

THE MICROBIAL RETTING ENVIRONMENT OF *Hibiscus cannabinus* AND ITS IMPLICATIONS IN
BROADER APPLICATIONS

David K. Visi

Dissertation Prepared for the Degree of
DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

May 2015

APPROVED:

Robert Benjamin, Major Professor
Michael Allen, Committee Member
Aaron Roberts, Committee Member
Lee Hughes, Committee Member
Stevens Brumbley, Committee Member
Art Goven, Chair of the Department of
Biological Sciences
Costas Tsatsoulis, Interim Dean of the
Toulouse Graduate School

Visi, David K. *The Microbial Retting Environment of Hibiscus cannabinus and Its Implications in Broader Applications*. Doctor of Philosophy (Molecular Biology and Biochemistry), May 2015, 142 pp., 6 tables, 26 figures, bibliography, 135 titles.

Fiber-yielding plants is an area of increased interest due to the potential use in a variety of green-based materials. These biocomposites can be incorporated into multiple uses; for example, to replace building materials and interior vehicular paneling. The research here aims to focus in on the crop *Hibiscus cannabinus* for utilization into these functions. *H. cannabinus* is economically attractive due to the entire process being able to be accomplished here in the United States. The plant can be grown in a relatively short growth period (120-180 days), and then processed and incorporated in a biocomposite. The plant fiber must first be broken down into a useable medium. This is accomplished by the retting process, which occurs when microbial constituents breakdown the heteropolysaccharides releasing the fiber.

The research aims to bridge the gap between the primitive process of retting and current techniques in molecular and microbiology. Utilizing a classical microbiological approach, which entailed enrichment and isolation of pectinase-producing bacteria for downstream use in augmented microbial retting experiments. The tracking of the bacteria was accomplished by using the 16S rRNA which acts as “barcodes” for bacteria. Next-generation sequencing can then provide data from each environment telling the composition and microbial diversity of each tested variable. The main environments tested are: a natural environment, organisms contributed by the plant material solely, and an augmented version in which pectinase-producing bacteria are added. In addition, a time-course experiment was performed on the augmented environment providing data of the shift to an anaerobic environment. Lastly, a

drop-in set was performed using each isolate separately to determine which contributes to the shift in microbial organization. This research provided a much needed modernization of the retting technique. Previous studies have been subject to simple clone libraries and growth plate assays and next-generation sequencing will bring the understanding of microbial retting into the 21st century.

Copyright 2015

by

David K. Visi

ACKNOWLEDGEMENTS

I would first and foremost like to thank the University of North Texas, which gave me the opportunity to work as a graduate student in the department of Biological Sciences. I had an interesting experience with lab members in that I had such a large group of people that helped guide me through my graduate career. I would like to specifically thank Leslie Perry for being my guide and friend in my first year of graduate school and teaching me all the foundational aspects of bench top work and never showing frustration at my lack of experience. I was fortunate to have another lab at the University of North Texas Health Science Center and would like to specifically mention our two amazing post-docs in Yan Zhang, who worked tirelessly with me to get the bioinformatics pipeline we had desired for so long up and running and Marnie Rout, who provided me with invaluable advice, countless hours of editing, and help with understanding the intricacies of diversity estimates. Richard Hale, my friend and my colleague, who worked with me to finally get the kinks out of the whole-genome sequencing aspect.

My wife and family have been such a supportive structure in the hardships of my graduate career. My wife has made this journey with me being with me from the first day of graduate school in being a light-hearted happy go lucky fellow to being exactly the same today, but with a Ph.D. I can't miss the man that had one of the biggest influences on who I am today as a person and scientist, Dr. Michael Allen. He had a full lab when I applied, but took the chance on me anyway. I sincerely thank you for this opportunity and I hope that I have fully met your expectations of what it means to be a graduate student and scientist.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	iii
LIST OF TABLES.....	viii
LIST OF FIGURES.....	ix
LIST OF ABBREVIATIONS	xi
CHAPTER 1 INTRODUCTION.....	1
1.1 Movement Away from Petroleum-Based to a Greener Economy.....	1
1.2 Introduction to <i>H. cannabinus</i>	2
1.3 Mechanisms of Retting.....	3
1.3.1 Traditional Retting.....	6
1.3.2 Chemical Retting.....	7
1.3.3 Enzymatic Retting.....	7
1.3.4 Augmented Retting.....	9
1.4 Metagenomics for Understanding Microbial Communities.....	10
1.4.1 Communities at the Gene Level.....	10
1.4.2 Benefits and Negatives of NGS.....	11
1.4.3 Shotgun Environment DNA Metagenomics.....	16
1.4.4 16S Phylogenetic Studies.....	16
1.5 Microbial Communities in Retting.....	19
1.6 General Experimental Design and Hypothesis.....	20
CHAPTER 2 ISOLATION OF PECTINASE PRODUCING BACTERIA.....	22

2.1	Introduction	22
2.2	Material and Methods	23
2.2.1	Enrichment Process	23
2.2.2	Assay for Detection of Degraded Pectin.....	23
2.2.3	Full-Length 16S Amplification of the Isolates	24
2.3	Results and Discussion	25
2.4	Future Work with Organisms.....	27
CHAPTER 3 MICROBIAL COMMUNITIES OF DIFFERENT RETTING TECHNIQUES		29
3.1	Introduction	29
3.2	Materials and Methods.....	30
3.2.1	Controls and Augmented Retting Conditions.....	30
3.2.2	Extraction of DNA from Plant-Associated Bacteria and Amplification of 16S V4 Region	33
3.2.3	Quantification and PCR Amplification	35
3.2.4	Ion Torrent Workflow	38
3.2.5	Data Analysis.....	41
3.2.6	Rarefaction Curves.....	42
3.2.7	16S TA clone library of Experiment Day 4	42
3.3	Results.....	43
3.3.1	Overall Microbial Diversity of Control and Experimental Retting Environments.....	46
3.3.2	“Natural” Microbial Retting – C2D4.....	46
3.3.3	Composition of the Plant-Associated Microbial Flora – C1D4.....	47

3.3.4	Augmented Microbial Retting with Pectinolytic Isolates – E1D4	47
3.3.5	Environment-Related Differences in Microbial Richness	48
3.3.6	TA Clone Library	50
3.4	Discussion.....	50
3.4.1	Firmicutes as a Dominant Phylum in Microbial Retting Communities.....	52
3.4.2	Nitrogen as a Possible Factor and Nitrogen-Fixing Clostridia.....	53
3.4.3	PCR Bias Associated with 16S rRNA Analysis	54
3.4.4	Rarefaction is Good, but Special Attention should be Given to Others ...	55
3.4.5	Final Thoughts.....	56
CHAPTER 4 TIME-COURSE STUDY ON THE EXPERIMENT ENVIRONMENT.....		57
4.1	Introduction	57
4.1.1	Time Course Study of Metagenomic Samples	57
4.2	Materials and Methods.....	58
4.2.1	Preparation of Kenaf.....	58
4.2.2	Setup of Time-Course Retting.....	58
4.3	Results.....	60
4.3.1	Overall Phylogenetic Diversity at the Phylum Level	60
4.3.2	Time-Course at the Family Level.....	60
4.3.3	Diversity Estimates of the Time-Course Retting Experiment	64
CHAPTER 5 DROPOUT STUDIES ON THE MICROBIAL RETTING ENVIRONMENT.....		66
5.1	Introduction	66
5.2	Materials and Methods.....	67

5.2.1	Long Amplicon Implementation and Update to Ion Torrent Workflow...	67
5.2.2	Extraction of DNA and Amplification of 16S V5 Region.....	67
5.2.3	Setup of High-Performance Computer Server using TALON and Mothur	69
5.2.4	Diversity Analysis of Data.....	73
5.3	Results.....	76
5.3.1	General Sequencing Results from the Ion Torrent Server.....	76
5.3.2	Overall Phylogenetic Data for Dropout Experiments	85
5.3.3	Diversity Estimates of Overall.....	90
CHAPTER 6 CONCLUSION		96
6.1	Overview	96
6.1.1	Drop Out Experiments in Relation to Previous Studies	97
6.1.2	Microbial Changes at the Community Level.....	97
6.2	Future Directions	99
6.2.1	Testing of endophytes	99
6.2.2	Whole Genome Sequencing of Isolates.....	99
6.3	Final Thoughts.....	101
APPENDIX A 16S SEQUENCES FROM SEQUENCING OF PECTINASE PRODUCING BACTERIA.....		103
APPENDIX B MOTHUR TALON SCRIPT. EXAMPLE SCRIPT FOR USE WITH THE TALON HPC		109
APPENDIX C BARCODE OLIGOS FILE REQUIRED FOR MOTHUR.....		117
APPENDIX D ION TORRENT BARCODED PRIMERS FOR LONG-AMPLICON 16S PCR		120
APPENDIX E SUMMARY.SEQS COMMAND OUTPUTS FROM MOTHUR.....		122
BIBLIOGRAPHY		132

LIST OF TABLES

	Page
Table 3.1. The general experimental design of the microbial diversity experiments.....	30
Table 3.2. Primers used in the experiments.....	31
Table 5.1. Table shows summary of the various barcodes and their subsequent reads and mean read lengths.	80
Table 5.2. Rarefaction values placed in a table for the ease of the reader.	91
Table 5.3 Inverse-Simpson values for all samples.....	91
Table 5.4 UniFrac analysis of phylogenetic tree.....	94

LIST OF FIGURES

		Page
Figure 1.1.	Composite board composed of <i>H. cannabinus</i>	4
Figure 1.2.	Unretted fresh fibers from <i>H. cannabinus</i>	5
Figure 1.3.	Composite board created by retted fibers of <i>H. cannabinus</i>	5
Figure 1.4.	Pictured are retted fibers of <i>H. cannabinus</i>	6
Figure 1.5.	An example of a chimeric sequence.	15
Figure 2.1.	Isolated organisms grown on YEP plates and treated with 1% cetrimide solution showing halos around the colonies indicating pectin degradation.....	24
Figure 2.2.	Phylogenetic tree of all organisms isolated from enrichments from <i>H. cannabinus</i> and <i>Phoenix dactylifera</i>	27
Figure 3.1.	General workflow for the batch/replicate experiments.	32
Figure 3.2.	Initial experimental design for the tested environments.....	33
Figure 3.3.	Phylogenetic compositions of different retting environments at day 4	44
Figure 3.4.	Phylogenetic composition at the Order level.	45
Figure 3.5.	Rarefaction curves.	49
Figure 3.6.	Phylogenetic analysis of Clostridium isolates.....	50
Figure 4.1.	Phylogenetic comparison of time-course study.	62
Figure 4.2.	Phylum	63
Figure 4.3.	Rarefaction curves set at 97% similarity indicate of species level comparisons..	64
Figure 5.1.	Figure shows the output heat map from the Ion Torrent PGM for the dropout experiments.	78
Figure 5.2.	General outputs from the Ion Torrent.....	79
Figure 5.3.	Phylum level assignments for all samples.	81

Figure 5.4.	Family level assignments of the <i>Bacillus</i> DP1 time-course dropout experiments.	82
Figure 5.5.	Family level assignments of the <i>Paenibacillus</i> DP2 time-course dropout experiments.	83
Figure 5.6.	Family level assignments of the <i>Bacillus</i> K1 time-course dropout experiments..	84
Figure 5.7.	Rarefaction curves of <i>Bacillus</i> DP1 samples of day 2 and 4	87
Figure 5.8.	Rarefaction curves of <i>Paenibacillus</i> DP2 samples of day 2 and 4.....	88
Figure 5.9.	Rarefaction curves of <i>Bacillus</i> K1 samples of Day 2 and 4	89
Figure 5.10.	Principle coordinate analysis plots for the drop out tests.....	93

LIST OF ABBREVIATIONS

bp	Basepair
°C	Degrees Celsius
ddH ₂ O	Distilled deionized water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
EDTA	Ethylediamine tetraacetic acid
ePCR	Emulsion polymerase chain reaction
ES	OneTouch Enrichment System
EtBr	Ethidium bromide
EtOH	Ethanol
ISPs	Ion sphere particles
Kb	Kilobase
LB	Lysogeny broth
M	Molar
m	Minutes
Mb	Megabase
MG-RAST	Metagenomic rapid annotations using subsystems technology
mΩ	Milli ohm
mL	Milliliter
mM	Millimolar

μg	Microgram
μL	Microliter
ng	Nanogram
nM	Nanomolar
NaOH	Sodium hydroxide
NaOCl	Sodium hypochlorite
NGS	Next-generation sequencing
NIH	National Institutes of Health
NSF	National Science Foundation
N	Normal
OTU	Operational taxonomic units
PCR	Polymerase chain reaction
PGM	Ion torrent personal genome machine
pmol	Picomole
PSI	Pounds per square inch
rDNA	Ribosomal DNA
RDP	Ribosomal database project
rpm	Rotations per minute
rRNA	Ribosomal RNA
s	Seconds
sff	Standard flowgram format
TAE	Tris acetate EDTA

UNT	University of North Texas
USDA	US Department of Agriculture
YEP	Yeast extract pectin

CHAPTER 1

INTRODUCTION

1.1 Movement Away from Petroleum-Based to a Greener Economy

Transitioning to a green, bio-based economy necessitates changing more than just our sources of petroleum, and it encompasses far more than merely energy sources, e.g. wind, solar, geothermal, petroleum, nuclear or otherwise. Recent socio-political turmoil implores a movement away from petroleum-based products toward more renewable alternatives. In addition to hydrocarbon fuels, plastics along with commodity and specialty compounds must be replaced with sustainable alternatives. An example is plastic, traditionally a petroleum by-product, now has a plant-derived alternative, PlantBottle™, Coca Cola Corp., where plant sugars are converted to polyethylene terephthalate (PET) plastic (1). This research aims to target a specific area of green materials.

Composites are a material typically made up of at least two different types, which yield a difference in properties when used in conjunction. Fiberglass and carbon fiber materials are examples of composites. Fiberglass materials not only include petroleum-based epoxy or resin, but requires a large energy input when producing the initial glass fibers to be incorporated. Biocomposites are composites composed of green materials, e.g. switching out the fiberglass for plant fibers and using some biologically-derived resins (Figure 1.1).

The research focus of the current project assesses the viability of fibers found in the plant *Hibiscus cannabinus*, commonly known as kenaf, for use as an industrial biocomposite. The traditional plant-fiber extraction process is known as retting, during which pectin, lignocellulose, and other heteropolysaccharides are removed from the bast fiber (or also known

simply as plant fiber). This can consist of various mechanisms, including enzymatically, chemically, and microbially-driven retting processes. This project investigated and evaluated the retting process via exploiting natural and enriched bacterial communities using modern molecular techniques.

Applications of natural fibers in advanced and high-value composite materials depend upon generating high quality fibers in conjunction with industrial-scale production. Achieving this balance during the retting process is a challenge that biotechnology and microbiology are poised to address. A brief background on the plant, the four main approaches to retting (traditional, chemical, enzymatic, and augmented), and the relevance of bioinformatics and microbiology to retting are discussed in detail below.

1.2 Introduction to *H. cannabinus*

Kenaf (*H. cannabinus*) is a fiber-bearing plant and has been explored by the USDA as a potential cash crop for many regions of the United States. *H. cannabinus* plant belongs to the Malvaceae family and is a warm season annual fiber crop closely related to cotton, *Gossypium hirsutum* L., Malavaceae, the latter being a highly successful cash crop through the southeastern states. Kenaf is a fast-growing plant, capable of attaining heights ranging from 1 – 4 meters in a single growing season. The woody stem of the kenaf plant can be separated into two main parts of the plant, the bast and the core. An example of the bast fiber bundles is shown in Figure 1.2. The core component of the kenaf plant is composed largely of cellulose, hemicellulose, and xylan (2). Retting of kenaf typically involves stripping the bark from the harvested material, and then soaking the former in water (*e.g.* a pond or river) allowing

endogenous microbes to break down the heteropolysaccharides binding the fiber bundles to facilitate their separation (3). The resultant plant fiber finds use in traditional areas such as cordage and fabrics and have been in use for these products for upwards of six millennia (4). Research in the United States explored kenaf utilization to address increased needs for cordage material for the war effort during World War II. These efforts eventually led to initiatives in increasing crop yield (5). More recent applications include their incorporation into “green” composites for the auto and aerospace industries (6–10) and is shown in Figure 1.1 and 1.3. The movement towards increased use of plant fibers such as kenaf as a replacement for interior components is accelerating. The biocomposite materials have been incorporated into vehicles produced by BMW and Ford, and also aircraft, where weight reduction is an essential goal (11).

1.3 Mechanisms of Retting

Traditional retting refers to soaking harvested material in bodies of water (lakes or rivers) and allows the natural microbial communities to release the fibers. Mechanical separation of fibers and chemical retting using high pH or other harsh treatments are also employed. While these industrial processes are easier to scale, they result in poor quality fibers (9, 12). Investigations using enzymatic retting have reported high quality fibers, but the process is likely cost-prohibitive for large-scale applications due to the high cost of purified pectinase enzymes (13, 14). Traditional microbial retting therefore offers the best chance for large-scale production of high quality fibers. However, little is known of the microbes involved or the process dynamics that occur during traditional retting. Knowledge of these areas may lead to improved process design and/or microbial assemblages that improve industrial-scale

production. The four types of retting have been described and will be further discussed, specifically, traditional microbial, chemical, enzymatic, and augmented retting. Specifically, an example of retted kenaf fibers are shown in Figure 1.4.

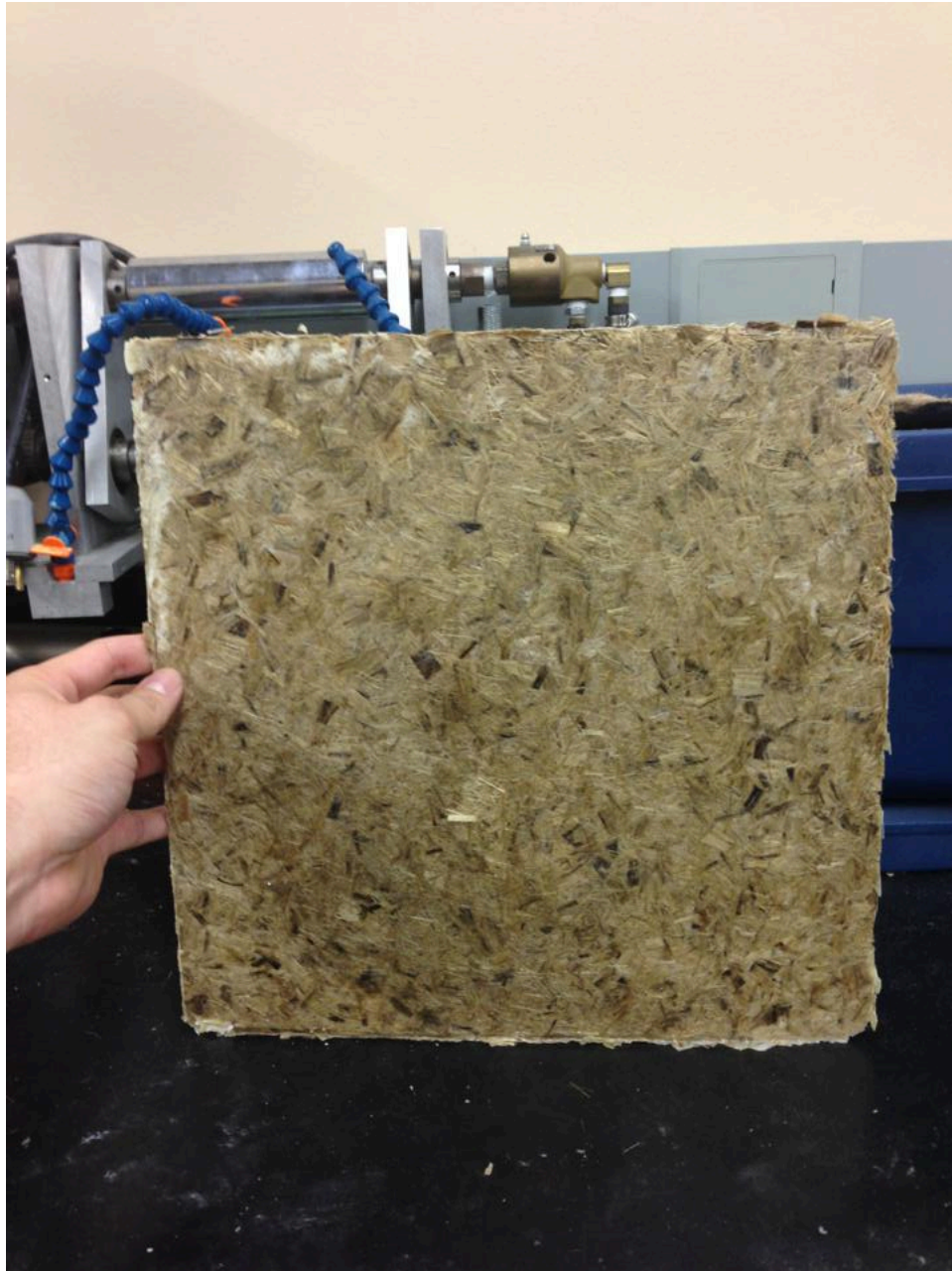


Figure 1.1. Composite board composed of *H. cannabinus*.



Figure 1.2. Unretted fresh fibers from *H. cannabinus*. These fibers have been removed from the core and are pre-retted material.

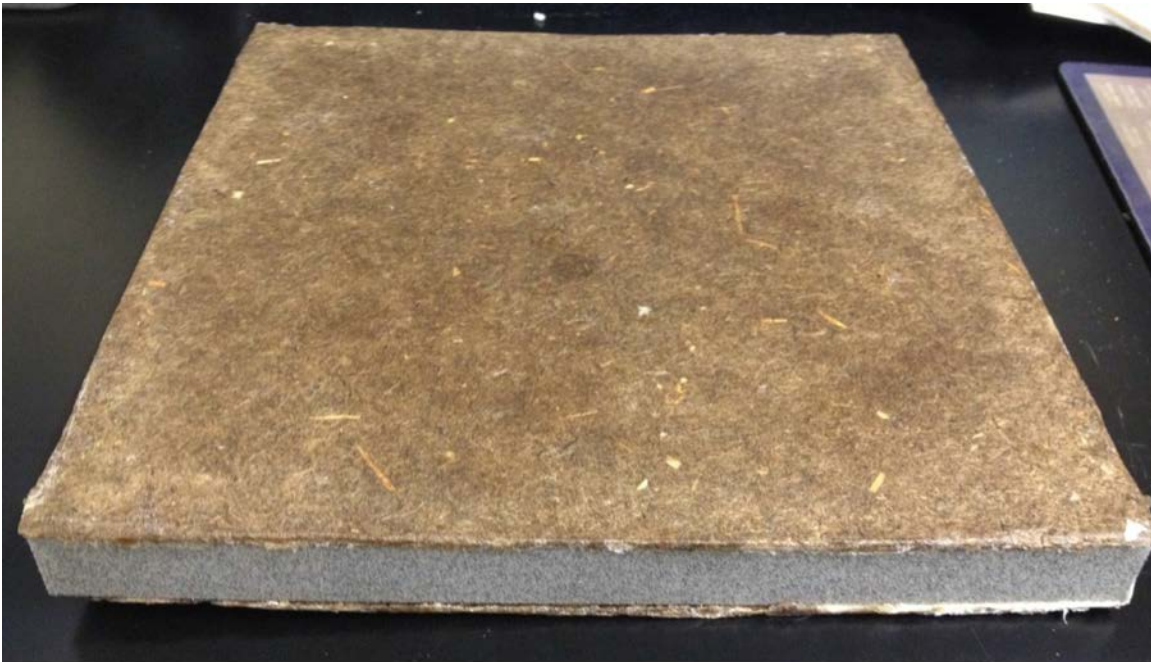


Figure 1.3. An example of a composite board created by retted fibers of *H. cannabinus*.



Figure 1.4. Pictured are retted fibers of *H. cannabinus*.

1.3.1 Traditional Retting

The traditional retting process is a system that has been used for thousands of years and is a simple way of removing fibers from the heteropolysaccharide matrix, which holds them together. During this process plant material is, often but not always, physically removed from the core (stalk), and submerged into bodies of water, e.g. lakes, lagoons, rivers, or large tanks, where microorganisms found naturally on the plant and in the water, break down the pectin that binds the plant fibers (15).

Multiple variables can affect the final quality of the fiber, including microbial burden and type, climate, temperature, plant traits, and how all these factors influence each other. These networks of interactions all contribute to a process that is crude and uncontrolled, which can dramatically affect the final fiber quality and possible incorporation into the composite. Over-

retting or under-retting can also dramatically influence the final biocomposite quality and must be addressed. Jute fibers, for example, have been shown to be subject to over-retting, which leads to weakened fibers due to increased loss of polysaccharides, and under-retting results in the fiber not binding as efficiently to the matrix, thus producing a suboptimal composite (16).

1.3.2 Chemical Retting

Chemical retting involves a strong alkaline treatment; often a saturated solution of a strong base such as sodium hydroxide (17–19). Advantages of this type of treatment include reproducibility, rate (on the order of hours versus days/weeks with other retting methods), ease of scaling, and comparatively small costs to implement. But the alkaline treatment of bast fibers is a harsh reaction and unlike the biological forms of retting, indiscriminately removes reactive polysaccharides from the fiber, resulting in a smoother surface on the produced fibers, which has been shown to deleteriously impact tensile properties (17). Supporting this, recent research suggests that structural weakening and decreased porosity of *H. cannabinus* fibers occurs with NaOH treatment (20).

1.3.3 Enzymatic Retting

With the advancements in biotechnology made in the last decade, there has been an increase in accessibility, lowered cost, and universal recognition of enzymes for industrial and business applications. Enzymes can target precise bonds coupling organic molecules. Unlike the previously described application of a strong base, which acts indiscriminately to plant material, enzymes are highly specific. Since enzymes are directed to a specific substrate, numerous

unwanted side reactions common with chemical retting are avoided. Additionally, commercial enzymes are typically extracted from mesophilic microorganisms, which means they thrive under ambient conditions, adding another benefit to the enzymatic technique over chemical retting protocols that often require extreme temperatures, pH, and/or pressure. Enzymatic retting has proven successful in small-scale lab settings, but has not been applied on an industrial scale (21).

A variety of fiber sources have been used as case studies testing the benefits of enzymatic retting. These have typically involved the enzyme pectinase. In hemp fibers, enzymatic retting was found to impart increased tensile strength, elasticity, and flexural strength, which are all benefits to the overall performance of composites (22). Other fibers have yielded mixed results with regards to mechanical properties imparted by enzymatic retting (23, 24).

An aspect that has not been mentioned previously, but has broad impacts on the quality of the fiber is the impact of the starting plant material, which can have dramatic effects on the final fiber quality. Consideration must be given to growing conditions, species variation, ecotypes, as well as particular plant sections (e.g. top vs. base). The plant sections in regards to the top vs. base encompass old vs. new growth. The new growth that is towards the top of the kenaf plant has less complex heteropolysaccharides binding the fibers resulting in easier removal as compared to the “older” sections towards the bottom of the plant. Limiting fluctuations in these parameters may also increase the efficacy of the enzymatic retting process facilitating greater process optimization. Variation among commercial enzyme preparations being produced is a reality, as most are synthesized by a variety of biotech companies and any

alterations or contaminants in their structure can result in altered or ineffective activity.

Isolation and identification of new pectinase enzymes continues to be a vigorous field of study (25–32) and new enzymatic formulations and procedures for their manufacture may eventually make it an economically viable answer for not only retting applications, but also give it an integral role in high quality paper production, tea and coffee fermentation, botanical oil extractions, and treatment of pectic waste water (33).

1.3.4 Augmented Retting

Augmented retting approaches provide a bridge between the uncontrolled parameters inherent in the natural retting process by utilizing the processes of enzymatic retting reactions. Supplementing the retting solution with bacterial isolates that produce pectinases that target the plant cellular structure should increase the rate at which the process occurs. Augmented retting will, ideally, exploit the benefits from traditional and enzymatic retting. This is accomplished through the introduction of a bacterial inoculum into the retting solution. The foremost advantage of such a methodology is that the organism(s)/bacterial isolate(s) employed excrete the enzymes for retting *in situ* using the degraded pectins and hemicellulosic constituents as their energy (carbon) sources negating the need for introduction of purified enzymes. This methodology results in the benefits of enzymatic retting, albeit with slightly less control, but at a fraction of the cost since the only added cost are the bacteria growth requirements.

1.4 Metagenomics for Understanding Microbial Communities

1.4.1 Communities at the Gene Level

With the exponential increases in data production from large-scale sequencing platforms and the concomitant decrease in cost, even outpacing Moore's law, it has become economically viable to analyze a multitude of biologically relevant questions with the help of next-generation sequencing. This approach is taken due to the inherent difficulty in analyzing bacterial communities using traditional methods, especially since large numbers of bacteria cannot be cultivated in a traditional microbiological sense and must be analyzed through the use of DNA sequences (34–36). Modern analysis of prokaryotic phylogenetic data by means of the 16S rRNA gene was pioneered by Carl Woese and George Fox (37). This work introduced a new taxonomic system which included a third branch of life in what would come to be known as the domain Archaea; it was an unorthodox means that utilized a molecular technique to change the way phylogenetics and taxonomy were accomplished (38, 39). The 16S rRNA sequence was chosen for phylogenies due to slow rates of evolution in the gene, which is under high-selective pressures and its ubiquity among prokaryotes. All life contains some semblance of a 16S rRNA gene, which provides a unique target for phylogenetic studies. The 16S rRNA is a part of the 30S small ribosomal subunit of prokaryotic ribosomes. The ribosome is a multicomponent holoenzyme necessary for protein synthesis. The 16S rRNA encodes an anti-Shine-Dalgarno sequence, which directs the small ribosomal subunit to the corresponding sequence 5' of the start codon of the mRNA and is involved in regulating mRNA translation efficiency (40–43). The 16S rRNA gene is a 1.5 kb sequence of DNA, which is short enough for molecular characterization by Sanger sequencing, but contains conserved regions which allow

for short fragments to be amplified for phylogenetic analysis (44). The full-length 16S rRNA can be analyzed by means of BLAST using the typical nucleotide collections (NCBI) (45). Additionally, there are many specific microbial 16S databases that are specifically tailored to quick analysis of small fragments of the 16S gene (45–48). These other databases, i.e. RDP and Greengenes, allow large-scale 16S metagenomic datasets to be analyzed quickly and efficiently.

1.4.2 Benefits and Negatives of NGS

There is some loss of phylogenetic discriminatory power due to reduced information in analyses using smaller gene sequence fragments compared with the full-length 16S gene.

However, this is tolerated since next-generation sequencing platforms typically produce read-lengths much smaller (100-400 bp) than that of traditional Sanger sequence (~700 bp read length), but with dramatically higher output. The trade-offs between next-generation sequencing (NGS) and Sanger sequencing is partially ameliorated by the ability to target amplicons to highly informative regions. One caveat to the strict limitation of only sequencing small fragments by NGS is that the technology is rapidly improving as evidenced by the increases in read lengths since the introduction of the various platforms, e.g. Ion Torrent 100-400 bp, Illumina 150-600 bp and Roche 454 GS FLX+ approaching 1000 bp (49, 50).

The approach of using 16S small fragments as a tag for bacterial organisms seeks to understand the underlying microbial communities without utilizing a culture-dependent techniques. This is accomplished by extraction of DNA from the sample, PCR amplification of informative regions of the 16S gene, sequencing on a massively parallel NGS platform, and followed by analysis with increasingly advanced high-performance computing resources. This

results in millions to billions of reads for a single sample per run, which can elucidate the phylogenetic structure and to some extent, the function of complex communities. This approach is taken here because current bioinformatics and NGS platforms cannot readily handle total metagenomics of complex environments, i.e. bulk DNA extraction, fragmentation, and sequencing of total DNA, due to the large complexities associated with even modest levels of bacterial diversity.

1.4.2.1 DNA Extraction

As with any PCR associated experiment there are aspects such as primer specificity, DNA extraction, chimeras, and other associated bias that must be addressed. General workflow within any metagenomic analysis starts with a DNA extraction performed prior to DNA sequencing. DNA extraction is most often performed using either alkaline lysis or a mechanical bead-beating method. The former utilizes a high pH to disrupt and lyse the bacterial cells while the mechanical method uses a shaking in the presence of hard, inert, inorganic particles (e.g. silica sand) to physically shear the bacteria. There is an assumption that all bacterial cells are lysed equally, but this is not the case. There are two types of cell walls in bacteria: Gram-positive cells have a thick peptidoglycan wall and Gram-negatives have a cell wall consisting of a thin inner-peptidoglycan and more complex outer membrane. The general understanding of the lysis efficiencies of the different cell wall types are that Gram negative cells are easier to lyse via the alkaline method while Gram positive cells are more difficult. Endospores are also present in some gram-positive bacteria, which further complicate DNA extraction methods. Endospores are dormant, non-reproductive states containing a variety of associated proteins

that are designed to protect the DNA and organism. This specifically creates a problem to lysis methods because these endospores are extremely hardy. In addition to lysis methods, storage and archiving actions can have effects on ultimate DNA extraction. Repeated free-thaw cycles can lyse cells, but does not extend to endospore-forming bacteria, which are inherently resistant (51).

Even within various DNA extraction kits there are differences associated with total abundance of DNA extracted that can lead to errors associated with ribosomal copy number and microbial community composition. There are differences in the types of bacteria extracted, Feinstein *et al.* found that the relative abundances of sequences from rarely cultivated groups including Acidobacteria, Gemmatimonades, and Verrucomicrobia were higher in the first extraction, but that the reverse was true for Proteobacteria and Actinobacteria. Many of these associated bias were alleviated by doing triplicate extractions and pooling together for downstream analysis (52).

1.4.2.2 PCR Associated Errors

Multi-template PCR is another type of bias to be addressed. Potential variations occur within multi-template PCR. 16S associated PCR utilizes degenerate primers, that is primer-associated nucleotides that have a variety of binding efficiencies to other nucleotides, which can result in variations in amplification of template DNA (53). There are associated biases with annealing of primers to DNA templates in which increased cycles do not increase copy number and the reannealing aspect inhibits the template-primer interaction (54). Errors associated with the misincorporation of bases during amplification and sequencing errors can further hinder

processing of phylogenetic data. Researchers have sought to determine the “best” DNA polymerase for use in amplicon-based 16S phylogenetic studies, but there has not been a definitive study (55, 56). Specifically, high fidelity is important in the final sequencing data that is to be analyzed, but issues with proper annealing temperatures associated with different polymerases have not been taken into account in the prior studies (57, 58).

1.4.2.3 Chimeras Associated with PCR

Chimeras are an inherent problem with PCR-associated analysis. A chimera is a fusion DNA molecule that originates from two different amplicons in a PCR (59–63) (see Figure 1.5). During the annealing and extension phases of PCR, instead of the primers attaching and amplifying template DNA the two amplicons act as primers to each other and amplify. This results in a combined chimeric amplicon that is now a conglomerate of the two initially separate amplicons. The issue with chimeras, in the context of 16S phylogenetic analysis, is that it artificially inflates the diversity estimates of a microbial community (64). There are no specific ways to reduce chimeras during the amplification, sans lowering the overall number of amplification cycles, but there are multiple examples of software (65) that utilize a variety of techniques to reduce the number of chimeras in the final quality-controlled data. DECIPHER is a search-based approach to chimera detection in 16S rRNA studies (66). Chimera Slayer is another example in which it uses a database of known microbial genes as well as a test chimeric batch (67). The presence of chimeras has potential effects of making comparisons between communities difficult as well as inflating estimates of diversity which makes reduction of them essential in the final conclusions drawn from 16S phylogenetic data.

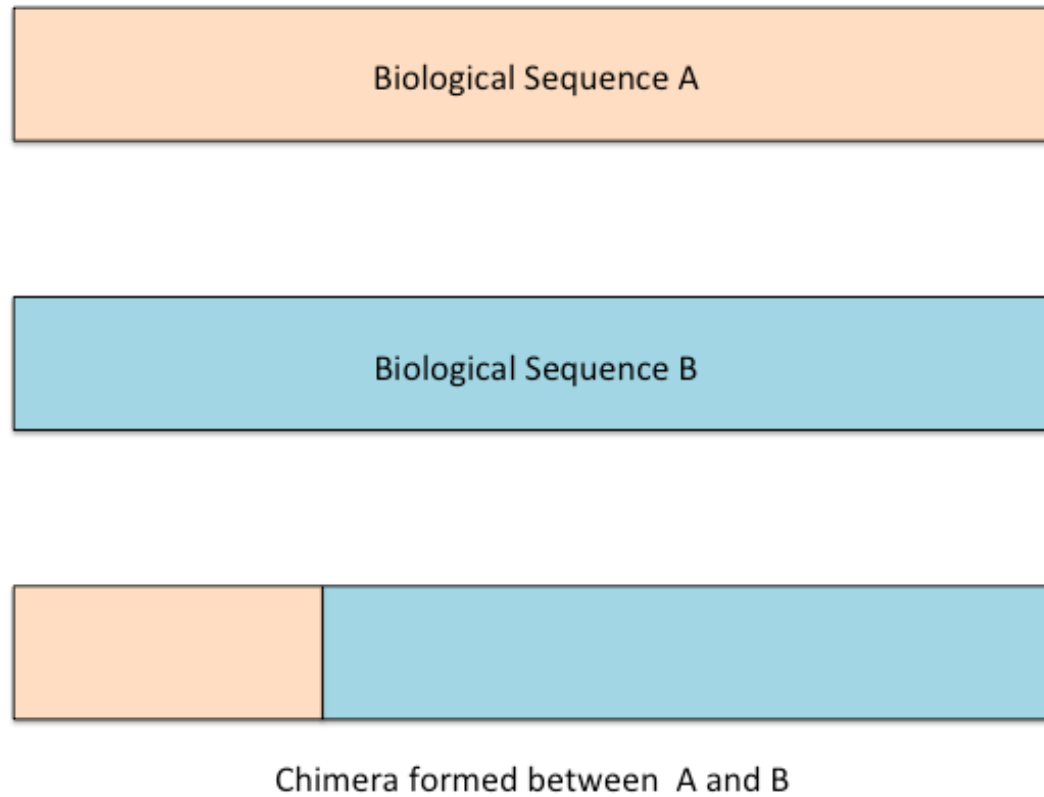


Figure 1.5. An example of a chimeric sequence. Two sequences are merged into one sequence producing a chimera.

1.4.3 Shotgun Environment DNA Metagenomics

Shotgun environmental DNA metagenomics, i.e. sequencing all environmental DNA, has been accomplished within very simple environments, e.g. acid mine drainage, but a more readily accessible technique involves using 16S PCR amplicons to identify complex environments thereby simplifying the highly complex bioinformatics required for whole environmental DNA metagenomics (68). As mentioned before, there are large inherent complexities with analyzing whole DNA from an environment. First and foremost, there is an issue of databases in that a comprehensive database is typically lacking. The computational pipelines for analysis are coming and some are in existence like MG-RAST (69). MG-RAST is a functional open source web application server that helps automate phylogenetic and functional analysis of metagenomes. The system relies on clusters of nodes, which allow rapid annotation of 16S and metagenomic data. Additionally, there is a move within the system to allow more data to be open source in that the priority of analysis of data is based on a timeline of when data will be available to the public. Highest priority goes to those that immediately make the data available for other researchers to use. Pertaining to the actual system now, it automatically produces functional assignments to the inserted DNA/mRNA sequences within the metagenomes by performing sequence comparisons to databases in both nucleotide and amino-acid levels. This results in both a phylogenetic/taxonomic match with organisms and also functional assignments to sequences by means of nucleotide and protein alignments (70).

1.4.4 16S Phylogenetic Studies

Each tag, or small 16S fragment, is representative of one organism. Thus, there is an

assumption that each bacterial organism, on average, has one 16S rRNA gene, which when amplified, ignoring associated PCR biases, results in one organism per 16S tag. Instead of the whole environmental DNA metagenomics, i.e. sequencing all environmental DNA, there is an “enrichment” process, the small 16S fragment is extracted from the “noise” or environmental DNA by means of primers and PCR (71, 72). Post clean-up of the extracted and enriched DNA, what remains is the small 16S fragment, which can then be sequenced via an NGS platform.

This metagenomic approach outlined above has been successfully applied to a variety of projects, including current studies of the *Amblyomma americanum* (lone star tick) microbiome, the cloacal microbiomes of captive-bred and wild Attwater’s prairie chicken, *Tympanuchus cupido attwateri*, and larger projects like the Human Microbiome Project.

In the lone star tick, microbial survivorship and diversity were moderately reliant upon environmental factors and the succession of blood feeding, molting, and aging (73). Their findings indicate that the continuance and/or advent of disease-causing bacteria may be dependent in part on temporal changes in the microbial community of the tick microbiome. This research has broader impacts on tracking of pathogens in associated vectors, with possible environmental factors, including increased incidences of pathogens correlated with drought.

Similarly, the Attwater’s prairie chicken research showed that diet, environment, rearing and age, all have an effect on microbial community (74). Antibiotics, specifically enrofloxacin, do not appear to cause prolonged significant changes in the microbial community structure of the Attwater’s prairie chicken. Additionally, microbial communities are highly variable between individuals even when controlling for diet and environment. This research provided at least some insight into a critically endangered species (74, 75).

The Human Microbiome Project is a large-scale research project initiated by the NIH in order to understand the microbiomes of humans, including what organisms comprise a normal versus diseased state, and normal patterns of biogeography, ecology, metabolism, and function (76). The work spawned large interest in the functional roles of microbiomes in humans, including identification of novel taxa, detecting alterations in bacterial communities due to low-income and disease states, and association of microbes with reflux disorders and esophageal adenocarcinoma (77–79). Microbiome studies provide the means and vast amounts of information of the underlying taxa, but also facilitate inference of metabolic capacity that culture-dependent experiments cannot accomplish.

Current understanding of the microbial retting process will be expanded upon by taking the next step in the innovation process: moving away from simple culture-dependent studies and small-scale clone libraries toward applying the 16S metagenomic techniques to the microbial retting communities. This approach has already been applied in a variety of areas such as the Deep Water Horizon oil spill, which was one of the largest environmental disasters in the United States. The 16S analysis of microbial communities associated with the surface sediments in 64 sites around the oil spill provided multiple uncultured bacteria, specifically Gammaproteobacteria and a *Colwellia* species, which had large similarities associated with the deep-sea plum (80). Specifically, the data showed that that the indigenous sediment communities provide ways for bioremediation of oil to occur within the system. Understanding the microbial retting communities can at least begin to be understood at the organismal level and inferences can be made from them with regards to how retting occurs.

1.5 Microbial Communities in Retting

In spite of many decades of research, surprisingly little is known of the microbial community involved in retting. A number of studies conducted over the last half century have focused on culturable isolates, most often identifying *Bacillus*, *Clostridium*, and *Pseudomonas* spp. (81–84). In several studies, pectinolytic retting isolates grown separately and re-introduced to the retting vats have shown the ability to accelerate the process rate, while also improving fiber quality (85–87). Reports employing more modern molecular techniques to retting, however, are sparse and little is known of the actual microbes involved, the dynamic behavior of the community over time, or the suitability of microbial amendments to the retting solution.

Numerous pectinase-producing bacteria have also been described, particularly members of the genera *Bacillus* (88–91), *Paenibacillus* (26, 27, 92), and *Clostridium* (93–95). Members of these bacterial genera have been isolated from retting solutions, as well as inoculated into retting systems for process optimization. However, the fate of these bacteria in retting solutions remains unclear.

The general view of the retting process involves the colonization by aerobic bacterial such as *Bacillus* and *Paenibacillus* during the initial phase of retting, followed by the subsequent displacement by *Clostridium* spp. as conditions become anaerobic (84). In another study, the retting solution of jute was analyzed by creating a 16S clone library and performing amplified ribosomal DNA restriction analysis (ARDRA) on the fragments (96). Their results indicated large amounts of Proteobacteria (41%), in addition to Firmicutes (7%), Verrucomicrobia (5%), Acidobacteria (5%), Chlorobiales (5%), and Actinobacteria (2%) at two different jute-retting locations of Krishnanagar and Barrackpore in India.

1.6 General Experimental Design and Hypothesis

This project aims to increase our understanding of the underlying microbial communities involved in kenaf retting. Understanding what is occurring during the retting process is essential to obtaining a better-retted fiber, while also optimizing for low production cost and balancing the benefits of the natural retting environment. This study took a multi-part approach:

1. Classical microbiology approach: Entails isolation of organisms and identification of favorable qualities
2. Phylogenetic approach: Understanding the underlying microbial community, determining what it is, how it can be changed, and determining whether nutrients can further change the community.

Classical microbiological techniques use isolation of pure cultures, followed by manipulations and studies of single organisms operating outside of their respective natural environments. This research harnessed the classical microbiological approaches by specifically targeting bacteria capable of producing pectinase; the action of which should liberate the kenaf fiber when produced in the retting solution. Organisms that produce large amounts of pectinase should in theory, process the plant fibers more quickly, thus introducing an organism in large concentrations would potentially increase the rate at which retting occurs. The phylogenetic approach to be applied in this work takes a known technique, 16S rRNA gene sequencing, and determines what organisms are present in the retting milieu, and derive understanding of the process through comparisons between controls and experimental/augmented retting conditions. This research yielded data useful for optimization

of the retting process, and ultimately the production of stronger fibers at faster rates. These fibers in turn were incorporated into a bio-composite material for green building applications. Finally, the results described here have increased our understanding of the basic concepts of regulating the structure and function of microbial communities, with far reaching implications for biotechnology and medicine.

CHAPTER 2

ISOLATION OF PECTINASE PRODUCING BACTERIA*

2.1 Introduction

Classical microbiology practice hinges on the ability to isolate an organism in order to study. This has always been a cornerstone in microbiology beginning with the initial experiments conducted in the labs of Robert Koch with his isolation of causative agents of tuberculosis, cholera, and anthrax through the use of selective and differential media. A pure culture was one of his greatest contributions to microbiology and the medical field in that an organism could be grown, isolated, and streaked onto a petri dish so that it could be further studied (97). Isolation of an organism for study was and is still an important aspect of the field of microbiology.

As mentioned previously, there are estimates of 90-99% of bacterial organisms are unculturable, and while this research intends to also pursue culture independent techniques, it is still advantageous to search and isolate bacteria that provide favorable characteristics that can potentially enhance qualities and speed of processing of the plant material (98). The enrichment process allows the selection of pectinase-producing bacteria and a colorimetric assay allows visualization of the degradation of pectin in the media. The experiments will pursue the concept that on the plant material, typically the “starter” in the microbial retting process, will have organisms that produce pectinase. Isolating and understanding these

*Parts of this chapter have been previously published, either in part or in full, from D. K. Visi, N. D'Souza, B. G. Ayre, C. L. Webber III, and M. S. Allen, “Investigation of the bacterial retting community of kenaf (*Hibiscus cannabinus*) under different conditions using next-generation semiconductor sequencing,” *Journal of Industrial Microbiology and Biotechnology*, 40, 465-475 (2013). Reproduced with permission from Springer Publishing, Ltd.

organisms are an essential component for not only understanding the retting process, but also further enhancement of the downstream process by introducing an inoculum to the initial reaction.

2.2 Material and Methods

2.2.1 Enrichment Process

The enrichment process was accomplished by gradual growth of bacterial specifically able to degrade pectic substances. Yeast extract pectin media were prepared in 1-L batches using the following parameters: 5 g yeast extract, 5 g pectin, pH 7.2, and adjusted with 1 N NaOH as adapted from Tewari et al 2002 (28). Autoclave conditions were 15 PSI, 121°C, and 15 m. 5 g of plant material, (i.e. kenaf or date palm fronds), were placed into 250 mL of sterilized YEP media in a 500 mL Erlenmeyer flask and placed in a shaking incubator at 200 rpm and 37°C. After 24 hours, 2 mL of enriched culture were inoculated into 250 mL fresh YEP. This enrichment process was repeated three more times. The solution was serially diluted to 10^{-4} and subsequent dilutions were plated on YEP 1% agar plates and incubated for 48 hours at 37°C. This was supplemented with plates grown in anaerobic conditions at ambient temperatures for 48 hours. Colonies were picked based on morphological characteristics and then streaked onto a fresh YEP plate using the four-quadrant technique for isolation of pure culture and incubated for 48 hours at 37°C.

2.2.2 Assay for Detection of Degraded Pectin

Detection of degraded pectin was an integral component of determining whether an

isolate was in fact producing pectinase. The 1% cetrимide solution when used on a YEP plate binds to pectic substances, specifically long chain polygalacturonic acid, resulting in an opaque background. Clear areas represent zones of pectin degradation.

A 1% (weight/volume) cetrимide solution (25) was prepared. Additional YEP plates were prepared with the isolated colonies (replicates of the original), were streaked for isolation and allowed to incubate for a period of 48 hours. 1 mL of the 1% cetrимide solution was used on each plate. After a one-hour incubation at 37°C the production of pectinase was confirmed by zones of clearing surrounding the colonies. Examples of the assay are shown in figure 2.1 with *Bacillus* DP1 and *Paenibacillus* DP2.

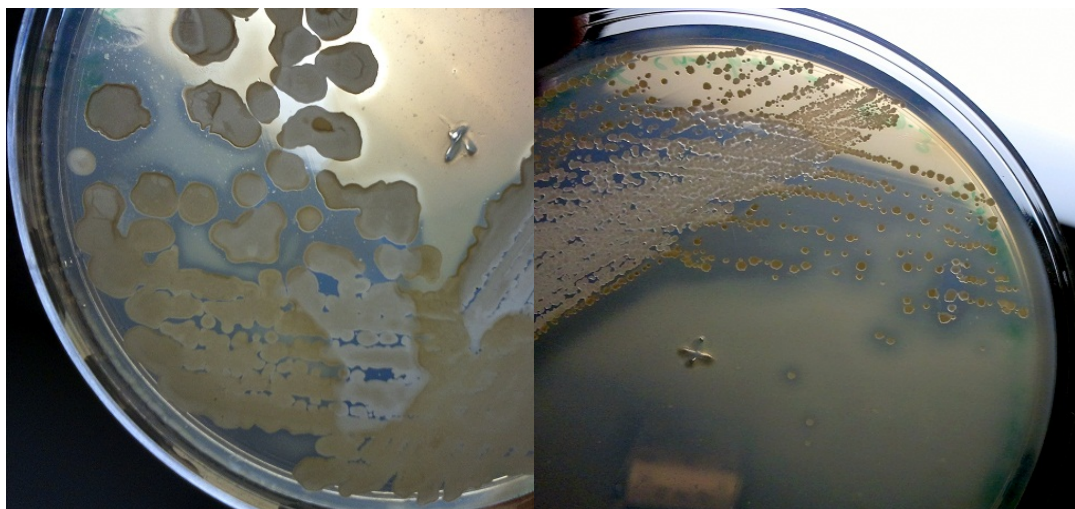


Figure 2.1. Isolated organisms grown on YEP plates and treated with 1% cetrимide solution showing halos around the colonies indicating pectin degradation. Left plate shows *Bacillus* DP1 and right is *Paenibacillus* DP2.

2.2.3 Full-Length 16S Amplification of the Isolates

After confirmation of pectinase production, the 16S gene sequences of each organism were amplified by colony PCR using universal bacterial primers 27F and 1492R (99). The colonies were picked with sterilized toothpicks and inserting the bacterial matter directly into

the PCR 16S amplification. The PCR was set up as the following: illustra PuReTaq Read-To-Go PCR Beads (GE Healthcare Life Sciences, Pittsburgh, PA), 250 nM of 27Fwd, 250 nM of 1492Rev, and autoclaved ddH₂O up to the manufacture's recommended amount (25 µL). The illustra PCR Beads provide the buffer, *Taq* polymerase, and nucleotides for the reaction. PCR parameters were 95°C for 5 min followed by 30 cycles of denaturing at 95°C for 15 s, annealing at 56°C for 30 s, and extension at 72°C for 90 s using a 48 Well MJ Mini Thermal Cycler (Bio-Rad Laboratories, Inc. Hercules, CA). Each PCR was confirmed by gel electrophoresis for a ~1.5 KB product. The gel electrophoresis was performed using TAE buffer, 1% agarose gel, stained using EtBr, ran for 30 m at 120 V, and imaged to visualize the bands. The PCR was cleaned using a MO BIO Ultra clean 15 DNA purification kit (MO BIO Laboratories Inc., Carlsbad, CA) using manufacturer's instructions and sent for sequencing to Eurofins MWG Operon, Huntsville, AL. After bidirectional sequencing, the two fragments for were assembled into one ~1.5 kb fragment and compared to the NCBI database using BLAST (45). Additionally, the fragments were aligned using MEGA 5 (100), and placed into a maximum-likelihood phylogenetic tree for comparison.

2.3 Results and Discussion

Classical microbiology requires isolation of the organisms and then the eventual study of the said organism in pure culture. There are estimations that only of 1-5% of known bacterial organisms are able to be cultivated, but nonetheless the aim of this study was to produce organism(s) that produce pectinase. This study began with the isolation of microorganisms that produce pectinase, which was performed by using subsequent enrichments of pectin media

and eventual plating and identification. Universal primers were used to target the 16S region of the bacteria, which gives a phenetic match to the specific genus and possibly species.

Five organisms were isolated from the aerobic enrichments from the Date Palm fronds and three organisms isolated from *H. cannabinus*. The full-length 16S gene sequences are shown in Appendix I. The nomenclature selected for these organisms were made by using the highest genus BLAST match, in all cases either *Bacillus* or *Paenibacillus*, followed by DP or K which indicated the source of the plant material (“DP” is date palm fronds and “K” is kenaf) and then a numerical indicator. For example, *Bacillus* DP1 indicates that it came from the date palm fronds (DP) and that it was the first selected for characterization through 16S in the group of date palm isolations.

Multiple organisms were used to “root” the phylogenetic tree. *Bacillus thuringiensis* and *Bacillus amyloliquefaciens* were used as similar organism to the *Bacillus* organism that were isolated from the date palm and kenaf plants. This was done because the genus *Bacillus* is inherently diverse in terms of phylogeny and we wanted the tree to help separate out the organisms that were isolated. Two isolates were used to help distinguish the *Paenibacillus* isolates, i.e. *Paenibacillus polymyxa* and *Paenibacillus jamilae*. The organisms isolated all clustered within the *P. jamilae*, but still maintaining within the overall cluster with the genus *Paenibacillus*. As mentioned previously the genus *Paenibacillus* has been shown to produce numerous enzymes that have activity against polysaccharides. The last two known organisms are *Clostridium hveragerdense* and *Escherichia coli* were used as out groups. This also confirmed the consistency of the phylogenetic tree in that *Paenibacillus* and *Bacillus* are both within the phylum Firmicutes. Within the context of phylogeny this means that the clustered

organism should be more similar to *Clostridium* than to *Escherichia*, which was shown within the branch lengths.

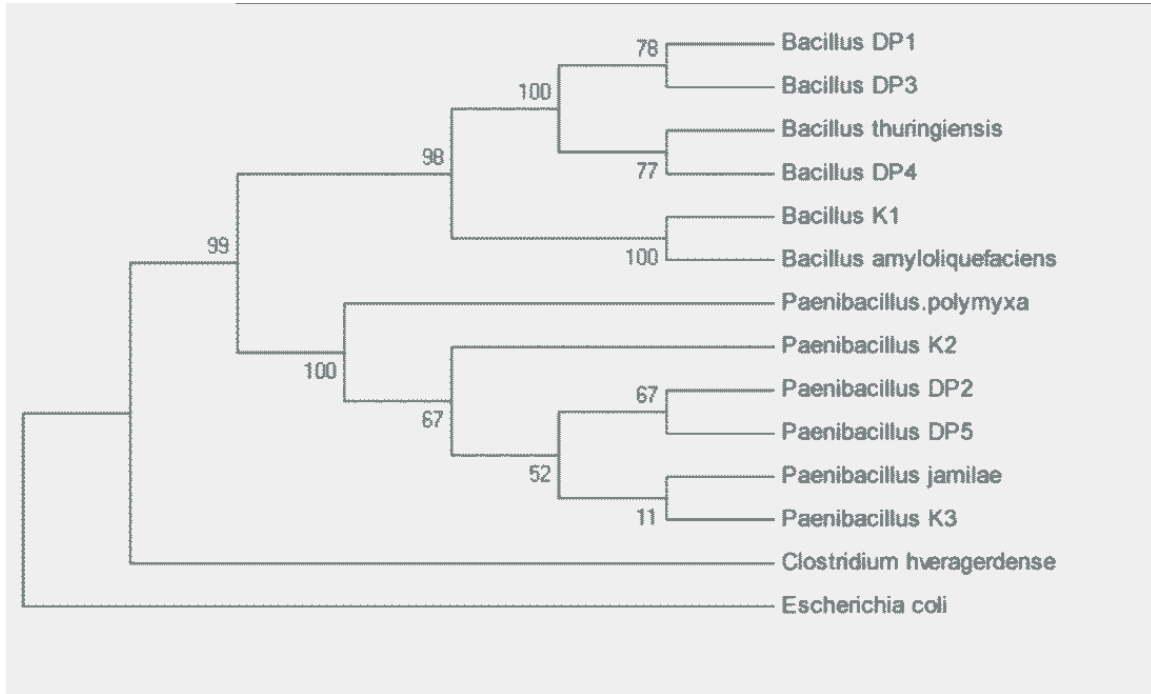


Figure 2.2. Phylogenetic tree of all organisms isolated from enrichments from *H. cannabinus* and *Phoenix dactylifera* (date palm). Additional organisms were inserted as reference organisms. *Bacillus thuringiensis*, *Bacillus amyloliquefaciens*, *Paenibacillus polymyxa*, and *Paenibacillus jamilae* were used as similar organisms while the *Clostridium hveragerdense* was used as a similar organism, i.e. within the phylum Firmicutes, and *Escherichia coli* was a further outgroup that is within a different phylum Proteobacteria.

2.4 Future Work with Organisms

Additional characterizations of these organisms will be performed to determine their optimal growth conditions. Variables that will be tested are the optimal temperature, pH, and general growth characteristics as compared to each other. However, we must keep this within the context of our project as a whole, i.e. determining the best protocol for the retting of kenaf. These are important details, but must be used sparingly as compared to the optimal

parameters of the retting of kenaf. Additionally there is a need to target the anaerobic organisms such as *Clostridium* and *Prevotella* as this process seems to point towards the bulk of the process occurring during the anaerobic stage.

CHAPTER 3

MICROBIAL COMMUNITIES OF DIFFERENT RETTING TECHNIQUES*

3.1 Introduction

Determining the microbial community involved in retting is essential to increasing efficiency. As described before, the typical understanding of the microbial retting environment is a shift from aerobic to anaerobic conditions as time progresses. To elucidate this process we designed three different retting environments to test the changes in microbial species composition (Table 3.1). Killed controlled (Control 1) consisted of autoclaved pond water was performed to assess the microbial community on the plant fibers themselves and estimate their contribution to the retting process. Pond control (Control 2) was performed to mimic a natural retting microbial process, including contributions from the plant material and pond water. Pond water was chosen as an “inoculum” due to its inherently complex nature and the presence of endogenous plant degraders. Finally, (augmented killed controlled) Experimental 1 was performed to determine the extent to which the pectinolytic isolates could become established, their influence on the bacterial community structure and transition, and their influence on the retting rate

* Parts of this chapter have been previously published, either in part or in full, from D. K. Visi, N. D'Souza, B. G. Ayre, C. L. Webber III, and M. S. Allen, “Investigation of the bacterial retting community of kenaf (*Hibiscus cannabinus*) under different conditions using next-generation semiconductor sequencing,” *Journal of Industrial Microbiology and Biotechnology*, 40, 465-475 (2013). Reproduced with permission from Springer Publishing, Ltd.

Retting Environment	Amount of Water (MilliQ)	Pond water (autoclaved = A, not autoclaved = N)	Bacterial Inoculum	Amount of Plant Material
Control 1	750 mL	250 mL A	N/A	20 g.
Control 2	750 mL	250 mL N	N/A	20 g.
Experimental 1	750 mL	250 mL A	Yes*	20 g.

Table 3.1. The general experimental design of the microbial diversity experiments. *Inoculum consisting of 50 mL of each centrifuged culture: *Bacillus* DP1, *Paenibacillus* DP2, *Bacillus* K1 as described in the text.

3.2 Materials and Methods

3.2.1 Controls and Augmented Retting Conditions

Kenaf was obtained from two sources, C. Webber (USDA) denoted as KOK and S. Shi denoted as KMS. KOK was grown at the USDA agricultural research station in Lane, OK. Stalks were harvested 185 days post-planting and the kenaf bark was manually stripped from the core. The first bottom meter of the stripped bark containing bast fibers from the kenaf material was cut into 2 cm pieces and immediately used in subsequent experiments. KMS was grown at Mississippi State University North farm, and grown for approximately 6 months, dried in the field, and stored for approximately three years until this experiment. This harvested kenaf was stripped from the core cut into 2 cm lengths (approximately 1 cm widths) pieces and mixed together.

Experimental set up of each of the retting environments, i.e. E1, C1, and C2 are shown in Table 3.1. KOK samples were performed first and are denoted by the type of environment they were subject to as well as the KOK name after. The KMS samples were subject to the E1 and C1 environment types and were performed in duplicates to confirm previous findings with the KOK samples.

The pond water used for the inoculate on the natural retting environment came from a water reservoir located in front of the Environmental Science Building on the UNT campus. This was chosen because it was a relatively stagnant water-source, which also included a large amount of decaying plant material which would presumably select for pectinase or similar enzymes to degrade heteropolysaccharides. The pond water was taken at one time and used for all of the subsequent experiments in this study.

The three bacteria used to inoculate the experimental treatments (E1) were isolated as described in Chapter 2. Fresh media was inoculated from -80°C freezer stocks and grown overnight in 250 mL of lysogeny broth (LB) in 500-mL Erlenmeyer flasks shaking at 37°C and 225 rpm (101). 50-mL of cell cultures were pelleted by centrifuged at 7 197 g for 30 m, suspended in milliQ water (18 MΩ), and used to inoculate the appropriate experimental tanks.

The retting environments were setup in tandem and started at the same time to reduce risks in possible systemic bias. Additionally all of the retting experiments were performed at ambient temperatures (30°C) in a stationary place.

Primers	Sequences
27F	AGAGTTTGATCMTGGCTCAG
1492R	ACCTTGTTACGACTT
IonA-E786F	CCATCTCATCCCTGCGTGTCTCCGACTCAGGATTAGATACCCTGGTAG
IonAs1-E786F	CCATCTCATCCCTGCGTGTCTCCGACTCAG<u>CT</u>GATTAGATACCCTGGTAG
IonAs2-E786F	CCATCTCATCCCTGCGTGTCTCCGACTCAG<u>T</u>AGATTAGATACCCTGGTAG
IonAs3-E786F	CCATCTCATCCCTGCGTGTCTCCGACTCAG<u>A</u>AAGATTAGATACCCTGGTAG
IonAs4-E786F	CCATCTCATCCCTGCGTGTCTCCGACTCAG<u>C</u>AGATTAGATACCCTGGTAG
Ion P1-E989Rev	CCTCTCTATGGGCAGTCGGTGATCTTGTGCGGGCCCCCGTCAATTC

Table 3.2. Primers used in the experiments. Universal primers 27F and 1492R used for full-length 16S amplification. “Ion” refers to the Ion Torrent specific nested PCR primers. Bolded sequences are the Ion Torrent specific adapter sequences. “A” refers to the Ion-specific sequence of the forward primer. “1-4” denote the four different forward barcoded 786 primers and the underlined portion are the actual two nucleotide barcode sequences used in this study. IonP1-E989Rev was used in all nested PCR reactions.

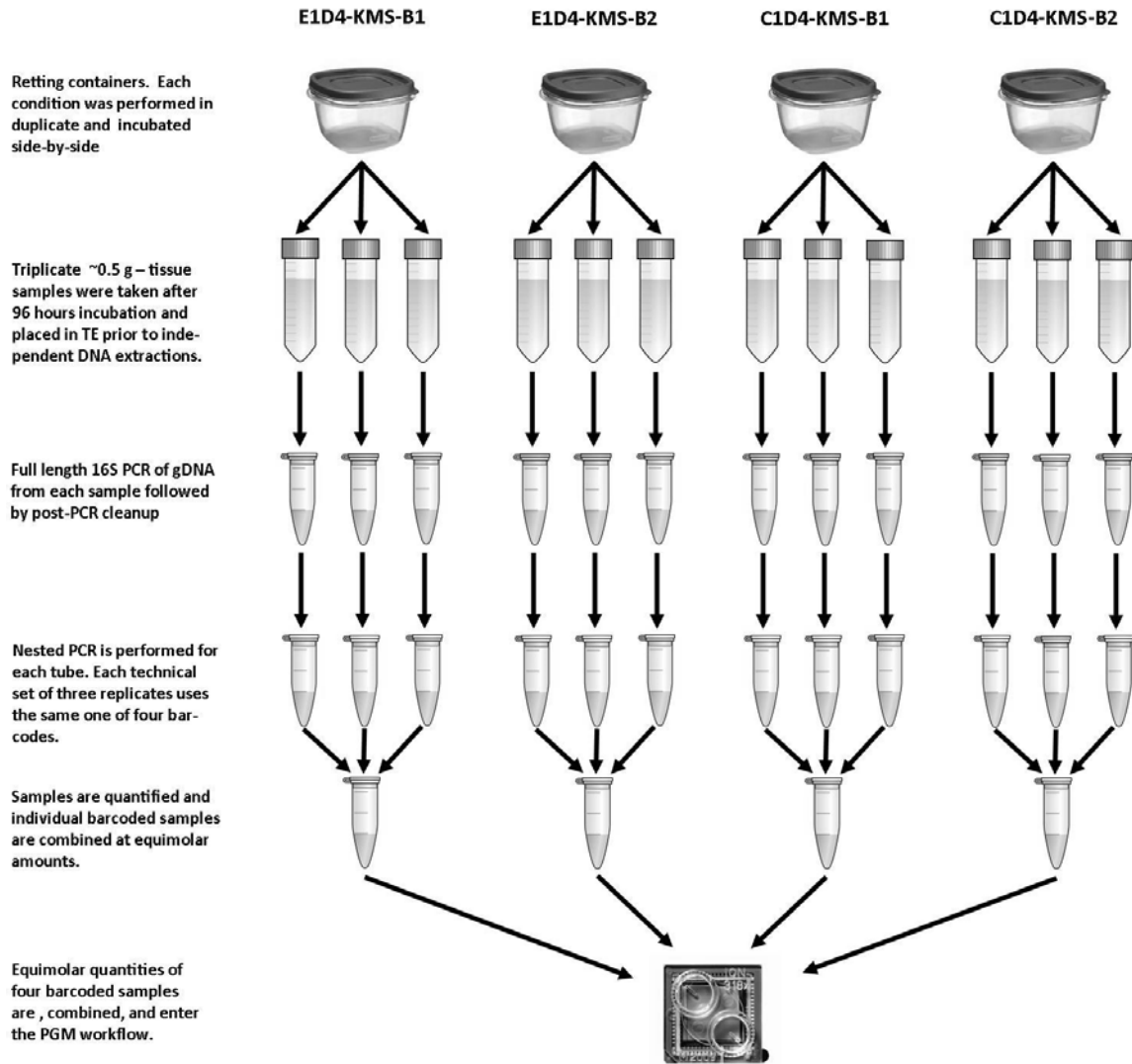


Figure 3.1. General workflow for the batch/replicate experiments.

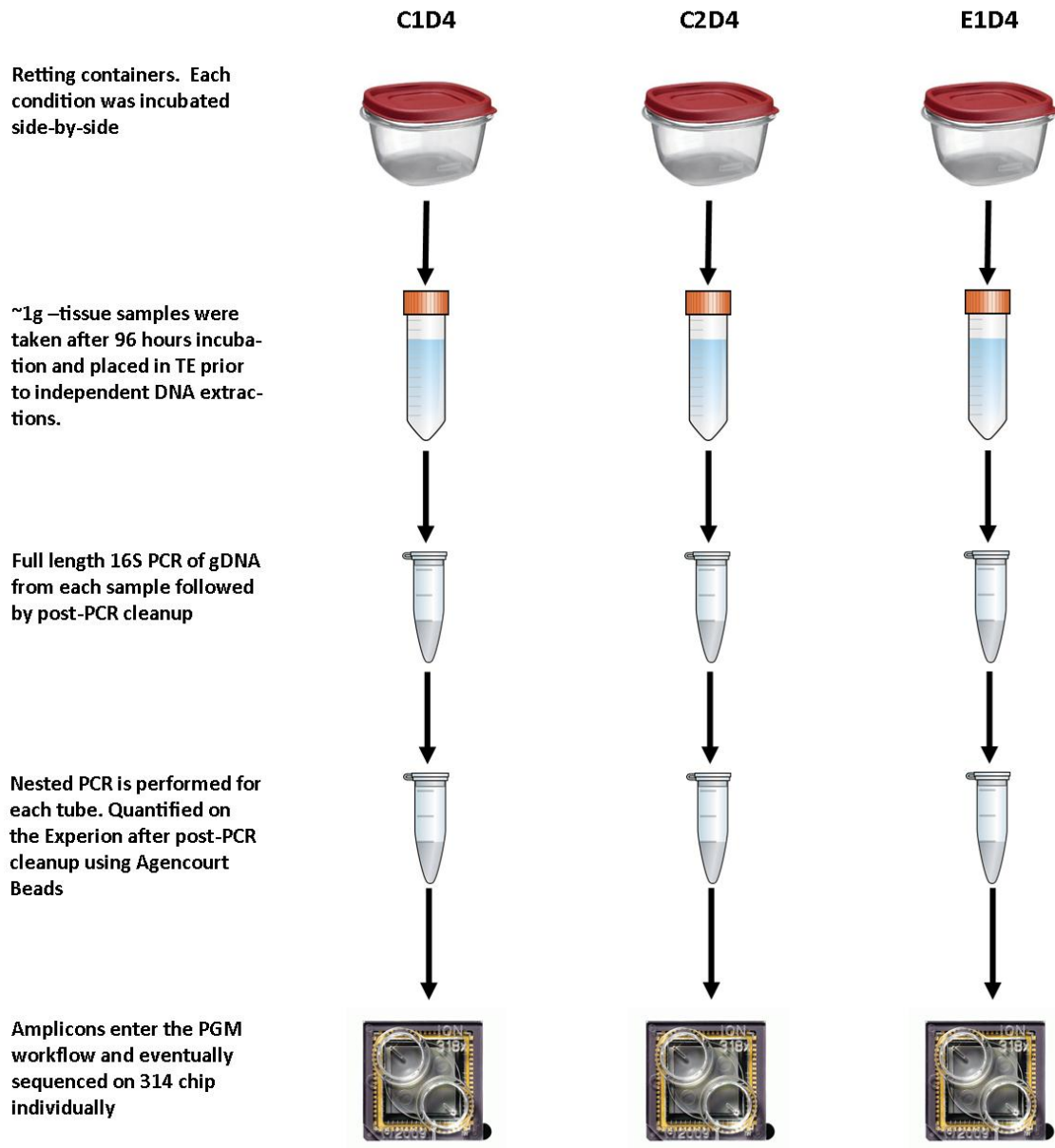


Figure 3.2. Initial experimental design for the tested environments.

3.2.2 Extraction of DNA from Plant-Associated Bacteria and Amplification of 16S V4 Region

Samples of bark fiber were removed at day 4 (96 hours), rinsed in milliQ water, and used for direct DNA extraction from adherent microbial species. The DNA extraction was accomplished by bead beating in suspension buffer using the MO BIO Fecal DNA kit (MO BIO Laboratories, Inc., Carlsbad, CA) and Fisher Vortex Mixer (Thermo Fisher Scientific, Waltham,

MA) as per the manufacturer's instructions. 1-3 g of plant material was added to the dry bead tubes (which includes garnet beads and is included with the DNA extraction kit) with 550 μL of bead solution. This solution is vortexed briefly. Then 60 μL of SDS solution is added (Solution S1) along with 200 μL of inhibitor removal solution (Solution IRS). The bead tubes were secured to a MO BIO Vortex adapter tube holder and shaken horizontally for 10 m at maximum speed to disrupt surface-associated bacteria from the plant material. The vortexing of the solution along with the presence of a disruption agent and collision of the beads with bacterial cells causes the cells to break open. The tubes were removed from the vortex and centrifuged at 10 000 x g for 30 s. This centrifugation step allows the cell debris to collect towards the bottom and the supernatant contains the gDNA from the bacterial cells. The supernatant was then transferred to a clean 2 mL collection tube. 250 μL of solution S2 was added, vortexed for 5 s, and incubated at 4°C for 5 m. This is an additional inhibitor removal step for which to precipitate non-DNA organic and inorganic materials including polysaccharides, cell debris, and proteins. These inhibitor removal steps are essential to the downstream processes, which involve PCR, because organic and inorganic matter can inhibit amplification. The tubes were then centrifuged for 1 m at 10 000 x g. The 450 μL of the supernatant was transferred to a clean 2.0 mL collection tube. The pellet at the bottom of the tube contains the various non-DNA organic and inorganic materials while the supernatant contains the DNA.

Solution S3 was shaken and then 900 μL was added to the supernatant and vortexed for 5 s. Solution S3 is a high concentration salt solution and since DNA binds tightly to silica at high salt concentrations, this will allow the binding of DNA to the silica filters which allowed the non DNA organic and inorganic materials to flow through when centrifuged. 650 μL of the

supernatant/Solution S3 solution was added to a spin filter and centrifuged for 10,000 x g for 1 m. Contaminants flow through the silica filter, while the DNA binds to the silica filter due to the high salt solution. The flow through was then discarded and another 650 μ L of the remaining solution was added to the spin filter and centrifuged again for 10,000 x g for 1 m. 300 μ L of solution S4 was added to the center of the spin filter and again centrifuged for 10,000 x g for 30 s. Solution S4 is a ethanol-based solution which further cleans the DNA on the silica filter and removes any residual salt and contaminants that are on the filter membrane. The flow through was discarded and again the spin filter assembly was centrifuged for 10,000 x g for 1 m to dry out the spin filter. This second centrifugation step is essential to the process of getting clean gDNA, because residual ethanol can disrupt downstream DNA processes like PCR and restriction digests. The spin filter was transferred to a new clean collection tube and 50 μ L of solution S5 was added to the center of the white filter membrane and centrifuged for 10,000 x g for 30 s. Solution S5 is a sterile elution buffer composed of 10 mM Tris which allows the DNA to be release from the silica filter membrane due to the absence of salt in the solution. Additionally, the solution should be added directly to the center of the membrane as this allows the release of the gDNA from the filter. After final centrifugation, the spin filter was removed and the flow through contained the gDNA from the plant extractions. The gDNA solutions were labeled and stored at -20°C.

3.2.3 Quantification and PCR Amplification

After extraction, DNA samples were quantified by spectrophotometry, with a Nanodrop 1000 (Thermo Fisher Scientific Inc., Waltham, MA) and used for downstream processes. PCR

amplification of the variable region five (V5) of the 16S rRNA was performed in a nested fashion, i.e. two-part amplification. The initial amplification was of the entire length of the 16S rRNA through the use of primers 27F and 1492R (102). This was required due to spurious bands being produced from the initial tests of the Ion Torrent specific primers, which were removed when the initial full-length PCR was accomplished. The general workflows are described in Figures 3.1 and 3.2.

PCR tubes were set up using the following parameters: 10 μ L of 5X HF Buffer, 200 μ M of dNTPs, 250 nM of 27Fwd, 250 nM of 1492Rev, 1.5 μ L of DMSO, 0.5 μ L of Phusion DNA polymerase 2 000 units/mL (New England BioLabs Inc., Ipswich, MA), 20 ng of gDNA, and was brought up to a final volume of 50 μ L. A \sim 1.5 Kb fragment was amplified from each of the respective sources using the following thermal cycler protocol: initial denaturation at 98 $^{\circ}$ C for 5 m, denaturation at 98 $^{\circ}$ C for 15 s, annealing of primers at 56 $^{\circ}$ C for 30 s, extension of fragments at 72 $^{\circ}$ C for 1 m, for 25 cycles, and a final extension at 72 $^{\circ}$ C for 5 m.

After confirmation on a 1% agarose gel, the PCR product was cleaned using Agencourt Ampure XP beads (Beckman Coulter Genomics, Danvers, MA) as per the manufacturer's instructions and quantified on a Nanodrop 1000. The diluted, cleaned product served as the template for the second Ion Torrent-specific nested 16S PCR amplification using IonA-E786Fwd and IonP1-E989Rev (44). These primers, i.e. IonA-E786Fwd and IonP1-E989Rev not only included the specific regions to V5, but also included Ion Torrent specific tags necessary for sequencing (Table 3.2).

50- μ L PCR reactions were set up using the following parameters: 10 μ L of 5X HF Buffer, 200 μ M of dNTPs, 250 nM of IonA-E786 Fwd, 250 nM of IonP1-E989Rev, 1.5 μ L of DMSO, 0.5 μ L

of Phusion DNA polymerase (NEB), 20 ng full length 16S DNA, and was brought up to a final volume of 50 μ L. Ion Torrent specific primers produced a 210 bp band using the following 2-step PCR thermal cycler protocol: 98°C initial denaturation for 3 min, 98°C denaturation for 15 s, 61°C annealing and extension for 15 s, and repeated for 25 cycles, with a final extension for 5 min. The resultant PCR products were purified using Agencourt Ampure XP beads (Beckman Coulter Genomics, Danvers, MA) following the protocol as outlined by Ion Torrent, and quantified on a Nanodrop 1000. Product quality was confirmed on an Experion Automated Electrophoresis System (Bio-Rad, Hercules, CA). The Experion system is an automated electrophoresis system, which utilizes microfluidics to determine both the concentration and distribution of a DNA sample. This is essential for the downstream Ion Torrent workflow because of a need for an accurate and precise concentration and check for the size of the fragment.

For samples E1D4-KMS-B1, E1D4-KMS-B2, C1D4-KMS-B1, and C1D4-KMS-B2 we used custom barcoded forward primers as shown in Table 3.2. The PCR protocol remained the same as the previous non-barcoded primers (*i.e.* IonA-E786Fwd). Specific extraction, replications, and technical replicates are shown in Figure 3.1.

Cleaned PCR products were diluted to the appropriate nanomolar concentration (8.4). Samples were amplified by emulsion PCR using the Ion Torrent OneTouch System with the Ion OneTouch System Template Kit as per the manufacturer's instructions. The resultant beads were enriched on the Ion Torrent ES prior to loading on to 314 10Mb sequencing chips. Each environment was sequenced separately, except for the barcoded batch samples, which were confirmatory experiments to the previous sequenced samples. These samples were diluted to

the appropriate nanomolar concentration (8.4 nM), pooled together to maintain equimolarity, and then diluted to the appropriate library dilution concentration (~26 pM) prior to loading on the OneTouch emulsion PCR system. Following enrichment was performed on the same system and reagents as previous runs.

3.2.4 Ion Torrent Workflow

The Ion Torrent PGM system was selected as the preferred choice for sequencing the 16S amplicons produced. The Ion Torrent works on the premise of the release of a proton during normal DNA synthesis. The release of a proton results in a net decrease in pH thus indicating an addition of a nucleotide into the new synthesized DNA strand. Sequencing by synthesis is used by most of the NGS platforms like Roche 454 and Illumina systems and it is the same mechanism for the Ion Torrent. However, where the Ion Torrent differs is that it uses no modified nucleotides as compared to other platforms and requires no optics. The measurement of the added nucleotide comes from the complementary metal-oxide semiconductor-based pH detectors within the wells of the sequencing chip that measure the addition of the protons to the well environment.

3.2.4.1 Emulsion PCR

The general workflow is ligating appropriate Ion Torrent specific tags onto target DNA to be sequenced, emulsion PCR (ePCR), and then sequencing on the PGM. Emulsion PCR is performed on the Ion Torrent OneTouch instrument. Once the target DNA is diluted to the appropriate amount as recommended by the manufacturer, the diluted DNA can be loaded into

the Ion Torrent OneTouch. The OneTouch system is comprised of the emulsion PCR aspect (OneTouch) and the Ion OneTouch ES (enrichment system). The OneTouch system first creates an emulsion of the DNA in which clonal amplification can occur. A reaction filter is used to create the emulsion, i.e. it creates a nano-scale droplet that is immersed in oil, in which the aqueous portion that, contains all the necessary aspects for PCR: the template DNA, buffer, polymerase, dNTPs, primers, and the Ion Sphere Particles [ISPs]). Prior to the ePCR a dilution must be applied to the input DNA, e.g. in the manufacturer's protocol there is a dilution of 1:650 of an 8.4 nM solution, but ultimately it is to achieve the final concentration of 26 pM in order for 10-30% of ISPs to be positive with clonal DNA. This dilution is an essential step in the ePCR because the dilution is required for the optimal amount of DNA to be clonally amplified. If the dilution is under the required amount there will be a large amount of ISPs that have no DNA thus lowering the eventual efficiency of sequencing. Excess DNA results in polyclonal reads, in the ideal dilution one DNA sequencing target should be in one droplet during the ePCR so that it can be clonally amplified during the replication step: one template DNA to one ISP. When the dilution is too high, this results in multiple DNA targets being in one aqueous droplet. Amplification of the resultant polyclonal droplet produces a hybrid signal during the sequencing step, specifically when dNTPs are incorporated there is a certain level in change of pH that is standardized during the initialization step, but if half of the sequences are incorporating an adenine and the other half is not, then the pH change will be around half of the surrounding droplets. This results in sequences being lost during the final phase of data analysis due to being filtered out by the Ion Torrent PGM Torrent Server version 2.0.1. After completion on the

OneTouch, which accomplishes the ePCR aspect of the workflow, the ISPs are enriched on the OneTouch ES.

3.2.4.2 Enrichment Using the Ion Torrent ES

The enrichment step utilizes a magnetic bead based protocol in which streptavidin beads attach to the biotinylated ISPs. In the prior step, ISPs have been incorporated with the desired target DNA to be sequenced, in addition to the Ion Torrent specific tags there is a biotin binding site on the said tags, so positive ISPs will bind to the streptavidin beads and the negative ISPs, with no clonally amplified DNA, will be washed out. This enrichment step is performed on the Ion Torrent ES. The streptavidin beads are added to the mixture of ISPs and allowed to bind. Then a magnetic separation of the beads occurs leaving behind the streptavidin beads that have no DNA template attached. This is done in multiple washing steps on the ES until there is an enrichment solution of positive ISPs, in the range of 90-100%. The last step is a breaking step, consisting of a high pH solution that removes the positive ISPs from the streptavidin beads. Finally, there is a neutralization solution added to the positive ISPs. After enrichment and breaking of the beads the ISPs can be sequenced on the Ion Torrent PGM.

3.2.4.3 Sequencing on the PGM

The sequencing platform in the Ion Torrent workflow is the PGM and utilizes a semiconductor sequencing by synthesis technology. This system is non-optically based in that it is detecting a natural addition of a dNTP, which releases a hydrogen proton, which decreases pH. The technology is based around the chip in that it is a high-density array with 3.5 μm well that

can only contain a single 2 μm ISP bead. Below the well layer there are two additional aspects of the chip, a pH meter, and a voltage detector, which one detects the release of the proton, i.e. an increase in pH, and the voltage detector, both working together to determine when dNTPs are incorporated into the synthesizing strand of DNA. The scalability of the system also relies on the chip as increases in data output has partially to do with increased read lengths, but also the density at which the wells are incorporated onto the chip. Specifically, the 314 is rated at around 10 Mb of sequencing data, 316 is around 100 Mb, and 318 goes up to 1000 Mb, but surprisingly the data output in the lab resulted in much higher outputs than the rated amount.

3.2.5 Data Analysis

SFF files were converted to FASTA and quality files using Galaxy (103). The SFF file is a standard flowgram format, which has quality information, nucleotide sequence, and the flowgram, similar to the outputs found on the 454 platform. The resulting FASTA and quality files were then inserted into RDP Pyrosequencing Pipeline using Pipeline Initial Process (<http://pyro.cme.msu.edu/>) (46). This program removed the forward and reverse primers from each of the sequence fragments, any sequences under 100 base pairs, and any sequences with ambiguous nucleotides (N). This quality control ensured that only sequences that had the correctly sequenced primer, both forward and reverse were included in the downstream process. Additionally, the removal of the primers and quality checking in terms of accuracy help with the downstream processes, e.g. classification and diversity estimates. The processed FASTA files were placed into RDP Classifier and set at 50% confidence (104).

3.2.6 Rarefaction Curves

Sequences were aligned using Infernal aligner (105). Infernal aligner or inference of RNA alignment, allows the search of DNA sequence databases for not only sequence similarities, but also account for the secondary structure inherently with 16S RNA molecules. A combination of sequence alignment and RNA secondary structure help identify the RNA homologs better than simply doing a BLAST type analysis in which only sequences are accounted for. An additional benefit to utilizing the Infernal program is that there is a reduction in time of alignment, or in other words a lowering of computational power needed for completion of a task. Specifically it uses an accelerated profile hidden Markov Model and HMM-banded CM alignment methods, lowering the needed time for calculation in some situations at 10 000-fold (106). Aligned sequences from each retting condition were generated with their individual cluster files based on the RDP pyrosequencing pipeline. The cluster files produced were used to generate rarefaction curves that defined the number of operational taxonomic units (OTUs) at 97% similarity level with respect to the total number of reads for each sample.

The concept of operational taxonomic units is the basis for diversity comparisons. The most basic definition is how to determine different groups in a set environment. This definition can be further expanded to how inclusive or exclusive the OTU threshold can be set. The general consensus for bacterial species is a similarity of 97% at the 16S rRNA level (107). This threshold can be increased to show differences at a higher level of taxonomy,

3.2.7 16S TA clone library of Experiment Day 4

PCR was repeated using the extract gDNA from experiment day 4 (E1D4-KOK). Using

PuReTaq Read-To-Go PCR beads (GE Healthcare, Buckinghamshire, UK) reactions were set up using the following parameters: 250 nM of 27Fwd (Table 2), 250 nM of 1492Rev (Table 2), 20 ng of E1D4-KOK gDNA, was brought up to a final volume of 25 μ L, and the following thermal cycler protocol: initial denaturation at 95°C for 10 min, denaturation at 95°C for 15 s, annealing at 56°C for 30 s, extension at 72°C for 90 s, for 30 cycles, and a final extension at 72°C for 10 m. The final extension time on the amplification protocol was to ensure an aspect of the segments of amplified DNA to include necessary overhanging adenines and is required in the TA cloning reaction. TA cloning was performed following instructions as outlined by Invitrogen. Colonies were picked from LB agar with 50 μ g/mL kanamycin plates and grown in 10 mL LB with 50 μ g/mL kanamycin. Cells were harvested by centrifugation and plasmids were extracted from each pellet using the 5 Prime FastPlasmid Mini Kit (5 Prime, Gaithersburg, MD). The DNA was quantified as described and sequenced by Eurofins MWG Operon. Sequences were identified by BLAST, aligned using MEGA, and placed into a maximum likelihood phylogenetic tree for comparison (45, 100).

3.3 Results

Data yield from each of the runs ranged from 278 626 to 591 627 sequences pre-quality control. The highest run was the barcoded sequences including E1D4-B1-KMS through C1D4-B2-KMS. After applying RDP Pyrosequencing initial processor sequences were dropped for either having incorrect forward and reverse primers, having ambiguous nucleotides, being under the 100 bp cutoff, or any combination of the previous, which resulted in a range from 28 549 to 148 850 sequences (110 ± 1.86 to 113 ± 0.73).

Phylogenetic Composition of Different Retting Environments at Day 4

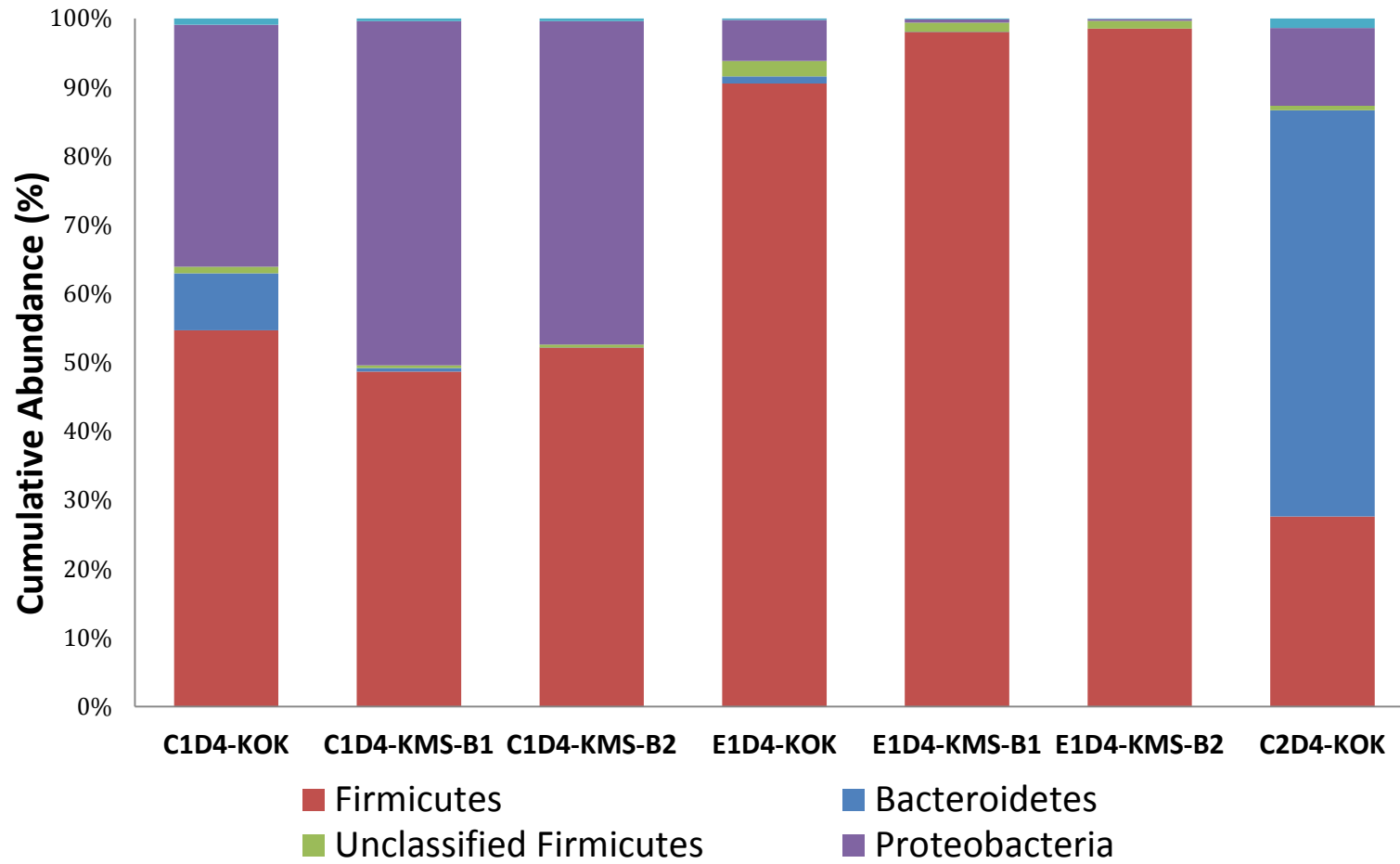


Figure 3.3. Phylogenetic compositions of different retting environments at day 4.

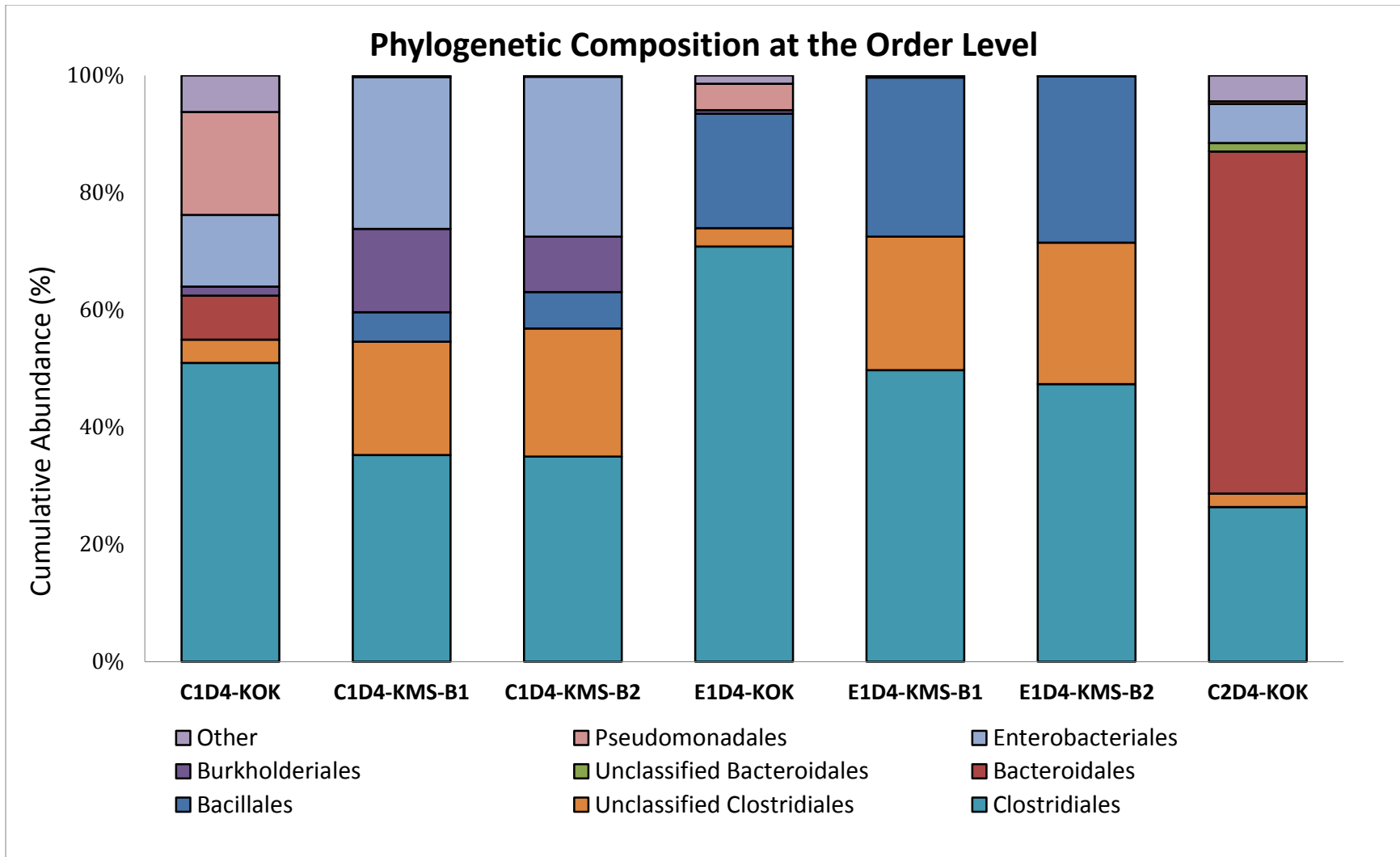


Figure 3.4. Phylogenetic composition at the order level.

3.3.1 Overall Microbial Diversity of Control and Experimental Retting Environments

The microbial communities in all three retting environments were dominated by the domain Bacteria. Three predominant phyla across all samples were the Firmicutes, Proteobacteria, and Bacteroidetes and were comprised of the following orders: Aeromonadales, Bacillales, Bacteroidales, Burkholderiales, Clostridiales, Enterobacteriales, Lactobacillales, Pseudomonadales, Rhodocyclales, Sphingobacteriales, Selenomonadales, Sphingomonadales, and Xanthomonadales (Figure 3.3 and 3.4). However the predominant organisms present in most of the retting environments were represented by the orders Clostridiales, Enterobacteriales, Bacilliales, and Bacteriodetes

3.3.2 “Natural” Microbial Retting – C2D4

C2D4-KOK was performed to characterize a “natural” microbial retting community. As expected, this pond water inoculum resulted in an environment with markedly increased diversity as compared to the other retting experiments. Phylum Bacteroidetes dominated this environment at 59% (Figure 3.3). This discrepancy was clearer at the order level where Bacteroidales comprised 58% of the population, while the next closest retting environment C1D4-KOK had only 8%, and none (or less than 0.5%) being detected in the others. Clostridiales (26%) and Enterobacteriales (7%) were also present, but at substantially lower amounts as compared to the other environments (Figure 3.4).

3.3.3 Composition of the Plant-Associated Microbial Flora – C1D4

All of the C1D4 experiments containing autoclaved pond water (i.e. C1D4-KOK, C1D4-KMS-B1, and C1D4-KMS-B2) had large amounts of Firmicutes: 55, 49, and 52%, respectively, with the remainders being predominately composed of the Proteobacteria at 35, 50, and 47% (Figure 3.3). C1D4-KOK had large amounts of the order Clostridiales at 51%, but showed additional diversity not found in the E1D4-KOK set: 18% Pseudomonadales, 12% Enterobacteriales, and 8% Bacteroidales, among others (Figure 3.4). The KMS samples of C1D4 had the same hierarchical order found in C1D4-KOK, but showed slight changes in terms of Clostridiales (35% for KMS samples), Enterobacteriales (26% and 27%), unclassified Clostridiales (19 and 22%), Burkholderiales (14 and 9%), and finally one taxonomic order Bacillales (5 and 6%) that was not present in the C1D4-KOK sample.

3.3.4 Augmented Microbial Retting with Pectinolytic Isolates – E1D4

E1D4 was performed to determine the effect of a large initial inoculum of pectinolytic organisms on retting efficiency as well as community structure and composition. In addition to 250 mL of autoclaved pond water, E1D4 contained 750 mL of MilliQ water and the three pectinolytic organisms: *Bacillus* DP1, *Paenibacillus* DP2, and *Bacillus* K1. All sets of E1D4 were dominated by the phylum Firmicutes ranging from 91 to 99% of total diversity (Figure 3.3). Slight changes in composition were noted and were likely due to the different origins, treatments of the plants, and fresh versus stored (Oklahoma and Mississippi). E1D4-KOK had 91% composition Firmicutes, which were further sub-divided at the order level into Clostridiales (71%), Bacillales (20%), Pseudomonadales (4%), and unclassified Clostridiales (3%). The other

set that included replicates of the plant samples from Mississippi, E1D4-KMS-B1 and E1D4-KMS-B2, had Clostridiales at 50 and 47%, unclassified Clostridiales at 23 and 24%, and Bacillales at 27 and 28%, respectively (Figure 3.4).

3.3.5 Environment-Related Differences in Microbial Richness

Microbial richness was defined based on the number of operational taxonomic units (OTUs at 97% sequence similarity) identified in each of the different retting solutions.

Sequences from each sample were subjected to alignment and then complete linkage clustering using a max distance of 3% and placed into rarefaction curves (Figure 3.5). OTUs ranged from 832 to 2683. C2D4-KOK had almost 2.5-fold higher microbial richness than found in any of the KMS samples.

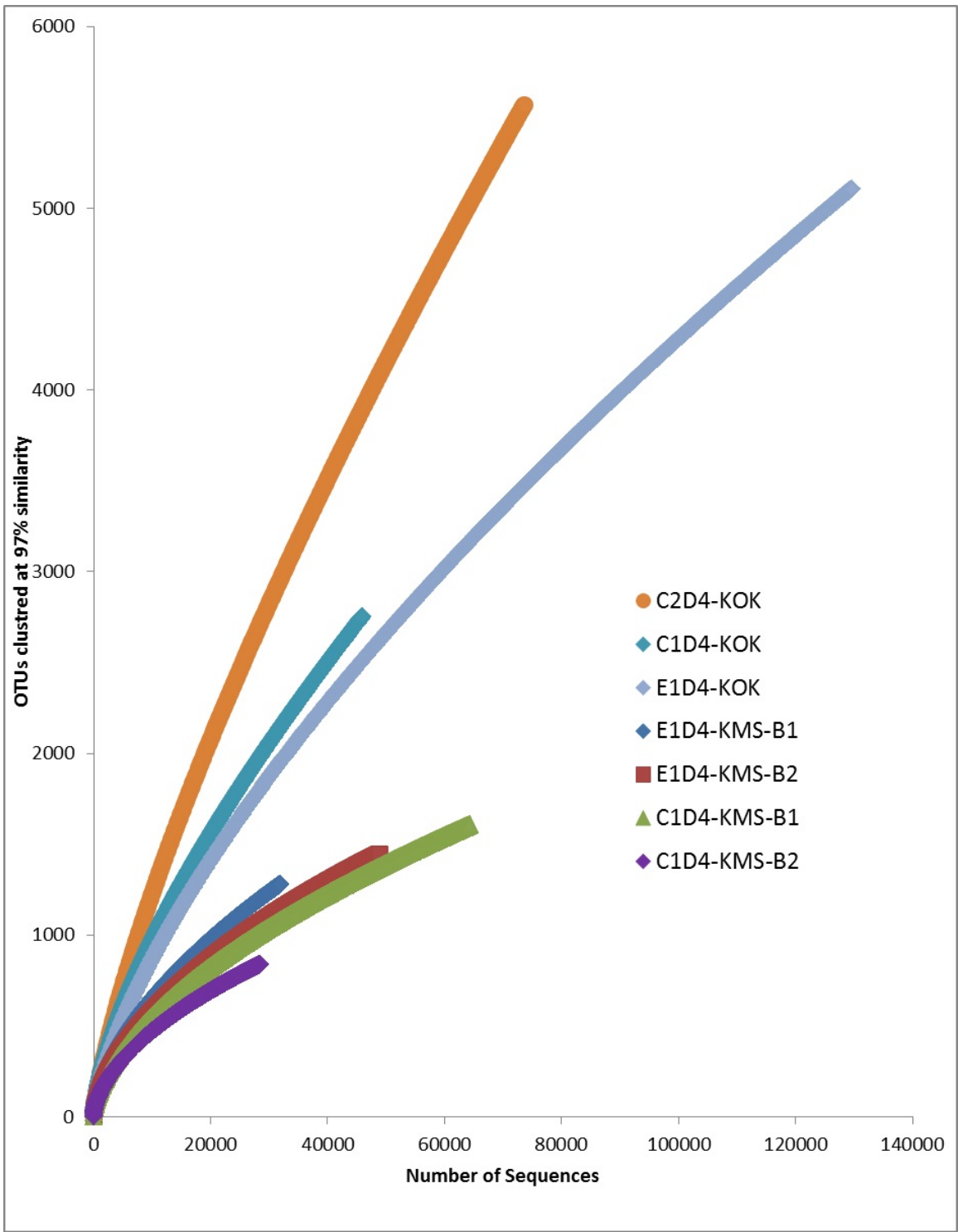


Figure 3.5. Rarefaction curves.

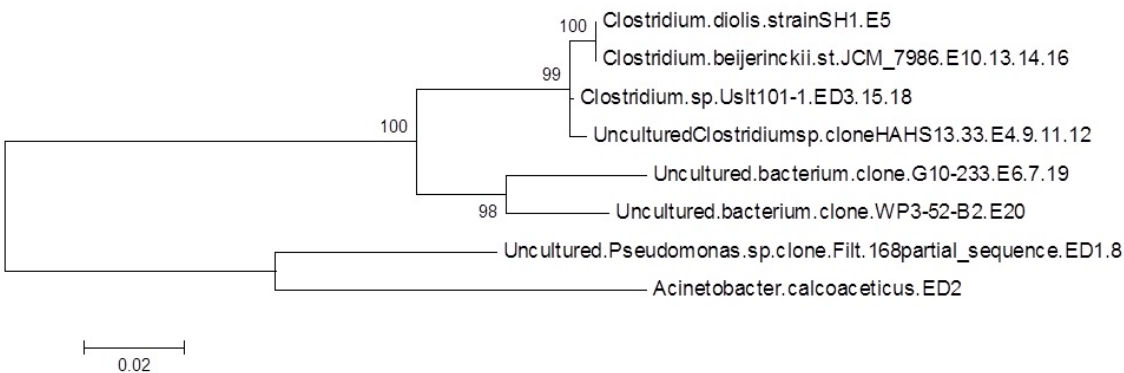


Figure 3.6. Phylogenetic analysis of Clostridium isolates.

3.3.6 TA Clone Library

Twenty clones generated from a near full length 16S clone library representing E1D4-KOK were prepared and sequenced. Sequences were analyzed using BLAST, and the full-length sequence of the closest match was used to generate a maximum likelihood phylogenetic tree (Figure 3.6). Sixteen of the nineteen successful sequences were found to cluster within the genus *Clostridium*. Among these, four were identified as likely *C. beijerinckii*, and three most closely matching *Clostridium* sp. UsIt101-1. These anaerobic bacteria are notable for their ability to fix nitrogen (108). Also present were members of the genera *Pseudomonas* and *Acinetobacter* (2 and 1, respectively).

3.4 Discussion

Understanding the microbial community in the retting environment is an essential step to improving retting process efficiency. While the crude process has been performed for thousands of years, molecular techniques have only recently been applied to understanding the

microbial community involved. Selection of day four in the retting solutions was based on previous experiments, which showed significant breakdown of pectin and lignin surrounding the fibers. Additionally, this was sufficient time to allow the solution to go anaerobic. A small number of studies have used Sanger sequencing of 16S clone libraries generated from plant retting solutions to explore aspects of retting microbial communities, but next-generation sequencing of 16S amplicons has not previously been reported.

Three basic retting experiments were initially conducted. C2D4 was performed to mimic a natural retting microbial process, including contributions from the plant material and pond water. Pond water was chosen as an “inoculum” due to its inherently complex nature and the presence of endogenous plant degraders. C1D4 with autoclaved pond water was performed to assess the microbial community on the plant fibers themselves and estimate their contribution to the retting process. Finally, E1D4 was performed to determine the extent to which the pectinolytic isolates could become established, their influence on the bacterial community structure and transition, and their influence on the retting rate.

The second set of the C1D4 experiments, i.e. KMS, had very similar compositions, but the overall microbial communities as compared to the original KOK experiments remained relatively unchanged, i.e. the major constituents are present, independent of the differences in source location and treatment. C2D4-KOK was the only set that used fresh plant material and fresh pond water, which most closely represented a “traditional” microbial retting environment and had the highest microbial richness. The microbial community was found to be markedly different from the experimental inoculation E1D4 and the autoclaved pond water C1D4 in both composition and overall diversity.

The order Clostridiales were found to be a major component across all retting environments except C2D4-KOK, which had a lesser percentage (Figure 3.4). Members of this group have been found to produce numerous pectinases (93–95) which are favorable to the microbial retting process. The other large constituent in the process was the order Bacteroidales, but this group was only found in the C2D4-KOK samples. Specifically, the strongest matches indicate the presence of members of the genus *Prevotella*, which has not typically been described as being involved in microbial retting process. However, members of this genus have been found to produce pectinases and identified as members of the ruminant gut (109, 110).

3.4.1 Firmicutes as a Dominant Phylum in Microbial Retting Communities

The phylum Firmicutes dominated samples C1D4 and E1D4 including both sets of plant materials KOK and KMS. When comparing the C1D4 and E1D4 samples, while order Clostridiales is a large component of both, in all of the E1D4 experiments the percentage of Firmicutes increased by a large margin (91-98%) as compared to the C1D4 (KOK and KMS) samples (49-55%). As described above, the Firmicutes are prolific pectinase producers and many have been isolated based on this property. These results were surprising given that the only difference between C1D4 and E1D4 was the addition of the isolates *Bacillus* DP1, *Paenibacillus* DP2, and *Bacillus* K1.

The typical retting solution starts as an aerobic environment, which is favorable to aerobes or facultative anaerobes like *Bacillus* and *Paenibacillus*, respectively (84). However, as metabolism proceeds, anaerobic conditions are established. This process has been

characterized in other retting experiments which showed a later colonization by *Clostridium* by the shift from initial aerobic organisms, specifically *Bacillus licheniformis* and *B. subtilis*, to anaerobic, *Clostridium acetobutylicum* and *C. felsineum* (111). This was confirmed by the presence of large amounts of Clostridiales in the E1D4 and C1D4 sets and to a lesser extent C2D4-KOK.

The order Bacillales (Figure 3.4) represented a large amount of the diversity in the E1D4 retting environments. Typically with short length sequences there is a loss of fidelity when going lower in the taxonomic arrangement (*i.e.* more sequences start being unclassified), however looking at this specific order, *i.e.* Bacillales at a more specific taxon most of the organisms are assigned to the genus *Paenibacillus*. This leads to the conclusion that this is one of our starting isolates used to inoculate the retting environment, *i.e.* *Paenibacillus* DP2 since it was not present in the others, *i.e.* C1D4 & C2D4. *Paenibacillus* has been a promising organism in the field of microbial retting due to the fact that numerous isolates have been found to produce varying kinds of pectinases showing activity against highly methylated pectin (26), pectate lyase (27) and hydrolytic enzymes(92). Additionally *Paenibacillus* has been shown to produce antibiotics which may have led to the changes in the microflora of the different microbial retting environments (112).

3.4.2 Nitrogen as a Possible Factor and Nitrogen-Fixing Clostridia

In a confirmatory experiment, a small clone library of near full-length 16S fragments was generated and 20 random clones were sequenced for higher level identification and as a quality control check for our methodology. The results identified an especially interesting group of

organisms that clustered within the group of nitrogen-fixers related to *Clostridium* sp. UsIt101-1. This group of nitrogen fixing bacteria was previously described by Miyamoto and Minamisawa (108, 113). They reported that as oxygen levels decreased, the anaerobic nitrogen-fixing consortia (ANFICOs) were able to establish and begin fixing nitrogen. Nondiazotrophic bacteria found in their experiment were *Bacillus* sp. The addition of *Bacillus* and *Paenibacillus* might have selected or supported the growth of ANFICOs in the retting solution, thus explaining why the experimental retting communities of E1D4 were so different from C1D4 and C2D4. The identification of this group and their apparent dominance in the experimental reaction, which lacked the addition of the highly eutrophic pond water (autoclaved), suggests that this retting solution was nitrogen limited. If so, this finding suggests that manipulation of nitrogen levels could serve as a primary mechanism to control and manipulate the microbial community structures during the retting process. This conclusion is consistent with the findings of Banik et al., who showed that the retting process could be accelerated by the addition of nitrogen (85). The results further suggest that this group of bacteria is naturally associated with kenaf, either as surface associated or potentially as endosymbionts. Given their apparent importance during retting, this line of research warrants further investigation.

3.4.3 PCR Bias Associated with 16S rRNA Analysis

KMS plant sets elucidated the impact of PCR-based error associated with random changes within the PCR reactions. The general outline of our experiments is shown in Figure 3. E1D4-KMS-B1 and E1D4-KMS-B2 were biological as were C1D4-KMS-B1 and C1D4-KMS-B2. Each retting container had three pseudo-replicates removed, DNA extracted, PCR amplification of

the full-length 16S, nested PCR with barcodes for each biological replicate, quantification, and final pooling for insertion into the Ion Torrent workflow. When looking at the phylogenetic diversity at the phylum level we see a distinct similarity between the two individual biological replicates for each group, i.e. E1D4-KMS and C1D4-KMS (Figure 3.3 and 3.4). The results indicate that PCR-based error was effectively minimized in the process, which we found to be robust among the major group of bacteria. However, it should be noted that PCR-based *bias* was not expressly investigated here. Additionally, when comparing the original KOK samples with E1D4, there is still the large presence of the Firmicutes leading to the conclusion that regardless of the natural flora on the plant or its treatment, the detected community structure remains highly similar.

3.4.4 Rarefaction is Good, but Special Attention should be Given to Others

Microbial diversity is essential to understanding a system. But it is not only reliant on taxonomic or phylogenetic studies and special attention should be given to the diversity estimates such as the rarefaction curves. The threshold was set at 97%, which is currently the standard for describing differences at the species level. Rarefaction curves while providing information with regards to OTUs and sampling amount doesn't show the complete picture as well as the 97% thresholds have limitations in themselves.

The 97% threshold with regards to the 16S rRNA delimitates the line between where the science community distinguishes a species. Our previous understanding of bacterial isolates have been phenotypic based, i.e. biochemical tests, but as more molecular tools have become available DNA-DNA hybridization and the like have been invaluable contributors. The idea is

that DNA is allowed to create a hybrid with other like organisms' DNA in which the binding percentage determines how similar the organisms are at a DNA level. Mechanistically this means similarity in gene content and position as well as similarity in nucleotide compositions in the genes. Other means of understanding microbial diversity such as UniFrac analysis, inverse-Simpson, and principle coordinate analysis will be utilized in the future.

3.4.5 Final Thoughts

For the production of high-value, green composite materials, what is ultimately required is the cost-efficient production of fibers with consistent, uniform properties. These separated fibers can be used in downstream applications such as composites. Composites in this context require fibers be embedded in a resin. Each resin has its own properties, and therefore reacts differently with different fibers (20). Future work will investigate if different microbes can be used to generate fibers with modified properties (*e.g.* surface hydrophobicity) that can be tuned for specific resins and/or applications. Inclusion of bio-based resins made from polylactic acid or polyhydroxyalkanoates would result in completely green and biodegradable composites for a number of applications (8). Current efforts on river water retted kenaf in these polymers have been reported.

CHAPTER 4

TIME-COURSE STUDY ON THE EXPERIMENT ENVIRONMENT

4.1 Introduction

Bacterial retting of fiber yielding plants has recently been an area of increased study. The retting process is when microbial constituents breakdown the heteropolysaccharides from harvested material thereby releasing the fibers. The fibers can be incorporated in a variety of applications, including as an addition to biodegradable composites. Previous studies have been analyzed the retting solution for its bacterial constituents, but none have tracked the microbial community through the retting process. In the previous chapter, the microbial retting environment was described as different test environments, some recreating a natural environment, one testing the microbial constituents on the plant fiber and then an augmented version wherein pectinase producing bacteria were added. This aspect of the dissertation aims to take the augmented version of the microbial retting environment further and track the progress through initial day to completion of retting, i.e. day 1 through 4.

4.1.1 Time Course Study of Metagenomic Samples

Here we describe efforts to follow the development of the bacterial retting community through time by 16S rRNA amplicon sequence analysis. Experiments were performed at lab scale with the addition of three previously isolated pectinolytic bacterial isolates: *Bacillus* DP1, *Paenibacillus* DP2, and *Bacillus* K1. Plant material was removed at Day 1 through Day 4 and DNA extraction was performed on the surface adhering bacteria. A full length 16S PCR was performed using 27F and 1492R, which in turn provided the template for a nested PCR with

primers 786F and 939R. The resulting product was subject to the Ion Torrent PGM workflow including emulsion PCR and sequencing.

4.2 Materials and Methods

Setup of the following experiments follows generally the previous chapter's experimental design (E1). With the exception of excluding the various control retting vessels, other aspects remained the same. The main goal of this experiment was to elucidate the change in bacterial constituents in the retting solution as compared to a temporal change.

4.2.1 Preparation of Kenaf

Kenaf was provided by C. Webber (USDA) from fields grown at the USDA agricultural research station in Lane, OK. Stalks were harvested at approximately 185 days post-planting and the harvested kenaf bark was manually stripped from the core. The first bottom meter of the kenaf was used for subsequent experiments and was cut into 2 cm pieces. The kenaf retting vessels were generally set up as follows: 750 mL of MilliQ water, 250 mL of autoclaved pond water, 20 g of plant material and the bacterial inoculum which consists of 50 mL each of fresh culture of *Bacillus* DP1, *Paenibacillus* DP2, and *Bacillus* K1 spun down and resuspended in MilliQ water.

4.2.2 Setup of Time-Course Retting

50 mL cultures were prepared using LB, which consisted of 10 g of Tryptone, 5 g yeast extract, and 10 g NaCl. The LB was then adjusted to pH 7.0 using 1 M NaOH. 50 mL of LB were

aliquoted into 250 mL erlenmeyer flasks and autoclaved for sterilization with the following conditions: 121 °C, 15 PSI, and 15 m. Each flask was inoculated separately with each of the isolates listed above and incubated for 18 h in a shaking incubator at 225 rpm and 37°C.

The kenaf retting vessels were incubated in an incubator at 30°C. Sterile forceps were used to remove samples at 24 hour increments, rinsed in milliQ water, and subsequently used for direct DNA extraction from adherent microbial species by bead beating in suspension buffer using the Mo-BIO fecal DNA kit and Fisher Vortex Mixer, as per the manufacturer's instructions. After extraction, DNA samples were quantified by spectrophotometry using a Nanodrop and used for downstream processes. Full length 16S PCR was performed using standard protocols with 27F and 1492R. ~1.5 Kb fragment was amplified from each of the respective sources. After confirmation, the PCR product was cleaned using Agencourt Ampure XP beads and quantified on a Nanodrop 1000. The diluted, cleaned product served as the template for the second Ion Torrent-specific nested 16S amplification using fusion primers IonA-E786 Fwd and IonP1-E989Rev derived from Baker et. al 2003 (44).

PCR were set up using the previous parameters replacing the gDNA with newly synthesized full-length 16S amplicons. Ion Torrent-specific primers produced a 210 bp band using the following 2-step PCR thermal cycler protocol: 98 °C initial denaturation for 3 m, 98 °C denaturation for 15 s, 61 °C annealing and extension for 15 s, and repeated for 25 total cycles, with a final extension for 5 m. The resultant PCR products were purified following the protocol as outlined by Ion Torrent. Product quality was confirmed on an Experion Automated Electrophoresis System. Cleaned PCR products were diluted to the appropriate nanomolar concentration (8.4) and inserted into the emulsion PCR. Each sample was run separately

through the workflow on the Ion Torrent, i.e. each day was on a separate 314 chip. Samples were prepared and amplified by emulsion PCR using the Ion Torrent OneTouch System per the manufacturer's instructions. The resultant beads were enriched on the Ion Torrent ES prior to loading on model 314 10Mb sequencing chips with 100 bp chemistry. After quality control and removal of the primers through RDP Pyrosequencing Pipeline, 100,000 sequences from each sample were inserted into RDP Classifier to give a taxonomic match with a bootstrap cutoff set at 50%.

4.3 Results

4.3.1 Overall Phylogenetic Diversity at the Phylum Level

After quality control for accuracy for forward primer reads, the resulting FASTA was inserted into RDP Classifier and placed at 50% confidence level. Microbial community analysis showed the phylum Firmicutes as the major constituent throughout the entire process, Days 1 through 4 (55 to 94%), phylum Proteobacteria showed a progressive loss from Day 1 to 4, 36% to 5%, while phylum Bacteroidetes showed a consistent trend throughout except day 4, 0.1% as compared to 1% for all other days (Figure 4.2).

4.3.2 Time-Course at the Family Level

Upon examination at a finer taxonomic level, there showed a rapid loss of *Bacillus* DP1 and K1, and a more gradual loss of the family *Paenibacillaceae* 1 as the time course progressed (Figure 4.1). These groups were replaced by members of the family *Clostridiaceae* 1. This study showed the loss of introduced *Bacillus* spp. likely corresponding to the shift from aerobic to

anaerobic conditions in the microbial retting process. Loss of facultatively anaerobic *Paenibacillus* may be the result of nitrogen limitation, as suggested by the rise in nitrogen fixing *Clostridia*. Understanding these changes in the retting process will facilitate continued optimization of process efficiency and fiber quality.

Overall trend in the environment shows a shift from a relatively diverse group of organisms collapsing into domination by the single phylum Firmicutes, which is represented by the two families of *Clostridiaceae* 1 and *Paenibacillaceae* 1. Particularly interesting is the trend that occurred on the day 3 to 4 transition in that there was a loss of the family *Paenibacillaceae*, 2% shifting upwards back to 21% on day 4. This shift is concomitant with sudden appearance and disappearance of members of the *Comamonadaceae* on day 3. Further effort will be required to understand underpinnings of this result.

Percent Comparison of Experiment Days 1 - 4 at the Family Taxonomic Level

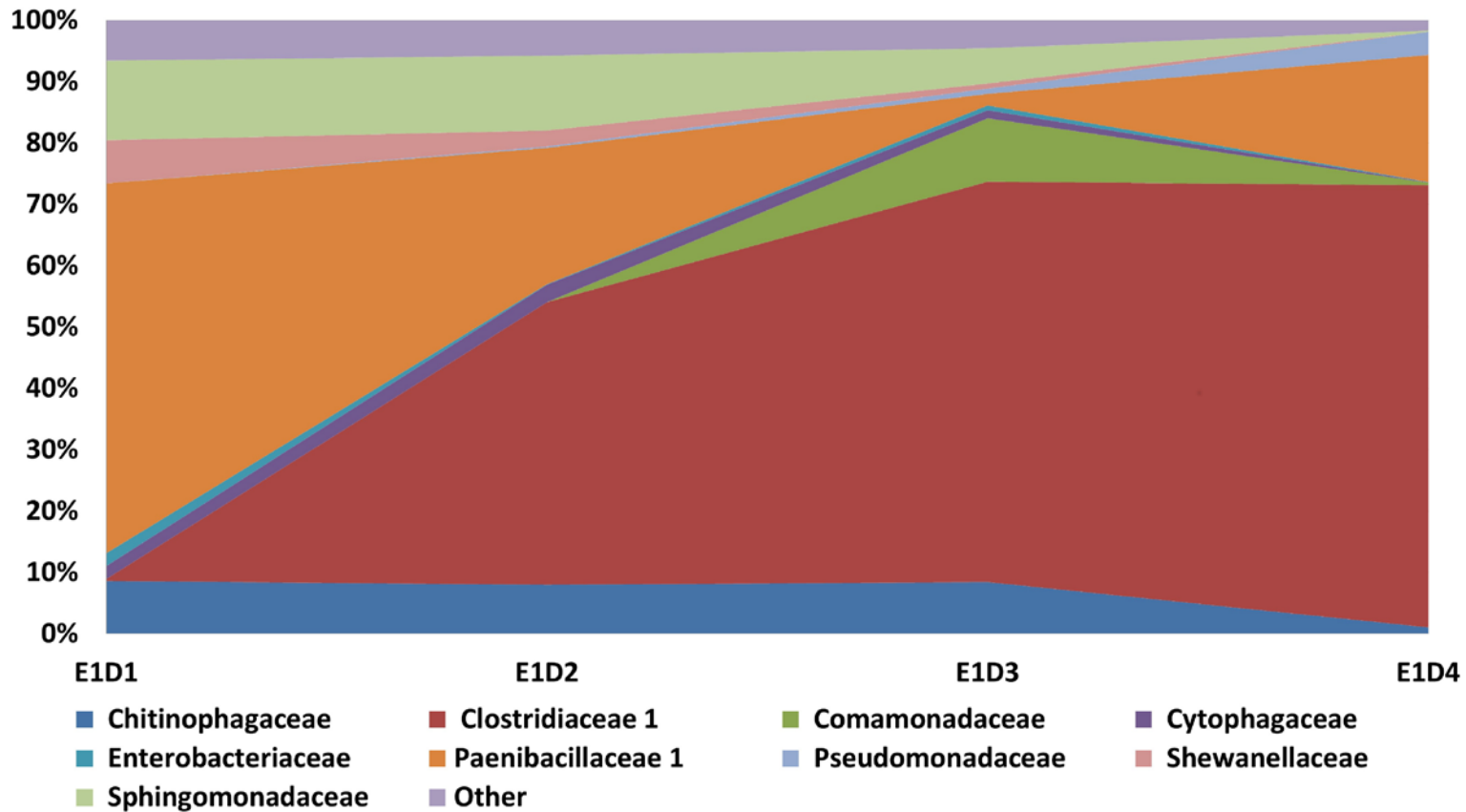


Figure 4.1. Phylogenetic comparison of time-course study.

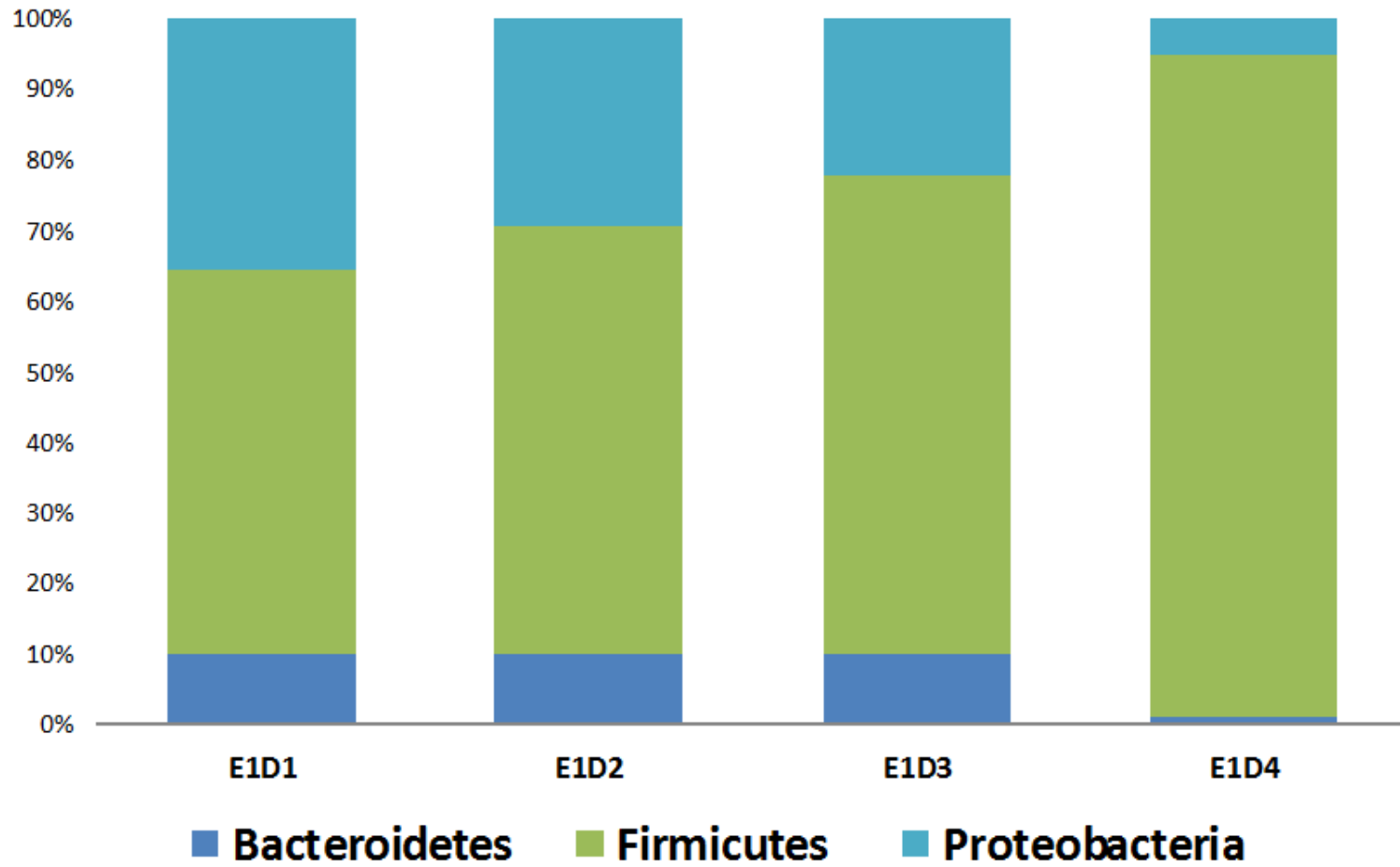


Figure 4.2. Phylum

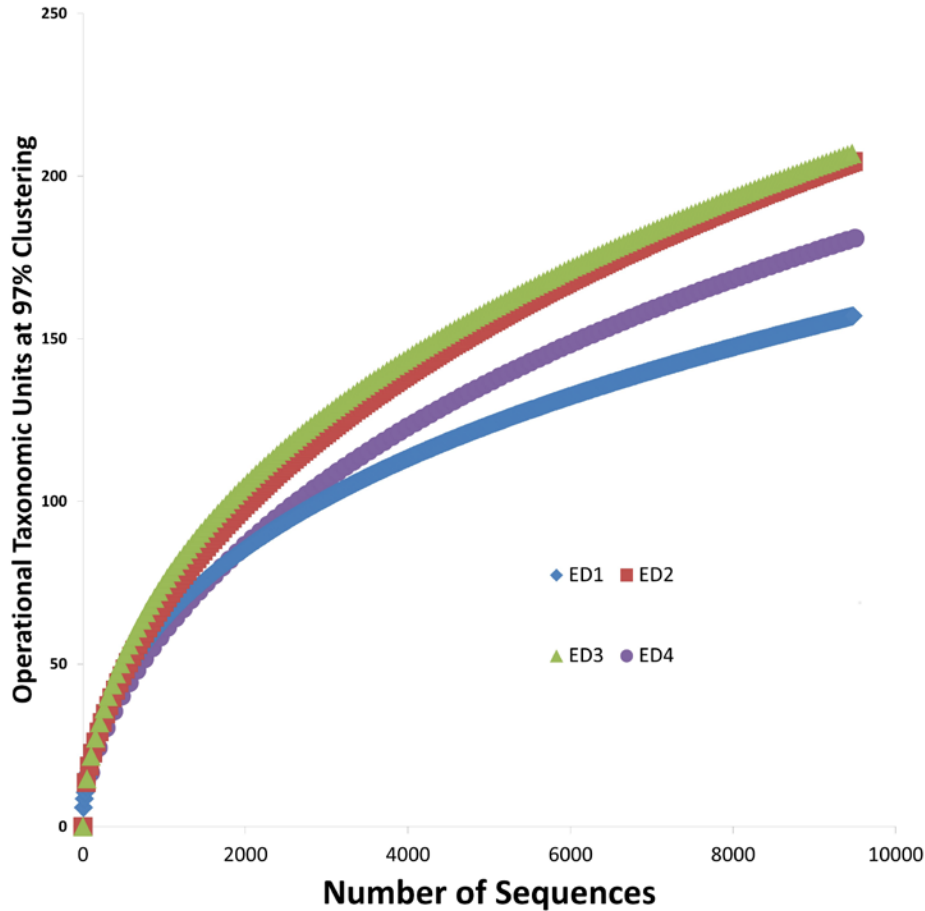


Figure 4.3. Rarefaction curves set at 97% similarity indicate of species level comparisons.

4.3.3 Diversity Estimates of the Time-Course Retting Experiment

Also of interest are the diversity estimates offered by the rarefaction curves, which compare number of sequences and operational taxonomic units (OTUs). ED1 has the lowest value in terms of OTUs at 156, while the largest is represented by ED3 at 206 (Figure 4.3). This is likely the result of limited colonization of the bark ribbons after only one day. Both ED2 and ED3 had the largest amounts of OTUs as well as some of the largest percentages of *Clostridia*. The continued dominance of this group along with the relative decrease in OTU diversity by day 4

suggest that a relatively small group of bacteria have begun dominating the environment by this time.

The presence of high numbers of *Clostridia* throughout the experiment (which were not part of the original inoculum) suggest the organisms were readily present on the or on the plant material. Experiment confirms shift from aerobic to anaerobic environment. This was a similar finding in previous research, but this experiment showed the actual time-course change during the course of the retting. Additional testing of the *Clostridium* isolates from day 4 suggests that nitrogen fixation may be occurring after this transition. The pectin-producing consortium appears to be pushing the population to lower diversity dominated by the phylum Firmicutes as described in other control experiments (Figure 4.2). This was confirmed by analysis of the full-length 16S gene, which showed a similar match to other nitrogen-fixing bacteria within the genus *Clostridium*. *Clostridium* species seem to be an important component of the kenaf-retting environment. Future work will attempt to confirm their presence on the plant material. Increased performance of naturally retted kenaf composites vs. standard materials warrants further investigation of the retting environment.

The overall shift in the communities of the time-course retting process yields the questions of which organisms are pushing the communities to be highly dependent on the genus *Clostridium*. Specifically, a dropout experiment can be performed in which each isolate, i.e. *Bacillus* DP1, *Paenibacillus* DP2, and *Bacillus* K1, can be added independently and the communities could be checked to see if the similar shift is conserved or whether it requires the combination of the isolates to have this drastic shift in the day 4 communities.

CHAPTER 5

DROPOUT STUDIES ON THE MICROBIAL RETTING ENVIRONMENT

5.1 Introduction

Global environmental concerns have led to a growing interest in renewable resources such as plant-based fibers. Beyond textiles and cordage, plant fibers have the potential for incorporation into renewable, bio-based composite materials for the building and manufacturing sectors. Successful commercialization of fiber production requires optimization of fiber extraction. Retting is the traditional method of fiber extraction, whereby endogenous microorganisms break down heteropolysaccharides to release fiber bundles. Previous studies have analyzed the retting solution for its bacterial constituents, but none have followed changes in the microbial community through the retting process. This research aims to track the bacterial components of the retting community through time, and determine the effects of bacterial augmentation with isolated pectinolytic bacteria using next generation sequencing of 16S rRNA gene amplicons.

Specifically, this area of research targets understanding the contribution of the different isolates separately, that is to determine whether one isolate is pushing the microbial retting process or that it is in some conjunction of the three. Three experimental vessels will be setup in the same fashion as the previous experimental conditions, but with each isolate being used separately. The isolates used will be *Bacillus* DP1, *Paenibacillus* DP2, and *Bacillus* K1.

5.2 Materials and Methods

5.2.1 Long Amplicon Implementation and Update to Ion Torrent Workflow

With the advent of increasing sequencing length with updates to the Ion Torrent workflow the need to implement longer reads to gain increased information from the phylogenetic studies is a necessity. Specifically, the introduction of 200 bp average sequencing length gave the ability to sequence longer amplicons, that is use of 505F and 806R which was taken from the same source as the previous primer sets(44) . Much like the previous chapters the Ion Torrent specific tags needed to be added when the primers were synthesized and the actual primer sequences are shown in Appendix 4. The longer read lengths allowed for the introduction of IonXpress barcode tags and were included into the primer set, which allows barcoding sequences up to 96 unique samples.

5.2.2 Extraction of DNA and Amplification of 16S V5 Region

Samples of bark fiber were removed at day 2 and 4 (48 and 96 h, respectively), rinsed in milliQ water, and used for direct DNA extraction from adherent microbial species. The DNA was accomplished by bead beating in suspension buffer using the MO BIO Fecal DNA kit (MO BIO Laboratories, Inc., Carlsbad, CA) and Fisher Vortex Mixer (Thermo Fisher Scientific, Waltham, MA) as per the manufacturer's instructions. The gDNA solutions were labeled and stored at -20°C.

After extraction, DNA samples were quantified by spectrophotometry, with a Nanodrop 1000 (Thermo Fisher Scientific Inc., Waltham, MA) and used for downstream reactions. PCR amplification of the variable region four (V4) of the 16S rRNA was performed in a direct fashion

as compared to the previous experiments in which it was a nested two-step PCR. This was accomplished by both the increase in length that the Ion Torrent PGM provided going from 100 bp read lengths to 200 and the use of primers that covered the V4 region, 515F and 806R.

25- μ L PCRs were set up using the following parameters: 5 μ L of 5X HF Buffer, 200 μ M of dNTPs, 250 nM of 515F, 250 nM of 806R, 1.5 μ L of 10X BSA, 0.25 μ L of Phusion DNA polymerase (NEB), 1 μ L of gDNA, and was brought up to a final volume of 25 μ L with sterile water. Ion Torrent specific primers produced a 210 bp band using the following 2-step PCR thermal cycler protocol: 98°C initial denaturation for 3 m, 98°C denaturation for 10 s, 52°C annealing for 15 s, 72°C extension for 15 s, and repeated for 25 cycles, with a final extension at 72°C for 5 m. Each sample was performed in triplicate. The resultant PCR products were purified using Agencourt Ampure XP beads (Beckman Coulter Genomics, Danvers, MA) following the protocol as outlined by Ion Torrent, and quantified on a Nanodrop 1000. Product quality was confirmed on an Agilent 2100 Bioanalyzer (Agilent Inc., Santa Clara, CA). The Bioanalyzer system is an automated electrophoresis system, which utilizes microfluidics to determine both the concentration and distribution of a DNA sample. The Bioanalyzer High Sensitivity DNA Analysis kit was used which allows detection of samples down to 5 pg/ μ L and a size range of 50 – 7 000 bp. This is essential for the downstream Ion Torrent workflow because of a need for an accurate and precise molarity and distribution of the size of the fragment. The Bioanalyzer showed a ~360 bp fragment, which is the correct size when accounting for the 16S fragment as well as the Ion Torrent specific tags.

Cleaned PCR products were diluted to the appropriate concentration, 8.4 nM. Each barcoded sample was pooled together prior to insertion into the Ion Torrent workflow. Samples

were amplified by emulsion PCR using the Ion Torrent OneTouch System per the manufacturer's instructions. The resultant beads were enriched on the Ion Torrent ES prior to loading on to model 314 sequencing chip. This process was performed independently for each condition.

5.2.3 Setup of High-Performance Computer Server using TALON and Mothur

Bioinformatics is an essential component of this research. As was previously described, the general workflow of the small fragment 16S analysis, utilized a variety of bioinformatics suites that could adequately handle the smaller 100 bp fragments of the 16S rRNA. However, as data sets increase in size, i.e. throughput, and also length, 100-200 bp increase, additional remedies must be made to the bioinformatics pipeline currently utilized. We utilized the Talon HPC computing provided by UNT in order to quickly analyze large datasets.

Recent updates to the Ion Torrent server required a conversion of the raw BAM files to SFF, but it can be accomplished on the server. The BAM format is a binary form of the SAM format. The SAM format is a form of sequence data that can be stored using a series of tab delimited ASCII columns. When the files are converted they can be uploaded to the server on Talon utilizing SSH. Specifically, we used Filezilla for SSH transfer to due its ease of use, it being free and open-source, as well as its ability to be used on a multitude of platforms such as Linux, Mac OS, and Windows. Once uploaded, analysis can occur on the Talon server.

Mothur is one of the most-highly cited bioinformatics programs utilized for 16S analysis. Previously, only the command line style mothur was available and it covered all the major operation systems available, but more recently there are graphical user interface programs in

the current releases. The idea behind mothur is to have a single program that can be used to analyze the complete spectrum of 16S data that is from classification to the end diversity estimates. Additional programs must be used to visualize the data, but mothur provides all the output files that can be easily converted and used in programs like SigmaPlot and Excel.

The script utilized in the 16S pipeline is in Appendix II. Sffinfo converts the files from the uploaded sff to FASTA and quality files required for downstream processing. The flow variable is set to F indicating false, which does not result in a flowgram output from mothur as this is not a component in the Ion Torrent. Trim.flows takes the files outputted from the previous command and starts the quality control process on the files from sequencing. The oligos file is essential to processing because it tells mothur what each of the barcode files are and the name of each sample. This file must be formatted exactly correct, i.e. tabs in between the columns, and also have the exact Ion Torrent specific oligo tags to correctly sort the FASTA file from the analysis. Additionally, the reverse primer has been left off since the current sequencing length of the Ion Torrent do not always reach the end of the reads thus the reverse primer is left off. Ambiguous nucleotides (N) are set to the standard zero, which kicks out any sequences that have an N in the sequence data, maximum homopolymers are set to 8 due to the inherent difficult that next-generation sequencing techniques have with homopolymers, barcode difference is set to one change in the nucleotide out of 10, and primer difference is set to two changes again for quality control. Quality control of the sequences must be all of the above criteria or it will be removed from analysis. The output files can be extremely large from the trim.seqs output and even to the point that HPC can be slowed down and analysis can take days, thus there is a unique sequences command (unique.seqs), which concatenates the data to only having sequencing

that are unique in the downstream analysis. This can reduce datasets on the order of 5-fold and lowers the downstream analysis time from 5-fold to 25-fold in this example. Alignments are then applied to the outputted and quality checked sequences. The alignment program in mothur (align.seqs) aligns the output of the QC commands, i.e. the user-supplied fasta-formatted sequence file, to the fasta-formatted template alignment file. The general approach that mothur takes is to find the closest template for each candidate using kmer searching, blastn or suffix tree searching. Then a pairwise alignment is made between the candidate and de-gapped template sequences using a variety of alignment methods. Finally, the program reinserts the gaps to the candidate and template pairwise alignments using the NAST algorithm so that the candidate sequence alignment is compatible with the original template alignment. Mothur requires a user provided fasta-formatted template alignment file and we have specifically choose the SILVA database for the alignments made in this project. There are numerous alignments available such as greengenes and ARB, but the SILVA database is suggested due to its ability to quickly align files and is not as inaccurate as the greengenes database (114). The default parameters on the align.seqs command were utilized which are kmer searching with 8mers and the Needleman-Wunsch pairwise alignment method with a reward of +1 for a match and penalties of -1 and -2 for a mismatch and gap. Even with the utilization of multiple cores on the alignments, the finished aligned outputs can take hours to a day to finish.

Screen.seqs allows the user to determine what sequences are kept after analysis. Post-alignment there is a user-directed step that seeks to keep as many long sequence reads as possible, but also maintain the exclusion of small sequences, without unnecessarily removing

large amounts of sequences. The alignment position number that was determined to adequately meet these criteria was a start position of 13 862 and an end position of 22 549. Filter.seqs command is then ran to remove a “.” which indicates missing data and “vertical=T” will remove any columns that contain exclusively gaps within. Pre.cluster command again, similar to the unique.seqs, reduces the data so that will split the sequences by group and then within each group it will pre-cluster those sequences that are with 1 or 2 bases of a more abundant sequence.

A chimera check is initiated on the sequences removing any presumptive chimeras produced by the initial 16S PCR. The chimera check utilizes the UCHIME program, which will first divide the sequences by group and then check each sequence within a group using the more abundant sequences as reference sequences (115). There are benefits to this in that this approach allows the more abundance sequences to be used as a reference database as the idea is that the chimeras should be rarer than the more abundant parent sequences. If the program determines a sequence to be a chimera in one group, then it considers it a chimera in all samples and will be removed.

Classifying sequences is the main way that the microbial communities are understood. The classify.seqs command accomplishes this by means of a naïve Bayesian classifier utilizing the RDP database as the taxonomic match (116). The classifier can rapidly provide taxonomic placement based on the 16S rRNA sequence data and accurately classify bacterial and archaeal sequences down from domain to genus with a confidence estimate for each assignment. The naïve Bayesian classification refers to the idea that the data attributes are independent and even when the independency is violated it is still efficient at assignments (117). The outputs

from this classification can be edited in Excel and allowed to remove contaminants such as mitochondrial and chloroplast DNA, which were added to the script.

5.2.4 Diversity Analysis of Data

The final steps of the analysis are with regards to splitting up the sequences into OTUs and then doing diversity estimates. This is accomplished by using the `dist.seqs` command which will generate a distance matrix using a cutoff determined by the user, in this case 0.03, which means it will remove any pairwise distance larger than 0.03. This cutoff is usually attributed to the sequence difference to determine different species, i.e. a 16S rRNA sequence that has above 97% sequence similarity is considered to be the same species. After the distance matrix has been computed from the `dist.seqs` command the actual clustering of the OTUs is accomplished using the `cluster` command in `mothur`. During diversity estimates and analysis there it must be taken into account the sampling amount for each variable. During sequencing runs, even if the sequences have been diluted to equimolar amounts there are still discrepancies in the amount of sequences sequenced from barcode to barcode. This can potentially have a large range of differences between samples so a subsampling must be applied before diversity estimates are made. After clustering, the samples are subsampled using the lowest number of sequences for a given barcode, and is applied to all the barcodes, normalizing the data to the lowest number. This step is essential to getting diversity estimates that are valuable because if a particular sample is high in terms of sampling, i.e. number of sequences, it can artificially inflate the diversity estimates of the “high” sample.

5.2.4.1 Alpha Diversity

There are multiple ways to looking at diversity estimates. There are two main genres of determining diversity and richness of OTUs within different environments composed of the alpha and beta diversity. The alpha diversity estimates is the average OTUs in an environment at a local scale (118, 119). The local scale within these experiments would be the different microbial retting environment being tested. The beta diversity is the differences between the different environments, i.e. site A compared to site B. In the previous chapters the main measure of diversity was through the use of rarefaction curves. Rarefaction curves gives a means of comparing the richness observed across different environments. Generally, as the sampling effort increase, that is the number of sequences, the number of OTUs will increase, but we aimed in the previous studies and the current to keep the sampling efforts uniform across samples. When the rarefaction curves reaches parallel with the x-axis the environment has been sampled adequately, within reason. In another phrasing the environment has been appropriately sampled to where an increase in sampling efforts would not yield an increase in novel OTUs detected.

There are two additional alpha diversity estimates that will be used: Chao1 richness estimate and inverse-Simpson diversity estimator. Chao1 is accomplished within the Mothur suite by utilizing the command collect. Single. This richness estimate gives the minimum richness within a community and with increasing sampling efforts the value will continue to rise. This value utilizes the sample concept of OTUs, but only takes into account the overall number of OTUs and the number of OTUs with only one sequence as well as two sequences in the calculations.

The second alpha diversity used is the inverse-Simpson diversity estimate, which is another way to describe microbial diversity within a sample (120). Additionally, this value remains relatively stable, thus is a better way of comparing different environments as compared to the Chao1 estimates. The inverse-Simpson diversity estimate is an indication of the richness within a community with uniform evenness that would have the same level of diversity. In another iteration it is the probability that upon randomly observing an individual from a group, the individual has already been selected. With a calculation of the Simpson diversity estimate, with an increasing value this means that the sample is less diverse, which is not intuitive thus the reciprocal of the value is taken (121, 122).

5.2.4.2 Beta Diversity

Beta diversity is the differentiation of OTUs among environments (118). The main indices that will be used for beta diversity analysis is UniFrac and principle coordinate analysis (PCoA). Principle coordinate analysis is a way for multidimensional analysis of the similarities or dissimilarities of data.

5.2.4.3 Weighted vs. Unweighted UniFrac Analysis

The UniFrac methods of analysis are a means to test whether two or more communities have the same structure (123, 124). It calculates the distance between community structures using phylogenetic information. Specifically, this is accomplished by randomizing the sample category of each taxa on the phylogenetic tree and generating a distribution of UniFrac distance values. Using these distance values, the statistical significance can be calculated using the

distances between samples. The main difference between weight and unweighted forms of UniFrac is that the weighted aspect utilizes the relative abundance of each of the taxa within the specific communities. There is a move to utilize a more generalized form of UniFrac that uses both the weight and unweighted forms, which result in a lessened impact of large abundances of taxa and rare taxa.

5.3 Results

5.3.1 General Sequencing Results from the Ion Torrent Server

Number of sequences per barcode from the analysis ranged from 9 641 to 33 610 with a total number of sequences of overall sequences being 310 638 from a 314 chip. Initial characterization at the phylum level assigned most sequences to the phylum Firmicutes (96-99%), while to a much lesser extent to the phylum Proteobacteria and unclassified (3-4%). During the sequencing procedure, barcode 17 failed, i.e. BK1-D4-2. The chart is formatted to with the following variables: the initial component of the variables refer to the isolate used in the drop out experiment, i.e. BDP1 refers to *Bacillus* DP1, PDP2 is *Paenibacillus* DP2, and BK1 is *Bacillus* K1. The second component refers to the day that the DNA was extracted from and as described previously we chose day 2 (D2) and day 4 (D4). The last component, which is a numerical value, refers to the specific DNA extraction performed in that there were three separate DNA extractions performed for each day and dropout environment.

As shown in figure 5.1 loading density was around 78% for the Ion Torrent chip. From our experiences in the lab a loading density of around 60-80% is the typical load. There were areas of blue, which indicate lower than optimal loading, which can be indicative of bubbles

during the loading process. However, this is a normal aspect of the loading protocol as this must be done by hand and it is inherently difficult to completely negate bubbles from forming within the injection of the ISPs into the chip. During the multiple centrifugations and pipetting there is always the chance of losing ISPs, but the loss of sample is negligible. 68% of the final reads totaling 657 143 were inserted into the mothur pipeline for downstream analysis. 26% of reads were determined to be polyclonal. Polyclonals are caused by possible oversaturation of amplicons during the initial ePCR step of the Ion Torrent workflow. Because the concentration is low (26 pM) only 30% of the ISPs are ideally targeted to be positive post-ePCR. If this factor is changed significantly there is a higher chance of multiple amplicons amplifying on the same ISP within the same aqueous droplet. This results in mixed signals or pH changes, given off during the sequencing by synthesis. In addition the polyclonal bin can be used for generic sequencing errors not only those exclusively caused by polyclonal amplification in the ISPs.

The barcoded samples shown in Table 5.2 show the general statistics with regards to pre and post quality check by means of mothur. There were varying averages in the pre-quality check, which seemed to be as a result of smaller fragments being present in the samples. However, after the sequences were checked for sequencing errors, homopolymers, and quality scores the average length of amplicons was an average of 214 bp, which is the targeted length of the 16S amplicon.

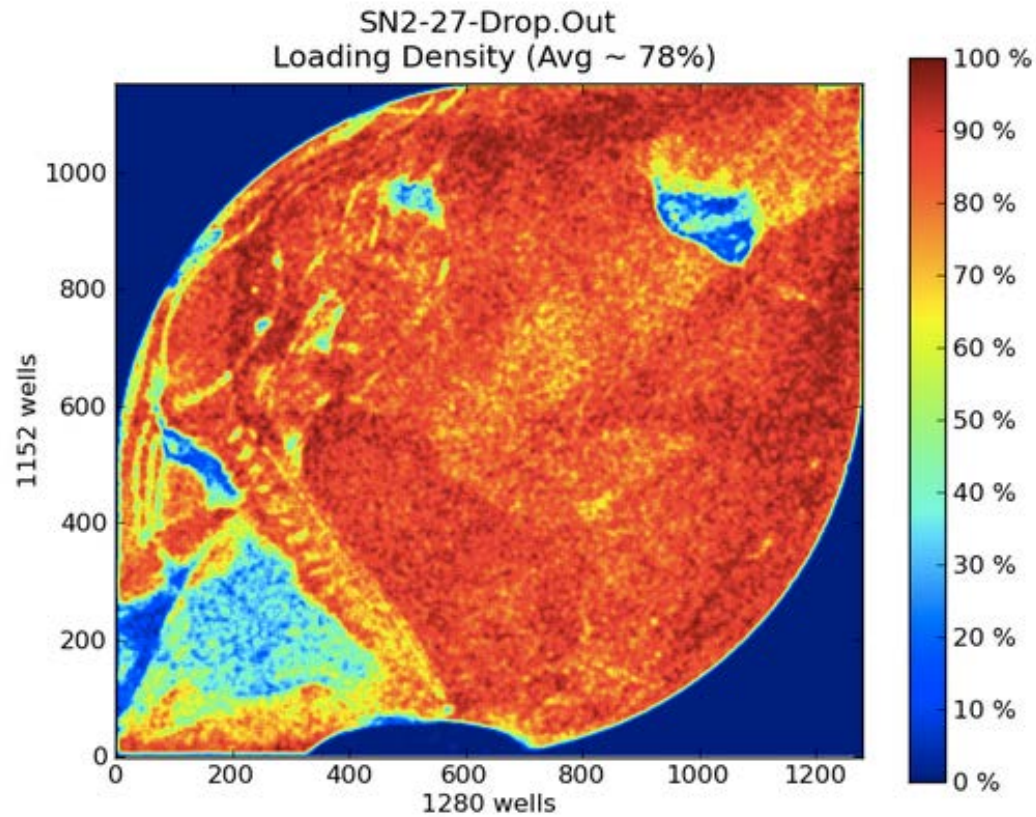


Figure 5.1. Figure shows the output heat map from the Ion Torrent PGM for the dropout experiments. There was an average of 78% load on the sequencing chip. Red indicates a high load (100%) and ranges down to blue, which indicates no live ISPs.

	Count	Percentage
Total Addressable Wells	1,262,519	
▸ Wells with ISPs	984,974	78%
▸ Live ISPs	980,025	99%
▸ Test Fragment ISPs	15,862	2%
▸ Library ISPs	964,163	98%

	Count	Percentage
Library ISPs / Percent Enrichment	964,163	99%
▸ Filtered: Polyclonal	253,249	26%
▸ Filtered: Primer dimer	12	<1%
▸ Filtered: Low quality	53,759	6%
▸ Final Library Reads	657,143	68%

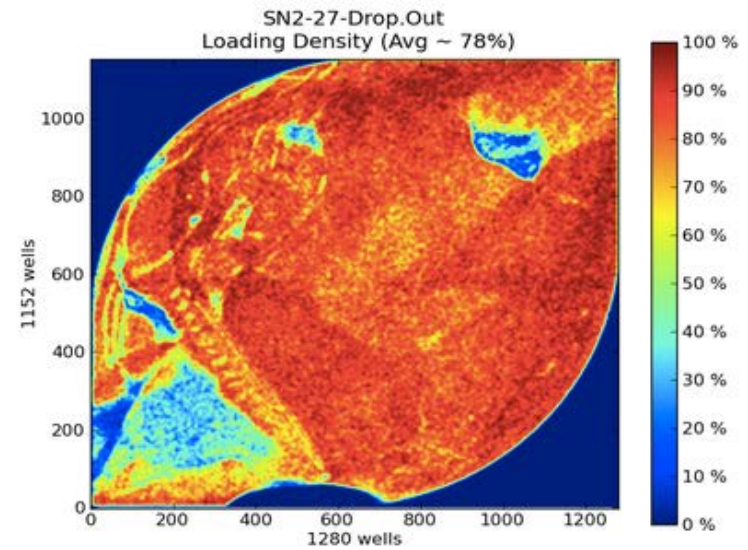


Figure 5.2. General outputs from the Ion Torrent. Includes the initial reads and also the filtered reads base on either polyclonal, primer dimer, or low quality.

Sample	Barcode	Total Number of Reads	Read Length (bp)	Reads for Analysis	Mean Length for Analysis
BDP1-D2-1	1	42 959	151	16 391	214
BDP1-D2-2	2	46 595	175	22 339	214
BDP1-D2-3	3	30 456	196	16 610	214
BDP1-D4-1	4	18 914	198	9 642	214
BDP1-D4-2	5	36 133	202	18 197	214
BDP1-D4-3	6	31 611	213	18 112	214
PDP2-D2-1	7	25 828	201	15 033	214
PDP2-D2-2	8	52 649	221	33 610	214
PDP2-D2-3	9	29 633	204	17 638	214
PDP2-D4-1	10	26 863	212	14 978	214
PDP2-D4-2	11	44 426	156	16 619	214
PDP2-D4-3	12	53 483	201	29 820	214
BK1-D2-1	13	26 834	202	13 492	214
BK1-D2-2	14	47 789	203	24 658	214
BK1-D2-3	15	33 266	200	16 894	214
BK1-D4-1	16	21 563	217	12 840	214
BK1-D4-2*	17	-	-	-	-
BK1-D4-3	18	23 094	218	13 985	214

Table 5.1. Table shows summary of the various barcodes and their subsequent reads and mean read lengths. *BK1-DK-2 failed during the sequencing run, but the specific environment tested, that is Bacillus K1 Day 4, still had two replicates.

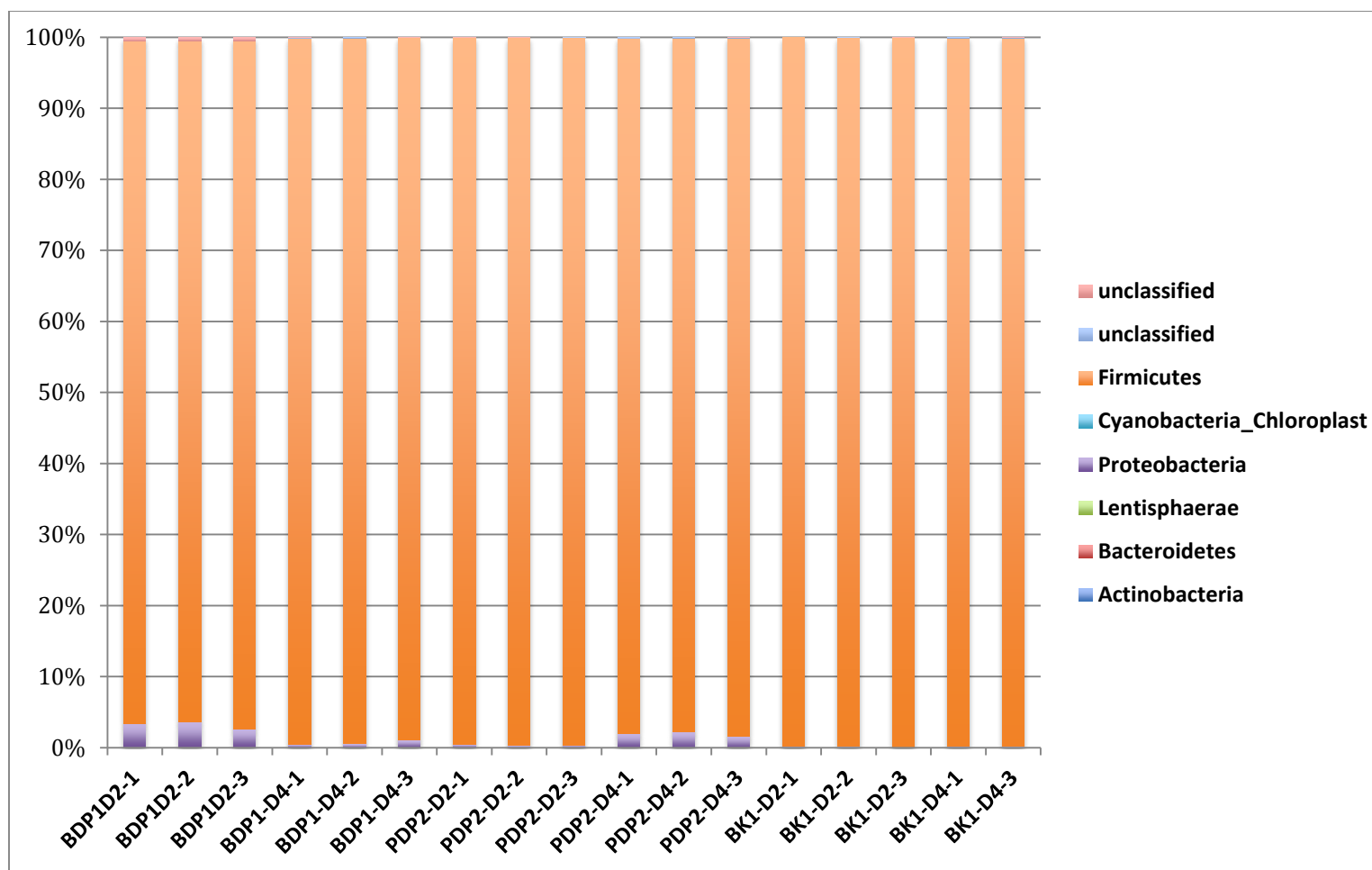


Figure 5.3. Phylum level assignments for all samples. Note that BK1-D4-2 failed during sequencing. The range of the phylum Firmicutes ranged from 96-100%.

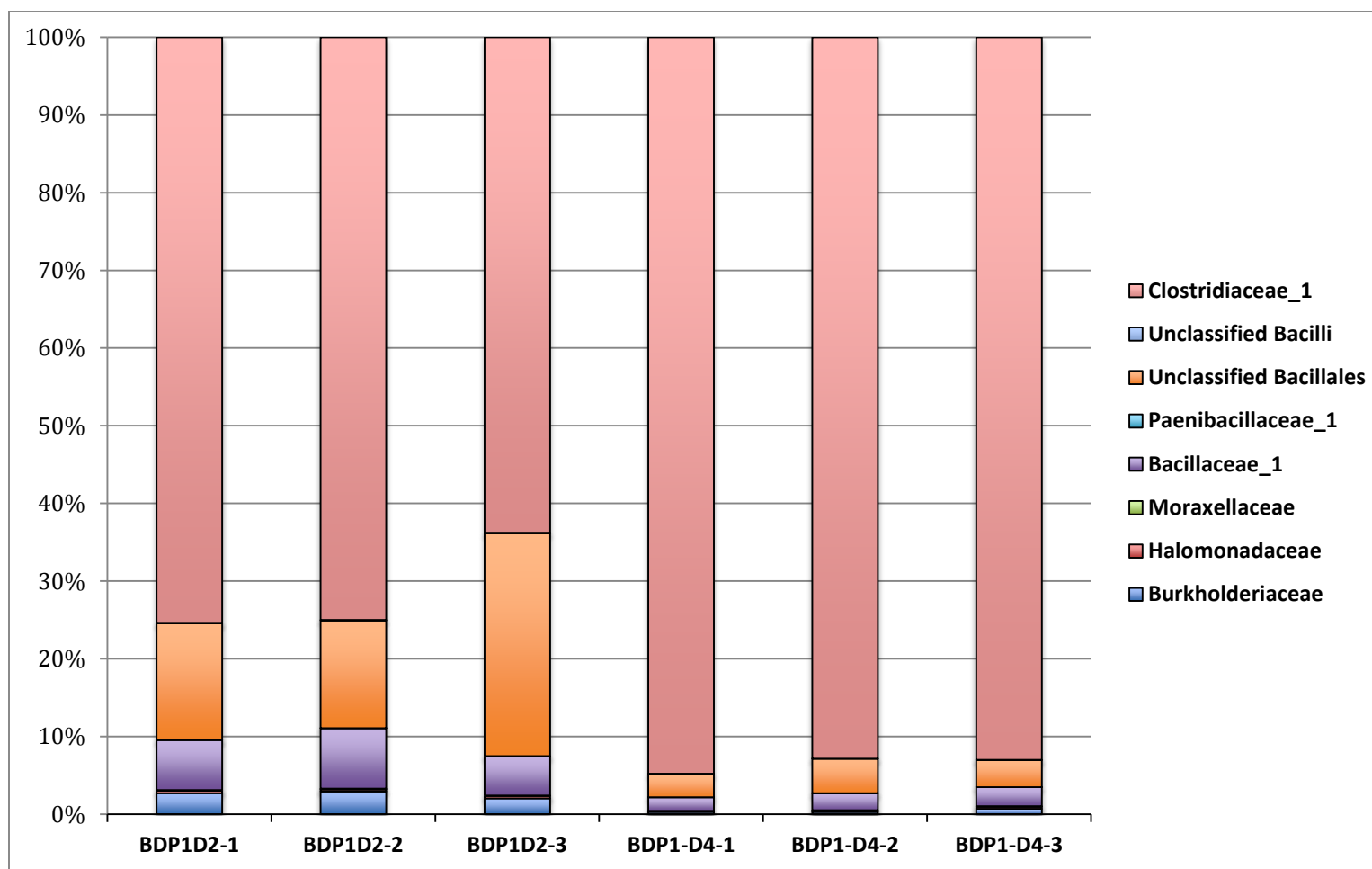


Figure 5.4. Family level assignments of the Bacillus DP1 time-course dropout experiments.

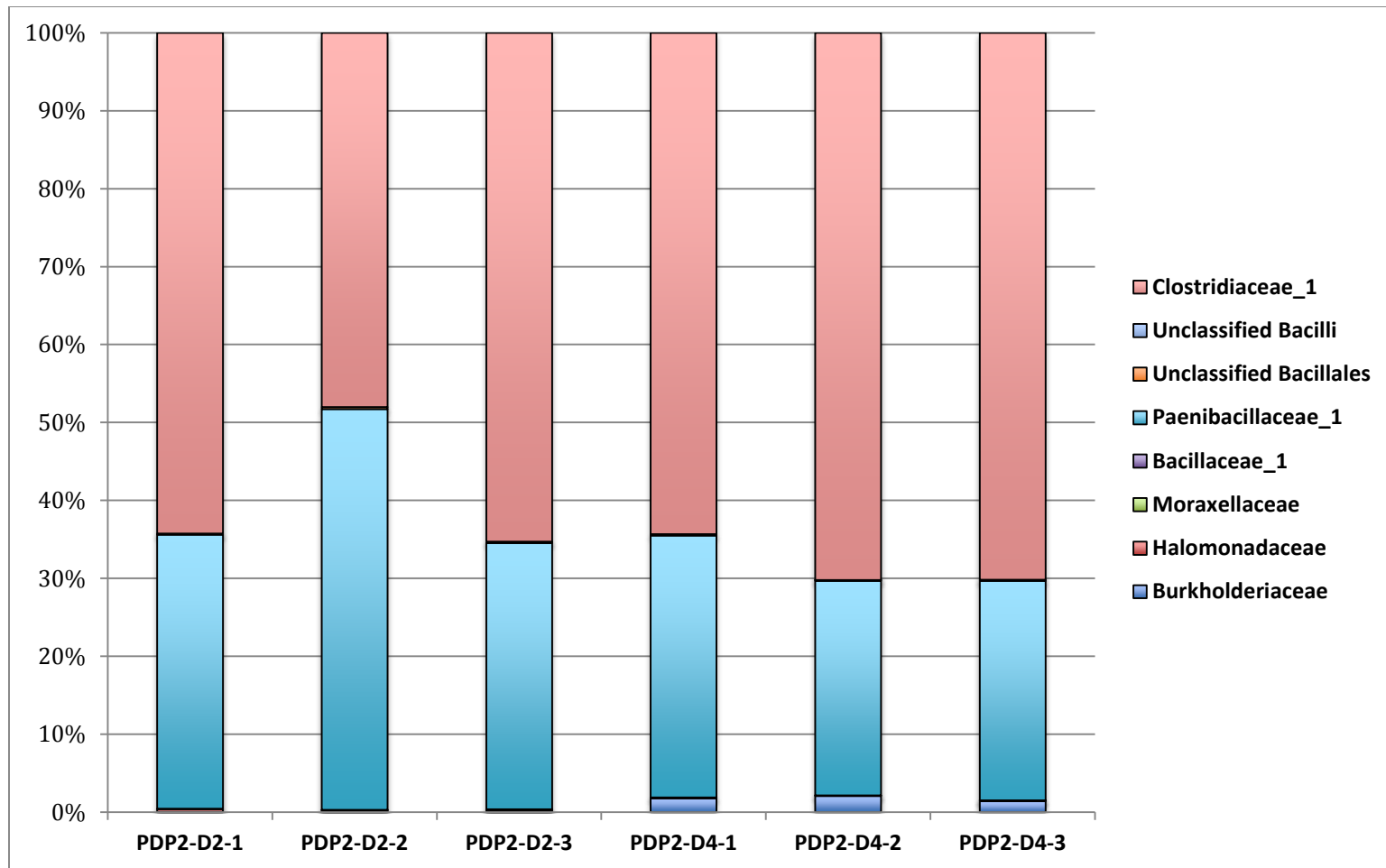


Figure 5.5. Family level assignments of the Paenibacillus DP2 time-course dropout experiments.

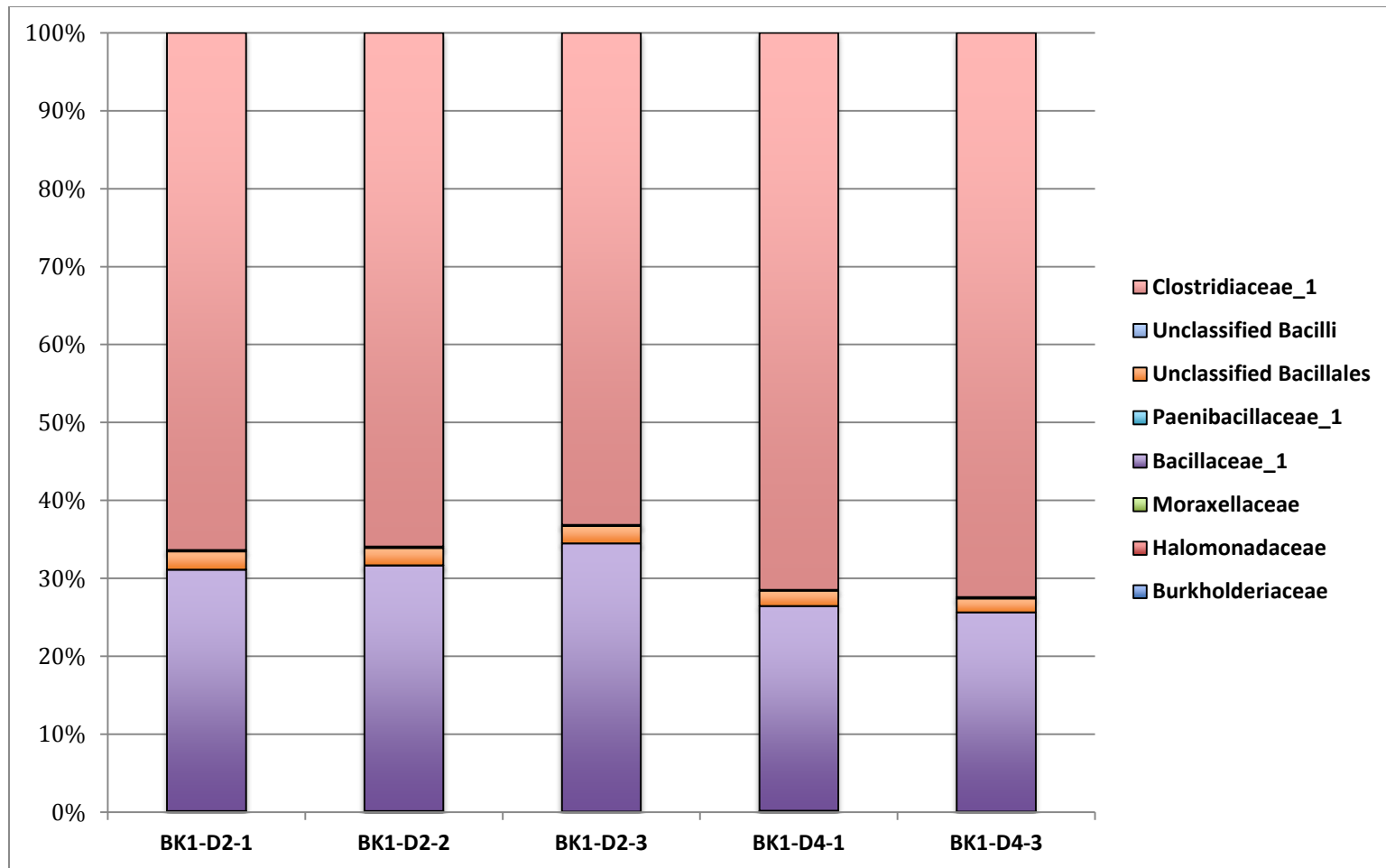


Figure 5.6. Family level assignments of the Bacillus K1 time-course dropout experiments.

5.3.2 Overall Phylogenetic Data for Dropout Experiments

At the phylum level across all samples they were relatively consistent yielding a 96-100% of the phylum Firmicutes. The *Bacillus* DP1 (BDP1) resulted in slight presence of the phylum Proteobacteria (1-3%) which was also present in the day 4 of *Paenibacillus* DP2 at 2% showing a rebound from the previous 48 h in the sample, which was below 1%. The microbial communities predominant across all samples were of the family *Bacilliaceae*, *Clostridiaceae*, *Paenibacilliaceae*, and to a much lesser extent *Burkholderiaceae*. The main constituents were from the phylum Firmicutes that remained consistent across all samples.

5.3.2.1 *Bacillus* DP1 Drop out at the Family Taxonomic Level

Bacillus DP1 dropout was used to determine the contribution by *Bacillus* DP1 to the overall microbial retting community. The times tested were at 48 hours (D2) and 96 hours (D4) which was chosen due to a desire to test the initial colonization by the *Bacillus* DP1 as well as testing the typical end time point (96 h) as determined by previous experiments and to keep consistent with previous studies. The average percentage of *Bacilliaceae* was 6% and unclassified Bacilliales at 19% with a total of 25% contribution from organisms from the class Bacilli at time point 48 h (Figure 5.4). The unclassified Bacilliales (19%) and the *Bacilliaceae* (6%) was an interesting development, as in the previous studies these sequences were all located at either a higher taxonomical level, i.e. class *Bacilli* or within the order Bacilliales. The hypothesis with regards to this is that due to the sequencing length of the Ion Torrent PGM (200 bp flows) and the particular 16S fragment (515-806), enough taxonomic information was not achieved on some sequences resulting in the unclassified Bacilliales. The trend

continuing to day 4 (96 h) has the family *Clostridiaceae* 1 (93%) dominating the environment with the *Bacilli* at 6% (Figure 5.4).

5.3.2.2 *Paenibacillus* DP2 Drop out at the Family Taxonomic Level

Paenibacillus DP2 had a relatively homogenous phylogenetic diversity through the time course at 48 and 96 h. At time point 48 h 40% of the total percentage abundance was the family *Paenibacillaceae* while the *Clostridiaceae* made up the remainder at 59% (Figure 5.5). As the environment continued to the time point 96 h the *Clostridiaceae* increased in percent abundance to 68% while the *Paenibacillaceae* dropped to 28% and to a much lesser extent the *Burkholderiaceae* were present at 2%. Additionally, there was no presence of genus *Bacillus* within the experiment indicating that there was no contamination from the other *Bacillus*-focused retting containers and that the *Bacillus* does not appear to be a surface-associated bacterium within the *H. cannabinus* plant.

5.3.2.3 Diversity of *Bacillus* K1 Drop out at the Family Taxonomic Level

The family *Bacillaceae* remained within the system maintaining 32% at time point 48 h slightly lowering to 26% at 96 h (Figure 5.6). The *Clostridiaceae* also maintained a relatively stable population with 65% at time point 48 h increasing slightly to 72%. As mentioned in the previous phylogenetic data in the *Bacillus* DP1 study this species of *Bacilli* matched well taxonomically to the RDP database. There was only a slight presence of the unclassified Bacilliales at 2% that was maintained throughout. This seems to point towards the previous hypothesis that in the *Bacillus* DP1 isolate has a more ambiguous 16S fragment that leads to less than optimal taxonomical classification as compared to the one used in this dropout experiment, *Bacillus* K1.

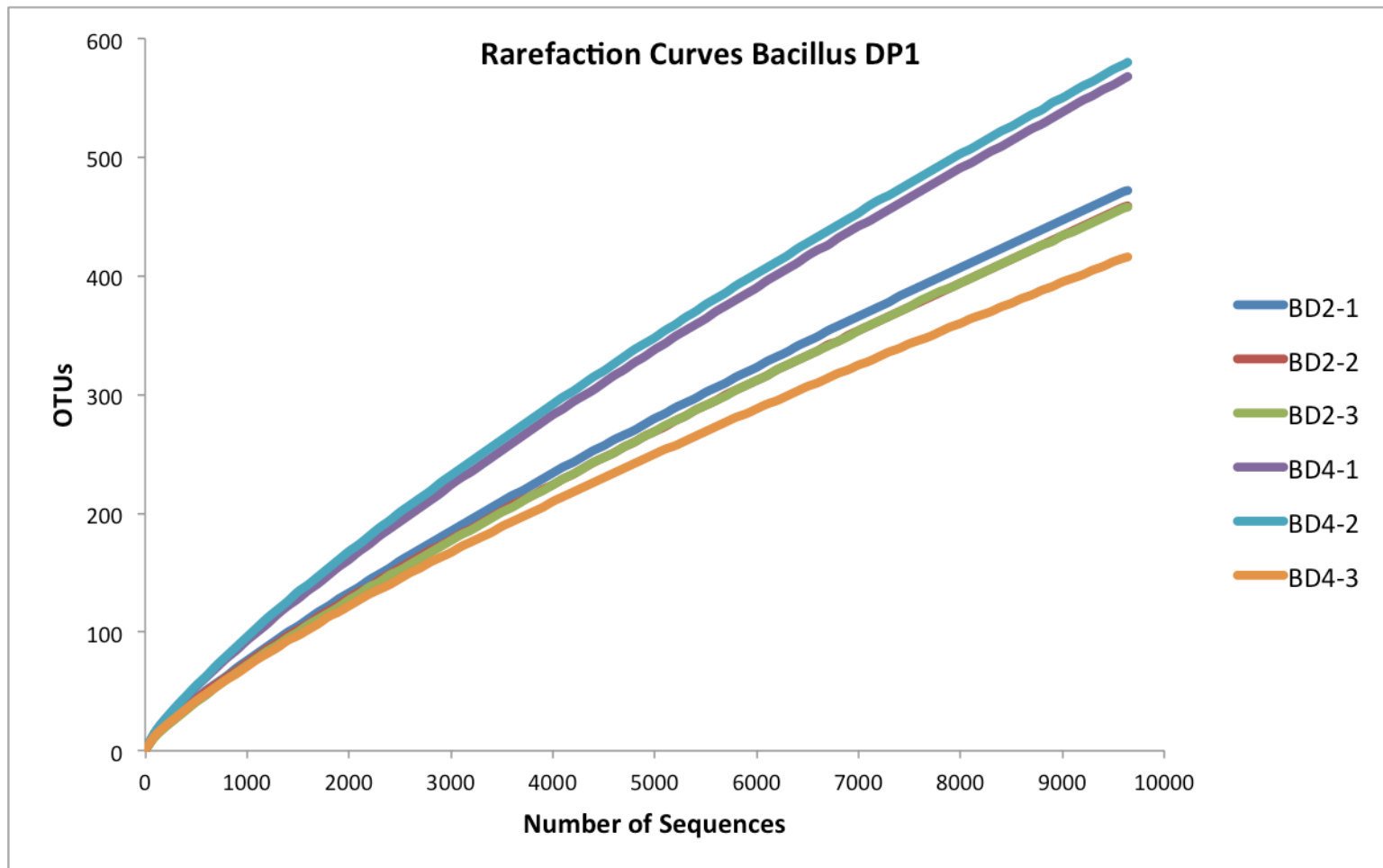


Figure 5.7. Rarefaction curves of *Bacillus* DP1 samples of day 2 and 4. D2 and D4 refers to the day and the final number refers to the specific replicate.

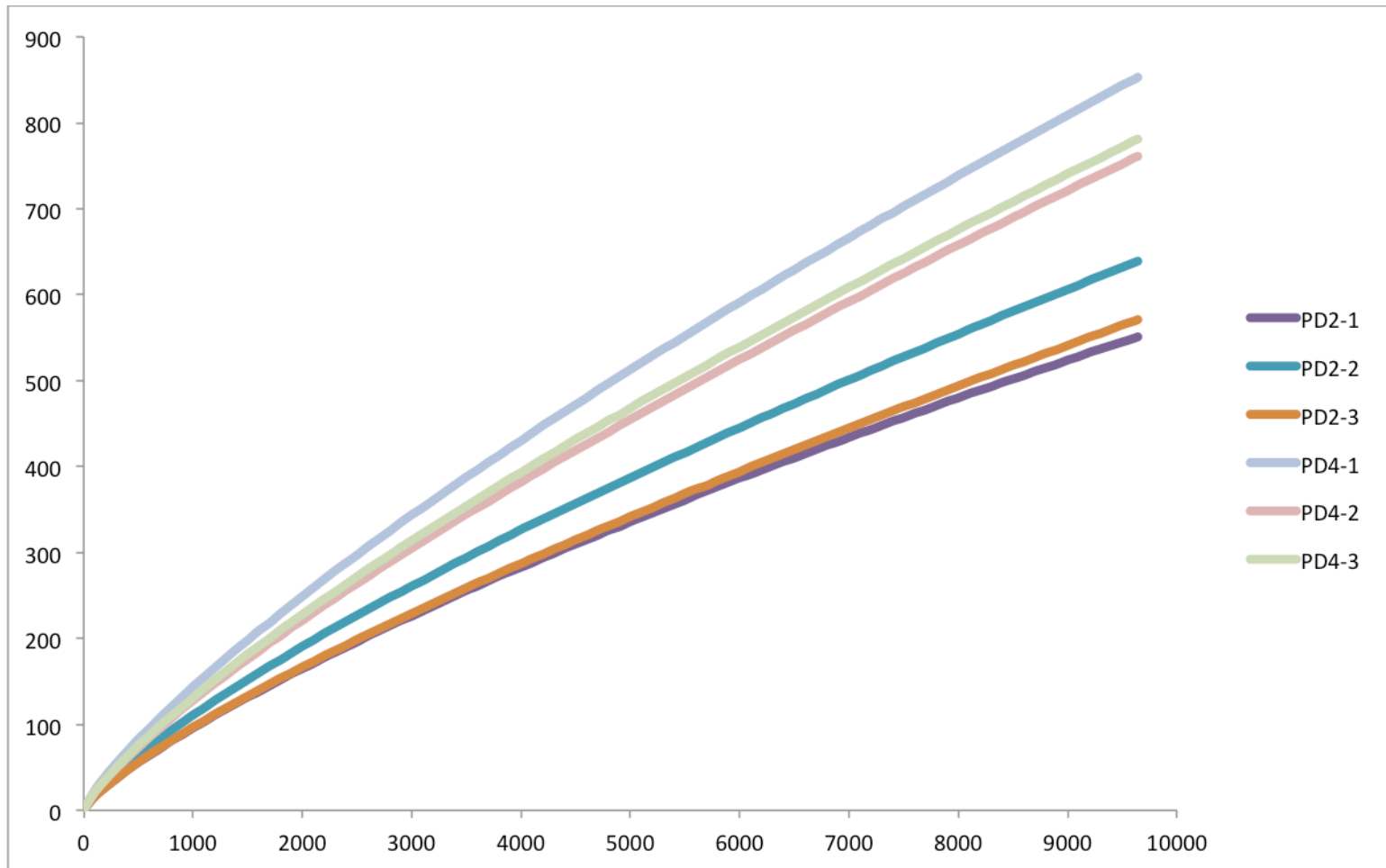


Figure 5.8. Rarefaction curves of *Paenibacillus* DP2 samples of day 2 and 4. D2 and D4 refers to the day and the final number refers to the specific replicate.

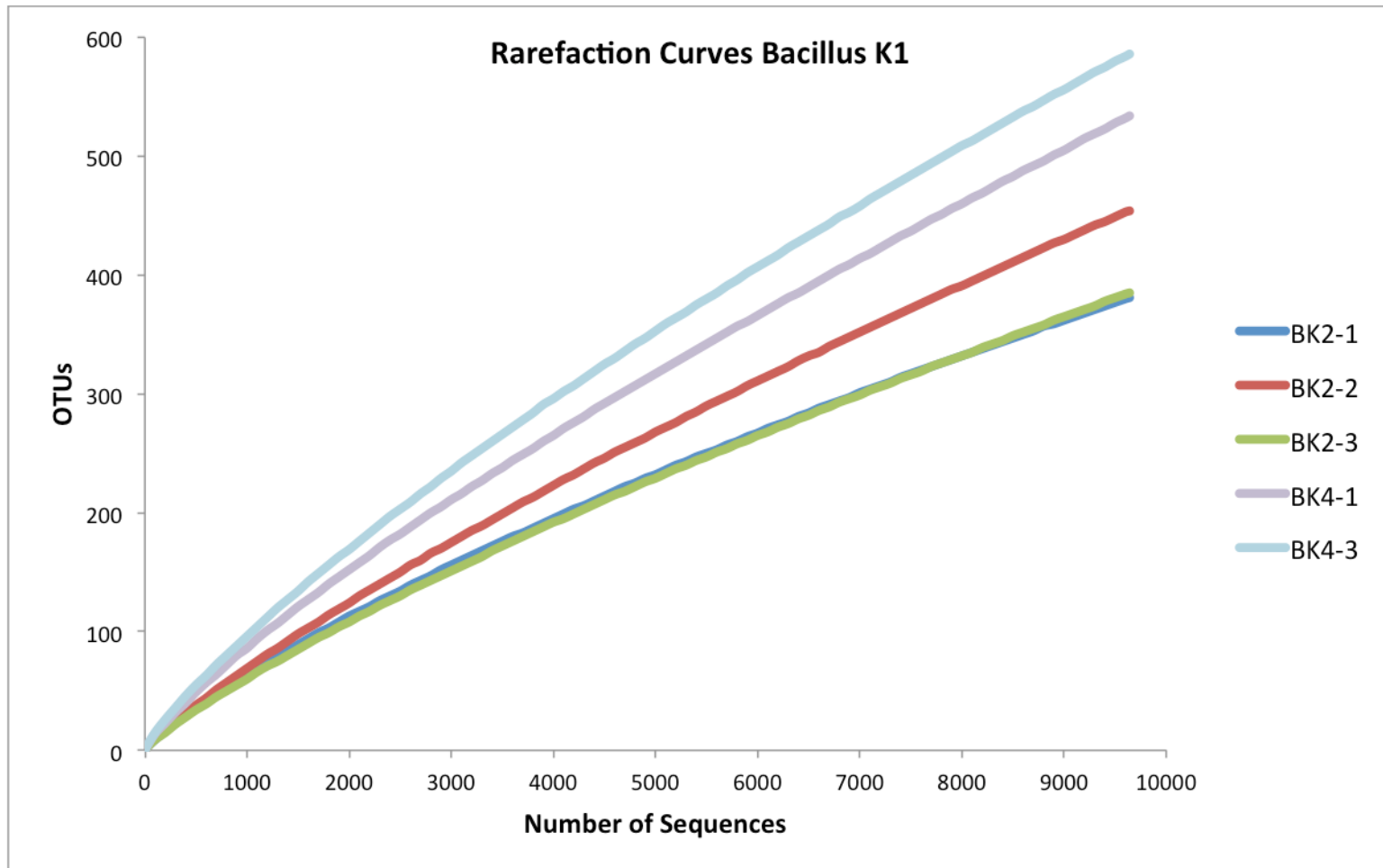


Figure 5.9. Rarefaction curves of *Bacillus* K1 samples of Day 2 and 4. D2 and D4 refers to the day and the final number refers to the specific replicate.

5.3.3 Diversity Estimates of Overall

Multiple means were used to determine the diversity estimates of the samples in the drop out experiments. Rarefaction curves provided a means to determine microbial richness across samples. Alpha diversity was utilized by means of inverse-Simpson diversity estimator, which indicates the richness in an environment with uniform evenness that would have the same level of microbial diversity. UniFrac analysis was also used to determine whether two or more communities have the same structure.

5.3.3.1 Rarefaction Curves

The rarefaction curves as mentioned before provide a means to compare microbial richness across samples. Mothur calculates the values by means of 1 000 randomizations to generate the rarefaction curve data which can then be plotted, which yields the end curves. The curves clustered together with their respective triplicates (Figure 5.7, 5.8, and 5.9). The most diverse samples were from *Paenibacillus* DP2 sets (PD4) with an average OTUs at 48 h were 587 and h 96 at 798 (Table 5.2). All environments tested showed an increase in OTUs from the two time points: 48 and 96 h. This was expected as an increase in time allows more organisms to be colonized. Particularly interesting is that the *Paenibacillus* DP2 sets had the highest microbial richness as compared to all other sets in that it was expected to have lower amounts due to a possible antibiotic action.

	OTUS	Average
BD2-1	472	
BD2-2	459	463
BD2-3	458	
BD4-1	568	
BD4-2	580	521
BD4-3	416	
PD2-1	551	
PD2-2	639	587
PD2-3	571	
PD4-1	853	
PD4-2	761	798
PD4-3	781	
BK2-1	381	
BK2-2	454	407
BK2-3	385	
BK4-1	534	
		560
BK4-3	586	

Table 5.2. Rarefaction values placed in a table for the ease of the reader. An average was taken from each OTUs totals with the sampling efforts normalized.

	sobs	invsimpson	Average
BD2-1	472	1.97773	
BD2-2	459	2.00409	2.111030333
BD2-3	458	2.351271	
BD4-1	568	1.86605	
BD4-2	580	1.710254	1.754499
BD4-3	416	1.687193	
PD2-1	551	2.517185	
PD2-2	639	3.125943	2.709325667
PD2-3	571	2.484849	
PD4-1	853	4.464015	
PD4-2	761	3.567234	3.816761
PD4-3	781	3.419034	
BK2-1	381	2.107099	
BK2-2	454	2.141149	2.138214
BK2-3	385	2.166394	
BK4-1	534	2.887233	
			2.977045
BK4-3	586	3.066857	

Table 5.3 Inverse-Simpson values for all samples. The sobs are the number of observed OTUs.

5.3.3.2 Alpha Diversity

Inverse-Simpson index was utilized to give a numerical value and understanding to the alpha diversity. The data points dovetailed with the other samples with the exception of the *Bacillus* DP1 sets in that there was a lessening of the inverse-Simpson diversity index while the rarefaction curves showed a slight increase. The inverse-Simpson diversity index has the advantage of maintaining a uniformed evenness when making the comparisons, while the rarefaction curves only takes into account the number of sequences sampled and the number of OTUs. However, the *Paenibacillus* DP2 and *Bacillus* K1 samples maintained the same trend as found in the rarefaction curves (Table 5.3).

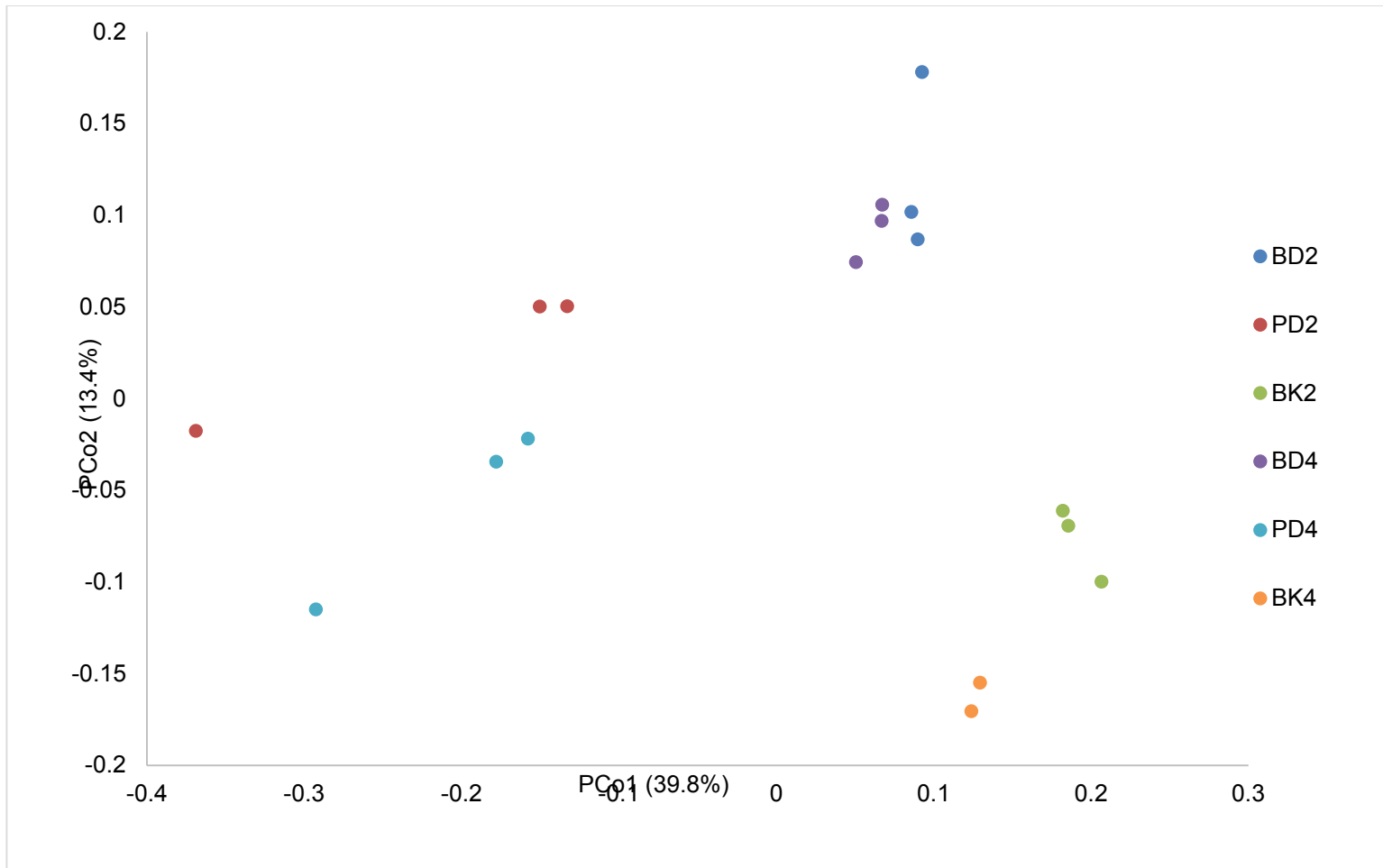


Figure 5.10. Principle coordinate analysis plots for the drop out tests. The PCo1 (x-axis) accounts for 39.8% of the total variation while PCo2 (y-axis) accounts for 13.4% of the total variation.

Tree#	Groups	Weighted	WSig	Unweighted	UWSig
1	BD2-BD4	0.584502	0.002	0.745232	0.36
1	BD2-BK2	1	<0.001	1	0.19
1	BD4-BK2	1	<0.001	1	0.341
1	BD2-BK4	1	<0.001	1	0.374
1	BD4-BK4	1	0.003	1	0.513
1	BK2-BK4	1	<0.001	1	0.112
1	BD2-PD2	1	<0.001	1	0.491
1	BD4-PD2	1	0.049	1	0.499
1	BK2-PD2	1	<0.001	1	0.442
1	BK4-PD2	1	<0.001	1	0.632
1	BD2-PD4	1	<0.001	1	0.525
1	BD4-PD4	1	0.058	1	0.506
1	BK2-PD4	1	<0.001	1	0.509
1	BK4-PD4	1	0.013	1	0.639
1	PD2-PD4	0.869861	<0.001	0.937364	0.549

Table 5.4 UniFrac analysis of phylogenetic tree. Both weighted and unweighted UniFrac values are shown. The weighted UniFrac values all were significantly different from all pairwise comparisons with the exception of BD4-PD4.

5.3.3.3 PCoA and UniFrac Values

Continuing with the beta diversity and principle coordinate analysis the samples all clustered together as expected (Figure 5.10). First beginning with the PCoA shown in Figure 5.10 is that the x-axis explained 39.8% of the total variation while the y-axis contributed to 13.4% of total variation. These two variables accounted for a total of 53.2% of total variation. The colors in the graph indicate the different replicates performed, but each of the different environments, i.e. *Bacillus* DP1, *Paenibacillus* DP2, and *Bacillus* K1, even with the differences in time points tested. There were slight deviations on the principle coordinate analysis with regards to *Paenibacillus* DP2 (light blue and red dots). The *Paenibacillus* sets did cluster together as compared to the other environments, but there was an increase in spread accounted in both time points, i.e. 48 and 96 h. Presumably this is due to the inherently

complexity of the environment and was probably hindered by the sample size. In all other diversity and richness estimates *Paenibacillus* DP2 was much higher than the rest of the environmental sets, rarefaction values: 587 and 798, inverse-Simpson: 2.7 and 3.8 (Table 5.2 and 5.3).

In the weighted UniFrac analysis all samples were determined to be significantly different and this is in contrast to the unweighted analysis in which there was no significance detected (Table 5.4). In the weighted version the abundance of OTUs are taken into account in the calculation for significance and the opposite is true for the unweighted version.

CHAPTER 6

CONCLUSION

6.1 Overview

Research into kenaf retting has been sparse as shown by the lack of microbial research into understanding the components in the retting process. This research aimed to bridge that gap between the retting process and the underlying microbial constituents as it relates. Beginning with a classical microbiological approach, organisms were isolated and analyzed that produce pectinase which is favorable for the microbial retting reaction. Three different retting environments were tested: recreating a “natural” retting process, microbes only contributed by the plant material, and an augmented retting environment with the addition of the previously isolated pectinase-producing microorganisms.

In short, we performed a 16S study on the microbial community of different retting environments using the next generation sequencing platform Ion Torrent PGM. The work showed markedly different constituents from the different microbial retting conditions. C1D4 likely represented a community based on the microbes inherently present on the plant material, while C2D4 inoculated with pond water may reflect the constituents in a “natural” microbial retting environment. E1D4 showed how well the basal community responds to an inoculum of pectinolytic bacteria. We expected differences to occur when using different sources of plant materials given the number of different variables in their pre-treatment (*i.e.* growth, harvest time, location of growth, age, etc.). Surprisingly, the results were strikingly similar, suggesting that the *Clostridium* spp. in question may be more closely associated with the kenaf plant than simple surface contaminants.

6.1.1 Drop Out Experiments in Relation to Previous Studies

The previous hypothesis was that *Paenibacillus* was the main driving force in the change in environment seen in time-course experiments. In the dropout experiments (Figure 5.5), matched up with the previous experiments both in the initial testing of differential microbial retting environments (Figure 3.3 and 3.4) and also the time course experiment (Figure 4.1 and 4.2). This is not necessarily negating that both the *Bacilli* are not contributing, but *Paenibacillus* DP2 seems to be the main driving force behind the push towards an environment dominated by the *Clostridia*.

With a better understanding of the microbial diversity found in the different retting environments, future studies will seek to understand more about the roles of bacterial groups in the breakdown of specific heteropolysaccharides through metagenomic and metatranscriptomic studies. These studies could help facilitate discoveries of novel heteropolysaccharases that could further decrease retting time or produce fibers with different properties.

6.1.2 Microbial Changes at the Community Level

Changing communities and augmenting structure of the bacterial constituents was a key point in this research. The added organisms provided vast shifts in the bacterial communities. This included shifting some communities to strongly be composed of the phylum Firmicutes, while in the initial control groups having much more diverse compositions of the phylum Proteobacteria. Being able to augment communities has large-scale applications than simply those described here, i.e. retting communities of kenaf. In one of the most promising fields,

human microbiome research, bacterial communities have been linked with a vast multitude of disease states (79, 125–129). Understanding the basis of what constitutes a “disease” state can then help understand and making changes to shift it to a “healthy” state. Similar affects have been seen in treating *Clostridium difficile* infections. *C. difficile* infections results in 14,000 American deaths per year (130). Fecal microbiota transplantation (FMT) has been utilized over the past 50 years in treatment for *C. difficile* infections (131). Some estimates put the success rate of FMT treatment for *C. difficile* infections as high as 90% (132). Case studies and trials are currently in development to combat Crohn’s disease and inflammatory bowel syndrome by means of FMT treatment (133). Extending to probiotics, *Saccharomyces boulardii* has been shown to have efficacy against Crohn’s disease. The patients tested were in clinical remissions and utilized *S. boulardii* in maintenance treatment of Crohn’s disease which had a lower of clinical relapses from 37.5% in the control group to 6.25% in the treatment group (134).

Many of this research has been driven by the same basic questions that our research has asked, what is the state of microbial communities under normal conditions and how can we alter it? The Human Microbiome Project as discussed before, was performed much in the same rational as the Human Genome Project that is basic research and understanding of the underlying complexity associated with each. Now that is not to say the projects did not have an end goal or future directions in mind, but was initially performed to get a basic understanding of the genome and microbiome. These same concepts connect the work here in that initially the work started out as understanding what the baseline communities of kenaf retting are. Then as the research progressed augmentations of the community resulted in favorable

conditions. These favorable conditions lead to increased fiber quality in the kenaf plants used in the biocomposites.

6.2 Future Directions

6.2.1 Testing of endophytes

Determining where the microorganisms are coming from and understanding is essential to understanding the process. This topic was covered in one of the previous chapters and it another potential area for additional research in the future. The main objective would be to test the *Clostridia* located on the plant material that had similarities found in two previous papers Miniamisawa et al. 2004 and Miyamoto et al. 2004 (108, 113, 135). In the papers they described a surface sterilization technique to only isolate those endophytes located within the plant, i.e. sterilization with 70% ethanol for 0.5 to 1 m and a 1 to 2% NaOCl wash for 0.5 to 15 m, washed with sterilized water, and then macerated with a mortar, a pestle, and sterilized quartz sand. This strategy could be utilized within our model plant, *H. cannabinus*, to determine whether this same *Clostridia* are present in the plants acting as some symbiotic relationship. Additionally, it is not out of the question that these *Clostridia* are present on the surface of the plant material, which could be confirmed by studies with the surrounding plant material and a 16S phylogenetic study could be done on those samples.

6.2.2 Whole Genome Sequencing of Isolates

Completing sequencing of the various isolates that have been used. As sequencing gets more advantageous in terms of cost, isolates can be sequenced at a faster pace. These isolates

are organisms that have been enriched from a process that has advantages to many industries like paper and even food processing. Pectinase production still remains an important aspect of textile and food production (33). Whole-genome sequencing of organisms gives much needed details in terms of gene structure, possible components in the genome that might affect environmental changes.

As previously described *Paenibacillus* DP2 is either the key player or a constituent in the push of the environment towards one dominated by *Clostridium*. We hypothesized that there were some antimicrobial action being produced by the *Paenibacillus* leading to the domination by the phylum Firmicutes in the experimental retting environments. In small lab-scale experiments there did not seem to be a specific antimicrobial product when testing the supernatant of the *Paenibacillus*, but there could be other mechanisms or the growth conditions were not optimal for antibiotic production. This leads to trying a culture-independent technique to figure out the mode of action in the shift.

Paenibacillus DP2 has been sequenced, but the bioinformatics on the sequences still need to be accomplished. The entire process of retting is inherently complex, as this research has shown. The 16S studies performed just provided a snapshot into what is happening in terms of community dynamics. Sequencing *Paenibacillus* DP1, *Bacillus* DP1, and *Bacillus* K1, should provide additional information of how they are changing the environment to a specific group or organisms, in this case the phylum Firmicutes. Specifically, even with the addition of the dropout tests it still isn't conclusive on why the environments are shifting. Antibiotic production is a possible mechanism of action, which could not be detected by conventional inhibition

assays, but could be detected from whole-genome studies. The WGS studies would provide a culture-independent way of understanding the underlying genes in the system.

6.3 Final Thoughts

Microbial retting is still a complex and inherently difficult process to understand, as this is a natural process in which plant material is degraded. Processes in nature are complex and diverse and we aimed to recreate this within the lab at a much smaller scale. This dissertation at least began to elucidate some of the processes that are underlying the retting process. Isolation of microorganisms that produce favorable enzymes helpful to this process and continuing with understanding the “normal” retting community, augmenting with our previous isolates, and trying to breakdown the community into simpler constituent parts. The research still has a long way to go in terms of completely understanding how the process works, but it adds to the scientific thought as a whole by completing a part of the picture in terms of the bacterial components of this process.

This 16S research performed was one of the first accomplished using the Ion Torrent PGM. At the time we were one of the few labs that had access to the PGM and we presented one of the first presentations at a national meeting showing a proof of concept that it was possible with the initial iteration of the Ion Torrent PGM with the 100 bp sequencing chemistry. This was even recognized by Life Technologies during the meeting. In this research there was progressive change including increasing lengths provided from the newly improved chemistry and allowed the integration of different primer sets to show deeper levels in taxonomy, that is more specifics in terms of the organisms present in the environment.

This research showed a journey not only with myself as the researcher, but as a lab as a whole. Beginning with simple classical microbiology, we have extended beyond integrating various parts and techniques into our understanding as a group and growth. This lab came from doing *no* next-generation sequencing to one that is almost completely dominated by it. Integrating slowly components like increases in read lengths, expanding of our understanding of large-scale computational projects, constant changes in protocols associated with sequencing, and then teaching others within and from outside the technique and understanding of the underlying technology and benefits from it. The research was accomplished through the dedicated work of the lab and the principle investigator.

APPENDIX A

16S SEQUENCES FROM SEQUENCING OF PECTINASE PRODUCING BACTERIA

These fragments were sequenced using Sanger technique and utilized the 27F and 1492R for a forward and reverse sequencing reaction.

Bacillus DP1

AAC ACG TGG GTA ACC TGC CCA TAA GAC TGG GAT AAC TCC GGG AAA CCG GGG CTA ATA CCG
GAT AAC ATT TTG AAC CGC ATG GTT CGA AAT TGA AAG GCG GCT TCG GCT GTC ACT TAT GGA
TGG ACC CGC GTC GCA TTA GCT AGT TGG TGA GGT AAC GGC TCA CCA AGG CAA CGA TGC GTA
GCC GAC CTG AGA GGG TGA TCG GCC ACA CTG GGA CTG AGA CAC GGC CCA GAC TCC TAC GGG
AGG CAG CAG TAG GGA ATC TTC CGC AAT GGA CGA AAG TCT GAC GGA GCA ACG CCG CGT GAG
TGA TGA AGG CTT TCG GGT CGT AAA ACT CTG TTG TTA GGG AAG AAC AAG TGC TAG TTG AAT
AAG CTG GCA CCT TGA CGG TAC CTA ACC AGA AAG CCA CGG CTA ACT ACG TGC CAG CAG CCG
CGG TAA TAC GTA GGT GGC AAG CGT TAT CCG GAA TTA TTG GGC GTA AAG CGC GCG CAG GTG
GTT TCT TAA GTC TGA TGT GAA AGC CCA CGG CTC AAC CGT GGA GGG TCA TTG GAA ACT GGG
AGA CTT GAG TGC AGA AGA GGA AAG TGG AAT TCC ATG TGT AGC GGT GAA ATG CGT AGA GAT
ATG GAG GAA CAC CAG TGG CGA AGG CGA CTT TCT GGT CTG TAA CTG ACA CTG AGG CGC GAA
AGC GTG GGG AGC AAA CAG GAT TAG ATA CCC TGG TAG TCC ACG CCG TAA ACG ATG AGT GCT
AAG TGT TAG AGG GTT TCC GCC CTT TAG TGC TGA AGT TAA CGC ATT AAG CAC TCC GCC TGG
GGA GTA CGG CCG CAA GGC TGA AAC TCA AAG GAA TTG ACG GGG GCC CGC ACA AGC GGT
GGA GCA TGT GGT TTA ATT CGA AGC AAC GCG AAG AAC CTT ACC AGG TCT TGA CAT CCT CTG
ACA ACC CTA GAG ATA GGG CTT CTC CTT CGG GAG CAG AGT GAC AGG TGG TGC ATG GTT GTC
GTC AGC TCG TGT CGT GAG ATG TTG GGT TAA GTC CCG CAA CGA GCG CAA CCC TTG ATC TTA
GTT GCC ATC ATT AAG TTG GGC ACT CTA AGG TGA CTG CCG GTG ACA AAC CGG AGG AAG GTG
GGG ATG ACG TCA AAT CAT CAT GCC CCT TAT GAC CTG GGC TAC ACA CGT GCT ACA ATG GAC
GGT ACA AAG AGC TGC AAG ACC GCG AGG TGG AGC TAA TCT CAT AAA ACC GTT CTC AGT TCG
GAT TGT AGG CTG CAA CTC GCC TAC ATG AAG CTG GAA TCG CTA GTA ATC GCG GAT CAG CAT
GCC GCG GTG AAT ACG TTC CCG GGC CTT GTA CAC ACC GCC NGT CAC ACC ACG AGA GTT TGT
AAC ACC CGA AGT CGG TGG GGT AAC CT

Paenibacillus DP2

TTT GCT TCT AAC TAA CCT AGC GGC GGN CGG NNN NNT AAC ACG TAG GCA ACC TGC CCA CAA
GAC AGG GAT AAC TAC CGG AAA CGG TAG CTA ATA CCC GAT ACA TCC TTT TCC TGC ATG GGA
GAA GGA GGA AAG GCG GAG NAA TCT GTC ACT TGT GGA TGG GCC TGC GGC GCA TTA GCT AGT
TGG TGG GGT AAA GGC CTA CCA AGG CGA CGA TGC GTA GCC GAC CTG AGA GGG TGA TCG GCC
ACA CTG GGA CTG AGA CAC GGC CCA GAC TCC TAC GGG AGG CAG CAG TAG GGA ATC TTC CGC
AAT GGG CGA AAG CCT GAC GGA GCA ACG CCG CGT GAG TGA TGA AGG TTT TCG GAT CGT AAA
GCT CTG TTG CCA GGG AAG AAC GTC TTG TAG AGT AAC TGC TAC AAG AGT GAC GGT ACC TGA
GAA GAA AGC CCC GGC TAA CTA CGT GCC AGC AGC CGC GGT AAT ACG TAG GGG GCA AGC GTT
GTC CGG AAT TAT TGG GCG TAA AGC GCG CGC AGG CGG CTC TTT AAG TCT GGT GTT TAA TCC
CGA GGC TCA TCT TNN GGG TCG CAC TGG AAA CTG GGG AGC TTT GAG TGC AGA ATA GGA AGA
GTG GAA TTT CCA CGT GTA GCG GGT GAA AAT GCG TAG AGA TGT GGA GGA ACA CCA GTG GCG
AAG GCA CTC TNT GGG CTG TAA CTG ACG CTG AGG CGC GAA AGC GTG GGG GAG CAA ACA
GGA TTA GAT ACC CTG GTA GTC CAC GCN GTA AAC GAT GAA TGN TAG GTG TTA GGG GTT TCG
ATA CCC TGG TGC CGA AGT TAA CAC ATT AAG CAT TTC CGC CTG GGG AGT NCG GTC GCA AGA
CTG AAA CTC AAA GGA AAT TGA CGG GGA CCC GCA CAA GCA GTG GAG TAT GTG GTT TAA TTC

GAA GCA ACG CGA AGA ACC TTA CCA GGT CTT GAC ATC CCT TTG ACC GGT CTA GAG ATA GAC
CTT TCC TTC GGG ACA GAG GAG ACA GGT GGT GCA TGG TTG TCG TCA GCT CGT GTC GTG AGA
TGT TGG GTT AAG TCC CGC AAC GAG CGC AAC CCT TAT GCT TAG TTG CCA GCA GGT CAA GCT
GGG CAC TCT AAG CAG ACT GCC GGT GAC AAA CCG GAG GAA GGT GGG GAT GAC GTC AAA TCA
TCA TGC CCC TTA TGA CCT GGG CTA CAC ACG TAC TAC AAT GGC CGG TAC AAC GGG AAG CGA
AGG AGC GAT NTG GAG CCA ATC CTA GAA AAG CCG NTC TCA GTT CGG ATT GTA GGC TGC AAC
TCG CCT ACA TGA AGT CGG AAT TGC TAG TAA TCG CGG ATC AGC ATG CCG CGG TGA ATA CGT
TCC CGG GTC TTG TAC ACC CCA NCC GNC ACA CCC CGA GAG TTT ACA ACA CCC GAA

Bacillus DP3

GCT TGC TCT TAT GAA GTT AGC GGC GGA CGG GTN NNT AAC ACG TGG GTA ACC TGC CCA TAA
GAC TGG GAT AAC TCC GGG AAA CCG GGG CTA ATA CCG GAT AAC ATT TTG AAC CGC ATG GTT
CGA AAT TGA AAG GCG GCT TCG GCT GTC ACT TAT GGA TGG ACC CGC GTC GCA TTA GCT AGT
TGG TGA GGT AAC GGC TCA CCA AGG CAA CGA TGC GTA GCC GAC CTG AGA GGG TGA TCG GCC
ACA CTG GGA CTG AGA CAC GGC CCA GAC TCC TAC GGG AGG CAG CAG TAG GGA ATC TTC CGC
AAT GGA CGA AAG TCT GAC GGA GCA ACG CCG CGT GAG TGA TGA AGG CTT TCG GGT CGT AAA
ACT CTG TTG TTA GGG AAG AAC AAG TGC TAG TTG AAT AAG CTG GCA CCT TGA CGG TAC CTA
ACC AGA AAG CCA CGG CTA ACT ACG TGC CAG CAG CCG CGG TAA TAC GTA GGT GGC AAG CGT
TAT CCG GAA TTA TTG GGC GTA AAG CGC GCG CAG GTG GTT TCT TAA GTC TGA TGT GAA AGC
CCA CGG CTC AAC CGT GGA GGG TCA TTG GAA ACT GGG AGA CTT GAG TGC AGA AGA GGA AAG
TGG AAT TCC ATG TGT AGC GGT GAA ATG CGT AGA GAT ATG GAG GAA CAC CAG TGG CGA AGG
CGA CTT TCT GGT CTG TAA CTG ACA CTG AGG CGC GAA AGC GTG GGG AGC AAA CAG GAT TAG
ATA CCC TGG TAG TCC ACG CCG TAA ACG ATG AGT GCT NAA GTG TTA GAG GGT TTC CGC CCT
TTA GTG CTG AAG TTA ACG CAT TAA GCA CTC CGC CTG GGG AGT ACN GGC CGC AAG GCT GAA
ACT CAA AGG AAT TGA CGG GGC CCG CAC AAG CNG GTG GAG CAT GTG GTT TAA TTT GAA GCA
ACG CGA AGA ACC TTA CCA GGT CTT GAC ATC CTT TGA CAA CCC TAG AGA TAG GGC TTC TCC
TTT GGG AGC AGA GTG ACA GGT GGT GCA TGG TTG TCG TCA GCT CGT GTC GTG AGA TGT TGG
GTT AAG TCC CGC AAC GAG CGC AAC CCT TGA TTT TAG TTG CCA TCA TTT AGT TGG GCA CTC
TAA GGT GAC TGC CGG TGA CAA ACC GGA GGA AGG TGG GGA TGA CGT CAA ATC ATC ATG CCC
CTT ATG ACC TGG GCT ACA CAC GTG CTA CAA TGG ACG GTA CAA AGA GCT GCA AGA CCG CGA
GGT GGA GCT AAT CTC ATA AAA CCG TTC TCA GTT CGG ATT GTA GGC TGC AAC TCG CCT ACA
TGA AGC TGG AAT CGC TAG TAA TCG CGG ATC AGC ATG CCG CGG TGA ATA CGT TCC CGG GCC
TTG TAC ACA CCN CCC GNC ACA CCA CGA GAG TTT GTA ACA CCC GAA GTC GGT GGG GTA ACC

Bacillus DP4

GCA AGT CGA GCG AAT GGA TTA AGA GCT TGC TCT TAT GAA GTT AGC GGC GGA CGG GNN NNN
AAC ACG TGG GTA ACC TGC CCA TAA GAC TGG GAT AAC TCC GGG AAA CCG GGG CTA ATA CCG
GAT AAC ATT TTG AAC TGC ATG GTT CGA AAT TGA AAG GCG GCT TCG GCT GTC ACT TAT GGA
TGG ACC CGC GTC GCA TTA GCT AGT TGG TGA GGT AAC GGC TCA CCA AGG CAA CGA TGC GTA
GCC GAC CTG AGA GGG TGA TCG GCC ACA CTG GGA CTG AGA CAC GGC CCA GAC TCC TAC GGG
AGG CAG CAG TAG GGA ATC TTC CGC AAT GGA CGA AAG TCT GAC GGA GCA ACG CCG CGT GAG
TGA TGA AGG CTT TCG GGT CGT AAA ACT CTG TTG TTA GGG AAG AAC AAG TGC TAG TTG AAT
AAG CTG GCA CCT TGA CGG TAC CTA ACC AGA AAG CCA CGG CTA ACT ACG TGC CAG CAG CCG
CGG TAA TAC GTA GGT GGC AAG CGT TAT CCG GAA TTA TTG GGC GTA AAG CGC GCG CAG GTG

GTT TCT TAA GTC TGA TGT GAA AGC CCA CGG CTC AAC CGT GGA GGG TCA TTG GAA ACT GGG
AGA CTT GAG TGC AGA AGA GGA AAG TGG AAT TCC ATG TGT AGC GGT GAA ATG CGT AGA GAT
ATG GAG GAA CAC CAG TGG CGA AGG CGA CTT TCT GGT CTG TAA CTG ACA CTG AGG CGC GAA
AGC GTG GGG AGC AAA CAG GAT TAG ATA CCC TGG TAG TCC ACG CCG TAA ACG ATG AGT GCT
AAG TGT TAG AGG GTT TCC GCC CTT TAG TGC TGA AGT TAA CGC ATT AAG CAC TCC GCC TGG
GGA GTA CGG CCG CAA GGC TGA AAC TCA AAN GAA TTG ACN GGG GGC CCG CAC AAG CGG
GTG GAG CAT GTG GTT TAA TTC GAA GCA ACG CGA AGA ACC TTA CCA GGT CTT GAC ATC CTC
TGA AAA CCC TAG AGA TAG GGC TTC TCC TTC GGG AGC AGA GTG ACA GGT GGT GCA TGG TTG
TCG TCA GCT CGT GTC GTG AGA TGT TGG GTT AAG TCC CGC AAC GAG CGC AAC CCT TGA TCT
TAG TTG CCA TCA TTA AGT TGG GCA CTC TAA GGT GAC TGC CGG TGA CAA ACC GGA GGA AGG
TGG GGA TGA CGT CAA ATC ATC ATG CCC CTT ATG ACC TGG GCT ACA CAC GTG CTA CAA TGG
ACG GTA CAA AGA GCT GCA AGA CCG CGA GGT GGA GCT AAT CTC ATA AAA CCG TTC TCA GTT
CGG ATT GTA GGC TGC AAC TCG CCT ACA TGA AGC TGG AAT CGC TAG TAA TCG CGG ATC AGC
ATG CCG CGG TGA ATA CGT TCC CGG GCC TTG TAC ACA CCG CCC GTC ACA CCA CGA GAG TTT
GTA ACA CCC GAA GTC GGT GGG GTA ACC TTT

Paenibacillus DP5

GCA GTC GGG CGG GGT TNN TTA GAA GCT TGC TTC TAT GTA ACC TAG CGC CGG AGG GGT NNN
TAA NNC NNA GGC AAC CTG CCC ACA AGA CAG GGA TAA CTA CCG GAA ACG GTA GCT AAT ACC
CGA TAC ATC CTT TTC CTG CTT GGG AGA AGG AGG AAA GGC GGA GCA ATC TGT CAC TTG TGG
ATG GGC CTG CGG CGC ATT AGC TAG TTG GTG GGG TAA AGG CCT ACC AAG GCG ACG ATG CGT
AGC CGA CCT GAG AGG GTG ATC GGC CAC CCT GGG ACT GAG ACA CGG CCC AGA CTC CTA CGG
GAG GCA GCC GTA GGG AAT CTT CCG CAA TGG GCG AAA GCC TGA CGG AGC AAC GCC GCG TGA
GTG ATG AAG GTT TTC GGA TCG TAA AGC TCT GTT GCC AGG GAA GAA CGT CTT GTA GAG TAA
CTG CTT CNA TAG TGA CGG TAC CTG ANA AGA AAG CCC CGG CTA ACT ACG TGC CAG CAG CCG
CGG TAA TAC NTA GGG GGC AAG CGT TGT CCG GAA TTA TTG GGC GTA AAG CGC GCG CAG GCG
GCT CTT TAA GTC TGG TGT TTA ATC CCG AGG CTC AAC TTC GGG TCG CAC TGG AAA CTG GGG
AGC TTG AGT GCA GAA GAG GAG AGT GGA ATT CCA CGT GTA GCG GAG AAA TGC GTA GAG ATG
TGG AGG AAC ACC NGN GGC GAA GGC GAC TTC TCT

Bacillus K1

TGC AAG TCG AGC GGA CAG ATG GGA GCT TGC TCC CTG ATG TTA GCG GCG GAC GGG TGA GTA
ACA CGT GGG TAA CCT GCC TGT AAG ACT GGG ATA ACT CCG GGA AAC CGG GGC TAA TAC CGG
ATG GTT GTT TGA ACC GCA TGG TTC AGA CAT AAA AGG TGG CTT CGG CTA CCA CTT ACA GAT
GGA CCC GCG GCG CAT TAG CTA GTT GGT GAG GTA ACG GCT CAC CAA GGC GAC GAT GCG TAG
CCG ACC TGA GAG GGT GAT CGG CCA CAC TGG GAC TGA GAC ACG GCC CAG ACT CCT ACG GGA
GGC AGC AGT AGG GAA TCT TCC GCA ATG GAC GAA AGT CTG ACG GAG CAA CGC CGC GTG AGT
GAT GAA GGT TTT CGG ATC GTA AAG CTC TGT TGT TAG GGA AGA ACA AGT GCC GTT CAA ATA
GGG CGG CAC CTT GAC GGT ACC TAA CCA GAA AGC CAC GGC TAA CTA CGT GCC AGC AGC CGC
GGT AAT ACG TAG GTG GCA AGC GTT GTC CGG AAT TAT TGG GCG TAA AGG GCT CGC AGG CGG
TTT CTT AAG TCT GAT GTG AAA GCC CCC GGC TCA ACC GGG GAG GGT CAT TGG AAA CTG GGG
AAC TTG AGT GCA GAA GAG GAG AGT GGA ATT CCA CGT GTA GCG GTG AAA TGC GTA GAG ATG
TGG AGG AAC ACC AGT GGC GAA GGC GAC TCT CTG GTC TGT AAC TGA CGC TGA GGA GCG AAA
GCG TGG GGA GCG AAC AGG ATT AGA TAC CCT GGT AGT CCA CGC CGT AAA CGA TGA GTG CTA

AGT GTT AGG GGG TTT CCG CCC CTT AGT GCT GCA GCT AAC GCA TTA AGC ACT CCG CCT GGG
GAG TAC GGT CGC AAG ACT GAA ACT CAA AGG AAT TGA CGG GGG CCC GCA CAA GCG GTG GAG
CAT GTG GTT TAA TTC GAA GCA ACG CGN NAA CCT TAC CAG GTC TTG ACA TCC TCT GAC AAT
CCT AGA GAT AGG ACG TCC CCT TCG GGG GCA GAG TGA CAG GTG GTG CAT GGT TGT CGT CAG
CTC GTG TCG TGA GAT GTT GGG TTA AGT CCC GCA ACG AGC GCA ACC CTT GAT CTT AGT TGC
CAG CAT TCA GTT GGG CAC TCT AAG GTG ACT GCC GGT GAC AAA CCG GAG GAA GGT GGG GAT
GAC GTC AAA TCA TCA TGC CCC TTA TGA CCT GGG CTA CAC ACG TGC TAC AAT GGA CAG AAC
AAA GGG CAG CGA AAC CGC GAG GTT AAG CCA ATC CCA CAA ATC TGT TCT CAG TTC GGA TCG
CAG TCT GCA ACT CGA CTG CGT GAA GCT GGA ATC GCT AGT AAT CGC GGA TCA GCA TGC CGC
GGT GAA TAC GTT CCC GGG CCT TGT ACA CAC CGC CCG TCA CAC CAC GAG AGT TTG TAA CAC
CCG AAG TCG GTG AGG TAA CCT

Paenibacillus K2

TGC AGT CGA GCG GGG TTA TTT AGA AGC TTG CTT CTA AAT AAT CTA GCG GCG GAC GGG TGA
GTA ACA CGT AGG CAA CCT GCC CAC AAG ACA GGG ATA ACT ACC GGA AAC GGT AGC TAA TAC
CCG ATA CAT CCT TTT CCT GCA TGG GAG AAG GAG GAA AGG CGG AGC AAT CTG TCA CTT GTG
GAT GGG CCT GCG GCG CAT TAG CTA GTT GGT GGG GTA ATG GCC TAC CAA GGC GAC GAT GCG
TAG CCG ACC TGA GAG GGT GAT CGG CCA CAC TGG GAC TGA GAC ACG GCC CAG ACT CCT ACG
GGA GGC AGC AGT AGG GAA TCT TCC GCA ATG GGC GAA AGC CTG ACG GAG CAA CGC CGC GTG
AGT GAT GAA GGT TTT CGG ATC GTA AAG CTC TGT TGC CAG GGA AGA ACG TCT TGT AGA GTA
ACT GCT ACA AGA GTG ACG GTA CCT GAG AAG AAA GCC CCG GCT AAC TAC GTG CCA GCA GCC
GCG GTA ATA CGT AGG GGG CAA GCG TTG TCC GGA ATT ATT GGG CGT AAA GCG CGC GCA GGC
GGC TCT TTA AGT CTG GTG TTT AAT CCC GAG GCT CAA CTT CGG GTC GCA CTG GAA ACT GGA
GAG CTT GAG TGC AGA AGA GGA GAG TGG AAT TCC ACG TGT AGC GGT GAA ATG CGT AGA GAT
GTG GAG GAA CAC CAG TGG CGA AGG CGA CTC TCT GGG CTG TAA CTG ACG CTG AGG CGC GAA
AGC GTG GGG AGC AAA CAG GAT TAG ATA CCC TGG TAG TCC ACG CCG TAA ACG ATG AAT GCT
AGG TGT TAG GGG TTT CGA TAC CCT TGG TGC CGA AGT TAA CAC ATT AAG CAT TCC GCC TGG
GGA GTA CGG TCG CAA GAC TGA AAC TCA AAG GAA TTG ACG GGG ACC CGC ACA AGC AGT GGA
GTA TGT GGT TTA ATT CGA AGC AAC GCG AAG AAC CTT ACC AGG TCT TGA CAT CCC TCT GAC
CGG TCT AGA GAT AGG CCT TTC CTT CGG GAC AGA GGA GAC AGG TGG TGC ATG GTT GTC GTC
AGC TCG TGT CGT GAG ATG TTG GGT TAA GTC CCG CAA CGA GCG CAA CCC TTA TGC TTA GTT
GCC AGC AGG TCA AGC TGG GCA CTC TAA GCA GAC TGC CGG TGA CAA ACC GGA GGA AGG TGG
GGA TGA CGT CAA ATC ATC ATG CCC CTT ATG ACC TGG GCT ACA CAC GTA CTA CAA TGG CCG
GTA CAA CGG GAA GCG AAG CCG CGA GGT GGA GCC AAT CCT AGA AAA GCC GGT CTC AGT TCG
GAT TGT AGG CTG CAA CTC GCC TAC ATG AAG TCG GAA TTG CTA GTA ATC GCG GAT CAG CAT
GCC GCG GTG AAT ACG TTC CCG GGT CTT GTA CAC ACC GCC CGT CAC ACC ACG AGA GTT TAC
AAC ACC CGA AGT CGG TGA GGT AAC CGC AAG GAG CCA

Paenibacillus K3

CTT GCT TCT AAN TAA CCT AGC GGC GGA CGG GTG AGT AAC ACG TAG GCA ACC TGC CCA CAA
GAC AGG GAT AAC TAC CGG AAA CGG TAG CTA ATA CCC GAT ACA TCC TTT TCC TGC ATG GGA
GAA GGA GGA AAG GCG GAG CAA TCT GTC ACT TGT GGA TGG GCC TGC GGC GCA TTA GCT AGT
TGG TGG GGT AAA GGC CTA CCA AGG CGA CGA TGC GTA GCC GAC CTG AGA GGG TGA TCG GCC
ACA CTG GGA CTG AGA CAC GGC CCA GAC TCC TAC GGG AGG CAG CAG TAG GGA ATC TTC CGC

AAT GGG CGA AAG CCT GAC GGA GCA ACG CCG CGT GAG TGA TGA AGG TTT TCG GAT CGT AAA
GCT CTG TTG CCA GGG AAG AAC GTC TTG TAG AGT AAC TGC TAC AAG AGT GAC GGT ACC TGA
GAA GAA AGC CCC GGC TAA CTA CGT GCC AGC AGC CGC GGT AAT ACG TAG GGG GCA AGC GTT
GTC CGG AAT TAT TGG GCG TAA AGC GCG CGC AGG CGG CTC TTT AAG TCT GGT GTT TAA TCC
CGA GGC TCA ACT TCG GGT CGC ACT GGA AAC TGG GGA GCT TGA GTG CAG AAG AGG AGA GTG
GAA TTC CAC GTG TAG CGG TGA AAT GCG TAG AGA TGT GGA GGA ACA CCA GTG GCG AAG GCG
ACT CTC TGG GCT GTA ACT GAC GCT GAG GCG CGA AAG CGT GGG GAG CAA ACA GGA TTA GAT
ACC CTG GTA GTC CAC GCC GTA AAC GAT GAA TGC TAG GTG TTA GGG GTT TCG ATA CCC TTG
GTG CCG AAG TTA ACA CAT TAA GCA TTC CGC CTG GGG AGT ACG GTC GCA AGA CTG AAA CTC
AAA GGA ATT GAC GGG GAC CCG CAC AAG CAG TGG AGT ATG TGG TTT AAT TCG AAG CAA CGC
GAA GAA CCT TAC CAG GTC TTG ACA TCC CTN TGA CCG GTC TAG AGA TAG ACC TTT CCT TCG
GGA CAG AGG AGA CAG GTG GTG CAT GGT TGT CGT CAG CTC GTG TCG TGA GAT GTT GGG TTA
AGT CCC GCA ACG AGC GCA ACC CTT ATG CTT AGT TGC CAG CAG GTC AAG CTG GGC ACT CTA
AGC AGA CTG CCG GTG ACA AAC CGG AGG AAG GTG GGG ATG ACG TCA AAT CAT CAT GCC CCT
TAT GAC CTG GGC TAC ACA CGT ACT ACA ATG GCC GGT ACA ACG GGA AGC GAA GGC GCG AGG
TGG AGC CAA TCC TAG AAA AGC CGG TCT CAG TTC GGA TTG TAG GCT GCA ACT CGC CTA CAT
GAA GTC GGA ATT GCT AGT AAT CGC GGA TCA GCA TGC CGC GGT GAA TAC GTT CCC GGG TCT
TGT ACA CAC CGC CCG TCA CAC CAC GAG AGT TTA CAA CAC CCG AAG TCG GTG AGG TAA CCG
CAA GGA GCC AGC C

APPENDIX B

MOTHUR TALON SCRIPT. EXAMPLE SCRIPT FOR USE WITH THE TALON HPC

Any line with an “#” denotes a comment that explains the script. The lines beginning with “./mothur” are the commands that are ran on the Talon2 HPC server. Certain commands, specifically with regards to the diversity estimates are done within an interactive shell on Talon, which are explained below. Outputs from these commands have been included with the script and are in bold.

```
#!/bin/bash
```

```
#$ -V
```

```
#$ -cwd
```

```
#$ -q serial.q
```

```
#$ -m abe
```

```
#BSUB -M davidkvisi@gmail.com
```

#What you need to do is replace all of the word "27_drop" below with your starting name of your file. It will be something like 27_drop.sff. Make sure the spelling is exactly right or this script will fail on the first command. If you are using Gedit on Linux then you can simply click the Search option at the top and click on the Replace feature. Then simply type in "27_drop" in the first blank the name of your sff file in the second blank.

```
#Standard files that you need. barcode.oligos
```

```
./mothur "#sffinfo(sff=27_drop.sff, flow=F)"
```

```
#Name of output files: 27_drop.qual, 27_drop.fasta
```

```
./mothur "#trim.seqs(fasta=27_drop.fasta, oligos=barcode.oligos, qfile=27_drop.qual,  
minlength=200, maxambig=0, maxhomop=8, bdiffs=1, pdiffs=2, processors=16)"
```

#Name of output files: 27_drop.trim.qual, 27_drop.trim.fasta, 27_drop.scrap.qual,
27_drop.scrap.fasta, 27_drop.groups

```
./mothur "#unique.seqs(fasta=27_drop.trim.fasta)"
```

#Name of output files: 27_drop.trim.unique.fasta

```
./mothur "#align.seqs(fasta=27_drop.trim.unique.fasta, reference=silva.bacteria.fasta,  
processors=32)"
```

#Name of output files: 27_drop.trim.unique.align, 27_drop.trim.unique.flip.accnos,
27_drop.trim.unique.report/

```
./mothur "#screen.seqs(fasta=27_drop.trim.unique.align, name=27_drop.trim.names,  
group=27_drop.groups, start=13862, optimize=end, criteria=95, processors=32)"
```

#Name of output files: 27_drop.trim.unique.good.align, 27_drop.trim.unique.bad.accnos,
27_drop.trim.good.names, 27_drop.good.groups

```
./mothur "#filter.seqs(fasta=27_drop.trim.unique.good.align, vertical=T, trump=.,  
processors=32)"
```

#Output files: 27_drop.trim.unique.good.filter.fasta

```
./mothur "#unique.seqs(fasta=27_drop.trim.unique.good.filter.fasta,  
name=27_drop.trim.good.names)"
```

```
#Output files: 27_drop.trim.unique.good.filter.unique.fasta,  
27_drop.trim.unique.good.filter.names
```

```
./mothur "#pre.cluster(fasta=27_drop.trim.unique.good.filter.unique.fasta,  
name=27_drop.trim.unique.good.filter.names, group=27_drop.good.groups, diffs=2)"
```

```
#Output files: agri.trim.unique.good.filter.unique.precluster.fasta,  
agri.trim.unique.good.filter.unique.precluster.names
```

```
./mothur "#chimera.uchime(fasta=27_drop.trim.unique.good.filter.unique.precluster.fasta,  
name=27_drop.trim.unique.good.filter.unique.precluster.names, group=27_drop.good.groups,  
processors=32)"
```

```
#Output files: agri.trim.unique.good.filter.unique.precluster.uchime.chimeras,  
agri.trim.unique.good.filter.unique.precluster.uchime.accnos
```

#Also make sure that the file uchime is actually in the same directory. Also need to check that the privileges are set, i.e. that others can read and execute the program or it will fail.

```
./mothur
```

```
"#remove.seqs(accnos=27_drop.trim.unique.good.filter.unique.precluster.uchime.accnos,  
fasta=27_drop.trim.unique.good.filter.unique.precluster.fasta,  
name=27_drop.trim.unique.good.filter.unique.precluster.names, group=27_drop.good.groups,  
dups=T)"
```

```
./mothur "#classify.seqs(fasta=27_drop.trim.unique.good.filter.unique.precluster.pick.fasta,  
name=27_drop.trim.unique.good.filter.unique.precluster.pick.names,  
group=27_drop.good.pick.groups, template=trainset9_032012.pds.fasta,  
taxonomy=trainset9_032012.pds.tax, cutoff=80, processors=64)"
```

```
./mothur "#remove.lineage(fasta=27_drop.trim.unique.good.filter.unique.precluster.pick.fasta,  
name=27_drop.trim.unique.good.filter.unique.precluster.pick.names,  
group=27_drop.good.pick.groups,  
taxonomy=27_drop.trim.unique.good.filter.unique.precluster.pick.pds.wang.taxonomy,  
taxon=Mitochondria-Chloroplast-Archaea-Eukaryota-unknown)"
```

```
#Output files: agri.trim.unique.good.filter.unique.precluster.pick.pick.names,  
agri.trim.unique.good.filter.unique.precluster.pick.pick.fasta, agri.good.pick.pick.groups,  
agri.trim.unique.good.filter.unique.precluster.pick.pds.wang.pick.taxonomy.
```



```
./mothur "#dist.seqs(fasta=27_drop.trim.unique.good.filter.unique.precluster.pick.pick.fasta,  
cutoff=0.03, processors=32)"
```

#Output files

#As far as cutoff being set at 0.03 refers to the species level. You can adjust it to various amounts if you would like.

```
./mothur "#dist.seqs(fasta=27_drop.trim.unique.good.filter.unique.precluster.pick.pick.fasta,  
output=phylip, processors=32)"
```

```
./mothur
```

```
"#clearcut(phylip=27_drop.trim.unique.good.filter.unique.precluster.pick.pick.phylip.dist)
```

#Outputs phylogenetic tree using clearcut

```
./mothur "#cluster(column=27_drop.trim.unique.good.filter.unique.precluster.pick.pick.dist,  
name=27_drop.trim.unique.good.filter.unique.precluster.pick.pick.names)"
```

```
./mothur
```

```
"#make.shared(list=27_drop.trim.unique.good.filter.unique.precluster.pick.pick.an.list,  
group=27_drop.good.pick.pick.groups, label=0.03)"
```

#Note these steps are typically done calling up an interactive shell within the Talon server,
"qlogin -q test.q". Then the commands can be typed directly into the mothur.exe: by typing in
"./mothur"

```
./mothur "#count.groups()"
```

Using 27_drop.trim.unique.good.filter.unique.precluster.pick.pick.an.shared as input file for the shared parameter.

**1 contains 16389.
10 contains 14976.
11 contains 16619.
12 contains 29819.
13 contains 13492.
14 contains 24658.
15 contains 16893.
16 contains 12840.
17 contains 3.
18 contains 13984.
2 contains 22215.
3 contains 16520.
4 contains 9641.
5 contains 18195.
6 contains 18112.
7 contains 15032.
8 contains 33608.
9 contains 17634.**

Total seqs: 310630.

**Output File Names:
count.summary**

```
./mothur  
"#sub.sample(shared=27_drop.trim.unique.good.filter.unique.precluster.pick.pick.an.shared,  
size=9641)"
```

**17 contains 3. Eliminating.
Sampling 9641 from each group.
0.01**

Output File Names:

27_drop.trim.unique.good.filter.unique.precluster.pick.pick.an.0.01.subsample.shared

mothur >

```
collect.single(shared=27_drop.trim.unique.good.filter.unique.precluster.pick.pick.an.0.01.subsample.shared, calc=chao-invsimpson, freq=100)
```

mothur >

```
rarefaction.single(shared=27_drop.trim.unique.good.filter.unique.precluster.pick.pick.an.0.01.subsample.shared, calc=sobs, freq=100)
```

```
mothur > summary.single(calc=nseqs-coverage-sobs-invsimpson, subsample=9641)
```

Using

27_drop.trim.unique.good.filter.unique.precluster.pick.pick.an.0.01.subsample.shared as input file for the shared parameter.

APPENDIX C

BARCODE OLIGOS FILE REQUIRED FOR MOTHUR

The reverse primer is not included due to the sequencing reaction (200 bp) not typically reaching the end of the sequence.

forward GTGCCAGCMGCCGCGGTAA

reverse

barcode	CTAAGGTAAC	1
barcode	TAAGGAGAAC	2
barcode	AAGAGGATTC	3
barcode	TACCAAGATC	4
barcode	CAGAAGGAAC	5
barcode	CTGCAAGTTC	6
barcode	TTCGTGATTC	7
barcode	TTCCGATAAC	8
barcode	TGAGCGGAAC	9
barcode	CTGACCGAAC	10
barcode	TCCTCGAATC	11
barcode	TAGGTGGTTC	12
barcode	TCTAACGGAC	13
barcode	TTGGAGTGTC	14
barcode	TCTAGAGGTC	15
barcode	TCTGGATGAC	16
barcode	TCTATTCGTC	17

barcode	AGGCAATTGC	18
barcode	TTAGTCGGAC	19
barcode	CAGATCCATC	20
barcode	TCGCAATTAC	21
barcode	TTCGAGACGC	22
barcode	TGCCACGAAC	23
barcode	AACCTCATTC	24
barcode	CCTGAGATAC	25
barcode	TTACAACCTC	26
barcode	AACCATCCGC	27
barcode	ATCCGGAATC	28
barcode	TCGACCACTC	29
barcode	CGAGGTTATC	30
barcode	TCCAAGCTGC	31
barcode	TCTTACACAC	32

APPENDIX D

ION TORRENT BARCODED PRIMERS FOR LONG-AMPLICON 16S PCR

Forward	Primer A-key	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG{barcode}{template-specific-primer}-3'	
Reverse	Primer P-key	5'-CCTCTCTATGGGCAGTCGGTGATTCAG{ template-specific-primer}-3'	
Forward Primers			barcode
1ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGTGCCAGCMGCCGCGGTAA-3'	CTAAGGTAAC
2ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGTGCCAGCMGCCGCGGTAA-3'	TAAGGAGAAC
3ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGTGCCAGCMGCCGCGGTAA-3'	AAGAGGATTC
4ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGTGCCAGCMGCCGCGGTAA-3'	TACCAAGATC
5ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACGTGCCAGCMGCCGCGGTAA-3'	CAGAAGGAAC
6ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTCGTGCCAGCMGCCGCGGTAA-3'	CTGCAAGTTC
7ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCGTGCCAGCMGCCGCGGTAA-3'	TTCGTGATTC
8ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGGATAACGTGCCAGCMGCCGCGGTAA-3'	TTCCGATAAC
9ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGCGGAACGTGCCAGCMGCCGCGGTAA-3'	TGAGCGGAAC
10ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACCGAACGTGCCAGCMGCCGCGGTAA-3'	CTGACCGAAC
11ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTCGAATCGTGCCAGCMGCCGCGGTAA-3'	TCCTCGAATC
12ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGGTGGTTCGTGCCAGCMGCCGCGGTAA-3'	TAGGTGGTTC
13ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAACGGACGTGCCAGCMGCCGCGGTAA-3'	TCTAACGGAC
14ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGTGTCGTGCCAGCMGCCGCGGTAA-3'	TTGGAGTGTC
15ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGAGGTCGTGCCAGCMGCCGCGGTAA-3'	TCTAGAGGTC
16ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGATGACGTGCCAGCMGCCGCGGTAA-3'	TCTGGATGAC
17ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTATTTCGTGCCAGCMGCCGCGGTAA-3'	TCTATTTCGT
18ionA-515F		5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGAGGCAATTGCGTGCCAGCMGCCGCGGTAA-3'	AGGCAATTGC
Reverse Primer			
IonP1-806R		5'-CCTCTCTATGGGCAGTCGGTGATGGACTACHVGGGTWTCTAAT-3'	

APPENDIX E

SUMMARY.SEQS COMMAND OUTPUTS FROM MOTHUR

Shows the output files and summary statistics for each iterations of the quality-check, alignments, and screening outputs.

mothur > summary.seqs(fasta=27_drop.fasta)

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	8	8	0	1	1
2.5%-tile:	1	31	31	0	2	16471
25%-tile:	1	118	118	0	3	164705
Median:	1	240	240	0	3	329409
75%-tile:	1	278	278	0	4	494113
97.5%-tile:	1	292	292	0	6	642347
Maximum:	1	367	367	0	22	658817

of Seqs: 658817

mothur > summary.seqs(fasta=27_drop.trim.fasta)

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	200	200	0	2	1
2.5%-tile:	1	207	207	0	3	8321
25%-tile:	1	244	244	0	3	83202
Median:	1	249	249	0	4	166404
75%-tile:	1	257	257	0	5	249605
97.5%-tile:	1	264	264	0	6	324486
Maximum:	1	338	338	0	8	332806

of Seqs:332806

```
mothur > summary.seqs(fasta=27_drop.trim.unique.fasta,  
name=27_drop.trim.name)
```

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	200	200	0	2	1
2.5%-tile:	1	207	207	0	3	8321
25%-tile:	1	244	244	0	3	83202
Median:	1	249	249	0	4	166404
75%-tile:	1	257	257	0	5	249605
97.5%-tile:	1	264	264	0	6	324486
Maximum:	1	338	338	0	8	332806

of unique seqs: 94579

total # of seqs: 332806

```
mothur > summary.seqs(fasta=27_drop.trim.unique.align,  
name=27_drop.trim.names)
```

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1044	1046	1	0	1	1
2.5%-tile:	13862	22096	207	0	3	8321
25%-tile:	13862	22580	244	0	3	83202
Median:	13862	22588	249	0	4	166404
75%-tile:	13862	23959	257	0	5	249605
97.5%-tile:	13862	25283	264	0	6	324486
Maximum:	43116	43116	319	0	8	332806

of unique seqs: 94579

total # of seqs: 332806

```
mothur > summary.seqs(fasta=27_drop.trim.unique.good.align,  
name=27_drop.trim.good.names)
```

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	13855	22438	206	0	3	1
2.5%-tile:	13862	22529	225	0	3	7892
25%-tile:	13862	22583	245	0	4	78917
Median:	13862	22588	249	0	4	157834
75%-tile:	13862	23961	257	0	5	236751
97.5%-tile:	13862	25283	264	0	6	307776
Maximum:	13862	26807	319	0	8	315667

of unique seqs: 85349

total # of seqs: 315667

```
mothur > summary.seqs(fasta=27_drop.trim.unique.good.filter.unique.fasta,  
name=27_drop.trim.unique.good.filter.names)
```

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	422	203	0	3	1
2.5%-tile:	1	425	212	0	3	7892
25%-tile:	1	425	214	0	3	78917
Median:	1	425	214	0	3	157834
75%-tile:	1	425	214	0	5	236751
97.5%-tile:	1	425	216	0	6	307776
Maximum:	1	425	233	0	8	315667

of unique seqs: 28168

total # of seqs: 315667

```
mothur>summary.seqs(fasta=27_drop.trim.unique.good.filter.unique.precluster
.fasta, name=27_drop.trim.unique.good.filter.unique.precluster.names,
processors=32)
```

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	422	203	0	3	1
2.5%-tile:	1	425	213	0	3	7892
25%-tile:	1	425	214	0	3	78917
Median:	1	425	214	0	3	157834
75%-tile:	1	425	214	0	5	236751
97.5%-tile:	1	425	216	0	6	307776
Maximum:	1	425	233	0	8	315667

of unique seqs:14533

total # of seqs: 315667

./mothur

```
"#chimera.uchime(fasta=27_drop.trim.unique.good.filter.unique.precluster.fasta,  
name=27_drop.trim.unique.good.filter.unique.precluster.names,  
group=27_drop.good.groups, processors=32)"
```

Chimera check

It took 59 secs to check 16401 sequences. 731 chimeras were found.

The number of sequences checked may be larger than the number of unique sequences because some sequences are found in several samples.

mothur >

```
summary.seqs(fasta=27_drop.trim.unique.good.filter.unique.precluster.pick.fasta,name=27_drop.trim.unique.good.filter.unique.precluster.pick.names,processors=32)
```

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	422	203	0	3	1
2.5%-tile:	1	425	213	0	3	7774
25%-tile:	1	425	214	0	3	77739
Median:	1	425	214	0	3	155478
75%-tile:	1	425	214	0	5	233216
97.5%-tile:	1	425	216	0	6	303181
Maximum:	1	425	233	0	8	310954

of unique seqs: 13802

total # of seqs: 310954

BIBLIOGRAPHY

1. **Kriegel RM, Huang X, Schultheis MW.** 2009. Making polyester monomers from renewable raw materials; environmentally friendly; packaging. US20090246430 A1. USPTO.
2. **Ohtani Y, Mazumder BB, Sameshima K.** 2001. Influence of the chemical composition of kenaf bast and core on the alkaline pulping response. *J. Wood Sci.* **47**:30–35.
3. **D’Souza NA, Allen MS, Stevens K, Ayre B, Visi DK, Vidhate S, Ghamarian I, Webber III CL.** 2011. Biocomposites: The natural fiber contribution from bast and woody plants, p. 75–95. *In* Webber III, CL, Liu, A (eds.), *Plant Fibers as Renewable Feedstocks for Biofuel and Bio-based Products*. CCG International Inc., St. Paul, USA.
4. **Dempsey JM.** 1975. *Fiber Crops*. The University Presses of Florida, Gainesville.
5. **Nieschlag H, Nelson G, Wolff I, Perdue Jr. R.** 1960. A search for new fiber crops. *TAPPI* **43**:193–201.
6. **Anandjiwala R, Blouw S.** 2007. Composites from bast fibers: prospects and potential in the changing market environment. *J. Nat. Fibers* **4**:91–109.
7. **Mueller D, Krobjilowski A.** 2003. New discovery in the properties of composites reinforced with natural fibers. *J. Ind. Text.* **33**:111–130.
8. **Ogbomo S, Chapman K, Webber C, Bledoe R, D’Souza NA.** 2009. Benefits of low fenaf loading in biobased composites of Poly(L-Lactide) and the kenaf fiber. *J. Appl. Polym. Sci.* **112**:1294–1301.
9. **Parikh DV, Chen Y, Sun L.** 2006. Reducing automotive interior noise with natural fiber nonwoven floor covering systems. *Text. Res. J.* **76**:813–820.
10. **Webber III CL, Bledsoe VK, Bledsoe RE.** 2002. Kenaf harvesting and processing, p. 340–347. *In* J. Janick, AW (ed.), *Trends in New Crops and New Uses*. ASHS Press, Alexandria, VA.
11. **Schmiedel I, Barfuss GS, Nickel T, Pfeufer L.** 2014. Use of visible natural fibres in vehicle interiors. *ATZ Worldw.* **116**:20–23.
12. **Ramaswamy GN, Ruff CG, Boyd CR.** 1994. Effect of bacterial and chemical retting on kenaf fiber quality. *Text. Res. J.* **64**:305–308.

13. **Akin D, Foulk J, Dodd R, McAlister D.** 2001. Enzyme-retting of flax and characterization of processed fibers. *J. Biotechnol.* **89**:193–203.
14. **Evans JD, Akin DE, Foulk JA.** 2002. Flax-retting by polygalacturonase-containing enzyme mixtures and effects on fiber properties. *J. Biotechnol.* **97**:223–231.
15. **Barreveld WH.** 1993. Date Palm ProductsFAO Agricultural Services Bulletin No. 101. Rmoe, Food, and Agriculture Organization of the United Nations.
16. **Ahmed Z, Nizam SA.** 2008. Jute - microbiological and biochemical research. *Plant Tissue Cult. Biotechnol.* **18**:197–220.
17. **Kawahara Y, Tadokoro K, Endo R, Shioya M, Sugimura Y, Furusawa T.** 2005. Chemically retted kenaf fibers. *Soc. Fiber Sci. Technol. Japan* **61**:115–117.
18. **Khristova P, Kordsachia O, Khider T.** 2005. Alkaline pulping with additives of date palm rachis and leaves from Sudan. *Bioresour. Technol.* **96**:79–85.
19. **Taha I, Steuernagel L, Ziegmann G.** 2007. Optimization of the alkali treatment process of date palm fibres for polymeric composites. *Compos. Interfaces* **14**:669–684.
20. **Yang B, Nar M, Zhao S, Visi DK, Allen MS, Ayre BG., Webber III CL, Lu H, D'Souza NA.** 2014. Effects of alkaline and pectinase retted kenaf fibers on Poly(Hydroxybutyrate-CO-Valerate)/Poly(Butylene Adipate-CO-Terephthalate) polymer blend properties. *Compos. Part B Eng.* **56**:926–933.
21. **Mussig J.** 2010. *Industrial Applications of Natural Fibres: Properties and Technical Applications.*
22. **Saleem Z, Rennebaum H, Pudel F, Grimm E.** 2008. Treating bast fibres with pectinase improves mechanical characteristics of reinforced thermoplastic composites. *Compos. Sci. Technol.* **68**:471–476.
23. **Jacob N, Niladevi KN, Anisha GS, Prema P.** 2008. Hydrolysis of pectin: an enzymatic approach and its application in banana fiber processing. *Microbiol. Res.* **163**:538–44.
24. **Kapoor M, Beg QK, Bhushan B, Singh K, Dadhich KS, Hoondal GS.** 2001. Application of an alkaline and thermostable polygalacturonase from *Bacillus* sp. MG-CP-2 in degumming of ramie (*Boehmeria nivea*) and sunn hemp (*Crotalaria juncea*) bast fibres. *Process Biochem.* **36**:803–807.
25. **Beg QK, Bhushan B, Hoondal GS.** 2000. Production and characterization of thermostable xylanase and pectinase from *Streptomyces* sp. QG-11-3. *J. Ind. Microbiol. Biotech.* **24**:396–402.

26. **Boland WE, Henriksen ED, Doran-Peterson J.** 2010. Characterization of two pectate lyases from *Paenibacillus amylolyticus* strain 27C64 with activity on highly methylated pectin. *Appl. Environ. Microbiol.* **76**:6006–6009.
27. **Soriano M, Diaz P, Pastor J.** 2005. Pectinolytic Systems of Two Aerobic Sporogenous Bacterial Strains with High Activity on Pectin. *Curr. Microbiol.* **50**:114–118.
28. **Tewari R, Tewari R, Hoondal G.** 2005. Microbial pectinases, p. 191–208. *In* Barredo, JL (ed.), *Methods in Biotechnology, Vol. 17: Microbial Enzymes and Biotransformations.* Human Press Inc., Totowa, NJ.
29. **Gruben BS, Zhou M, Wiebenga A, Ballering J, Overkamp KM, Punt PJ, de Vries RP.** 2014. *Aspergillus niger* RhaR, a regulator involved in L-rhamnose release and catabolism. *Appl. Microbiol. Biotechnol.* **98**:5531–40.
30. **Kaur A, Singh A, Mahajan R.** 2014. Characterization of industrially-valuable xylanopectinolytic enzymes produced concurrently by a novel isolate of *Bacillus pumilus*. *Biotechnol. Lett.*
31. **Tu T, Bai Y, Luo H, Ma R, Wang Y, Shi P, Yang P, Meng K, Yao B.** 2014. A novel bifunctional pectinase from *Penicillium oxalicum* SX6 with separate pectin methylesterase and polygalacturonase catalytic domains. *Appl. Microbiol. Biotechnol.* **98**:5019–28.
32. **Bonnin E, Garnier C, Ralet MC.** 2014. Pectin-modifying enzymes and pectin-derived materials: applications and impacts. *Appl. Microbiol. Biotechnol.* **98**:519–32.
33. **Kashyap DR, Vohra PK, Chopra S, Tewari R.** 2001. Applications of pectinases in the commercial sector : a review. *Bioresour. Technol.* **77**:215–227.
34. **Amann R, Ludwig W, Schleifer K.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**:143–169.
35. **Borneman J, Skroch P, O’Sullivan K, Palus J, Rumjanek N, Jansen J, Nienhuis J, Triplett E.** 1996. Molecular microbial diversity of an agricultural soil in Wisconsin. *Appl. Environ. Microbiol.* **62**:1935–1943.
36. **Torsvik V, Øvreås L.** 2002. Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol.* **5**:240–5.
37. **Woese CR, Fox GE.** 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* **74**:5088–5090.
38. **Olsen G, Woese C.** 1993. Ribosomal RNA: a key to phylogeny. *FASEB J.* **7**:113–123.

39. **Woese C, Gutell R, Gupta R, Noller H.** 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol. Rev.* **47**.
40. **Shine J, Dalgarno L.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U. S. A.* **71**:1342–1346.
41. **Lee K, Holland-Staley C, Cunningham P.** 1996. Genetic analysis of the Shine-Dalgarno interaction: selection of alternative functional mRNA-rRNA combinations. *Rna* **2**:1270–1285.
42. **Jin H, Zhao Q, Gonzalez de Valdivia E, Ardell DH, Stenström M, Isaksson LA.** 2006. Influences on gene expression in vivo by a Shine – Dalgarno sequence. *Mol. Microbiol.* **60**:480–492.
43. **Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD, Gold L.** 1992. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Mol. Microbiol.* **6**:1219–1229.
44. **Baker GC, G, Smith JJ, J, Cowan DDA.** 2003. Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Methods* **55**:541–555.
45. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
46. **Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM.** 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37**:D141–5.
47. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner F.** 2013. The SILVA Ribosomal RNA Gene Database Project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**:D590–D596.
48. **DeSantis T, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**:5069–5072.
49. **Junemann S, Sedlazeck F, Prior K, Albersmeier A, John U, Kalinowski J, Mellmann A, Goesmann A, von Haeseler A, Stoye J, Harmsen D.** 2013. Updating benchtop sequencing performance comparison. *Nat. Biotechnol.* **31**:294–296.
50. **Tonge D, Pashley C, Grant T.** 2014. Amplicon-based metagenomic analysis of mixed fungal samples using proton release amplicon sequencing. *PLoS One* **9**.

51. **Moré MI, Herrick JB, Silva MC, Ghiorse WC, Madsen EL.** 1994. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Appl. Environ. Microbiol.* **60**:1572–80.
52. **Feinstein LM, Woo JS, Blackwood CB.** 2009. Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Appl. Environ. Microbiol.* **75**:5428–5433.
53. **Polz MF, Cavanaugh CM.** 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **64**:3724–30.
54. **Suzuki MT, Giovannoni SJ.** 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol.* **62**:625–30.
55. **Lahr DJ, Katz LA.** 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *Biotechniques* **47**:857–66.
56. **Fazekas A, Steeves R, Newmaster S.** 2010. Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques* **48**:277–85.
57. **Showalter AK, Lamarche BJ, Bakhtina M, Su M, Tang K, Tsai M.** 2006. Mechanistic comparison of high-fidelity and error-prone DNA polymerases and ligases involved in DNA repair. *Chem. Rev.* **106**:340–60.
58. **Ahn J, Kim B, Song J, Weon H.** 2012. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *J. Microbiol.* **50**:1071–4.
59. **Smyth RP, Schlub TE, Grimm A, Venturi V, Chopra A, Mallal S, Davenport MP, Mak J.** 2010. Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene* **469**:45–51.
60. **Thompson JR, Marcelino LA, Polz MF.** 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by “reconditioning PCR”. *Nucleic Acids Res.* **30**:2083–8.
61. **Meyerhans A, Vartanian JP, Wain-Hobson S.** 1990. DNA recombination during PCR. *Nucleic Acids Res.* **18**:1687–91.
62. **Judo MSB, Wedel AB, Wilson C.** 1998. Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res.* **26**:1819–1825.

63. **Odelberg SJ, Weiss RB, Hata A, White R.** 1995. Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. *Nucleic Acids Res.* **23**:2049–57.
64. **Liesack W, Weyland H, Stackebrandt E.** 1991. Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. *Microb. Ecol.* **21**:191–8.
65. **Qiu X, Wu L, Huang H, McDonel PE, Palumbo A V, Tiedje JM, Zhou J.** 2001. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Appl. Environ. Microbiol.* **67**:880–7.
66. **Wright E, Yilmaz L, Noguera D.** 2012. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl. Environ. ...* **78**:717–725.
67. **Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward D V, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methé B, DeSantis TZ, Petrosino JF, Knight R, Birren BW.** 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* **21**:494–504.
68. **Yelton AP, Comolli LR, Justice NB, Castelle C, Deneff VJ, Thomas BC, Banfield JF.** 2013. Comparative genomics in acid mine drainage biofilm communities reveals metabolic and structural differentiation of co-occurring archaea. *BMC Genomics* **14**:485.
69. **Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA.** 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**:386.
70. **Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F.** 2010. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc.* **2010**:1–10.
71. **Mullis KB, Faloona FA.** 1987. Specific Synthesis of DNA in vitro via a Polymerase-Catalyzed Chain Reaction. *Methods Enzymol.* **155**:335–350.
72. **Saiki RK, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA, Gelfand DH.** 1988. Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science (80-.).* **239**:487–491.
73. **Menchaca AC, Visi DK, Strey OF, Teel PD, Kalinowski K, Allen MS, Williamson PC.** 2013. Preliminary Assessment of Microbiome Changes Following Blood-Feeding and Survivorship in the *Amblyomma americanum* Nymph-to-Adult Transition using Semiconductor Sequencing. *PLoS One* **8**:e67129.

74. **Simon S.** 2014. Cloacal microbiota of captive-Bred and wild attwater's prairie chicken, *Tympanuchus cupido attwateri*. University of North Texas.
75. **Hammerly SC, Morrow ME, Johnson JA.** 2013. A comparison of pedigree- and DNA-based measures for identifying inbreeding depression in the critically endangered Attwater's Prairie-chicken. *Mol. Ecol.* **22**:5313–28.
76. **Gevers D, Pop M, Schloss PD, Huttenhower C.** 2012. Bioinformatics for the Human Microbiome Project. *PLoS Computational Biol.* **8**:e1002779.
77. **Yang L, Chaudhary N, Baghdadi J, Pei Z.** Microbiome in reflux disorders and esophageal adenocarcinoma. *Cancer J.* **20**:207–10.
78. **Wylie KM, Truty RM, Sharpton TJ, Mihindukulasuriya KA, Zhou Y, Gao H, Sodergren E, Weinstock GM, Pollard KS.** 2012. Novel bacterial taxa in the human microbiome. *PLoS One* **7**:e35294.
79. **Pop M, Walker AW, Paulson J, Lindsay B, Antonio M, Hossain M, Oundo J, Tamboura B, Mai V, Astrovskaya I, Bravo H, Rance R, Stares M, Levine MM, Panchalingam S, Kotloff K, Ikumapayi UN, Ebruke C, Adeyemi M, Ahmed D, Ahmed F, Alam M, Amin R, Siddiqui S, Ochieng JB, Ouma E, Juma J, Mailu E, Omere R, Morris J, Breiman RF, Saha D, Parkhill J, Nataro JP, Stine O.** 2014. Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition. *Genome Biol.* **15**:R76.
80. **Mason OU, Scott NM, Gonzalez A, Robbins-Pianka A, Bælum J, Kimbrel J, Bouskill NJ, Prestat E, Borglin S, Joyner DC, Fortney JL, Jurelevicius D, Stringfellow WT, Alvarez-Cohen L, Hazen TC, Knight R, Gilbert JA, Jansson JK.** 2014. Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. *ISME J.* **8**:1464–75.
81. **Ahmad M.** 1963. Studies on jute retting bacteria. *J. Appl. Microbiol.* **26**:117–126.
82. **Ali MM.** 1958. Aerobic bacteria involved in the retting of jute. *Appl. Microbiol.* **6**:87–89.
83. **Offonry S, Achi O.** 1998. Microbial populations associated with the retting of melon pods (*Colocynthis citrullus L.*) during seed recovery. *Plant Foods Hum Nutr* **52**:37–47.
84. **Tamburini E, Leon AG, Perito B, Mastromei G.** 2003. Characterization of bacterial pectinolytic strains involved in the water retting process. *Environ. Microbiol.* **5**:730–736.
85. **Banik S, Basak M, Sil S.** 2007. Effect of inoculation of pectinolytic mixed bacterial culture on improvement of ribbon retting of jute and kenaf. *J. Nat. Fibers* **4**:33–50.

86. **Di Candilo M, Bonatti PM, Guidetti C, Focher B, Grippo C, Tamburini E, Mastromei G.** 2010. Effects of selected pectinolytic bacterial strains on water-retting of hemp and fibre properties. *J. Appl. Microbiol.* **108**:194–203.
87. **Yu H, Yu C.** 2007. Study on microbe retting of kenaf fiber. *Enzyme Microb. Technol.* **40**:1806–1809.
88. **Bin Z, Guo-Cheng D, Wei S, Jian Z, Jian C.** 2008. Expression of *Bacillus subtilis* pectate lyase gene in *Pichia pastoris*. *Biochem. Eng. J.* **40**:92–98.
89. **Kobayashi T, Higaki N, Yajima N, Suzumatsu A, Hagihara H, Kawai S, Ito S.** 2001. Purification and properties of a galacturonic acid-releasing exopolygalacturonase from a strain of *Bacillus*. *Biosci. Biotechnol. Biochem.*, 2001/06/05 ed. **65**:842–847.
90. **Singh SA, Plattner H, Diekmann H.** 1999. Exopolygalacturonate lyase from a thermophilic *Bacillus* sp. *Enzyme Microb. Technol.* **25**:420–425.
91. **Takao M, Nakaniwa T, Yoshikawa K, Terashita T, Sakai T.** 2000. Purification and characterization of thermostable pectate lyase with protopectinase activity from thermophilic *Bacillus* sp. TS 47. *Biosci. Biotech. Biochem.* **64**:2360–2367.
92. **Nielsen P, Sorensen J.** 1997. Multi-target and medium-independent fungal antagonism by hydrolytic enzymes in *Paenibacillus polymyxa* and *Bacillus pumilus* Strains from barley rhizosphere. *FEMS Microbiol. Ecol.* **22**:183–192.
93. **Schink B, Zeikus JG.** 1983. Characterization of pectinolytic enzymes of *Clostridium thermosulfurogenes*. *FEMS Microbiol. Lett.* **17**:295–298.
94. **Tachaapaikoon C, Kosugi A, Pason P, Waeonukul R, Ratanakhanokchai K, Kyu K, Arai T, Murata Y, Mori Y.** 2012. Isolation and characterization of a new cellulosome-producing *Clostridium thermocellum* strain. *Biodegradation* **23**:57–68.
95. **Tamaru Y, Doi RH.** 2000. Pectate Lyase A, an enzymatic subunit of the *Clostridium cellulovorans* cellulosome. *Proc. Natl. Acad. Sci.* **98**:4125–4129.
96. **Munshi TK, Chattoo BB.** 2008. Bacterial population structure of the jute-retting environment. *Microb. Ecol.* **56**:270–282.
97. **Blevins SM, Bronze MS.** 2010. Robert Koch and the “golden age” of bacteriology. *Int. J. Infect. Dis.* **14**:e744–51.
98. **Wade W.** 2002. Unculturable bacteria--the uncharacterized organisms that cause oral infections. *J. R. Soc. Med.* **95**:81–3.

99. **Lane J, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR, Lane DJ, Stahl DA, Sogin ML.** 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad.* **82**:6955–6959.
100. **Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Bio. Evol.* **28**:2731–2739.
101. **Bertani G.** 1951. The Mode of Phage Liberation by Lysogenic *Escherichia coli*. *J. Bact.* **62**:293–300.
102. **Weisburg W, Barns S, Pelletier D, Lane D.** 1991. 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* **173**:697–703.
103. **Blakenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, GalaxyTeam.** 2010. Manipulation of FASTQ data with Galaxy. *Bioinforma.* **26**:1783–1785.
104. **Claesson MJ, O’Sullivan O, Wang Q, Nikkila J, Marchesi JR, Smidt H, de Vos WM, Ross RP, O’Toole PW.** 2009. Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS One* **4**:e6669.
105. **Nawrocki EP, Kolbe DL, Eddy SR.** 2009. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**:1335–7.
106. **Nawrocki EP, Eddy SR.** 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**:2933–5.
107. **Stackebrandt E, Goebel BM.** 1994. Taxonomic Note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* **44**:846–849.
108. **Minamisawa K, Nishioka K, Miyaki T, Ye B, Miyamoto T, You M, Saito A, Saito M, Barraquio WL, Teumroong N, Sein T.** 2004. Anaerobic Nitrogen-Fixing Consortia Consisting of Clostridia Isolated from Gramineous Plants. *Appl. Environ. Microbiol.* **70**:3096–3102.
109. **Gardner RG., Wells JE., Fields MW., Wilson DB., Russell JB.** 1997. A *Prevotella ruminicola* B14 operon encoding extracellular polysaccharide hydrolases. *Curr. Microbiol.* **35**:274–277.
110. **Hong PY, Wheeler E, Cann IK, Mackie RI.** 2011. Phylogenetic analysis of the fecal microbial community in herbivorous land and marine iguanas of the Galápagos Islands using 16S rRNA-based pyrosequencing. *ISME J.* **5**:1461–70.

111. **Donaghy JA, Levett PN, Haylock RW.** 1990. Changes in microbial populations during anaerobic flax retting. *J. Appl. Bacteriol.* **69**:634–641.
112. **Huang E, Yousef AE.** 2012. Draft genome sequence of *Paenibacillus polymyxa* OSY-DF, which coproduces a lantibiotic, paenibacillin, and polymyxin E1. *J. Bact.* **194**:4739–4740.
113. **Miyamoto T, Kawahara M, Minamisawa K.** 2004. Novel endophytic nitrogen-fixing *Clostridia* from the Grass *Miscanthus sinensis* as Revealed by Terminal Restriction Fragment Length Polymorphism Analysis. *Appl. Environ. Microbiol.* **70**:6580–6586.
114. **Schloss PD.** 2009. A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS One* **4**:e8230.
115. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R, Quine C.** 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**:2194–2200.
116. **Wang Q, Garrity GM, Tiedje JM, Cole JR.** 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**:5261–5267.
117. **Domingos P, Pazzani M.** 1997. On the optimality of the simple Bayesian classifier under zero-one loss. *Mach. Learn.* **29**:103–130.
118. **Whittaker RH.** 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* **30**:279–338.
119. **Whittaker RH.** 1972. Evolution and measurement of species diversity. *Taxon* **21**:213–251.
120. **Simpson EH.** 1949. Measurement of diversity. *Nature* **163**:688.
121. **Magurran AE.** 1988. Ecological diversity and its measurement. Springer Netherlands, Dordrecht.
122. **Zhou J, Xia B, Treves DS, Wu L-Y, Marsh TL, O’Neill R V., Palumbo A V., Tiedje JM.** 2002. Spatial and Resource Factors Influencing High Microbial Diversity in Soil. *Appl. Environ. Microbiol.* **68**:326–334.
123. **Lozupone C, Knight R.** 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**:8228–35.
124. **Hamady M, Lozupone C, Knight R.** 2010. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J.* **4**:17–27.

125. **Looft T, Johnson T a, Allen HK, Bayles DO, Alt DP, Stedtfeld RD, Sul WJ, Stedtfeld TM, Chai B, Cole JR, Hashsham S a, Tiedje JM, Stanton TB.** 2012. In-feed antibiotic effects on the swine intestinal microbiome. *Proc. Natl. Acad. Sci. U. S. A.* **109**:1691–6.
126. **Turnbaugh PJ, Hamady M, Yatsunencko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI.** 2009. A core gut microbiome in obese and lean twins. *Nature* **457**:480–4.
127. **Lepage P, Leclerc MC, Joossens M, Mondot S, Blottière HM, Raes J, Ehrlich D, Doré J.** 2013. A metagenomic insight into our gut’s microbiome. *Gut* **62**:146–58.
128. **Balfour SR.** 2007. Bacteria in Crohn’s disease: mechanisms of inflammation and therapeutic implications. *J. Clin. Gastroenterol.* **41**:37–43.
129. **Kaakoush NO, Day AS, Huinao KD, Leach ST, Lemberg DA, Dowd SE, Mitchell HM.** 2012. Microbial Dysbiosis in Pediatric Patients with Crohn ’ s Disease.
130. Clostridium difficile Infection. CDC Heal. Infect.
131. **Borody TJ, Khoruts A.** 2012. Fecal microbiota transplantation and emerging applications. *Nat. Rev. Gastroenterol. Hepatol.* **9**:88–96.
132. **Bakken JS, Borody T, Brandt LJ, Brill J V, Demarco DC, Franzos MA, Kelly C, Khoruts A, Louie T, Martinelli LP, Moore TA, Russell G, Surawicz C.** 2011. Treating *Clostridium difficile* infection with fecal microbiota transplantation. *Clin. Gastroenterol. Hepatol.* **9**:1044–9.
133. **Brandt LJ.** 2012. Fecal transplantation for the treatment of *Clostridium difficile* infection. *Gastroenterol. Hepatol. (N. Y).* **8**:191–4.
134. **Guslandi M, Mezzi G, Sorghi M, Testoni PA.** *Saccharomyces boulardii* in Maintenance Treatment of Crohn’s Disease. *Dig. Dis. Sci.* **45**:1462–1464.
135. **Saito A, Kawahara M, Ikeda S, Ishimine M, Akao S, Minamisawa K.** 2008. Broad distribution and phylogeny of anaerobic endophytes of cluster XIVa *Clostridia* in plant species including crops. *Microbes Environ.* **23**:73–80.