's evolutionary distance (Woese & Fox, 1977).

In order to carry out this method a particular gene for amplification and sequencing must be selected. The process of selecting a gene to determine evolutionary relationships can be streamlined by focusing on genes that perform a central function and are intimately involved in the cell's activity (Olsen, Lane, Giovannoni, Pace, & Stahl, 1986). The selected gene must also provide enough appropriate information for analysis, be present in all cells, evolve at a relatively constant rate, have enough variable regions so that differences can be seen, be capable of natural replication *in situ*, and not be transferred across organisms (Olsen et al., 1986). In most cases, the goal of efforts similar to this one is to identify the properties and makeup of a consortium of microorganisms present in a particular environment, such as hydrocarbon fuels. Therefore, the gene chosen must meet all of the above criteria and most importantly be evolutionarily linked to its relatives and variable enough to distinguish between them (Clarridge, 2004; Woese, 1987). Several genes fit this description: rRNA, RNA polymerase, elongation factor G, proton-translocating ATPases, and others (Olsen & Woese, 1993). The gene chosen by most researchers is rRNA (Clarridge, 2004). For purposes of clarity, it should be noted that rRNA is oftentimes used synonymously with rDNA, although their functions are quite dissimilar.

rRNA is a critical element of a cell's protein synthesis process, and thus is functionally and evolutionarily homologous in all organisms (Clarridge, 2004). In bacteria there are 3 different rRNAs: 5S which is ~120 nucleotides, 16S which is ~1550 nucleotides and 23S which is ~3000 nucleotides (Clarridge, 2004; Olsen et al., 1986; Woese, 1987). The exact nucleotide length varies in organisms, and the aforementioned lengths are averages. The 5S and 23S rRNAs were found to be inappropriate molecular tools for the analysis of microbial communities (Olsen et al., 1986). The 5S rRNA was not long enough to provide adequate information or detail to make an accurate comparison tool (Woese, 1987). The 23S rRNA was too large a molecule, and little research has been completed using it for genetic analysis (Olsen et al., 1986). Therefore neither has been chosen in typical research methodologies (Olsen et al., 1986). The most widely studied gene is the 16S rRNA gene (Clarridge, 2004; Schloss & Handelsman, 2004, 2006b).

The 16S rRNA gene is large enough to have conserved sequences, which are identical or nearly identical in all bacteria, and variable/hyper-variable regions (Baker et al., 2003). The variable regions provide distinguishing and statistically valid measurements of evolutionary distances, and thereby of species or other levels of classifications of bacteria (Clarridge, 2004). Regions within the 16S rRNA gene are less affected by reconfiguration that occur in the genome, and maintain a highly conserved picture of the organism's evolutionary history (Olsen & Woese, 1993). This is largely due to the fact that rRNA is a critical component of the cell's function.

For descriptions of microbial communities, the 16S rRNA gene is used in two basic ways. The entire ~1550 base pair (bp) length is sequenced when relatively few

microbes are analyzed, or a smaller 5', 500 bp region is used when sampling larger and more diverse communities. For instance, in cases requiring detail, such as describing a new species, it is appropriate to sequence the entire 16S rDNA gene multiple times (Clarridge, 2004). Also for research to distinguish between specific taxa or strains, sequencing the entire gene would be appropriate (Clarridge, 2004). However, when initially sampling an extremely diverse community such as microorganisms in hydrocarbon fuels, the first 500 bp provide sufficient information to differentiate between organisms. Furthermore, the first 500 bp region has been shown to hold a higher percentage of diversity than any other region. Clarridge et al. compared 100 organisms using the 1550 bp sequence and the 500 bp sequences and found the relationships to be highly similar (Clarridge, 2004). Since the goal of this thesis project was to differentiate between organisms and not to identify new species the use of the 500 bp portion of the 16S gene was justified.

In 1977, Woese & Fox, used the rRNA gene to completely transform the nomenclature of living organisms (Woese & Fox, 1977). Traditionally, living organisms had been classified into two distinct domains: Prokaryotae and Eukaryotae. However, as molecular genetics became a more common area of research, living organisms' genomes were investigated, and the traditional nomenclature became obsolete (Olsen & Woese, 1993). The rRNA gene was used to classify living organisms into three new domains (Woese & Fox, 1977). The first was Eubacteria, which includes all typical bacteria. The second was Urkaryotes, which was defined by the 18S rRNAs of the eukaryotic cytoplasm. Both of these corresponded nicely to the traditional groupings of Prokaryote and Eukaryote. However, a third classification was also introduced, Archaebacteria. The

Archaebacteria appear to be no more related to the typical bacteria than they are to eukaryotes. Investigating the genetic makeup of organisms has unlocked an entirely new classification system (Woese & Fox, 1977). This classification system has become the basis for all current molecular studies, including this thesis effort.

*Limitations*

While the introduction of molecular-based strategies to the study of microbial populations has overcome many of the traditional limitations of culture-based methods they are not without limitations themselves. The 16S gene analysis methodology used in this thesis effort is subject to certain biases and limitations as described below. The methods and techniques used to negate or account for these limitations will also be discussed.

One limitation of the 16S rDNA gene analysis method is its inability to characterize bacterial taxa to the species level, a goal that many ecologists assume to be the gold standard. Researchers commonly overestimate the precision to which the 16S gene is capable of characterizing bacterial taxa. While using the 16S rDNA gene analysis method it has become commonplace for bench top scientists to classify sequences that are 97% - 99.5% similar as the same species (Chai, 2008; Hughes, Hellmann, Ricketts, & Bohannan, 2001). However, using the DNA-DNA hybridization method, another molecular method being applied to microbial populations, those same sequences have been classified as different species (Fox, Wisotzkey, & Jurtshuk, 1992). These results emphasize the important point that relative similarity of 16s rDNA sequences is not necessarily a sufficient criterion to guarantee species identity. These findings imply that

the problem at hand relates more to species definition than to genus definition. Therefore this thesis effort stopped short of characterizing the bacterial communities to the species level and focused on classifying the bacterial communities in aviation fuels at the phyla and genera levels.

Another limitation of the 16S gene is its size relative to the genome as a whole. The 16S rDNA gene represents only 0.05% of the genome of a prokaryotic cell (Rodríguez-Valera, 2002). Given that it is common to sequence only a third to a half of the 16S gene, it is nearly impossible to predict the activities (physiology), style of life (niche) or biotechnological properties of the organism based on 16S alone (Rodríguez-Valera, 2002). There are examples where bacterial strains relate by more than 97% similarity at the 16S rDNA level but behave very differently physiologically and ecologically (Achenbach & Coates, 2000).

It is also necessary to discuss the biases associated with the polymerase chain reaction. Although PCR has become a routine and accepted method of DNA amplification several problems arise when the method is applied to environmental microbial communities (Wintzingerode, Göbel, & Stackebrandt, 1997). It has been shown that a single species of bacterium can contain multiple copies of the 16S gene (Dahllof, Baillie, & Kjelleberg, 2000). Therefore, PCR, a method of systematically amplifying small sequences of DNA, can dramatically bias the frequency distribution of the final mixture relative to the original mixture (Suzuki & Giovannoni, 1996). This bias is strongly dependent on the number of cycles of replication (Suzuki & Giovannoni, 1996). A possible solution to this bias is to remove a portion of the sequences that may

cause the data to be skewed. This process is further explained in the methodology section.

The important aspect to take away from this discussion is that these methods are not without bias or limitation but they are currently the best that molecular biology has to offer. However, it should be noted that these biases and limitations are universal; therefore results and conclusions are relative and can be compared.

**Microbial Diversity Statistics**

Diversity, ecologically speaking, is often defined as species richness (Hughes et al., 2001). Richness is defined as the number of unique taxonomic units present in a community (Nubel, Garcia-Pichel, Kuhl, & Muyzer, 1999). Microorganisms are the most abundant and species-rich group of organisms on the planet making it impossible to sample a community exhaustively (Harwood, 2008) (Hughes et al., 2001). Therefore statistics must be used to estimate the true diversity of a microbial community. Percent genetic similarity has become an accepted method of defining phylogenies although the topic is heavily debated as to what level of similarity defines a unit (Schloss & Handelsman, 2005). Contrary to genetic similarity, genetic distance may also be used with equal significance. The genetic distance between two sequences is the percentage of nucleotides in one sequence that are different from those in another after correcting for multiple substitutions, for example, by computing the maximum-likelihood distance with the Jukes-Cantor nucleotide substitution model (Jukes, Cantor, & Munro, 1969; Schloss et al., 2004). Thus far, no explicit criteria have been published (Schloss & Handelsman, 2005). However, researchers have theoretically proposed that >99% similarity (1%

distance) relates to the species level, >97% (3% distance) relates to the genus level, >90% (10% distance) relates to the family level, and >80% (20% distance) relates to the phylum level (Schloss & Handelsman, 2005). Therefore if a DNA sequence is >97% similar to or 3% distant from another DNA sequence, the organisms from which the sequences originated are assumed to be of the same genera. These cutoff values are empirically derived from modern rRNA sequence data and are not yet a validated classification system (Schloss & Handelsman, 2005). However, while this criterion is not yet validated, it is possible to compare community richness as long as unit definition is consistent throughout a study and the individual researcher maintains an intuitive sense of what is being analyzed (Hughes et al., 2001) (Konstantinidis & Tiedje, 2005). Researchers often dismiss the desire to define an organism at a specific taxonomic level and instead assign organisms to operational taxonomic units (OTUs). OTUs are basic groupings determined by sequence similarity. OTUs are then used for comparison of richness at the various phylogenetic levels in a metagenomic analysis.

This thesis effort uses the aforementioned cutoff values to characterize microbial contaminants in the various aviation fuels at the genus ($OTU_{0.03)}$ and phylum ($OTU_{0.20}$) levels. The genus level was selected instead of the species level based on the notion that the 16S gene does not provide enough information to classify at the species level (Chai, 2008; Fox et al., 1992; Konstantinidis & Tiedje, 2005).

A variety of statistical approaches have been developed to compare and estimate species richness from samples of macroorganisms (Chao, 1984; Chao & Lee, 1992; Chazdon, Colwell, Denslow, & Guariguata, 1998; Good, 1953; Gotelli & Colwell, 2001; Heck, Belle, & Simberloff, 1975). Studies have shown that these approaches may be

applied to the microbial world despite its greatly increased diversity (Hughes et al., 2001). In the following sections, four approaches used to investigate microbial diversity in this study are introduced. While species will often be referred to as the measured unit of diversity in these approaches they can be applied to any level of taxonomy with equal success (Hughes et al., 2001).

*Rarefaction*

In any community, the number of types of organisms observed increases with sampling effort until all types are observed. The relationship between number of types observed and sampling effort gives information about the total diversity of the sampled community (Hughes et al., 2001). This information can be plotted on an accumulation curve. An accumulation curve is a plot of the cumulative number of types observed versus sampling effort (Hughes et al., 2001) (Gotelli & Colwell, 2001). Because all communities contain a finite number of species, if sampling continued indefinitely, the curves would eventually reach an asymptote at the actual community richness. Due to the extremely high diversity in microbial communities it is nearly impossible to sample at this level; thus an asymptote will rarely be reached and the true richness will be remain unknown using this method (Hughes et al., 2001). However, the shape of the curve contains information as to how well the communities have been sampled (i.e., what fraction of the species in the community have been detected). The more concave-downward the curve, the better sampled the community (Hughes et al., 2001).

Rarefaction is a technique for comparing environments that have been unequally sampled (Hughes et al., 2001) (Heck et al., 1975). Rarefaction curves are randomized

species accumulation curves created by a repeated re-sampling algorithm (Gotelli & Colwell, 2001). Rarefaction curves can be standardized by proportions of DNA sequences sampled and number of OTUs observed. From these curves, richness can be compared as well as sampling effort considered (Hughes et al., 2001). Constructing rarefaction curves for the genus and phylum levels allowed for meaningful standardization and comparison of datasets but does not estimate true richness (Gotelli & Colwell, 2001).

*Coverage*

Coverage was first introduced and defined by I.J. Good in 1953 as an added indication of sampling effort. Good defined coverage (C) by the following formula:

$$\boldsymbol{C = 1 - \frac{n_1}{N}}$$

where N is defined as the community size and $n_1$ is defined as the number of species appearing only once (Good, 1953). Good's coverage has been defined as a "non-parametric estimator of the proportion of organisms in a community of infinite size that would be represented in a smaller community" (Kemp & Aller, 2004). The coverage of a given sequence library describes the extent to which the sequences in the library represent the total population (Singleton et al., 2001). This parameter is presented as a percentage; therefore, the higher the percentage, the higher the coverage, or sampling effort, for that particular community.

*Chao1*

A non-parametric richness estimator was defined by Chao in 1984—Chao1. Chao1 estimates the total species richness by the formula:

$$\boldsymbol{S_{Chao1} = S_{obs} + \frac{n_1^2}{2n_2}}$$

where, $S_{obs}$ is the number of observed species, $n_1$ is the number of singletons, or species occurring only once, and $n_2$ is the number of doubletons, or species occurring twice (Chao, 1984; Hughes et al., 2001; Schloss, 2005). This estimator is particularly useful when data sets are skewed toward the low-abundance classes, as they are likely to be in microbial communities (Hughes et al., 2001). The above formula is used to calculate Chao1 only when $n_1=0$ and $n_2 \geq 0$. When $n_1>0$ and $n_2 \geq 0$ and when $n_1=0$ and $n_2=0$ the following formula is used (Colwell & Coddington, 1994; Hughes et al., 2001; Schloss, 2005):

$$\boldsymbol{S_{CHAO1} = S_{obs} + \frac{n_1(n_1 - 1)}{2(n_2 + 1)}}$$

*ACE*

A second non-parametric richness estimator was defined by Chao in 1992—ACE (Chao & Lee, 1992). The abundance-based coverage estimator (ACE) incorporates data from all OTUs with fewer than 10 individuals. This includes more than just the singletons and doubletons as in the Chao1 estimator. ACE estimates OTU richness as:

$$\boldsymbol{S_{ACE} = S_{abund} + \frac{S_{rare}}{C_{ACE}} + \frac{F_1}{C_{ACE}} \gamma_{ACE}^2}$$

where, $S_{abund}$ is the number of abundant species (>10 observed) and $S_{rare}$ is the number of rare species (≤10 observed). Note that $S_{rare} + S_{abund}$ equal the total number of observed

species. $\boldsymbol{C_{ACE} = 1 - \frac{F_1}{N_{rare}}}$ estimates sample coverage where $F_1$ is the number of species with one individual and $N_{rare}$ is the number of rare sequences in the community. Finally,

$$\boldsymbol{\gamma^2_{ACE} = \left[\frac{S_{rare}}{C_{ACE}} \frac{\sum_{i=1}^{10} i(i-1)n_i}{N_{rare}(N_{rare}-1)} - 1, 0\right]}$$

which estimates the coefficient of variation of the $F_1$'s (Hughes et al., 2001).

Chao1, ACE and rarefaction values can be graphed as a function of the number of sequences analyzed resulting in asymptotic richness curves to investigate community richness and/or sampling effort (Chazdon et al., 1998). Asymptotic richness estimators provide lower-bound estimates for species-rich groups such as microorganisms, in which observed richness rarely reaches an asymptote, despite intensive sampling (Gotelli & Colwell, 2001). Both the ACE and the Chao 1 estimators underestimate true richness at low sample sizes, which most microbial samples are expected to be, and are therefore looked at as lower bounds of estimated microbial diversity (Hughes et al., 2001). These estimators are automatically calculated over the various similarity levels by the metagenomics programs discussed in the sections to follow.

**Metagenomic Analysis Programs**

Metagenomics is the genomic analysis of populations of microorganisms from an environmental sample (Handelsman, 2004). Numerous diversity estimators and comparative analysis software programs have been published over the years to facilitate the use of metagenomics to pursue statistically sound genome based ecological analyses (Schloss & Handelsman, 2008) (Gotelli & Colwell, 2001). Many of the available programs are capable of using sequence data from the 16S rDNA sequencing method to

comprehensively characterize a microbial community in ways that were not possible just a few years ago. The following sections will describe the programs used in this study to include their background and purpose, as well as required inputs and resulting outputs.

*RDP*

The Ribosomal Database Project (RDP) is a web-based sequence repository that provides ribosome related data and services to the scientific community, including data analysis, sequence alignment and a host of other tools in support of a robust metagenomic analysis of 16S rDNA sequences (Cole et al., 2007; Cole et al., 2005). The RDP was developed by the Center for Microbial Ecology and the Department of Microbiology and Molecular Genetics at Michigan State University. As of December, 2008 the RDP maintained 715,637 unique rRNA sequences available for sequence comparison and classification (http://rdp.cme.msu.edu/).

RDP has several functions that are available to the online user. Studies have used RDP primarily to classify sequences using its Classifier function. The RDP Classifier uses a naive Bayesian classifier to assign sequences to the RDP Taxonomy (Wang, Garrity, Tiedje, & Cole, 2007). The RDP Taxonomy is trained on the new phylogenetically consistent higher-order bacterial taxonomy proposed in the most recent update of the Taxonomic Outline of Bacteria and Archaea (TOBA) (Cole et al., 2007; Garrity, 2007; Wang et al., 2007). The classifier assigns a rDNA sequence to the lowest taxonomic level possible within a certain degree of confidence (80% default); genus being the lowest level available through RDP based on literature supporting the theory that the 16S gene does not provide sufficient phylogenetic basis to classify a sequence at

the species level (Cole et al., 2007; Konstantinidis & Tiedje, 2005). The RDP Classifier interface has been designed to make it relatively simple to work with large numbers of DNA sequences (Cole et al., 2007). In 2006, Kuske et al. used the RDP classifier to classify DNA sequences from soil samples to four pathogenic bacteria, including *Bacillus anthrasis* (Anthrax), and identified closely related species in over a third of soil samples (Kuske, Barns, Grow, Merrill, & Dunbar, 2006). This research has a significant impact on the ability to positively detect biological threat agents in environmental samples (Kuske et al., 2006).

The most recent update to RDP included the addition of MyRDP Space (Cole et al., 2007). MyRDP allows researchers to upload and maintain their own private sequence collection on the RDP servers for easy manipulation and grouping of sequences. Uploaded sequences are automatically aligned with the RDP public alignment using the RDP's modified version of RNACAD (Brown, 2000), a stochastic context-free grammar based aligner trained with the secondary structure model of Robin Gutell and colleagues (Cannone et al., 2002; Cole et al., 2007). Sequence alignment is a way of arranging the sequences of DNA in order to identify regions of similarity that may be a consequence of phylogenetic (functional, structural, or evolutionary) relationships between the sequences (Cole et al., 2005).

RDP also enables researchers to download their sequences in formats ready for input to a wide variety of third-party metagenomic tools (Cole et al., 2007). In this project, following sequence alignment and phylogenetic classification using the RDP classifier, the RDP download function was used to construct a distance matrix using a Jukes Cantor correction for multiple substitutions. The distance matrix will be formatted

similar to the output of DNADist from the Phylip package (Felsenstein, 2008), and should work in most programs that require DNAdist-formatted matrices (Cole et al., 2007). The RDP distance matrix is based on evolutionary distances between the sequences and is used as an input to the DOTUR and ∫-LIBSHUFF programs.

*DOTUR*

DOTUR is a freely distributed computer program that assigns large numbers of sequences to operational taxonomic units (OTUs) using either the nearest, average, or furthest neighbor clustering algorithms for all possible evolutionary distances. OTUs are sequence groupings determined by phylogenetic similarity. The furthest neighbor algorithm is the preferred method for 16S rDNA gene sequence analysis and consequently the most often used (Schloss & Handelsman, 2005). The furthest neighbor clustering algorithm generates OTUs so that all sequences within an OTU are at most X% distant from other sequences within the OTU (Schloss & Handelsman, 2005). Once sequences are assigned to OTUs, the program calculates several known diversity estimators and rarefaction data at various distance levels (Schloss & Handelsman, 2005).

This project used DOTUR version 1.53 to calculate rarefaction data, ACE and Chao1 richness estimators, and sample coverage data. DOTUR provides 23 output files that can be opened in spreadsheet format. Each file provides information to graph rarefaction curves, diversity estimator curves, or other classification data useful to researchers. This information can be used to compare the relative richness, the number of different OTUs in a community, and to determine if sampling effort was adequate.

*∫-LIBSHUFF*

∫-LIBSHUFF (an abbreviation of LIBrary SHUFFling) is a computer program that implements the integral form of the Cramer-von Mises test statistic (Anderson, 1962) to determine if two libraries are drawn from the same population and if one is a subset of the other. It builds upon work done by Singleton in the program LIBSHUFF (Singleton et al., 2001). An ∫-LIBSHUFF analysis compares two libraries to determine if they are significantly different from one another ($p \leq 0.05$). Significantly different libraries are assumed to have been derived from microbial communities of different composition (Schloss, 2008). Statistical methods help to determine whether differences in library composition are due to under sampling or to actual differences in the communities from which they were derived (Schloss et al., 2004).

The analysis begins by describing the two libraries in terms of coverage as described by Good (Good, 1953). The coverage (C) of a given sequence library describes the extent to which the sequences in the library represent the total population (Kemp & Aller, 2004). In order to calculate the coverage of a library, the criterion for what constitutes a unique sequence must first be decided. Rather than select a single arbitrary value as the criterion for uniqueness, the ∫-LIBSHUFF analysis calculates the coverage of a library for all values of evolutionary distance (D) ranging from 0.0 to 0.5 in increments of 0.01 (Schloss et al., 2004). An evolutionary distance of 0.0 represents identical sequences. An evolutionary distance of 0.50 is close to the maximal distance encountered in rRNA sequences within a prokaryotic domain (Singleton et al., 2001).

These values can then be used to plot a coverage curve (C vs. D) that describes how well the library represents the total community given varying criteria of uniqueness.

The equation for calculating the coverage for a single sample (X) is, $C_X = 1 - (N_X/n)$ where $N_X$ is the number of unique sequences in the sample and n is the total number of sequences. The value of $C_X$ will change based on the value of D selected (as the number of sequences in $N_X$ depends on the definition of "unique"). Small values of D tend to have correspondingly low coverage values in microbial communities (i.e., most sequences in a library appear unique when the criterion for uniqueness is based on very high sequence similarity). Higher values of D tend to produce correspondingly higher coverage values (i.e., when the criterion for uniqueness is very low sequence similarity, fewer sequences will be considered unique). Because each sequence in the library is compared to the other sequences within the same library, coverage values determined in this manner are referred to as "homologous coverage values", or "$C_X$", and the coverage curve generated from these data is referred to as a "homologous coverage curve", or "$C_X(D)$." By itself, the homologous coverage curve contains useful information about the library. For instance, if the library contains representatives of only a few of the bacterial genera in the original community, the coverage would be expected to be low at $D \leq 0.03$. Similarly, if most of the phyla present in the natural community are represented in the library, the coverage would be expected to be high at $D \leq 0.20$. In this fashion, the homologous coverage curve provides some insight into how well the microbial community was sampled (Singleton et al., 2001).

In order to compare two libraries, the ∫-LIBSHUFF analysis determines the coverage of one library (X) by a second library (Y) (Schloss et al., 2004). To accomplish this, each sequence in X is individually compared to all of the sequences in Y, and it is determined whether or not that sequence would be considered unique were it a part of Y

for a given value of D (Singleton et al., 2001). The resulting coverage values from this analysis are referred to as "heterologous coverage values", or "$C_{XY}$" and the resulting curve of $C_{XY}$ vs. D is called a "heterologous coverage curve", or "$C_{XY}(D)$". The equation for heterologous coverage is, $C_{XY} = 1 - (N_{XY}/n)$ where $N_{XY}$ is the number of sequences in the sample X that are not found in sample Y and n is the number of sequences in X (Singleton et al., 2001). Similar to the homologous coverage ($C_X$), $C_{XY}$ will vary based on the value of D selected because $N_{XY}$ will change based on the criterion for what determines a "unique" sequence (Singleton et al., 2001). The homologous and heterologous coverage curves can then be compared to determine the extent of difference between the two libraries, if any. Libraries derived from similar sources should have very similar homologous and heterologous coverage curves (Singleton et al., 2001).

The difference between the two curves may be compared using a statistical technique called the Cramer von Mises test statistic. The Cramer von Mises statistic is traditionally used to test the goodness of fit of a probability distribution (Pettitt, 1982). When applied to 16S rDNA gene sequence libraries the statistic measures the number of sequences that are unique to one library when two libraries are compared (Schloss et al., 2004; Singleton et al., 2001). The integral form of the statistic is more precise and accurate than the approximate form used in the original LIBSHUFF (Schloss et al., 2004). The integral formula for the Cramer von Mises statistic is the following:

$$\Delta \boldsymbol{C_{XY}} = \int_{\mathbf{0}}^{\infty} [\boldsymbol{C_X}(\boldsymbol{D}) - \boldsymbol{C_{XY}}(\boldsymbol{D})]^{\mathbf{2}} \boldsymbol{dD}$$

where, $C_X(D)$ and $C_{XY}(D)$ are measures of library coverage, and $D$ is the distance that is used to determine the level of coverage (Schloss et al., 2004). If the two libraries are identical, then $C_X(D)$ should be close to $C_{XY}(D)$ for all evolutionary distances D, yielding

a small difference, ΔC (Schloss et al., 2004; Singleton et al., 2001). Squaring the difference between $C_X(D)$ and $C_{XY}(D)$ makes ΔC sensitive to large differences between the homologous and heterologous curves (Singleton et al., 2001). By integrating over a range of evolutionary distances, the contributions of all differences between the homologous and heterologous curves are taken into account, yielding a more powerful test statistic than would have been obtained had only the largest difference between $C_X(D)$ and $C_{XY}(D)$ been considered (Singleton et al., 2001).

Once the difference between the two libraries or ΔC has been determined, it is necessary to determine whether or not the difference is statistically significant. Because the ΔC depends upon the community structure, the size of the library, as well as other complex factors, a Monte Carlo resampling approach is used to infer statistical significance (Singleton et al., 2001). To perform this resampling, ∫-LIBSHUFF shuffles the sequences of the two libraries together and randomly divides them into new libraries containing the same number of sequences as the originals (Singleton et al., 2001). The shuffled libraries are then analyzed identically to the originals and a ΔC value is calculated and recorded. The libraries are shuffled an additional 998 times, resulting in a total of 1000 ΔC values; one from the original libraries and 999 from randomly shuffled libraries. When all of the ΔC values are ordered from the highest to the lowest, the rank of the ΔC for the original libraries determines the probability of the two libraries being significantly different. When the ΔC value of the original libraries is greater than 95% of the ΔC values of the random shuffles, the libraries are considered significantly different with a p-value of 0.05 (Singleton et al., 2001). This procedure is motivated by the observation that the content of two libraries randomly sampled from the same population

of 16S rRNA genes will have approximately the same distribution (for large samples) as would be obtained by random shuffling (Singleton et al., 2001). When two libraries are dissimilar, the large majority of the random shuffles will have ΔC values less than the original libraries (Singleton et al., 2001).

*SONS*

A common goal in microbial ecology is to quantify the degree of overlap between the memberships and structures of two communities (Schloss & Handelsman, 2006a). For example, the fraction of OTUs that are shared between healthy and unproductive soils may indicate whether soil health is a function of community membership, structure, or both (Schloss & Handelsman, 2006a). If the memberships of two communities differ, then so will their structures (Schloss & Handelsman, 2006a). Also, if the richness of a community differs from that of another community, so will their memberships and structures (Schloss & Handelsman, 2006a). Yet if two communities have the same membership, then they will not necessarily have the same structure, and if the communities have the same richness, then they will not necessarily have the same membership (Schloss & Handelsman, 2006a).

SONS (an acronym for **S**hared **O**TUs a**n**d **S**imilarity) is a computer program that uses non-parametric estimators to estimate similarity between communities based on their membership and structure (Schloss & Handelsman, 2006a). SONS is essentially a carry-over from where ∫-LIBSHUFF left off. While ∫-LIBSHUFF reports the probability of statistical difference, or lack thereof, between two communities, it does not indicate at what phylogenetic levels those differences occur. Using output from DOTUR and an

indication of dissimilarity from ∫-LIBSHUFF, SONS utilizes non-parametric estimators, calculated across communities, to measure the fraction of OTUs shared by two communities as a function of genetic distance (Schloss & Handelsman, 2006a). SONS provides the capability to determine the abundance distribution of OTUs that are either endemic to or shared between communities using non-parametric estimators (Schloss & Handelsman, 2006a).

# Chapter III: Methodology

## Overview

The steps in this analysis are as follows: aviation fuel sample collection from various continental United States (CONUS) locations, DNA extraction from the fuel samples; Polymerase Chain Reaction (PCR) to amplify 16S rDNA sequences within the DNA extract, cloning of the amplified 16S rDNA products, then sequencing of the products, and finally, comparative analysis of the microbial sequences using various metagenomic applications.

The laboratory steps required to obtain the rDNA sequences used in this analysis were completed prior to the author taking part in the analysis effort. However, these steps have been included along with the sequence analysis steps in order to thoroughly explain the research methodology in its entirety. Similar procedures to the laboratory procedures provided here have been published in peer reviewed journals and should be referenced for further clarification (Denaro, 2005; Rauch et al., 2005; Vangsness et al., 2007).

## Sample Collection

Several military and civilian aircraft and storage tanks were sampled between 2005 and 2006. The JP-8 fuel samples were drawn from military aircraft and storage tanks at the following locations: Charleston AFB, South Carolina; Davis-Monthan AFB, Arizona; McGuire AFB, New Jersey; Mountain Home AFB, Idaho; Stewart AFB, New York; and Travis AFB, California. Jet A samples were collected from aircraft at commercial airbases in Victorville, California and Roswell, New Mexico. Note that

samples from commercial locations were taken from aircraft in long term storage ("moth-balled"). Biodiesel samples were taken from a single storage tank located at Dyess AFB, Texas. Locations of sample sites are depicted in Figure 5.

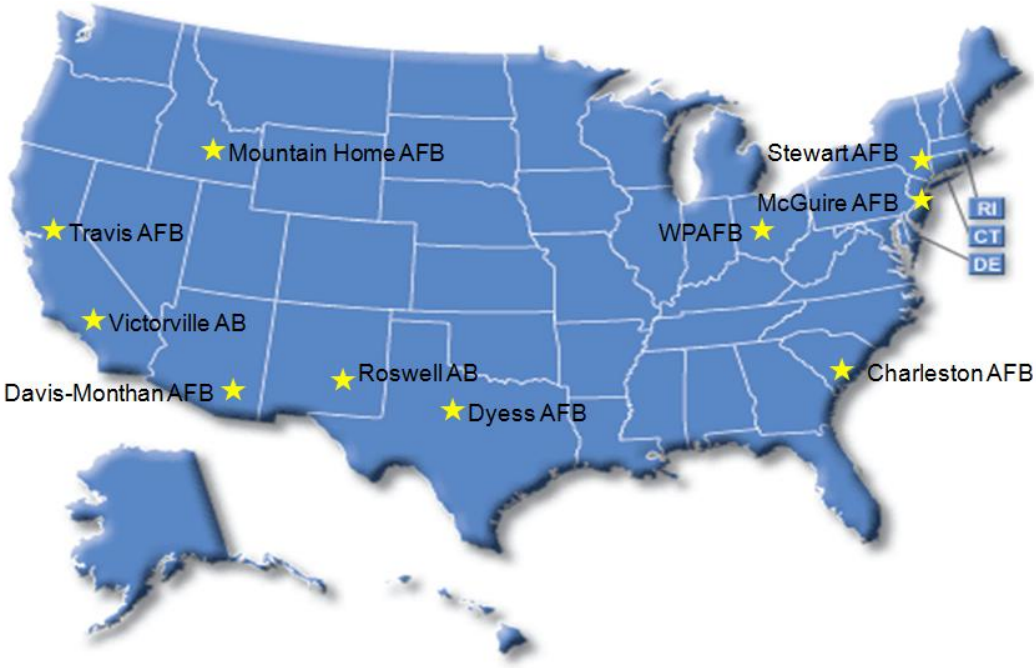



**Figure 5. Sample collection locations**

The sample collector drained fuel/water from the low point sumps in each wing and center body tank into HDPE 1L wide-mouth containers (Environmental Sampling Supply, Oakland, CA). Container preparation by the manufacturer included a non-phosphate detergent wash, multiple tap water and ASTM Type I de-ionized water rinses, 1:1 $HNO_3$ rinses, and oven drying. Two liters of fuel were collected from each sump and labeled with aircraft and tank identifiers. The sampling tools were sterilized with a 10% bleach solution and rinsed three times with sterile water between aircraft sampling.

The first liter of sample was shipped to UDRI by overnight air and was available for laboratory testing within 24 hours of sampling. The second liter of fuel/water was retained at the flight line for immediate analysis using a commercial adenosine triphosphate (ATP) test kit (Hy-LiTE®, MerckKGaA, 64271 Darnstadt, Germany).

**Microbial Extraction from Fuel Samples**

A mixed aliquot was selected for analysis from all samples. To prepare the mixed aliquot, samples were shaken by hand for a minimum of 30 seconds prior to sampling. 60 mL mixed fuel was collected in a sterile, disposable 60 mL syringe (Becton Dickinson and Company, Franklin Lakes, NJ). A sterile, hydrophobic 0.45 μm, 26 mm diameter, luer-lock tip filter (Corning, Corning, NY) was attached to the tip of the syringe and the fuel was filtered. The filter was removed from the syringe and placed in a laminar flow hood to dry. A new sterile 60 mL syringe was used to collect 60 mL sterile air. The filter was attached to the tip of the syringe and the air passed through the filter. This was repeated several times until the filter paper was dry. The filter was attached to the tip of a new syringe and 1.5 mL sterile water was collected through the filter into the syringe. The filter was removed and the water placed into a sterile 1.5 mL microcentrifuge tube. The filter was again attached to the tip of the same syringe and 0.7 mL sterile water was collected through the filter into the syringe. The filter was removed and the contents in the syringe placed into a new, sterile 1.5 mL microcentrifuge tube. At this point all samples were analyzed using direct PCR and rDNA sequencing.

**Direct PCR and DNA Sequencing**

A direct rDNA extraction method was used to eliminate the need for traditional microbial cultivation. 100 µL of sample was added to a 0.2 mL microtube and heated at 99°C for 10 minutes to liberate cellular DNA. Four microliters of lysed cell suspension were added to PCR reaction mixture containing forward primer, reverse primer, DNA polymerase and nucleotide solution in the amounts prescribed in the PCR protocol (Appendix C). Primer sequences and references are listed in Figure 4. A Primus thermocycler (MWG-Biotech, High Point, NC, USA) was used for PCR. The PCR profile consisted of initial denaturation at 94 °C for 2 min, 51 °C for 20 s, 72 °C for 30 s, followed by 30 cycles of 94 °C for 30 s, 51 °C for 20 s, and 72 °C for 30 s. PCR samples were analyzed by agarose gel electrophoresis to confirm amplification of the product. Bands were compared with 1 kB DNA ladder standard (Sigma-Aldrich Co.). Once amplification was verified by electrophoresis, the PCR amplimers were cloned into a plasmid vector using the TOPO TA Cloning kit (Invitrogen, Carlsbad, CA, USA) according to manufacturer's protocol (Appendix D). Viable, white colonies were subsequently picked and grown aerobically overnight at 37 °C in sterile Luria-Bertani (LB) broth, supplemented with 100 µg/ml ampicillin for plasmid selection. Colony PCR was performed as described in Appendix E to ensure the PCR insert had attached to the vector. Plasmid DNA purification was accomplished using the QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA, USA) as described in Appendix F. Purified DNA was digested with EcoRI restriction enzyme (Roche Biochemicals, Indianapolis, IN, USA) and the digested products were separated by agarose gel electrophoresis to confirm the presence of insert. Purification of plasmid DNA from 48 clones per plate and the DNA

sequencing reactions were performed by MWG Biotechnology Sequencing Laboratory (MWG-Biotech, High Point, NC). DNA sequencing was accomplished using M13 forward and reverse primers and output data was provided in FASTA format. At this point 3126 raw sequences were available for metagenomic analysis. This number was reduced following the trimming and sorting procedures described below.

**Sequence Trimming and Validation**

A thorough quality check procedure ensured only quality sequences were analyzed in this thesis effort. As a first step, all sequences less than 300 base pairs (bp) in length were automatically omitted because they did not provide a large enough region of the 16S rDNA gene to provide valid contribution to the project (Cole et al., 2007). During identification and deletion of sequences with less than 300 bp, sequences with numerous N's or repeated letters were also identified and removed. Repeated letters in sequences indicate possible contamination of the sample or a "stutter" in the DNA sequencer (Chai, 2008). N's sometimes appear in place of standard nucleotides letters (A, T, C, G), which indicate a point where any nucleotide could have been placed (Leon, 2008). Numerous N's indicate that the sample is not concentrated enough for the sequencer to produce a valid sequence (Leon, 2008). This step resulted in 1179 sequences being removed from further analysis. An example of this step of editing is summarized in Figure 6.

**>Sequence 1 785 bp**
**TTGNTAACGTACGGCCGAGTGAATTAGTAATACGACTCACTATAGGGCGAATTGGGCCT**
**ACTAGATGCATGCTCGAGCGGCCGCAGTGTGATGGATATCTGCAGAATTCGCCCTTGCT**
**GCTGGCACGTAGTTAGCCGGTGCTTATTCTGCGGGTACCGTCATTAGCGCCAGGTATTA**
**ACCGGCACCGTTTCGTTCCCGCCAAAAGTGCTTTACAACCCGAAGCCTTCTTCGCACAC**
**GGGCATTGCTGGATCAGGGTTGCCCCATTGTCCAAAATTCCCCACTGCTGCCTCCCGTA**
**GAGNTCTGGGCCGTGTCTCAGTCCCAGTGTGGGCTGGTCGTCCTCTCAAACCAGCTACG**
**GATCGAAGCCTTGGTGAGCCTTTACCTCACCAACTAGCTAATCCGATATCGGCCGCTCC**
**AATAGTGAGAGGTCTTGCGATCCCCCCCTTTCCCCCGTAGCGTTATCCGGTATTAGCTAC**
**GCTTTCGGTGTTTATCCCCCGCTACTGGGCACGTTCCGATACATTACTCACCCGTTCGCC**
**ACTCGCCACCAGGGTTGCCCCGTGCCTGCCGTTCGACTTGCATGTGTAAGGCATGCCGC**
**TAGCGTTCAATCTGAGCCAGGATCAAACTCTCCAAAGGCGAATCCAGCACACTGGCGGC**
**GTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGGCGTTAATCATGGGTCATAGCTGTT**
**TCCCTGTGTGAAATTGTTATCCGCTCACAATCCACACANATACGAGCCGGAGCATAAGT**
**GTAAGCCTGGGTGCAA**

**>Sequence 2 211 bp**
**TANTNTNNNNTNTTNGCGNGTNTTNGTNTNTNTNTNTNTTNTNNNNNNNNNNNTNCNTNN**
**ANNNNNNNNNNNNNNNNTNGNTNAAAAAAATANNAAAAAAAANAGGGGGGGGAGCCCCC**
**CCCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAAAAGNGGGGGGGGGGAAAAAAAA**
**AAAAAAAAAAATTTTTTTAAAAGGGGGGGTTT**

**>Sequence 3 665 bp**
**ATAAGTTGTTAAAGCAGGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTGGG**
**CCTACTAGATGCATGCTTCGAGCGGCCGCAGTGTGATGGATATCTGCAGAATTCGCCCT**
**TTGGAGAGTTTGATCCTGGCTCAGAGCGAACGTGGCGGCAGGCTAACACATGCAAGTCG**
**AACGAACTCTTCGGAGTTAGTGGCGGACGGGTGAGTAACACGTGGGAACGTNCCTTTAG**
**NTTCGGAATAACTCAGGGAAACTTGAGCTAATACCGGATGTGCCCTTCGGGAAAGATCT**
**ATCGCCTTTAGAGCGGCCCGCGTNCTGATTAGCTAGTTGGTGAGGTAAAGGCTCACCAA**
**GGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAAACACGGC**
**CCAAACTCCTACGGAGGCAGCAGTGGGGAATCTTGCGCAATGGGCGAAAGTGGACCGC**
**AGCCATGCCGCGTGAATGATGAAGGTCTTAGGATTGTAAAATTCTTTCACCGGGGACGA**
**TAATGACGGTACCCGGAGAAGAAGCCCCGGCTAAACTACGTGCCAGCAGCAAGGGCGA**
**ATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGGCGTT**
**AATCATGGTCATAGCTGG**

**FIGURE 6. Removal of low quality sequences**
**Sequence 1 and 3 represent quality sequences while Sequence 2 displays all the signs of a low quality sequence—short (less than 300 bp), N's, and repeated letters**

The next step of the sequence trimming process was to remove sequences that could introduce bias during sequence analysis. A decision was made by the researcher and sponsor to remove the M13 reverse primer sequences from analysis to prevent skewing the data towards sequences that were sequenced twice, with both the forward and reverse M13 primers. This decision was validated by the fact that preliminary classification using the RDP classifier resulted in almost identical classifications of both

sets of sequences. It should be noted that the forward primer group contained more sequences, meaning that if it were removed instead of the reverse primer group some diversity may have been lost. This step resulted in 714 sequences being removed from further analysis.

The final step in the sequence trimming process was to remove all irrelevant pieces of the rDNA sequences. Irrelevant pieces are those nucleotide chains preceding and following the restriction sites and primers that were designed to intentionally flank the variable and hyper-variable regions of the 16S gene, the region of interest in rDNA studies (Baker et al., 2003). Irrelevant pieces are a consequence of the sequencing reaction, whereby the DNA extension from the sequence primer could proceed past the PCR insert of interest and into the flanking EcoR1 restriction sequences and further plasmid sequences. The EcoR1 restriction sites provided a convenient means of locating these pieces for subsequent removal. The EcoR1 sites were identified using the program BioEdit (Hall, 1999). A screenshot from the BioEdit program depicting the EcoR1 restriction site (GAA TTC), the site at which all sequences were trimmed, is shown in Figure 7.

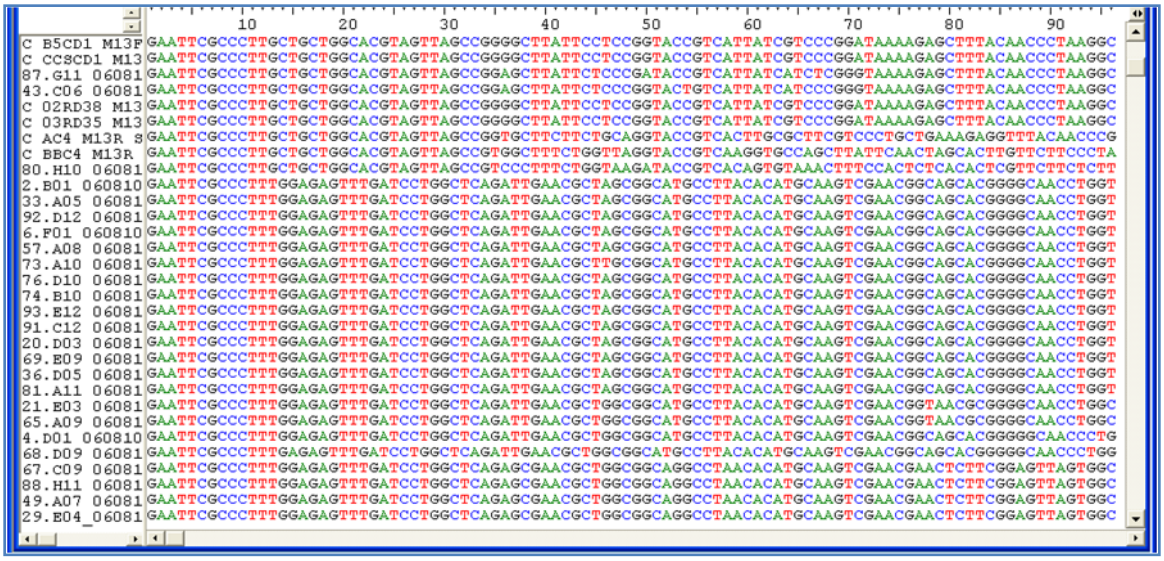



**Figure 7. Screenshot from BioEdit program**

## Nomenclature and Sorting

This analysis combined rDNA sequences from various locations, fuel types, airframes and sequencing labs. The samples were sequenced over the course of two years by various individuals. Consequently, a standardized nomenclature did not initially exist. A list of the nomenclature used to create the initial sequence identifiers during sampling, PCR and sequencing reactions is provided in Appendix A. Sequence identifiers allowed for each sequence to be uniquely identifiable.

In order to compare microbial communities from the different fuel types the data were sorted, using the nomenclature key and original sequence identifiers, into subsets based on location, fuel type, and airframe, and annotated accordingly; a process that took over two weeks to complete. At this point all sequence identifiers included the original identifier preceded by the additional sorting information. This step resulted in 33 sequences being removed from further analysis based on a lack of sufficient evidence for

placement. Figure 8 is a flow chart that describes the process the raw sequences underwent and the number of sequences removed in each phase.

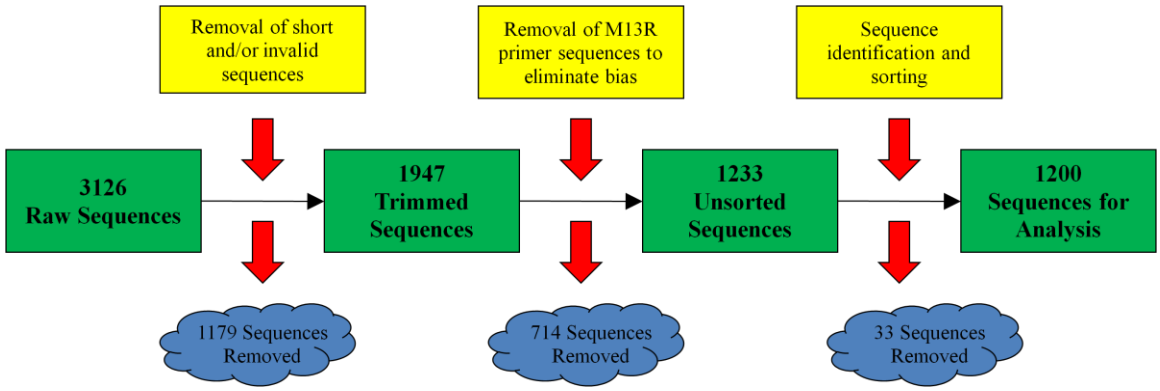



**Figure 8. Sequence validation flowchart**

The validation and sorting process outlined above was a crucial piece of this project. The sequences used for metagenomic analysis must have met all the criteria mentioned above and be correctly grouped into appropriate libraries for the library comparisons to be accurate and meaningful. The software packages used do not verify the input sequences provided to it; thus necessitating this extensive process. Ultimately this process resulted in 1200 sequences ready for further analysis. The 1200 validated sequences were divided into subsets of 828, 311, and 61 sequences from JP-8, Jet A, and biodiesel fuel samples, respectively.

## Analysis

The 1200 sequences remaining after trimming and sorting were uploaded to the RDP Release 10.7 using the MyRDP workspace. Sequences were uploaded in four groups: all sequences, JP-8, Jet A, and Biodiesel. Following automatic alignment to the

RDP taxonomy the sequences were prepped for phylogenetic analysis. Figure 9 is a flow chart that presents the order of the various analyses performed. These steps will be further explained within the subsequent paragraphs.

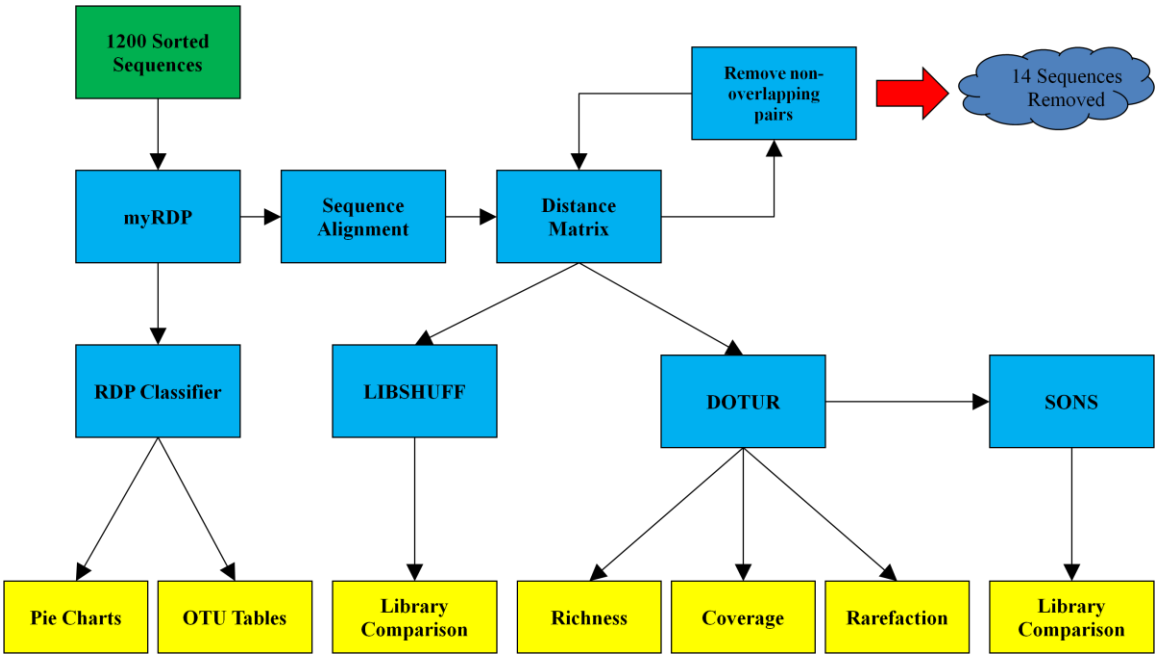



**Figure 9. Metagenomic analysis flowchart**

Sequences were analyzed by the RDP Classifier to determine the closest match to known 16S rDNA sequences within the RDP Hierarchy. Each rDNA sequence was assigned to the lowest level of taxonomy possible at an 80% confidence threshold. Assignments were shown on an interactive display where each node in the hierarchy listed the number of sequences assigned to that taxonomic rank (Figure 10). An 80% confidence estimate was generated for each assignment, and the assignments were displayed only when the estimate was above the specified confidence threshold.

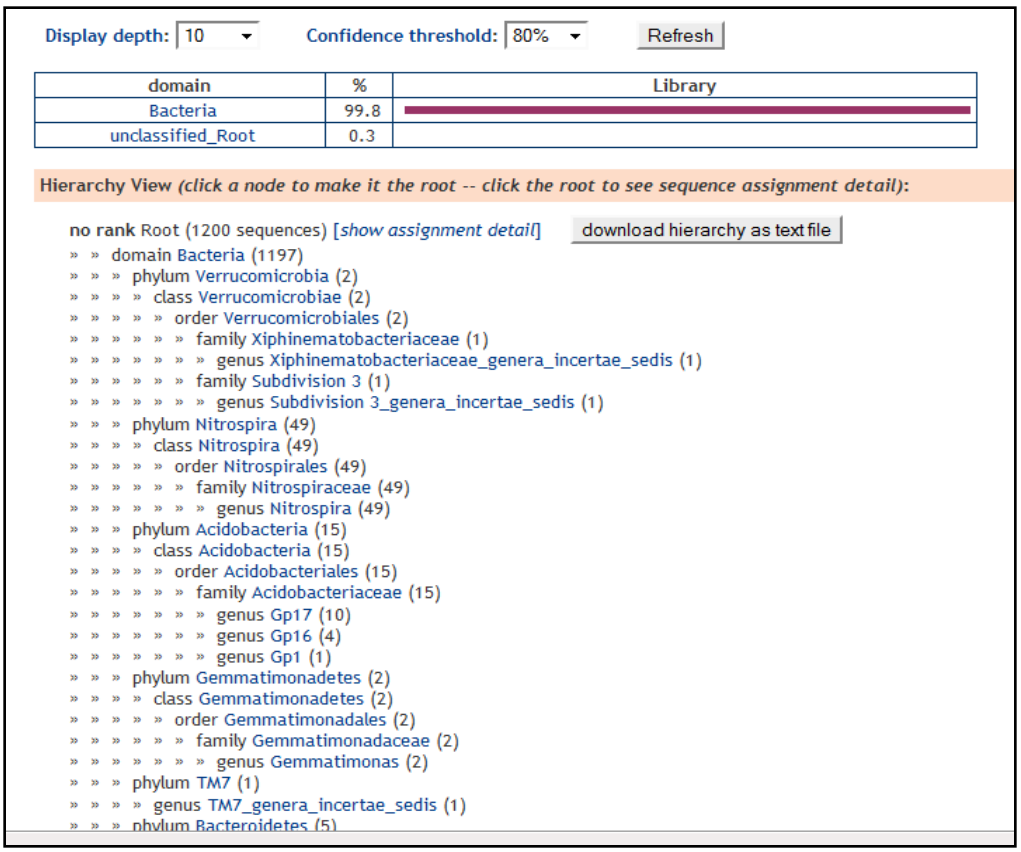



**Figure 10. RDP Classifier screenshot**

The assignment details shown above were downloaded as a text file and copied into an Excel spreadsheet for further examination. This was accomplished for all four sequence libraries. Pie graphs were constructed at the phylum level for each library to reveal the microbial composition of the communities. A summary table was also created to show the exact number of sequences present in each phylum. Taxonomic placement information was also used to create tables depicting the classification of the lower ranks (i.e. genera) for each library. The pie charts and tables are presented in Chapter IV under the phylogenetic classification section. The presence/absence information created in this step allowed for a qualitative assessment of the rDNA libraries at the various phylogenetic levels of interest.

The next step involved calculating the various diversity estimators using the DOTUR program. DOTUR requires a distance matrix for execution. Distance matrices for each of the four sequence libraries were downloaded using the RDP download function with Jukes-Cantor correction for multiple substitutions and specifying the use of ten character RDP sequence identifiers. RDP identifiers were used instead of the original identifiers due to DOTUR and ∫-LIBSHUFF program requirements. Once the distance matrices were created, the files were saved as distance files in the same folder as the DOTUR program executable file. Initial attempts to execute the DOTUR application resulted in errors. This was due to question marks in the distance matrices. Question marks in a distance matrix represent non-overlapping sequences (Chai, 2008). Non-overlapping sequences had to be omitted in order for the application to run properly. Benli Chai, an RDP co-founder and support staff member, developed a script for the Python Programming Language that reads in a distance matrix and removes all non-overlapping sequences, then outputs a new distance matrix without the non-overlapping sequences as well as a text file containing the sequence identifiers of the removed sequences (Chai, 2008). Running this script on each of the distance matrices resulted in 14 sequences being removed from further analysis. All 14 sequences came from the JP-8 sequence library, thus they were also removed from the "all sequences" library. A brief analysis of the removed sequences did not reveal why the sequences failed to overlap.

The "clean" distance matrices were used to run the DOTUR program. DOTUR, as well as ∫-LIBSHUFF and SONS, should be run from the command prompt rather than simply double-clicking the executable file. Instructions for DOTUR execution using the command prompt are provided in the DOTUR manual (Schloss, 2005). Successful

execution resulted in 20 files of output data for each library. These files were used to create graphs for the ACE, Chao1, rarefaction and coverage estimators as well as input to the SONS program which will be discussed later. DOTUR constructs ".c" files to plot accumulation curves (Schloss, 2005). These files are organized so that the first column is the number of sequences sampled. The next three columns represent the 99% evolutionary distance and provide the mean parameter as well as the parameter's upper and lower 95% confidence bounds (Schloss, 2005). The subsequent columns represent the further distance levels spaced at 1% increments. The 3% and 20% evolutionary distance level columns, as well as the sequence number column were copied into a new spreadsheet for further analysis.

Having isolated the appropriate information, accumulation curves were created for the ACE and Chao1 estimators at the genus and phylum levels. These graphs were used for comparison of relative diversity in order to address the research objectives of this thesis effort. Rarefaction data was analyzed graphically by plotting the proportion of observed richness as a function of the proportion of sequences sampled. This allowed for a standardized analysis of sampling effort and whether or not the communities were sampled at the appropriate level for a comprehensive analysis of the microbial diversity of the communities. The presence or absence of an asymptotic curve provides insight into this matter. Coverage was then determined using the formula for coverage presented in the literature review. The coverage was calculated for each fuel type and displayed in a bar chart at the genus and phylum levels. The charts constructed from DOTUR output are displayed in Chapter IV under the diversity analysis section.

The next step was used to statistically determine if two libraries came from the same population, different populations, or if one was a subset of the other. This step was carried out using the ∫-LIBSHUFF program and a distance matrix from RDP. A single distance matrix was used as an input to ∫-LIBSHUFF. The distance matrix was created using the RDP download function with a Jukes-Cantor correction for multiple substitutions and RDP sequence identifiers. The difference between this distance matrix and the distance matrices used as input to DOTUR is that ∫-LIBSHUFF is capable of comparing multiple libraries in a single execution. Therefore a single distance matrix was produced by selecting all three fuel type libraries from the MyRDP overview page and creating a distance matrix with the download function.

∫-LIBSHUFF was run from the command prompt line. Once the program was executed it required an input of the number of libraries in the distance matrix input file as well as the number of sequences in each library. The program automatically made pairwise comparisons between each of the three libraries resulting in 6 comparisons. Following execution of the program the associated p-values for the pairwise comparisons were printed in the command prompt. These values were recorded prior to closing the application. ∫-LIBSHUFF output a single ".coverage" file containing the coverage curve data required to construct graphs depicting the homologous and heterologous coverage curves. The graphs are presented in Chapter IV under the community membership and structure comparison section.

The last step was to determine the degree of overlap between the memberships and structures of the aviation fuel microbial communities. SONS was used to accomplish this objective. SONS was run from the command line prompt and required two input

files. The first input was a DOTUR-formatted ".list" file which contained the identity of the sequences in each OTU as a function of distance. The first column contained the distance used to define an OTU, the second was the number of OTUs at the respective distance, and the remaining columns included the identities of sequences in each OTU. This file was an output from the "all sequences" library execution during the DOTUR step of the analysis.

The second input was a ".names" file. This file was a tab-delineated Excel file containing the names of each sequence in the first column and the library designation in the second column. This file was created manually by selecting the "all sequences" library from the MyRDP overview page and downloading an ".ids" file using the SeqCart function of MyRDP. This file contained the RDP identifiers in the first column and the original sequence identifiers in the second. In order to synchronize with the ".list" file, which contains RDP identifiers, the first column was left intact. The second column was changed to designate the library from which the sequence came (JP-8, Jet A, Biodiesel). This was relatively simple because the sequences were already grouped by library.

In total, four ".names" files were created; the first as described above. The other three files were created similarly, however, using only two library designators; for example JP-8 and then "others." This was accomplished for all three fuel types in order to determine the region that overlaps between JP-8 and Jet A/Biodiesel, JP-8 and Biodiesel/Jet A, and Jet A and Biodiesel/JP-8.

After correctly formatting all SONS input files four executions of SONS were completed. Each execution determined the number of individuals in each community for each OTU as well as the fraction of shared OTUs between the communities and

accompanying shared richness estimators (Schloss & Handelsman, 2006a). Data from the various SONS outputs were used to create a three group (the three fuel types) Venn diagram which was used to easily visualize the community richness and membership overlap. This step was not trivial. Directions for creating the diagram are included in Appendix B.

# Chapter IV: Results and Discussion

## Overview

This chapter presents the results produced by the metagenomic analysis of rDNA sequences from samples of microbial contamination in the various aviation fuels. The emphasis of the results will be placed on information relevant to the research objectives presented in Chapter I which were to characterize the bacterial populations in the various aviation fuels by exploring community membership, and to investigate the effects of fuel type on microbial diversity and community structure. The following is a thorough assessment of the microbial communities present in the aviation fuels sampled for this thesis effort.

## Phylogenetic Classification of 16S rDNA Gene Libraries

Based on classification by the RDP Classifier, sequences similar to members of the Acidobacteria, Actinobacteria, Bacteroidetes, Chloroflexi, Cyanobacteria, Deinococcus-Thermus, Firmicutes, Gemmatimonadetes, Nitrospira, Plantomycetes, Proteobacteria, TM7, and Verrucomicrobia phyla were represented in the 16S rDNA sequence libraries from JP-8, Jet A, and biodiesel fuel samples (Table 5).

**Table 5. Phylum distribution of aviation fuel sequences**

| Phylum | JP-8 (n = 828) | Jet A (n = 311) | Biodiesel (n = 61) | Total (n = 1200) |
|---|---|---|---|---|
| Acidobacteria | 15 | 0 | 0 | 15 |
| Actinobacteria | 85 | 63 | 4 | 152 |
| Bacteroidetes | 5 | 0 | 0 | 5 |
| Chloroflexi | 7 | 0 | 0 | 7 |
| Cyanobacteria | 56 | 0 | 0 | 56 |
| Deinococcus-Thermus | 2 | 0 | 0 | 2 |
| Firmicutes | 83 | 99 | 2 | 184 |
| Gemmatimonadetes | 2 | 0 | 0 | 2 |
| Nitrospira | 49 | 0 | 0 | 49 |
| Plantomycetes | 2 | 0 | 0 | 2 |
| Proteobacteria | 459 | 149 | 55 | 663 |
| TM7 | 1 | 0 | 0 | 1 |
| Verrucomicrobia | 2 | 0 | 0 | 2 |
| Unclassified Bacteria | 57 | 0 | 0 | 57 |
| Unclassified Root | 3 | 0 | 0 | 3 |

Three sequences fell into an Unclassified Root category. Unclassified Root refers to sequences for which the RDP Classifier could not identify as bacterial 16S genes. They could have been non 16S genes, or 16S genes from non bacteria, or sequences of low quality (Cole et al., 2007). Further analysis of the three unclassified root sequences revealed that one may have come from the kingdom Archaea. Fifty-seven sequences fell into the Unclassified Bacteria category. Unclassified Bacteria referred to any sequence that was identified as Bacteria but did match a particular phylum with a confidence level of 80% or better. Figures 11 through 14 graphically depict the phylum distributions of each of the four libraries analyzed.

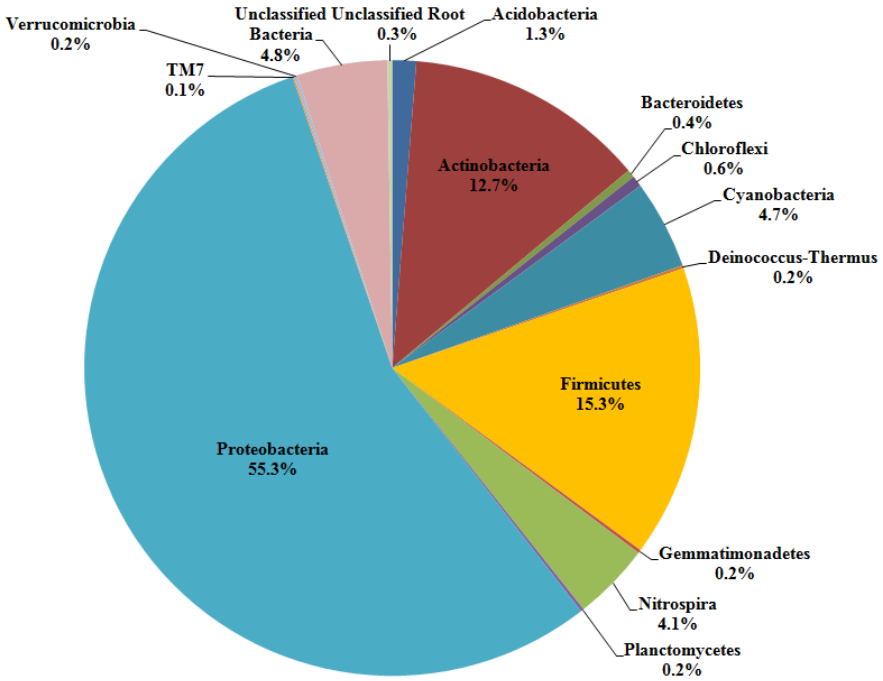



**Figure 11. Phylum distribution of all fuel sequences**

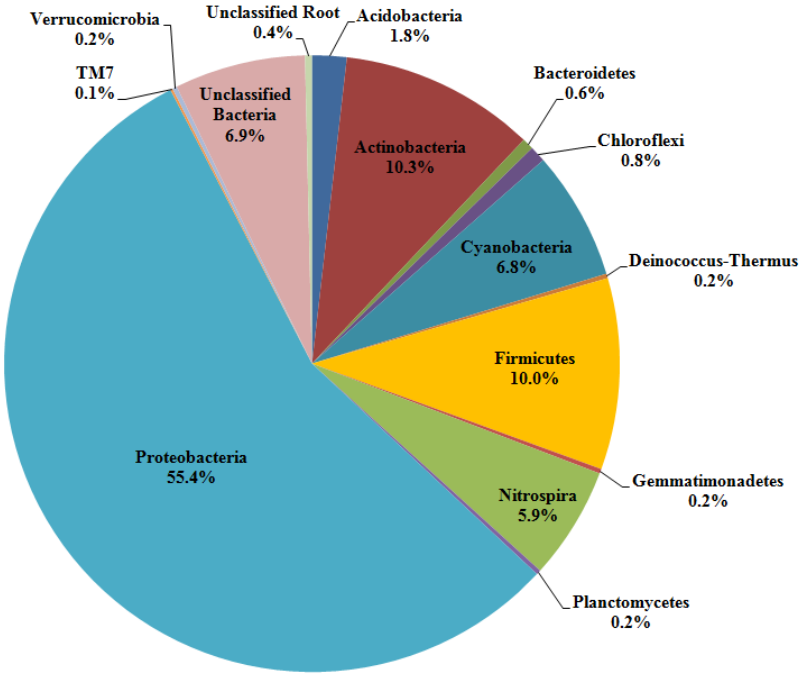



**Figure 12. Phylum distribution of JP-8 fuel sequences**

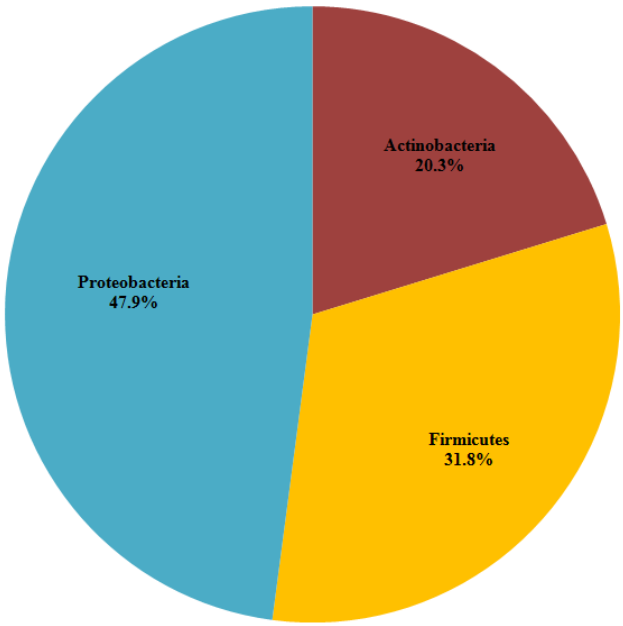



**Figure 13. Phylum distribution of Jet A fuel sequences**

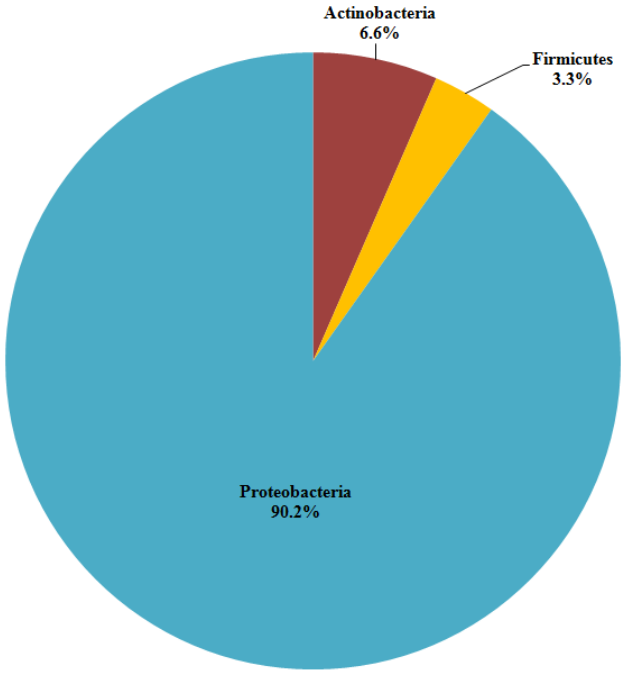



**Figure 14. Phylum distribution of biodiesel fuel sequences**

Proteobacteria dominated all three microbial communities with 55.8%, 47.9%, and 90.2% in the JP-8, Jet A, and biodiesel libraries, respectively. Members of the Proteobacteria, Firmicutes and Actinobacteria were represented in all three fuel types; in Jet A and Biodiesel they were the only phyla represented. Reasons for the lack of a phylum rich community in Jet A and biodiesel can be hypothesized. The Jet A samples were drawn from "moth-balled" aircraft. Rather than allowing time for the Jet A microbial community to thrive and diversify, species dominance may have set in and limited the number of identifiable species. In regard to biodiesel, the novelty of alternative fuels may promote improved fuel system maintenance which would result in less microbial growth. The identified microorganisms are summarized in Table 6.

**Table 6. Phylogenetic classification of aviation fuel sequences**

| Phylogenetic Classification | JP-8 (n = 828) | Jet A (n = 311) | Biodiesel (n = 61) | Total (n = 1200) |
|---|---|---|---|---|
| Acidobacteria | | | | |
| *Gp1* | 1 | 0 | 0 | 1 |
| *Gp16* | 4 | 0 | 0 | 4 |
| *Gp17* | 10 | 0 | 0 | 10 |
| Actinobacteria | | | | |
| *Actinomyces* | 0 | 1 | 0 | 1 |
| *Agromyces* | 1 | 0 | 1 | 2 |
| *Arthrobacter* | 2 | 12 | 0 | 14 |
| *Corynebacterium* | 2 | 0 | 0 | 2 |
| *Curtobacterium* | 3 | 0 | 0 | 3 |
| *Kytococcus* | 1 | 0 | 0 | 1 |
| *Microbacterium* | 6 | 15 | 0 | 21 |
| *Mycobacterium* | 0 | 7 | 0 | 7 |
| *Propionibacterium* | 17 | 6 | 1 | 24 |
| *Quadrisphaera* | 1 | 0 | 0 | 1 |
| *Rhodococcus* | 40 | 21 | 1 | 62 |
| *Rothia* | 1 | 1 | 0 | 2 |
| Unclassified Actinomycetales | 6 | 0 | 0 | 6 |
| Unclassified Corynebacterineae | 1 | 0 | 1 | 2 |
| Unclassified Microbacteriaceae | 2 | 0 | 0 | 2 |
| Unclassified Nocardiaceae | 1 | 0 | 0 | 1 |
| Unclassified Rubrobacterineae | 1 | 0 | 0 | 1 |

**Continued on next page**

**Table 6 – Continued**

| | | | | |
|---|---|---|---|---|
| Bacteroidetes | | | | |
| *Cloacibacterium* | 1 | 0 | 0 | 1 |
| *Hymenobacter* | 2 | 0 | 0 | 2 |
| Unclassified Sphingobacteriales | 2 | 0 | 0 | 2 |
| Chloroflexi | | | | |
| *Caldilinea* | 1 | 0 | 0 | 1 |
| Unclassified Anaerolineae | 5 | 0 | 0 | 5 |
| Unclassified Chloroflexi | 1 | 0 | 0 | 1 |
| Cyanobacteria | | | | |
| *Streptophyta* | 46 | 0 | 0 | 46 |
| Unclassified Cyanobacteria | 10 | 0 | 0 | 10 |
| Deinococcus-Thermus | | | | |
| *Deinococcus* | 1 | 0 | 0 | 1 |
| *Truepera* | 1 | 0 | 0 | 1 |
| Firmicutes | | | | |
| *Anaerotruncus* | 3 | 2 | 0 | 5 |
| *Bacillus a* | 0 | 11 | 0 | 11 |
| *Bacillus d* | 31 | 19 | 0 | 50 |
| *Bacillus f* | 0 | 1 | 0 | 1 |
| *Bacillus h* | 8 | 0 | 0 | 8 |
| *Clostridium* | 0 | 3 | 0 | 3 |
| *Staphylococcus* | 2 | 43 | 0 | 45 |
| *Streptococcus* | 1 | 3 | 1 | 5 |
| Unclassified Bacillaceae 2 | 2 | 0 | 0 | 2 |
| Unclassified Bacillales | 1 | 0 | 0 | 1 |
| Unclassified Bacilli | 0 | 0 | 1 | 1 |
| Unclassified Bacillus | 0 | 12 | 0 | 12 |
| Unclassified Clostridiales | 1 | 0 | 0 | 1 |
| Unclassified Ruminococcaceae | 34 | 5 | 0 | 39 |
| Gemmatimonadetes | | | | |
| *Gemmatimonas* | 2 | 0 | 0 | 2 |
| Nitrospira | | | | |
| *Nitrospira* | 49 | 0 | 0 | 49 |
| Planctomycetes | | | | |
| *Pirellula* | 2 | 0 | 0 | 2 |
| Proteobacteria | | | | |
| Alphaproteobacteria | | | | |
| *Bosea* | 11 | 0 | 0 | 11 |
| *Bradyrhizobium* | 2 | 1 | 0 | 3 |
| *Brevundimonas* | 23 | 0 | 6 | 29 |
| *Caulobacter* | 0 | 0 | 1 | 1 |
| *Hyphomicrobium* | 0 | 1 | 0 | 1 |
| *Methylobacterium* | 46 | 87 | 0 | 133 |
| *Phenylobacterium* | 0 | 0 | 1 | 1 |
| *Rhodocista* | 1 | 0 | 0 | 1 |
| *Sphingobium* | 3 | 0 | 1 | 4 |
| *Sphingopyxis* | 7 | 0 | 0 | 7 |
| Unclassified Alphaproteobacteria | 3 | 5 | 0 | 8 |
| Unclassified Bradyrhizobiaceae | 1 | 0 | 0 | 1 |
| Unclassified Caulobacteraceae | 3 | 0 | 1 | 4 |
| Unclassified Methylobacteriaceae | 1 | 0 | 0 | 1 |

**Continued on next page**

**Table 6 – Continued**

| | | | | |
|---|---|---|---|---|
| Unclassified Phyllobacteriaceae | 1 | 1 | 0 | 2 |
| Unclassified Rhizobiaceae | 0 | 1 | 0 | 1 |
| Unclassified Rhizobiales | 6 | 3 | 0 | 9 |
| Unclassified Rhodospirillaceae | 1 | 0 | 0 | 1 |
| Unclassified Sphingomonadaceae | 3 | 1 | 1 | 5 |
| Betaproteobacteria | | | | |
| *Acidovorax* | 0 | 1 | 0 | 1 |
| *Alcaligenes* | 1 | 0 | 0 | 1 |
| *Aquabacterium* | 2 | 0 | 0 | 2 |
| *Burkholderia* | 24 | 15 | 2 | 41 |
| *Comamonas* | 6 | 0 | 1 | 7 |
| *Cupriavidus* | 1 | 0 | 0 | 1 |
| *Delftia* | 29 | 0 | 1 | 30 |
| *Herbaspirillum* | 3 | 0 | 0 | 3 |
| *Janthinobacterium* | 11 | 0 | 1 | 12 |
| *Pandoraea* | 0 | 5 | 0 | 5 |
| *Pelomonas* | 1 | 0 | 0 | 1 |
| *Ralstonia* | 0 | 0 | 1 | 1 |
| Unclassified Alcaligenaceae | 70 | 0 | 29 | 99 |
| Unclassified Burkholderiaceae | 11 | 0 | 0 | 11 |
| Unclassified Burkholderiales | 1 | 0 | 0 | 1 |
| Unclassified Comamonadaceae | 1 | 12 | 2 | 15 |
| Unclassified Incertae sedis 5 | 17 | 0 | 3 | 20 |
| Unclassified Oxalobacteraceae | 1 | 0 | 0 | 1 |
| Unclassified Rhodocyclaceae | 5 | 1 | 0 | 6 |
| *Variovorax* | 0 | 1 | 0 | 1 |
| Gammaproteobacteria | | | | |
| *Acinetobacter* | 6 | 0 | 0 | 6 |
| *Alkanindiges* | 2 | 0 | 0 | 2 |
| *Citrobacter* | 1 | 0 | 0 | 1 |
| *Dyella* | 1 | 0 | 0 | 1 |
| *Flavimonas* | 2 | 0 | 0 | 2 |
| *Pseudomonas* | 91 | 10 | 1 | 102 |
| *Shigella* | 0 | 1 | 0 | 1 |
| *Stenotrophomonas* | 6 | 0 | 1 | 7 |
| Unclassified Enterobacteriaceae | 9 | 0 | 0 | 9 |
| Unclassified Gammaproteobacteria | 5 | 0 | 2 | 7 |
| Unclassified Pseudomonadaceae | 5 | 0 | 0 | 5 |
| *Yersinia* | 24 | 2 | 0 | 26 |
| Deltaproteobacteria | | | | |
| Unclassified Deltaproteobacteria | 2 | 0 | 0 | 2 |
| Epsilonproteobacteria | | | | |
| Unclassified Helicobacteraceae | 1 | 0 | 0 | 1 |
| *Wolinella* | 0 | 1 | 0 | 1 |
| Unclassified Proteobacteria | 8 | 0 | 0 | 8 |
| TM7 | | | | |
| TM7 genera *Incertae sedis* | 1 | 0 | 0 | 1 |
| Verrucomicrobia | | | | |
| Subdivision 3 genera *Incertae sedis* | 1 | 0 | 0 | 1 |
| Xiphinematobacteriaceae genera *Incertae sedis* | 1 | 0 | 0 | 1 |
| Unclassified Bacteria | 57 | 0 | 0 | 57 |
| Unclassified Root | 3 | 0 | 0 | 3 |

Phylogenetic analysis identified a total of 68 microbial genera with a confidence level of 80% or better, including 42 genera (61.8%) that were found in jet fuel for the first time according to the available literature (Tables 3 and 6). Those sequences that were classified to the genus level were based on similar sequences existing in the RDP hierarchy database. Additionally, 36 unclassified categories, encompassing 214 sequences (17.8%), were returned from the RDP Classifier. This figure does not include Unclassified Bacteria or Unclassified Root sequences which may be unclassified for other reasons. Unclassified categories are common for regions of less-well-studied bacterial diversity, which is the case with many environmental clone libraries, to include aviation fuel microbial communities (Cole et al., 2007). As stated previously, unclassified categories are a result of the RDP Classifier's inability to place a sequence in the hierarchy at the established confidence level. Such low confidence classification results may identify sequences where a more thorough phylogenetic analysis is warranted (Cole et al., 2007). Sequences similar to the genera *Propionibacterium*, *Rhodococcus*, *Streptococcus*, *Burkholderia*, *Pseudomonas* as well as Unclassified Sphingomonadaceae, and Unclassified Comamonadaceae were identified in all three sequence libraries.

**Diversity Analysis**

Diversity estimation and analysis is often limited by sampling effort. Therefore several approaches were used to calculate and compare sampling effort across rDNA sequence libraries prior to extrapolating results from the diversity estimators. Two methods used in this analysis are rarefaction and coverage as discussed in the literature

review section. The rarefied accumulation curves of the sequence libraries are depicted at the phylum and genus levels in Figure 15.

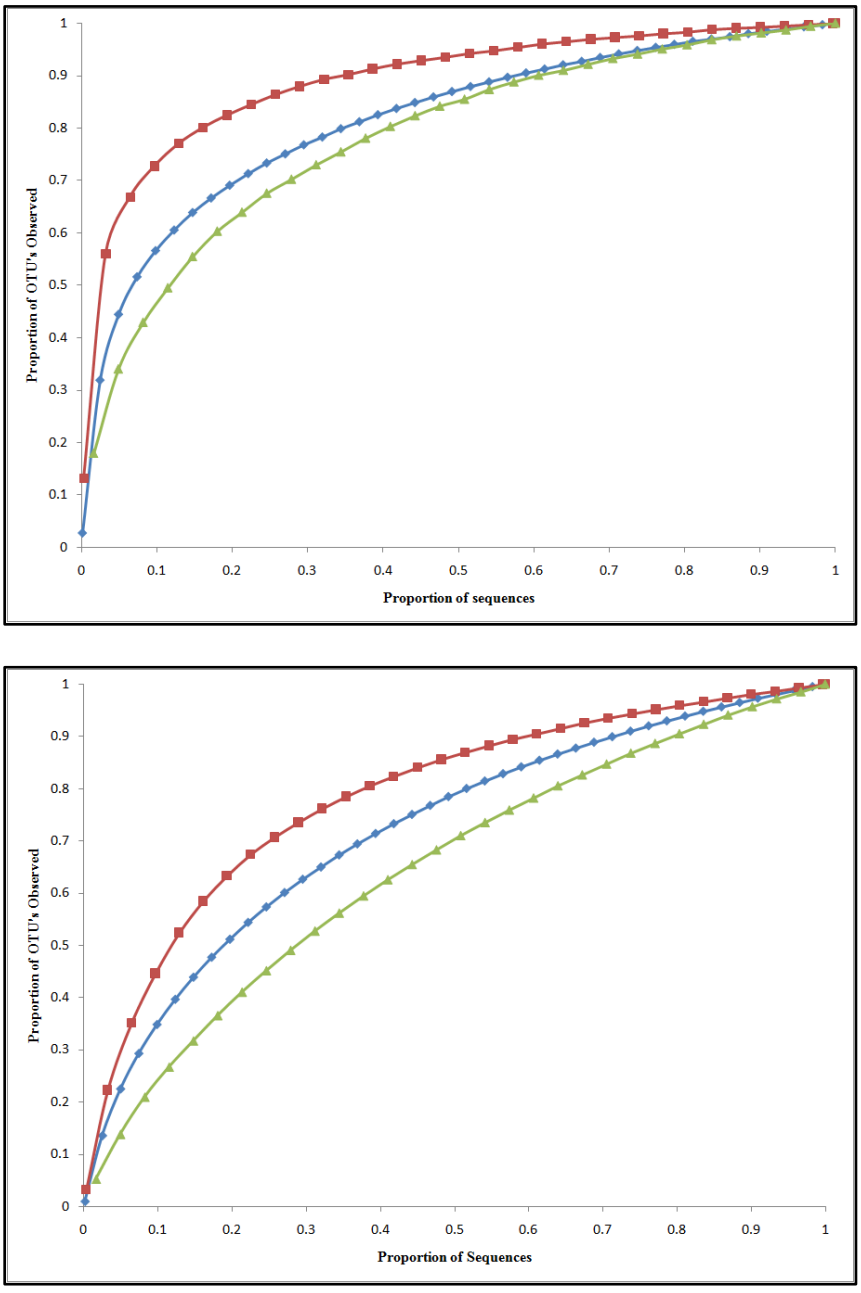



**Figure 15. Rarefied accumulation curves for fuel sequences libraries**
**Phylum level (Top) and genus level (Bottom) rarefaction curves for JP-8 (diamond), Jet A (square), and biodiesel (triangle)**

As seen in Figure 15, the three fuel libraries were each sampled more adequately at the phylum level. This is evident based on the greater amount of concave-downwardness in the resulting rarefaction curves. Relative to one another the Jet A library appears to have been sampled most adequately; JP-8 was more adequately sampled than biodiesel. Although sample size was much larger for JP-8 (828 sequences) than it was for Jet A and biodiesel (311 and 61, respectively) it was not the most adequately sampled. This alludes to the difference in relative diversities of the microbial communities. Given the larger sample size of JP-8, its richness must be much greater than either of the other two fuels in order to require more than a doubling of sampling effort to attain an equal relative sampling effectiveness. However, none of the curves reached a clear asymptote, indicating that the actual diversity of the libraries was only partially covered, especially at the genus level, and further sampling is likely to reveal additional taxa.

Coverage was also used to assess the completeness, or sampling effort, of the sequence libraries. Figure 16 summarizes the coverage values for all four sequence libraries.

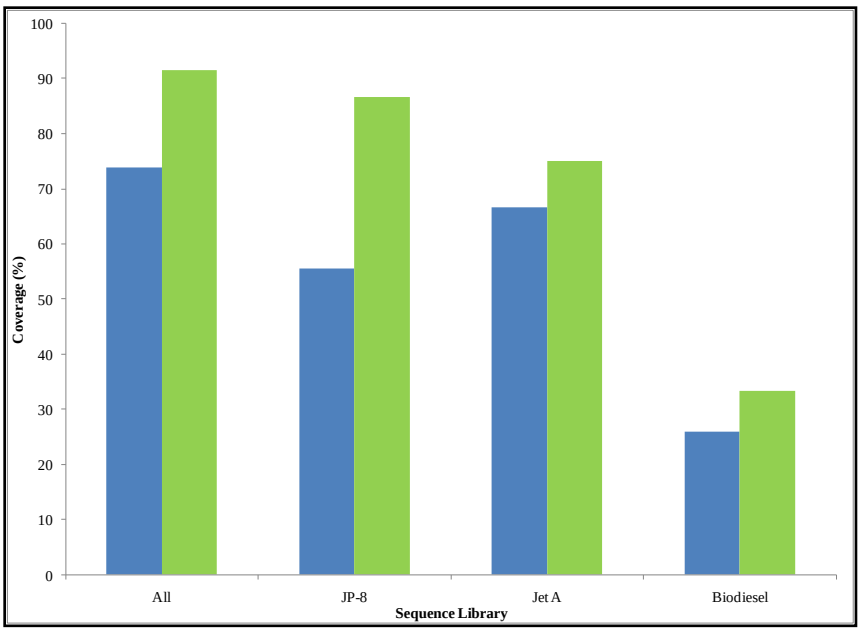



**Figure 16. Good's coverage**
**Dark bars represent the genus level. Light bars represent the phylum level.**

As noted in the rarefaction analysis, the phylum level coverage values were all higher than the genus level coverage values. Additionally, biodiesel was sampled less completely than the other two fuel types. Jet A was sampled more adequately than JP-8 and biodiesel at the genus level. However, Jet A was not shown to be more adequately sampled at the phylum level as was shown in the rarefaction analysis, suggesting a possible weakness in the two measures' ability to gauge sampling effort. The coverage values indicated that the microbial populations in aviation fuels are extremely diverse and a much larger sample size should be taken in order to obtain a representative sample, or complete coverage of the community, especially at the lower taxonomic ranks.

The next step was to examine diversity based on the ACE and Chao1 richness estimators using output from the DOTUR program. ACE and Chao1 richness estimates are displayed at the phylum and genus levels in Figure 17.

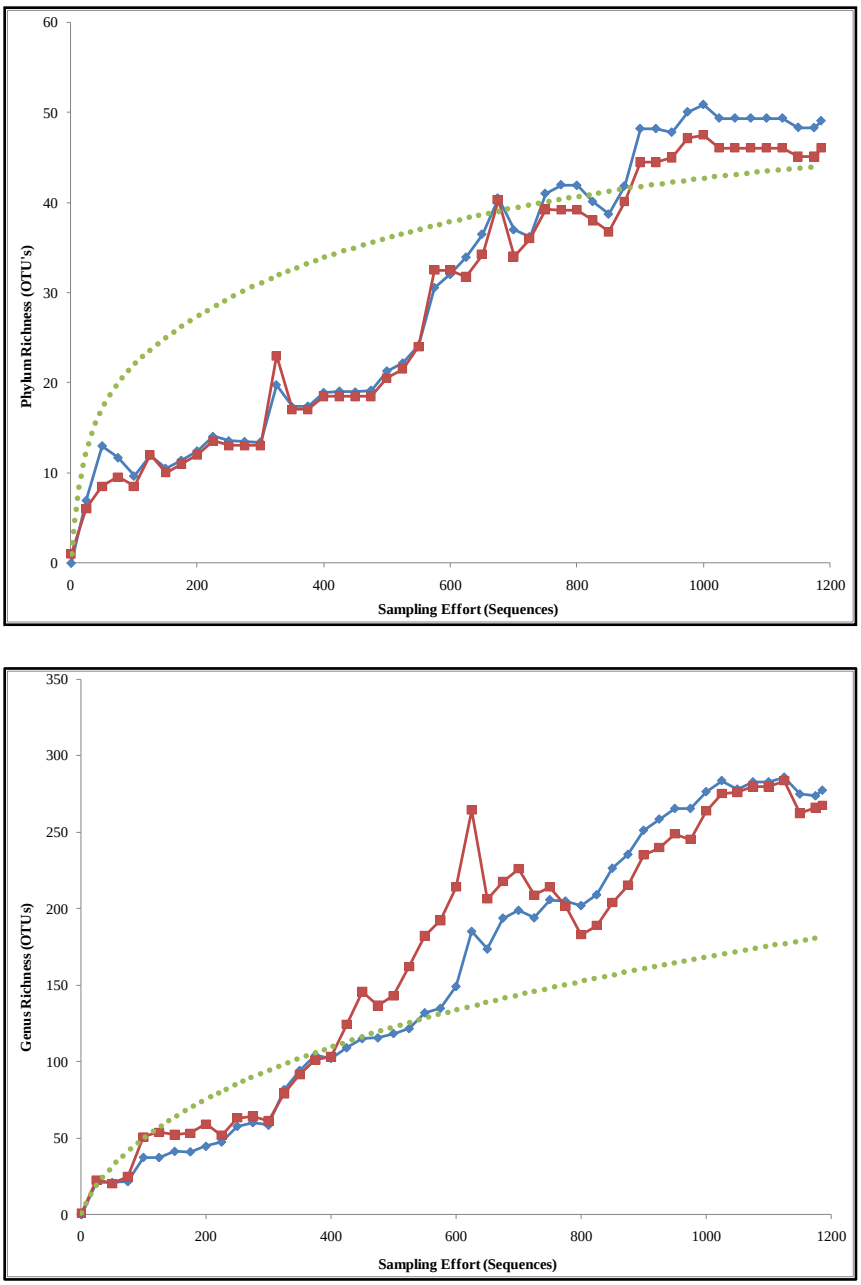



**Figure 17. Richness estimators for all fuel sequences**
**ACE (diamond) and Chao1 (square) richness estimators at the phylum level (Top) and genus level (Bottom). Rarefaction (dotted line) values based on observed OTUs.**

The estimators were more or less equivalent due to the large number of sequences analyzed. Total number of OTUs estimated at the phylum level is 46 and 49 for the ACE

and Chao1 estimators, respectively. Total number of OTUs at the genus level is 277 and 267 for the ACE and Chao1 estimators, respectively.

The estimators of total microbial diversity in aviation fuels are much greater than the actual number of taxa presented in the phylogenetic tables shown in the previous section. The observed richness, as depicted by the rarefaction curves, was always below either richness estimator. This is because rarefaction illustrates the observed richness of the samples while the ACE and Chao1 estimators estimate the richness in the community from the sequences available based on the equations stated in the literature review section. It is important to note that estimators are useful to compare relative diversities rather than attempt to reveal true diversity. Also note that diversity estimators are considered lower bound estimates of true diversity. Therefore it was necessary to graph the estimators from the individual fuel libraries and compare their relative diversities. Richness estimators at the phylum and genus levels are depicted in Figure 18. Estimators are graphed as a function of sampling effort with the Y-axis normalized in order to compare relative diversity of the microbial communities from the three fuel types.

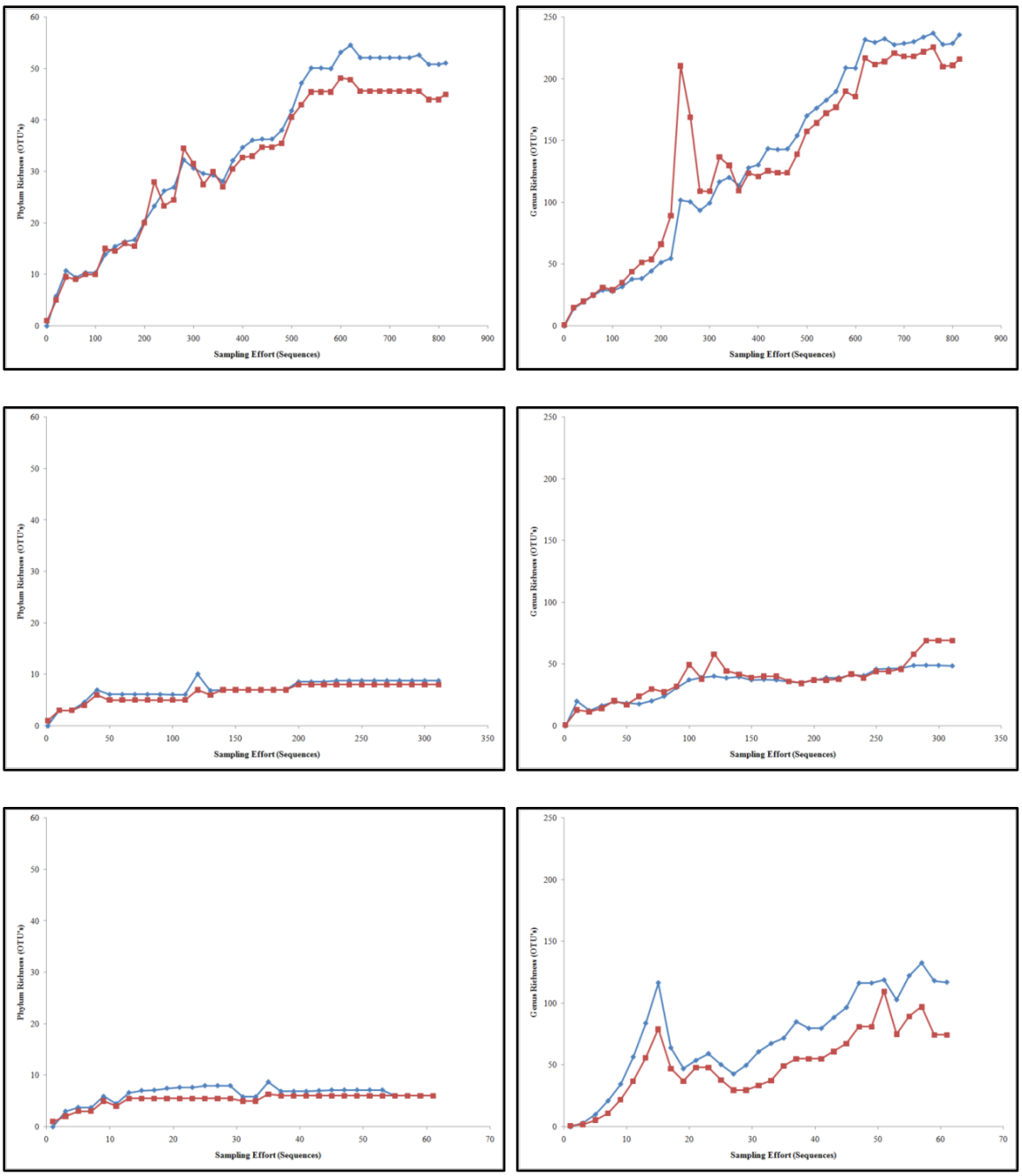



**Figure 18. Richness estimators for fuel type libraries**
**ACE (diamond) and Chao1 (square) richness estimators at the phylum level (left) and genus level (right) for JP-8 (top), Jet A (middle) and biodiesel (bottom)**

The richness estimators showed some interesting trends. The ACE estimator predicted the highest richness in most cases, with the exception of the genus level of Jet A. The JP-8 community had a vastly higher ACE and Chao1 estimate than the other two

communities as was predicted by the rarefaction analysis above. This conveyed that more OTUs are likely to be present in the JP-8 microbial community. Jet A and biodiesel had only slight differences in their estimator values, suggesting that the overall richness of the two communities is similar. Biodiesel had a slightly higher richness than Jet A (75 and 69, respectively). However, this does not suggest that the community membership or structure was similar or the same, only that richness was similar. Community membership and structure of the libraries was compared using the ∫-LIBSHUFF and SONS programs. Results based on the output from these programs are provided in the next section.

**Community Membership and Structure Comparison among Fuel Types**

∫-LIBSHUFF was used to statistically determine if two libraries were drawn from the same microbial community, or if one community was a subset of the other. Results from the pairwise comparisons of the three fuel type libraries are provided in Figure 19. The comparisons were graphed with coverage on the Y-axis as a function of evolutionary distance from zero to twenty percent. The p-values, representing statistical probability are also included. P-values less than 0.05 were considered to be significant. Significant differences meant that samples were indeed drawn from dissimilar microbial populations.

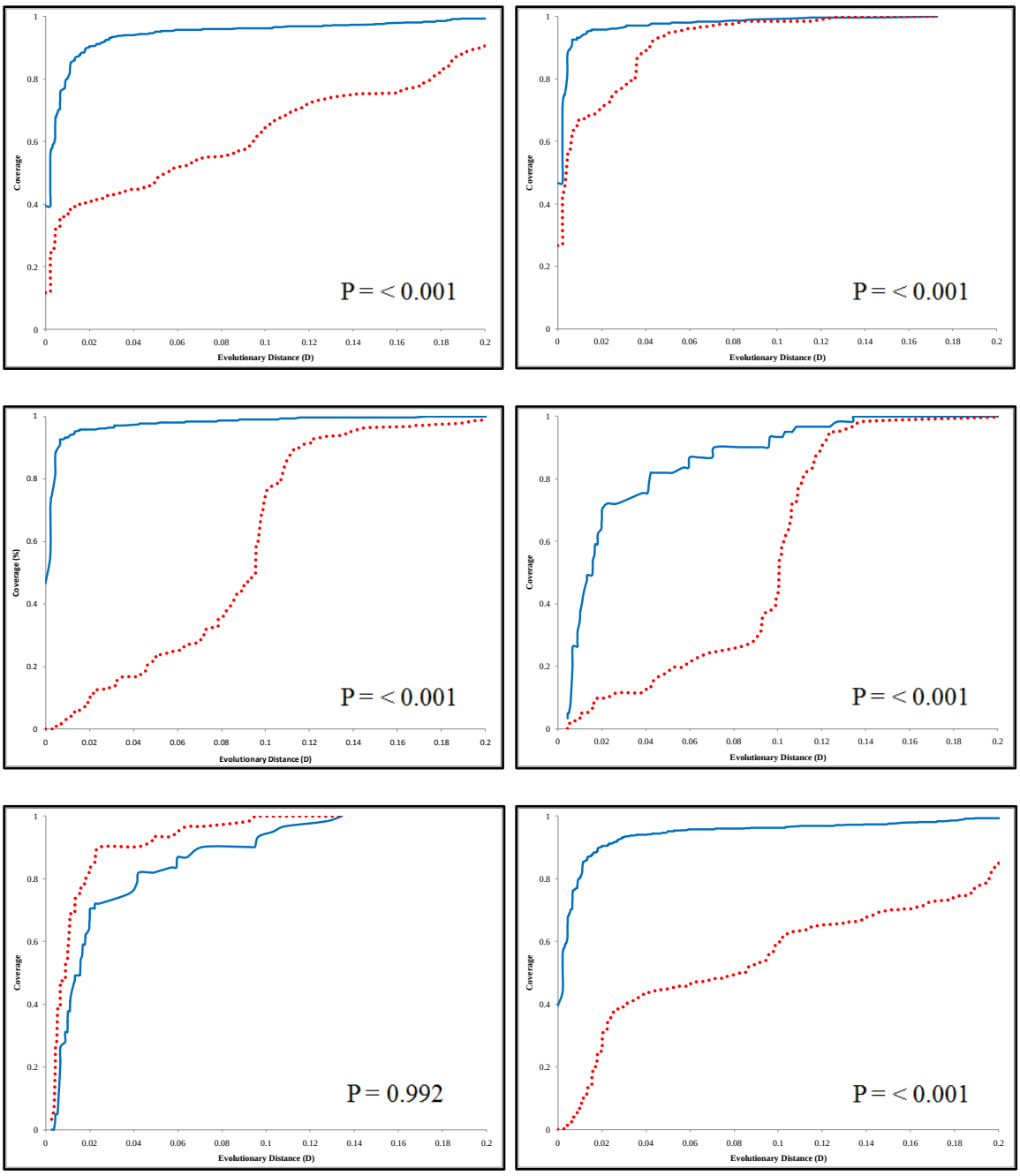



**Figure 19. Results of ∫-LIBSHUFF comparisons**
**Homologous (solid line) and heterologous (dotted line) curves for sequences libraries from aviation fuel samples. (Top) JP-8 vs. Jet A (left) and Jet A vs. JP-8 (right). (Middle) Jet A vs. biodiesel (left) and biodiesel vs. Jet A (right). (Bottom) Biodiesel vs. JP-8 (left) and JP-8 vs. biodiesel (right).**

The pairwise comparison of JP-8 and Jet A suggested that the two microbial communities were significantly different ($p<0.001$ in both cases). The same conclusion

was drawn from the pairwise comparison of Jet A and biodiesel ($p<0.001$ in both cases). However this was not the case between JP-8 and biodiesel. The significant p-value of JP-8 vs. biodiesel ($p<0.001$) suggests that the JP-8 library was drawn from a different microbial community than the biodiesel library. However, the insignificant p-value of biodiesel vs. JP-8 ($p=0.992$) suggested that the biodiesel microbial community is a subset of the JP-8 community. The insight gained from the ∫-LIBSHUFF program warranted a look at the microbial communities from a different perspective using the SONS program. The SONS program allowed for a visual representation of the overlap in community membership of the three sequence libraries. The results of SONS analyses are presented in a Venn diagram to show the shared membership and relative richness among JP-8, Jet A, and biodiesel at the genus level (Figure 20).

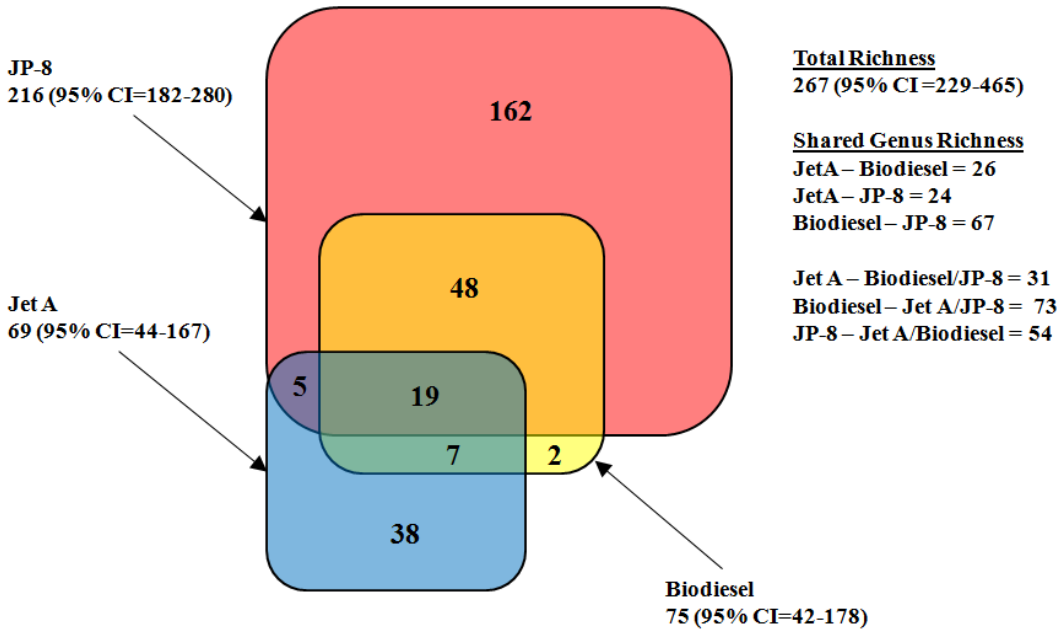



**Figure 20. Venn diagram showing genus richness and estimated community overlap**

A core membership of 19 genera ($OTU_{0.03}$) was estimated to be shared among the three aviation fuel microbial communities. Information from the phylogenetic classification of the sequences suggests that these genera may include: *Propionibacterium*, *Rhodococcus*, *Streptococcus*, *Burkholderia*, *Pseudomonas* as well as Unclassified Sphingomonadaceae, and Unclassified Comamonadaceae. Research suggests that shared populations may be responsible for essential support functions of a community (Schloss & Handelsman, 2006a). The biodiesel community was indeed shown as a subset of the JP-8 community as was alluded to by the ∫-LIBSHUFF analysis. JP-8 shared 24 and 67 OTUs with Jet A and biodiesel, respectively. The shared richness estimate between Jet A and biodiesel was 26 OTUs. The Chao1 richness estimates were 216, 69, and 75, for JP-8, Jet A, and biodiesel communities, respectively, and 267 for the combined data. These values all agreed with richness estimates from the DOTUR program output.

The majority of OTUs (75.7%), particularly from JP-8 and Jet A were endemic to a particular fuel type. Research suggests that endemic genera may serve as accessory populations, which are necessary to complement the core community in order to create the proper consortium of microorganisms to metabolize the various hydrocarbon fuel types and their differing chemical compositions (Schloss & Handelsman, 2006a).

# Chapter V: Conclusions

## Overview

This chapter summarizes the results from this study and provides significant conclusions and hypotheses. The research objectives are reviewed and the conclusions and insight gained from each are shared. This chapter also reviews the significance of this research and the contribution it made to the literature in this area. The chapter ends with suggestions for future research.

## Research Objective 1: Characterize the bacterial communities in the various aviation fuels by exploring community membership

In order to address this objective, the sample sequences were compared to a known database of 16s rDNA sequences, using the RDP Classifier, and classified into phyla and genera. Results showed that the sequences were classified into 13, 3, and 3 phyla for JP-8, Jet A, and biodiesel, respectively. Each phylum was further dissected into genera whenever possible, using an 80% confidence threshold for placement into the RDP hierarchy (Table 6). This type of information is useful for future researchers to fully explore the functional aspects, rather than the phylogenetic aspects, of the microbial communities brought to light by this thesis effort. Some examples of the type of research efforts that may develop from these findings are provided below.

While evidence of the problems associated with microbial contamination (biofilms, MIC, etc.) were not recorded at the time of sampling, it was initially presumed that the bacteria known to cause these problems were present in the communities. Therefore it was theorized that organisms from phyla and genera known to facilitate these effects would be present if/when a representative sample was taken from aviation fuel

systems. In regard to biofilms, numerous sequences were similar to microorganisms that have been shown to facilitate biofilm formation. Biofilm forming taxa identified in this study are: Plantomycetes, Actinobacteria, Sphingomonas, Rhizobiales, Enterobacteriaceae, *Staphylococcus*, *Clostridium*, *Mycobacterium*, *Bacillus*, *Deinococcus*, *Streptococcus*, *Burkholderia*, and *Pseudomonas* (MicrobeWiki, 2009). This list is not all-encompassing; however, it is an example of the usefulness of the microbial community characterization provided here. This research should be used as a stepping stone for future research endeavors.

Similar findings may enhance future research of organisms responsible for MIC. MIC is known to be enhanced by the presence of SRB such as *Desulfovibrio* sp.; however, none of the typical MIC-causing genera were revealed by this effort. It should be noted that SRB were identified in each of the previous studies using traditional culture methods (Table 3). *Desulfovibrio*, the organism identified most often, is a genus from the Phylum Proteobacteria , more specifically the DeltaProteobacteria class. Interestingly, only two sequences were placed into this classification and they were merely classified as Unclassified Deltaproteobacteria (Table 6). Given that the bacteria in question were most likely present in the current community, there are two possible explanations; either a representative sample was not obtained, resulting in only a partial picture of the natural microbial community in aviation fuel systems, or the molecular method applied in this research effort was not capable of isolating the organisms in question; perhaps due to primer specificity (Baker et al., 2003). Of note however, is that the phyla Nitrospira and Firmicutes were significantly represented in the JP-8 sequence library, and have been phenotypically linked to SRB (Bharathi, 2005). Further exploration of the classification

provided by the RDP classifier will allow for more insight into the community and may reveal valuable clues which future researchers may exploit.

**Research Objective 2: Investigate the effects of fuel type on microbial diversity and community structure**

This research objective was addressed using the DOTUR, ∫-LIBSHUFF, and SONS programs to analyze the sampled sequences and create graphs and charts in order to compare the microbial diversity and community structures of the various aviation fuel communities. Microbial communities are often extremely diverse and therefore the sequences analyzed by this research effort were a relatively small sample of the total microbial population. However, statistically speaking, the individuals present in the samples are likely to represent the dominant organisms in the natural community. Consequently, the metagenomic analysis provided many significant results but also highlighted some limitations that must be overcome in the future.

Sampling effort was considered prior to extrapolating results from richness estimation and composition analysis. Sampling effort was found to be lowest in the biodiesel sequence library and highest in Jet A. Additionally, sampling effort was higher at the phylum level than it was at the genus level, which was to be expected due to the vastness of the genus level. However, sampling effort was not an issue so severe as to prohibit significant conclusions from being drawn from the results, as only relative comparisons were required.

Richness estimators indicated that the richness of the JP-8 microbial community may be as much as three times higher than the richness of either Jet A or biodiesel. However, coupling this information with other dependent variables brings this conclusion

into question. Jet A was determined by rarefaction analysis to be the most adequately sampled of the three fuel types. Therefore it would be reasonable to state that Jet A supports a less diverse community of microorganisms than the other two fuel types. However, it should be noted that fuel sequences from Jet A were derived from samples of Jet A from "moth-balled" aircraft. The communities in these tanks may have had ample time for species dominance to occur, thereby limiting what would have been identified had the samples been taken from operational aircraft or fuel tanks being utilized on a daily basis. Additionally, Jet A samples were only drawn from two locations, both in the southwestern United States, which may have played an additional role in the outcome of this study.

Biodiesel was found to have a similarly low richness count which was counterintuitive to the common belief that these newer, alternative, biologically-friendly fuels are readily biodegradable and should therefore be more susceptible to microbial growth during storage (Robbins & Levy, 2004). This would be a significant finding were it not for some significant but yet unaccounted for variables. First, based on rarefaction analysis, biodiesel samples were grossly under sampled at the genus level. The biodiesel curve was nearly linear, meaning that nearly every sequence resulted in identification of a novel organism (Figure 15); further sampling is likely to reveal additional diversity. Second, similar to Jet A, samples were taken from a single source in the southwestern part of the United States, therefore a geographical bias may be present. Third, it should be noted that the novelty of alternative fuels may play an additional role in biodiesel's cleanliness, relative to JP-8. Biodiesel is still considered to be an experimental fuel and

is probably being extremely well-maintained compared to the conventional fuel types that have been in widespread use for decades.

**Significance of Research**

This research focused on characterizing the composition and diversity of aviation fuel microbial communities. Samples were taken from aircraft and storage tanks throughout the United States, sequenced in the lab and subjected to the metagenomic analysis described here. Classification of the 16s rDNA gene sequences resulted in a comprehensive analysis of the bacterial populations present in aviation fuel systems. As described in the literature review, microbial contamination has many deleterious effects on aviation fuel systems. This information provides a foundation for future researchers to work from in efforts to further isolate and study the genetics and behavior of the microbial contaminants commonly found in aviation fuel. Efforts to characterize the bacterial populations responsible for these effects are an on-going effort. This thesis effort is an essential prerequisite before a specifically targeted, permanent and reliable solution to a longstanding problem can be envisioned, and ultimately achieved.

**Future Research**

This study demonstrated the use of a molecular method to comprehensively characterize the microbial contamination in aviation fuel. While results were significant, this research merely hints at the true diversity of the microbial world. Subsequent studies should include additional sampling for metagenomic analysis purposes as well as more laboratory-based analyses to understand not only what microbes exist in jet fuel, but also

how they create biofilms, MIC, etc. and ultimately cause harm to fuel systems. Additional sampling across the vast geographical region in which the fuel types are dispersed would result in a more representative sample and should play a large role in the continuation of this type of research. Resulting rarefaction curves from larger sample sizes would be expected to approach asymptotic values, at which point the true diversity could be observed. Also, given the exponential growth currently being demonstrated by the various gene sequence databases, further illumination of the bacterial communities in aviation fuel systems is likely to result simply by re-classifying the sequences at a later date. These approaches will create avenues for further enrichment of the knowledge base and ultimately will aid in the development of successful mitigation strategies with which to attack the problem. The eventual goal is to prevent the initial formation of complex microbial communities in aviation fuel systems which may require novel, target specific biocides rather than the blanket approach being utilized by di-EGME and other biocides today.

## Appendix A: Initial Nomenclature for Sequence Identifiers

List of Sample ID's for CONUS sequences and what they correspond to:

**SL** followed by a number—these were sequenced by MWG, of High Point, NC. The SL number is just an MWG internal batch number, which we matched to a PO (purchase order) number. Each sample set we sent them had both an SL (MWG internal batch) number and a PO (our internal batch) number. The letters and numbers that follow the dash after an SL number correspond to the actual plate name, and give clues to the origin of the microbial DNA on that plate.

Here is what the letters mean:

**B** = Barksdale, LA AFB; B-52 aircraft, PCR product from standard plate culture

**bd** = Barksdale, but PCR product from direct PCR (DNA taken from fuel sample water filtrate, not cultured)

**cd** = Charleston, SC AFB; C-17 aircraft; PCR product from direct PCR

**d1-d4** = McGuire, NJ AFB; C-17 aircraft; PCR products from direct PCR

**d5-d11** = McGuire AFB; C-17 aircraft; PCR products from standard plate culture

**md2-8** = McGuire; KC-10 aircraft PCR products from direct PCR

**sd** = Wright Patterson AFB; S-13 storage tank; PCR products from direct PCR

**s1-s30** = Wright Patterson AFB; S-13 storage tank; PCR products from standard plate culture

**a1-a36** = Roswell, NM commercial air base; DC-9 aircraft; PCR products from standard plate culture, 2$^{nd}$ aircraft trip

**r1-r42** = Roswell, same as above, except 1$^{st}$ trip, and PCR products from standard plate culture

**rd33-rd40** = Roswell, 1$^{st}$ trip, PCR products from direct PCR

**s5d3** = Sacramento/Travis AFB, C-5 aircraft, PCR products from direct PCR

**scd1** = Sacramento/Travis AFB, KC-10 aircraft, PCR products from direct PCR

**std1-5** = Stewart, NY AFB; C-5 aircraft, PCR products from direct PCR

**v1-v25** = Victorville, CA commercial airbase; DC-9 aircraft, PCR products from standard plate culture

Samples run by other labs have different notations. Wright State uses:

**1051-36** = Davis Monthan, AZ, direct PCR, C-130 aircraft

**1052** = Davis Monthan, FTA (Fast Technology for Analysis DNA capturing) paper, C-130 aircraft

**A01-G07** = Dyess, TX, direct PCR, Biodiesel from storage tank

**MH101-296** = Mountain Home, ID AFB; direct PCR F-16 aircraft

**WR** = Warner Robbins, GA AFB; KC-135 aircraft, direct PCR products

## Appendix B: How to Create the SONS Diagram

The Venn diagram: This is a bit of a puzzle. There are probably other ways of doing this, but you can only do this with two or three libraries. Doing two treatments is simple: calculate the richness of each treatment and the overlapping region; scale each treatment and then overlap as appropriate. For the three libraries it is harder. You need to get the richness of each treatment and the richness of all three treatments together. Then you need to get the region that overlaps between A and B, A and C, and B and C. This is pretty simple using SONS with a names file stating those sequences from A, B, and C. Then you need to make a new set of names files so that you have A and BC, B and AC, and C and AB as the only two treatments in three separate files. Then you use SONS with each of these names files to get the region shared between A and BC, B and AC, and C and AB. To put it together see the sequence of Venn diagrams included along with the example below. The numbers used are from the "Shared Chao" column at the 0.03 distance level.

EXAMPLE:

For the three libraries we know:

A = 140

B = 274

C = 251

A-B = 88

A-C = 93

B-C = 122

A-B/C = 116

B-A/C = 153

C-A/B = 152

1. Draw three circles representing the three libraries and assign variable names to the overlapping regions (x, y, z, and m). See example Venn diagrams.

2. Determine a by subtracting the richness of A and B/C from the richness of A. Do likewise to determine b and c.

3. Next we need to determine the value for m. We know:

y+m=88 -> y=88-m

x+m=93 -> x=93-m

z+m=122 -> z=122-m

x+y+m1=116

y+z+m2=153

x+z+m3=152

We can plug the first three equations into the last three equations to get three possible values for m:

116=181-2m1+m1 -> m1=65

153=210-2m2+m2 -> m2=57

152=215-2m3+m3 -> m3=63

Considering the 95% CI ranges these values are all about the same. So I picked the lowest (m2=57) because it made everything else fit.

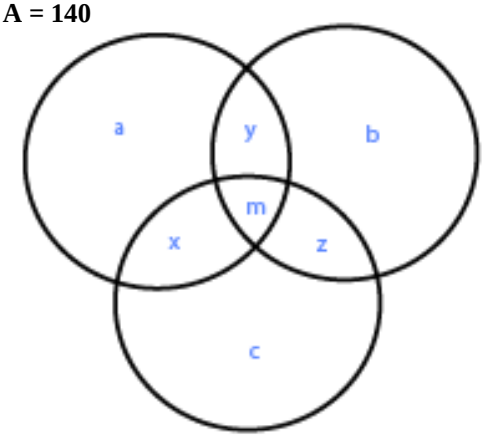
4. From those first equations in step 3 you can then determine the values of x (=36), y (=31), and z (=65).

5. To check and see how close everything is to "fitting" I then added all of the values in the Venn diagram to see if it was close to the total richness and I added all the values within a treatment to see if it was close to the value estimated for that treatment.

6. To draw the actual diagram I used PowerPoint and scaled the size of rounded rectangles to match the estimated richness. I tried circles, but it was difficult because it's hard to measure the overlapping area between circles. Note that if you scale a rectangle by 50% you are actually shrinking it to 25% of the original area.

1.

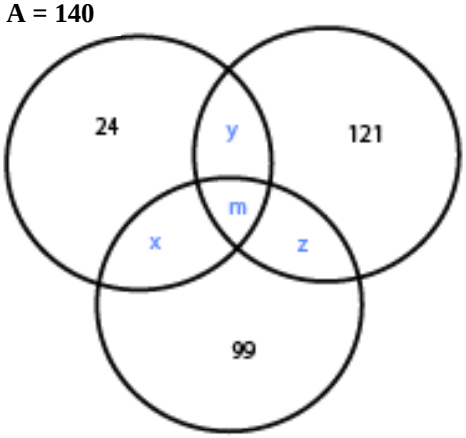
**A = 140** **B = 274**




**C = 251**

2

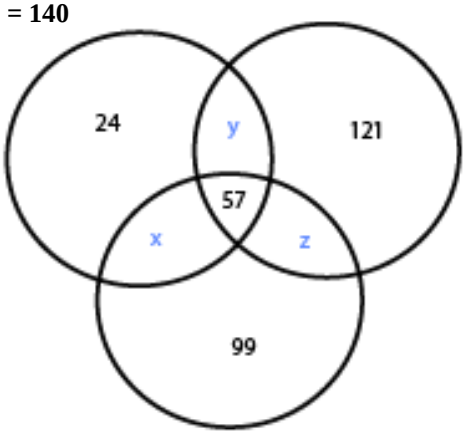
**A = 140** **B = 274**




**C = 251**

3.

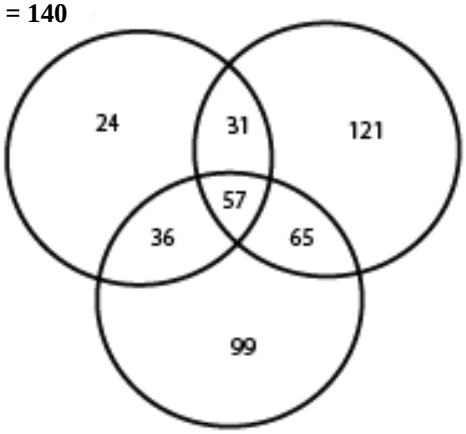
**A = 140** **B = 274**




**C = 251**

4.

**A = 140** **B = 274**




**C = 251**

## Appendix C: Direct PCR Protocol

1) Dilutions in preparation for direct PCR (Polymerase Chain Reaction)

-A fuel sample is filtered through a 0.45µm filter and the microbes are captured. They are washed off using sterile $H_2O$ and the neat sample is collected in a sterile tube.

-Gently agitate 0.2ml PCR tube with 100µl of neat sample. Briefly spin down tube in centrifuge for 30 seconds at 11,000 rpm. This ensures that all of the water/microbe mix is evenly dispersed and that no liquid is at the top of the opening which could be easily contaminated.

-Pipet 1µl of neat sample into a 0.2ml sterile PCR tube containing 99µl of sterile $H_2O$ and mix thoroughly. Repeat serial dilutions until you have a total of 5 tubes (1 neat sample and 4 dilutions).

-Tubes are then placed in a thermal cycler at 99°C for 10 minutes (this procedure lyses the cells and releases the DNA).

-Remove tubes from thermal cycler and briefly spin down at 11,000 rpm for 30 seconds.

2) DNA Protocol – Direct PCR (DNA amplification)

1. Usually, a 50 µl reaction is performed for dPCR; however, if sample is short, a 25µl reaction can be performed.
2. Spin all reagents before opening!
3. In sterile, well-labeled 0.2 mL microcentrifuge tubes, add the following for each sample in this order:

   17 µl sterile water (samples)
   2 µl 16S Forward Primer
   2 µl 16S Reverse Primer
   25 µl Red Taq Polymerase

   (for the positive control, add 19 µl of water instead of 17 µl and for the negative control add 21 µl of water)

   A master mix of the above reagents can be made by multiplying the amounts needed by the number of reactions to be done and mixing all reagents together prior to pipetting them into the individual tubes. If this is done, then 46 µl of MM is added to each sample tube, 48 µl MM is added to the positive control tube, and 50 µl is added to the negative control.

4. Add 4 µl of sample to each respective tube. Add 2µl of positive control as sample instead of 4 µl.

5. Place in thermal cycler and run on the "Fuelbug" cycle

    1. 94ºC, 2 minutes
    2. 94ºC, 30 seconds
    3. 51ºC, 20 seconds
    4. 72ºC, 30 seconds
    5. Go to 2., 30 times
    6. 72ºC, 5 minutes
    7. 10ºC, forever
    8. end.

6. After thermal cycler is complete, spin all samples for 30 seconds.
7. Electrophorese the sample by placing 10 µl of sample into each well of a 2% gel.
8. Electrophorese 2 hours at 60 volts (small gel) or 85 volts (large gel) in 1X Running Buffer (Tris-Acetate EDTA buffer).
9. Look for banding at the 500 bp band for bacterial confirmation.

-Depending on the number of samples, label on 0.2ml sterile PCR tube for each sample. Also include a positive control (DNA that has worked in the past) and a negative control (H2O is substituted for DNA, no DNA product should appear).

-Set up the PCR on ice using PCR reagents that will amplify the DNA. There is 50µl total/sample. Mix thoroughly by pipetting up and down.

-Place PCR reactions in thermal cycler and run appropriate protocol.

-Remove from thermal cycler and spin down briefly at 11,000 rpm for 30 seconds. Run 10µl of the PCR out of 50µl on a 1% agarose gel to see if any of the DNA product is visible. Store remaining PCR at -20°C.

-Sample is plated and sent off for sequencing.

**Definitions**

neat = not diluted or mixed with other samples.

thermal cycler = instrument that repeatedly cycles through various temperatures required for an iterative, temperature-dependent chemical process.

## Appendix D: Invitrogen® TOPO TA Cloning Protocol

### User Manual Version U (10 April 2006)

1. A master mix (MM) of reagents is made on ice – each reaction requires 1 µl of salt solution and 1 µl of TOPO pCR 2.1 Vector. Therefore add enough of each to the MM for all the samples testing. (ie: 2 samples, 2µl of each; 4 samples, 4µl of each) Mix gently, but well. Place on ice.
2. Setting the reaction up on ice, add 2µl of the MM to each well-labeled sterile 0.2 ml PCR tube. One tube for each sample being tested. Follow with 4µl of PCR product in the appropriate reaction tube.
3. Mix gently and incubate at room temperature (RT) for 5 minutes.
4. Place tubes on ice to stop reaction.
5. Thaw One Shot Chemically competent *E.coli* cells on ice – one tube for each test reaction. Cells are located in the -80°C freezer.
6. Add 2 µl of TOPO Cloning reaction to a vial of the cells (each reaction goes into a separate vial of cells). Mix gently and incubate on ice for 20 minutes. Place remaining reaction (4 µl) in the -20°C freezer.
7. Heat Shock the cells for 30 seconds at 42°C and place cells back on ice to stop reaction.
8. Add 250 µl of RT SOC Medium to the cells.
9. Cap the tubes and shake at 37°C for 1 hour at 200 RPMs.
10. Warm 2 LB with Kanamycin (50 µg/ml) and X-Gal (20 µg/ml) for each reaction. Label plates.
11. Spread 2 plates for each sample – one with 30 µl, one with 60 µl of transformed cells. Place remaining cells in the 4°C refrigerator overnight.
12. Place the plates in the 37°C incubator overnight.

## Appendix E: Colony PCR Protocol

1. Culture the TOPO plates prior to testing for colony PCR.
2. Using the TOPO Plates that were spread in the TOPO TA Cloning procedure, select as many **white** colonies as possible and subculture (or restreak) to another LB with Kanamycin (50 µg/ml) and X-Gal (20 µg/ml). This is done by only doing a small streak of each on the plate. A minimum of 48 colonies are needed to sequence, therefore, the more colonies that can be subcultured initially, the more likely it is to get 48 colonies for sequencing.
3. Incubate the plate overnight at 37°C.
4. Due to the size of the thermal cycler, colony PCR can be done on 95 colonies at a time, plus a negative control.
5. A Master Mix (MM) is made of the reagents – each reaction requires the following (for a 25µl reaction):

   2 µl 5µM M13 Forward primer
   2 µl 5 µM M13 Reverse primer
   3 µl Triton – 100, 1%
   12.5 µl Direct Load Master Mix (NEB)
   5.5 µl sterile $H_2O$

   For a reaction of 100X MM (100 reactions in the MM), the following amounts are needed:

   200 µl 5µM M13 Forward primer
   200 µl 5 µM M13 Reverse primer
   300 µl Triton – 100, 1%
   1250 µl Direct Load Master Mix (NEB)
   550 µl sterile $H_2O$

6. 25 µl of the MM is placed in a 0.2 µl PCR tube, and small amount of each white culture is added to the appropriately labeled PCR tube.
7. Tubes are placed in the thermal cycler and the "colony" protocol is run on the thermal cycler:

   1) 95°C, 2 min
   2) 95°C, 30 seconds
   3) 50°C, 45 seconds
   4) 72°C, 30 seconds
   5) Back to 2, 29 times
   6) 72°C, 5 min
   7) 10°C forever
   8) end

8. Remove the colony PCR from the thermal cycler and run all 25µl out on a large 2% gel (containing Ethidium Bromide), using a 100 bp ladder on each row. Run at about 75-85 Volts for 2 hours.
9. Those with inserts will band at about 700 bp; those without will band about 200 bp.

## Appendix F: QIAprep Spin Miniprep Protocol

1. Using sterile tubes, add 3 mL of LB broth with Kanamycin (50 µg/ml) to each tube. Inoculate the broth with 1 colony from the overnight isolated cultures done in "Colony PCR" procedure. Use only colonies that have been confirmed by colony PCR as containing inserts. Incubate overnight in 37°C incubator shaking at 200 RPM.
2. Transfer 2 mL of overnight culture to a sterile, labeled 2 mL microcentrifuge tube and centrifuge (at full speed ~14,000 rpm) for 8 minutes. Hold remaining 1 mL of culture overnight in the 4°C refrigerator.
3. Decant off the supernatant from microcentrifuge tube and "beat" tube against paper towels to get off any excess fluid.
4. Resuspend the pellet in 250 µl of Buffer P1*. Vortex tubes to make certain button is completely resuspended.
   ***NOTE:** Buffer P1 must have RNase A added to it and it should be refrigerated at all times.
5. Add 250 µl Buffer P2 and mix thoroughly by gently inverting the tubes 4-6 times. (do not allow the lysis rxn to proceed for more than 5 minutes)
6. Add 350 µl of Buffer N3 and mix immediately and completely by inverting the tubes 4-6 times. Solution should become cloudy (cell debris)
7. Centrifuge 10 minutes at full speed (14,000 RPM).
8. Decant and/or pipette off the supernatant into to the QIAprep spin column. Make sure both the spin column and the tube it sits in are labeled!
9. Centrifuge 1 minute and discard flow-through.
10. Wash spin column by adding 750 µl of Buffer PE and centrifuging for 1 minute.
11. Discard the flow-through and centrifuge the spin column again for 1 minute to remove any residual wash buffer.
12. Place the spin column into a sterile and labeled 1.5 mL microcentrifuge tube.
13. Elute plasmid DNA in 75 µl of Buffer EB. Place Buffer EB in the center of the spin column to elute. Let stand 1 minute and centrifuge for 1 minute.
14. Discard spin column, cap tube and freeze. Samples are now ready to be sequenced.

## Bibliography


Achenbach, L. A., & Coates, J. D. (2000). Disparity between bacterial phylogeny and physiology. *ASM News*(66), 714-715.

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev., 59*(1), 143-169.

Anderson, T. W. (1962). On the distribution of the two-sample Cramer-von Mises criterion. *Annals of Mathematical Statistics, 33*(3), 1148-1159.

Angell, P. (1999). Understanding microbially influenced corrosion as biofilm-mediated changes in surface chemistry. *Current Opinion in Biotechnology, 10*(3), 269-272.

Appenzeller, T. (1990). Democratizing the DNA sequence. *Science, 247*, 1030-1032.

ASTM. (1999). Standard Guide for Microbial Contamination in Fuels and Fuel Systems. *Designation D 6469-99*.

Atlas, R. M. (1981). Microbial degradation of petroleum hydrocarbons: an environmental perspective. *Microbiol. Mol. Biol. Rev., 45*(1), 180-209.

Bakanauskas, S. (1958). *Bacterial Activity in JP-4 Fuel*. Wright-Patterson Air Force Base, OH: Wright Air Development Center.

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

Balster, L. M., Chelgren, S. K., Strobel, E. M., Vangsness, M. D., & Bowen, L. L. (2006). Characterization of the bio-burden in United States aviation fuel: A comparison of FTA paper and other methods for obtaining genetic material from aviation fuel microbes. *American Chemical Society, 51*(3&4).

Beech, I. B., & Sunner, J. (2004). Biocorrosion: towards understanding interactions between biofilms and metals. *Current Opinion in Biotechnology, 15*, 181-186.

Bharathi, P. A. L. (2005). Synergy in Sulfur Cycle: The Biogeochemical Significance of Sulfate Reducing Bacteria in Syntrophic Associations. In N. Ramaiah (Ed.), *Marine Microbiology: Facets and Opportunities* (pp. 27-35). Dona Paula, Goa, India: National Institute of Oceanography.

Bradley, G. (2008, 26 Feb 08). Crop this: Virgin takes off with nut-fuel. *The New Zealand Herald*.

Brock, T. (1987). *The study of microorganisms in situ: Progress and problems.* Paper presented at the Symposia of the Society for General Microbiology, Cambridge, MA.

Brown, M. P. S. (2000). *Small subunit ribosomal RNA modeling using stochastic context-free grammars*. Paper presented at the Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology.

Bushnell, L. D., & Haas, H. F. (1941). The utilization of certain hydrocarbons by microorganisms. *J. Bacteriol., 41*(5), 653-673.

Cannone, J., Subramanian, S., Schnare, M., Collett, J., D'Souza, L., Du, Y., et al. (2002). The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs: Correction. *BMC Bioinformatics, 3*(1), 15.

Chai, B. (2008). Ribosomal Database Project Administrator. Michigan State University, East Lansing, MI. Personal Communication.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics, 11*, 265-270.

Chao, A., & Lee, S. M. (1992). Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc, 87*, 210-217.

Chazdon, R. L., Colwell, R. K., Denslow, J. S., & Guariguata, M. R. (1998). Statistical methods for estimating species richness of wood regeneration in primary and secondary rain forests of NE Costa Rica. In F. Dallmeier & J. A. Comiskey (Eds.), *Forest biodiversity research monitoring and modeling: conceptual background and Old World case studies* (pp. 285-309). Paris, France: Parthenon Publishing.

Chelgren, S. K. (2008). Microorganisms in Aviation Fuels. Air Force Research Laboratory Fuels Branch.

Chesneau, H. L. (1988). Fuel System Contaminants: An Introduction. In H. L. Chesneau & M. M. Dorris (Eds.), *Distillate Fuel: Contamination, Storage and Handling*. Philadelphia, PA: American Society for Testing and Materials.

Clarridge, J. E., III. (2004). Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clin. Microbiol. Rev., 17*(4), 840-862.

Cloud, J. L., Neal, H., Rosenberry, R., Turenne, C. Y., Jama, M., Hillyard, D. R., et al. (2002). Identification of Mycobacterium spp. by Using a Commercial 16S Ribosomal DNA Sequencing Kit and Additional Sequencing Libraries. *J. Clin. Microbiol., 40*(2), 400-406.

Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., et al. (2007). The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucl. Acids Res., 35*(suppl_1), D169-172.

Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam, S. A., McGarrell, D. M., et al. (2005). The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucl. Acids Res., 33*(suppl_1), D294-296.

Colwell, R. K., & Coddington, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. London, 345*, 101-118.

Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R., & Lappin-Scott, H. M. (1995). Microbial Biofilms. *Annual Review of Microbiology, 49*(1), 711-745.

Crum, M. G., Reynolds, R. J., & Hedrick, H. G. (1967). Microbial Penetration and Utilization of Organic Aircraft Fuel-Tank Coatings. *Appl. Environ. Microbiol., 15*(6), 1352-1355.

Daggett, D. L., Hendricks, R. C., Walther, R., & Corporan, E. (2007). *Alternate Fuels for use in Commercial Aircraft*. Seattle, WA: Boeing Corporation.

Dahllof, I., Baillie, H., & Kjelleberg, S. (2000). rpoB-Based Microbial Community Analysis Avoids Limitations Inherent in 16S rRNA Gene Intraspecies Heterogeneity. *Appl. Environ. Microbiol., 66*(8), 3376-3380.

Danigole, M. S. (2007). *Biofuels: An Alternative to US Air Force Petroleum Fuel Dependency.* Air War College, Air University, Maxwell Air Force Base, Alabama.

Davey, M. E., & O'Toole, G. A. (2000). Microbial Biofilms: from Ecology to Molecular Genetics. *Microbiol. Mol. Biol. Rev., 64*(4), 847-867.

Defense Energy Support Center. (1998). *Survey of Jet Fuels Procured by Defense Energy, 1990-1996*. Ft. Belvoir, VA.

Denaro, T. C., Sarah; Lang, Jara; Strobel, Ellen; Balster, Lori; Vangsness, Marlin. (2005). *DNA Isolation of Microbial Contaminants in Aviation Turbine Fuel Via Traditional Polymerase Chain Reaction (PCR) and Direct PCR (Preliminary Results)*: Air Force Research Laboratory; University of Dayton Research Institute.

Drancourt, M., Bollet, C., Carlioz, A., Martelin, R., Gayral, J.-P., & Raoult, D. (2000). 16S Ribosomal DNA Sequence Analysis of a Large Collection of Environmental and Clinical Unidentifiable Bacterial Isolates. *J. Clin. Microbiol., 38*(10), 3623-3630.

Edmonds, P. (1965). Selection of test organisms for use in evaluating microbial inhibitors in fuel-water systems. *Appl. Environ. Microbiol., 13*(5), 823-824.

Edmonds, P., & Cooney, J. J. (1967). Identification of Microorganisms Isolated from Jet Fuel Systems. *Appl. Environ. Microbiol., 15*(2), 411-416.

Edwards, T. (2003). Liquid Fuels and Propellants for Aerospace Propulsion: 1903 – 2003. *Journal of Propulsion and Power, 19*(6), 1089-1107.

Felsenstein, J. (2008). PHYLIP (Phylogeny Inference Package) (Version 3.68). Seattle, WA: Department of Genome Sciences, University of Washington.

Finefrock, V. H., Killian, L. N., & London, S. A. (1965). *Microbial Activity in Air Force Jet Fuel Systems*. Ft. Belvoir: Systems Research Labs Inc. Dayton, OH.

Finefrock, V. H., & London, S. A. (1966). *Microbial Contamination of USAF JP-4 Fuels*. Wright-Patterson AFB, OH: Air Force Propulsion Laboratory.

Fitch, W. M., & Margoliash, E. (1967). Construction of Phylogenetic Trees. *Science, 155*(3760), 279-284.

Fox, G. E., Wisotzkey, J. D., & Jurtshuk, P. J. R. (1992). How Close Is Close: 16S rRNA Sequence Identity May Not Be Sufficient To Guarantee Species Identity. *Int J Syst Bacteriol, 42*(1), 166-170.

Garrity, G. M. L., T.; Cole, J.; Harrison, S.; Euzeby, J.; Tindall, B. (2007). The Taxonomic Outline of Bacteria and Archea, *TOBA Release 7.7*. Michigan State University Board of Trustees.

Gaylarde, C. C., Bento, F. M., & Kelley, J. (1999). Microbial contamination of stored hydrocarbon fuels and its control. *Revista de Microbiologia, 30*, 01-10.

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika, 40*, 237-264.

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters, 4*(4), 379-391.

Graef, H. W. (2003). *An Analysis of Microbial Contamination in Military Aviation Fuel Systems.* Air Force Institute of Technology, Wright-Patterson AFB, OH.

Hagstrom, A., Pommier, T., Rohwer, F., Simu, K., Stolte, W., Svensson, D., et al. (2002). Use of 16S Ribosomal DNA for Delineation of Marine Bacterioplankton Species. *Appl. Environ. Microbiol., 68*(7), 3628-3633.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser, 41*, 95-98.

Handelsman, J. (2004). Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiol. Mol. Biol. Rev., 68*(4), 669-685.

Harwood, C. B., Merry. (2008). *The Uncharted Microbial World: microbes and their activities in the environment*. Washington, DC: American Academy of Microbiology.

Head, I. M., Saunders, J. R., & Pickup, R. W. (1998). Microbial Evolution, Diversity, and Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms. *Microbial Ecology, 35*(1), 1-21.

Heck, K. L., Belle, G. v., & Simberloff, D. (1975). Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology Letters, 56*, 1459-1461.

Hedrick, H. G., Carroll, M. T., Owen, H. P., & Pritchard, D. J. (1963). Viability of Selected Microorganisms in Hydrocarbon Fuels. *Appl. Environ. Microbiol., 11*(6), 472-475.

Hemighaus, G., Boval, T., Bacha, J., Barnes, F., Franklin, M., Gibbs, L., et al. (2006). *Aviation Fuels Technical Review (FTR-3)*: Chevron Corporation.

Hugenholtz, P., Goebel, B. M., & Pace, N. R. (1998). Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. *J. Bacteriol., 180*(18), 4765-4774.

Hughes, J. B., Hellmann, J. J., Ricketts, T. H., & Bohannan, B. J. M. (2001). Counting the Uncountable: Statistical Approaches to Estimating Microbial Diversity. *Appl. Environ. Microbiol., 67*(10), 4399-4406.

Jukes, T. H., Cantor, C. R., & Munro, H. N. (1969). *Evolution of Protein Molecules*: Academy Press.

Kemp, P. F., & Aller, J. Y. (2004). Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. *FEMS Microbiology Ecology, 47*(2), 161-177.

Kersey, P., & Apweiler, R. (2006). Linking publication, gene and protein data. *Nature Cell Biology, 8*(11), 1183-1189.

Konstantinidis, K. T., & Tiedje, J. M. (2005). Towards a Genome-Based Taxonomy for Prokaryotes. *J. Bacteriol., 187*(18), 6258-6264.

Kuske, C. R., Barns, S. M., Grow, C. C., Merrill, L., & Dunbar, J. (2006). Environmental Survey for Four Pathogenic Bacteria and Closely Related Species Using Phylogenetic and Functional Genes<sup>*</sup>. *Journal of Forensic Sciences, 51*(3), 548-558.

Leon, E. M. (2008). *Molecular Characterization of Wetland Soil Bacterial Community In Constructed Mesocosms.* Air Force Institute of Technology, Wright-Patterson AFB, OH.

Maurice, L. Q., Lander, H., Edwards, T., & Harrison, W. E. (2001). Advanced aviation fuels: a look ahead via a historical perspective. *Fuel, 80*, 747-756.

Merriam-Webster Online (Ed.) (2002) Merriam-Webster Collegiate Dictionary

Meshako, C. E., Bleckmann, C. A., & Goltz, M. N. (1999). Biodegradability and microbial toxicity of aircraft fuel system icing inhibitors. *Environmental Toxicology, 14*(4), 383-390.

MicrobeWiki. (2009). Biofilms. Retrieved 19 Feb 09.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Biotechnology, 24*, 17-27.

Neihof, R. A. (1988). Microbes in Fuel: An Overview with a Naval Perspective. In H. L. Chesneau & M. M. Dorris (Eds.), *Distillate Fuel: Contamination, Storage, and Handling* (pp. 6-14). Philadelphia, PA: American Society for Testing and Materials.

Neihof, R. A., & Bailey, C. A. (1978). Biocidal Properties of Anti-Icing Additives for Aircraft Fuels. *Appl. Environ. Microbiol., 35*(4), 698-703.

Nogales, B., Moore, E. R. B., Llobet-Brossa, E., Rossello-Mora, R., Amann, R., & Timmis, K. N. (2001). Combined Use of 16S Ribosomal DNA and 16S rRNA To Study the Bacterial Community of Polychlorinated Biphenyl-Polluted Soil. *Appl. Environ. Microbiol., 67*(4), 1874-1884.

Nubel, U., Garcia-Pichel, F., Kuhl, M., & Muyzer, G. (1999). Quantifying Microbial Diversity: Morphotypes, 16S rRNA Genes, and Carotenoids of Oxygenic Phototrophs in Microbial Mats. *Appl. Environ. Microbiol., 65*(2), 422-430.

Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R., & Stahl, D. A. (1986). Microbial Ecology and Evolution: A Ribosomal RNA Approach. *Annual Review of Microbiology, 40*(1), 337-365.

Olsen, G. J., & Woese, C. R. (1993). Ribosomal RNA: a key to phylogeny. *FASEB J., 7*(1), 113-123.

Pace, N. R. (1997). A Molecular View of Microbial Diversity and the Biosphere. *Science, 276*(5313), 734-740.

Pettitt, A. N. (Ed.) (1982) Encyclopedia of Statistical Sciences. New York, N.Y.: Wiley-Interscience.

Rahman, P. K. S. M., & Gakpe, E. (2008). Production, Characterization and Applications of Biosurfactants-Review. *Biotechnology, 7*(2), 360-370.

Rauch, M. E. (2008). *Threat of Fuel Degrading Microorganisms on US Military Fuel Supplies*. Maxwell Air Force Base, Alabama: Air University.

Rauch, M. E., Graef, H. W., Rozenzhak, S. M., Jones, S. E., Bleckmann, C. A., Kruger, R. L., et al. (2005). Characterization of microbial contamination in United States Air Force aviation fuel tanks. *Journal of Microbial Biotechnology, 33*, 29-36.

Robbins, J. A., & Levy, R. (2004). A review of the microbial degradation of fuel. In W. Paulus (Ed.), *Directory of Microbicides for the Protection of Materials: A Handbook* (pp. 177-201): Springer Netherlands.

Rodríguez-Valera, F. (2002). Approaches to prokaryotic biodiversity: a population genetics perspective. *Environmental Microbiology, 4*(11), 628-633.

Schloss, P. D. (2005). DOTUR Manual: Schloss Labs.

Schloss, P. D. (2008). Evaluating different approaches that test whether microbial communities have the same structure. *ISME J, 2*(3), 265-275.

Schloss, P. D., & Handelsman, J. (2004). Status of the Microbial Census. *Microbiol. Mol. Biol. Rev., 68*(4), 686-691.

Schloss, P. D., & Handelsman, J. (2005). Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. *Appl. Environ. Microbiol., 71*(3), 1501-1506.

Schloss, P. D., & Handelsman, J. (2006a). Introducing SONS, a Tool for Operational Taxonomic Unit-Based Comparisons of Microbial Community Memberships and Structures. *Appl. Environ. Microbiol., 72*(10), 6773-6779.

Schloss, P. D., & Handelsman, J. (2006b). Toward a Census of Bacteria in Soil. *PLoS Computational Biology, 2*(7), e92.

Schloss, P. D., & Handelsman, J. (2008). A statistical toolbox for metagenomics: assessing functional diversity in microbial communities. *BMC Bioinformatics, 9*, 34.

Schloss, P. D., Larget, B. R., & Handelsman, J. (2004). Integration of Microbial Ecology and Statistics: a Test To Compare Gene Libraries. *Appl. Environ. Microbiol., 70*(9), 5485-5492.

Singleton, D. R., Furlong, M. A., Rathbun, S. L., & Whitman, W. B. (2001). Quantitative Comparisons of 16S rRNA Gene Sequence Libraries from Environmental Samples. *Appl. Environ. Microbiol., 67*(9), 4374-4376.

Suzuki, M. T., & Giovannoni, S. J. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol., 62*(2), 625-630.

Swift, S. T. (1988). Identification and Control of Microbial Growth in Fuel Handling Systems. In H. L. C. M. M. Dorris (Ed.), *Distillate Fuel Contamination, Storage and Handling* (pp. 15-26). Philadelphia, PA: American Society for Testing and Materials.

Vaccari, D. A., Strom, P. F., & Alleman, J. E. (2006). *Environmental Biology for Engineers and Scientists*. Hoboken, NJ: John Wiley & Sons, Inc.

Van Hamme, J. D., Singh, A., & Ward, O. P. (2003). Recent Advances in Petroleum Microbiology. *Microbiol. Mol. Biol. Rev., 67*(4), 503-549.

Vangsness, M. D., Chelgren, S. K., Strobel, E. M., Balster, L. M., Bowen, L. L., & Mueller, S. (2007). *Microbial contamination studies in JP-8 fueled aircraft*. Paper presented at the IASH 2007, the 10th International Conference on Stability, Handling and Use of Liquid Fuels.

Vasanthakumar, A., Handelsman, J., Schloss, P. D., Bauer, L. S., & Raffa, K. F. (2008). Gut Microbiota of an Invasive Subcortical Beetle, Agrilus planipennis Fairmaire, Across Various Life Stages. *Environmental Entomology, 37*, 1344-1353.

Videla, H. A. (2001). Microbially induced corrosion: an updated overview. *International Biodeterioration and Biodegradation, 48*, 176-201.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol., 73*(16), 5261-5267.

Watkinson, R. J., & Morgan, P. (1990). Physiology of aliphatic hydrocarbon-degrading microorganisms. *Biodegradation, 1*(2), 79-92.

Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America, 95*(12), 6578-6583.

Wintzingerode, F. V., Göbel, U. B., & Stackebrandt, E. (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews, 21*(3), 213-229.

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Mol. Biol. Rev., 51*(2), 221-271.

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America, 74*(11), 5088-5090.

Zhang, K., Choi, H., Dionysiou, D. D., Sorial, G. A., & Oerther, D. B. (2006). Identifying pioneer bacterial species responsible for biofouling membrane bioreactors. *Environmental Microbiology, 8*(3), 433-440.

Zobell, C. E. (1946). Action of Microorganisms on Hydrocarbons. *Microbiol. Mol. Biol. Rev.*, *10*(1-2), 1-49.

## Vita

Captain Jerrod P. McComb graduated from Green Valley High School in Henderson, Nevada. He entered undergraduate studies at the United States Air Force Academy in Colorado Springs, Colorado where he graduated with a Bachelor of Science degree in Environmental Engineering in June 2004.

He was first assigned to the 50th Civil Engineer Squadron at Schriever AFB, Colorado Springs, Colorado where he served as the Base Pavements Officer in the Maintenance Engineering section. Captain McComb, then Lieutenant McComb, also served as the Simplified Acquisition of Base Engineering Requirements (SABER) Chief where he oversaw maintenance and construction projects that were vital to sustainment of base infrastructure and capabilities. While stationed at Schriever AFB he also served as the Prime Base Engineering Emergency Forces (Prime BEEF) Officer in charge of contingency training and exercise planning, execution and evaluation.

In August 2007, he entered the Graduate School of Engineering and Management, Air Force Institute of Technology, where he earned a Masters of Science degree in Engineering Management with a focus in Systems Analysis. Upon graduation, he will be assigned to the 92$^{nd}$ Civil Engineer Squadron at Fairchild AFB, WA.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 074-0188*



| **1. REPORT DATE** *(DD-MM-YYYY)* | **2. REPORT TYPE** | **3. DATES COVERED** *(From – To)* |
|---|---|---|
| 13-03-2009 | Master's Thesis | Aug 2007 – Mar 2009 |

**4. TITLE AND SUBTITLE**

A Metagenomic Analysis of Microbial Contamination in Aviation Fuels

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

McComb, Jerrod P, Captain, USAF

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)**
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

**8. PERFORMING ORGANIZATION REPORT NUMBER**

AFIT/GEM/ENV/09-M11

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
University of Dayton Research Institute
Energy and Environmental Engineering Division
Mr. Marlin Vangsness
300 College Park Ave
Dayton, OH 45469

**10. SPONSOR/MONITOR'S ACRONYM(S)**
UDRI

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**
Microbial contamination in aviation fuel arises due to the impracticality of keeping fuel tanks sterile and the inevitable presence of water from condensation. Microbial contaminants in aviation fuels are a concern because of their potential to degrade the fuel, accelerate corrosion within the fuel tank, and threaten flight safety. This research aids in mitigating those problems by comprehensively characterizing the microbial communities affecting aviation fuels. Advances in molecular biological techniques have allowed for the identification of microorganisms which were not identified by the traditional culture-based methodologies used in previous studies**.** This study employed a molecular method known as 16S rDNA gene analysis to describe the microbial communities in aviation fuel. The microbial communities in JP-8, Jet A, and biodiesel were evaluated at the phylum and genus levels of taxonomy. The JP-8 community was found to be much richer than both the Jet A and biodiesel community. The biodiesel community was found to be a subset of the JP-8 community. A small subset of microorganisms was found to exist across all three fuels while the majority of identified microorganisms were endemic to a single fuel type. Rarefaction analysis showed that further sampling is likely to reveal additional diversity.

**15. SUBJECT TERMS**
Microbial contamination, aviation fuel, metagenomics, phylogenetics, fuel microorganisms, JP-8, Jet A, biodiesel

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | Dr. Charles Bleckmann, AFIT/ENV |
| U | U | U | UU | 132 | **19b. TELEPHONE NUMBER** *(Include area code)* 937-255-3636 x4721 |

**Standard Form 298 (Rev. 8-98)**
Prescribed by ANSI Std. Z39-18