EVALUATION OF MICROBIAL COMMUNITIES FROM EXTREME ENVIRONMENTS AS INOCULA IN A CARBOXYLATE PLATFORM FOR BIOFUEL PRODUCTION FROM CELLULOSIC BIOMASS

A Dissertation

by

JULIA LEE COPE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Heather H. Wilkinson
Committee Members,	Terry J. Gentry
	Heath J. Mills
	Leland S. Pierson III
Head of Department,	Leland S. Pierson III

August 2013

Major Subject: Plant Pathology

ABSTRACT

The carboxylate biofuels platform (CBP) involves the conversion of cellulosic biomass into carboxylate salts by a mixed microbial community. Chemical engineering approaches to convert these salts to a variety of fuels (diesel, gasoline, jet fuel) are well established. However, prior to initiation of this project, little was known about the influence of inoculum source on platform performance. The studies in this dissertation test the hypothesis that microbial communities from particular environments in nature (e.g. saline and/or thermal sediments) are pre-adapted to similar industrial process conditions and, therefore, exhibit superior performances. We screened an extensive collection of sediment samples from extreme environments across a wide geographic range to identify and characterize microbial communities with superior performances in the CBP. I sought to identify aspects of soil chemistry associated with superior CBP fermentation performance. We showed that CBP productivity was influenced by both fermentation conditions and inocula, thus is clearly reasonable to expect both can be optimized to target desired outcomes. Also, we learned that fermentation performance is not as simple as finding one soil parameter that leads to increases in all performance parameters. Rather, there are complex multivariate relationships that are likely indicative of trade-offs associated within the microbial communities.

An analysis of targeted locus pyrosequence data for communities with superior performances in the fermentations provides clear associations between particular bacterial taxa and particular performance parameters. Further, I compared microbial

community compositions across three different process screen technologies employed in research to understand and optimize CBP fermentations. Finally, we assembled and characterized an isolate library generated from a systematic culture approach. Based on partial 16S rRNA gene sequencing, I estimated operational taxonomic units (OTUs), and inferred a phylogeny of the OTUs. This isolate library will serve as a tool for future studies of assembled communities and bacterial adaptations useful within the CBP fermentations.

Taken together the tools and results developed in this dissertation provide for refined hypotheses for optimizing inoculum identification, community composition, and process conditions for this important second generation biofuel platform.

DEDICATION

This work is dedicated to Robert Cope. Without his support and encouragement I would not have been capable of reaching for the stars.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Heather Wilkinson, and my committee members, Dr. Terry Gentry, Dr. Heath Mills, and Dr. Leland Pierson III, for their guidance and support throughout the course of this research. I would like to thank Dr. David Appel for his support and encouragement during my undergraduate term in the Department of Plant Pathology. I thank Dr. Karen-Beth Scholthof for her many conversations and encouragements. I would like to thank all of my lab mates, collaborators, and classmates; your camaraderie and questions are a source of constant inspiration. I thank the army of undergraduate workers and interns who have helped with everything from collecting to culturing in the last five years. I hope I have brought science to life for each of you.

Thank you to the Texas AgriLife Research Bioenergy Program and the Texas A&M University Office of the Vice President for Research Energy Resources Program for their financial support in the form of grants that supported my collections, supplies, experiments, and salary. Finally, my thanks go to the following permitting institutions and the many people who gave generously of their time to assist with collecting samples. We sampled the following locations in accordance with the permits listed: Grulla National Wildlife Refuge, Muleshoe National Wildlife Refuge permit number (PN): 21590-2008-003; Salt Plains National Wildlife Refuge PN: Permit:RSS-08-0910, Station:21630; Brazoria National Wildlife Refuge PN: 21543-8-101 and 21543-2-09-2; Bitter Lake National Wildlife Refuge PN: 22510-08-06; Bottomless Lakes State Park

PN: SU 09-02-01; San Francisco Bay National Wildlife Refuge PN: Permit:2009-014, Station:81640; Big Bend National Park PN: BIBE-2008_SCI-0036; Fish Springs National Wildlife Refuge PN: Permit:09-0002, Station:65540; Bear River Reserve National Wildlife Refuge PN: Permit:09-001, Station:65530; Savannah National Wildlife Refuge and Pinkney Island National Wildlife Refuge PN: 41620-09-031; Cape Romain National Wildlife Refuge PN: Permit:005, Station:42510; Laguna Boqueron National Wildlife Refuge, Laguna Cartagena National Wildlife Refuge, Cabo Rojo National Wildlife Refuge, and Jabos Bay Research Reserve National Wildlife Refuge PN: Permit:9010, Station:41521; Caladesy Island State Park, Honeymoon Island State Park, Charlotte Harbor State Park, and Collier-Seminole State Park PN: 04_09_1939; The Thousands Islands National Wildlife Refuge PN: Permit:2009-007, Station:42555; Valles Caldera National Preserve PN: VCT-2009_RIM-004; Yellowstone National Park PN: YELL-2009-SCI-5786; Stillwater National Wildlife Refuge PN: Permit:09-023, Station:84590; and finally Hapuna Beach State Park, Kekaha Kai State Park, Wailoa River State Park, and Akaka Falls State Park PN: H72152.

NOMENCLATURE

16S rDNA	16S Subunit of the Ribosome Genetic Material
AEQ	Acetic Acid Equivalent
BTEFAP	Bacterial Tag-encoded FLX Amplicon Pyrosequencing
CAT	Cellulose Agar for Thermophiles
CBP	Carboxylate Biofuels Platform
CC	Counter Current Exchange
CPDM	Continuum Particle Dispersal Model
DTAM	Drake's Thermophilic Acetogen Medium
F-data	Fermentation Data
HMW	High Molecular Weight
MANOVA	Multivariate Analysis of Variance
MGM	Modified Growth Medium
NCBI	National Center for Biotechnology Information
NMDS	Non-Metric Multidimensional Scaling
OTU(s)	Operational Taxonomic Unit(s)
PCR	Polymerase Chain Reaction
PLS	Partial Least Squares
RDP	Ribosomal Database Project
S-data	Sediment Data
TDS	Total Dissolved Solids

TABLE OF CONTENTS

Page

Page

Page

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

The carboxylate biofuels platform (CBP), developed at Texas A&M University involves the conversion of biomass into carboxylate salts by a mixed microbial community isolated from a marine environment (Fu & Holtzapple, 2010a; Hollister *et al.*, 2010b). As a 'second generation' biofuel process the CBP does not take sugar to ethanol, but rather uses cellulose as a substrate and moves through fermentation pathways to carboxylic acids (buffered to salts) and further on to gasoline, diesel fuel, and jet fuels. The products of the CBP work with the existing petroleum distribution infrastructure. To date, research to understand and optimize CBP fermentations have manipulated the substrates (Domke *et al.*, 2004; Fu & Holtzapple, 2011; Garlock *et al.*, 2011) and platform process parameters (Forrest *et al.*, 2010; Fu, 2007; Golub, 2012). This has led to the CBP being flexible to use a wide variety of waste product feed stocks or substrates. Use of a mixed microbial community for the breakdown of cellulose to fermentable substrates and fermentation within a single bioprocess makes the process highly desirable relative to a sugar platform fermentation, which requires both energy intensive sterile conditions and separate addition of expensive enzymes prior to fermentation.

Prior to initiation of this project, we knew little about the influence of inoculum source on performance of this biofuel platform. However, it should be noted, each of the few informed attempts to improve the community greatly increased yields (Fu &

Holtzapple, 2010a); I predict the proposed strategy will result in continued enhancement and advancements of the platform. Specifically, a switch to a marine community isolated from Galveston Island, TX, more than doubled the acid yields from the platform relative to the original terrestrial (non-saline) soil community inocula (Fu & Holtzapple, 2010a). Furthermore, a community from an even more saline environment, the Great Salt Lake, in Salt Lake City, UT, boosted the performance another 20% relative to the Galveston community (Fu, 2007). Thus, it is evident inocula from particular environments have the potential to exhibit superior performances in this process. It should be noted, in both cases researchers detected this superior performance during experiments run at 40°C, not at 55°C and acid production was the performance parameter of interest. With these previous attempts to improve process performance by manipulation of the microbial community in mind, it seems reasonable to predict that sampling microbial communities from a variety of extreme environments in nature with features that resemble process conditions (e.g. thermal temperature and high salt concentration) we will find communities that exhibit superior performances in the process. We reasoned that composition of the microbial community within the fermentations contributes to process performance. For this research the foremost goal was to screen microbial communities from extreme environments for superior conversion capacity at 55 °C under conditions wherein the salt concentration begins at 2 to 2.6%. In comparison, the salinity of seawater is 3.5%. The reasoning is that identification of communities that convert biomass well at 55 °C in the presence of reasonably high salt concentrations would be the best candidate to optimize the platform because they possess the potential to reduce

residence times. The residence time, in context to this project, is the amount of time biomass remains within the fermentation. Since fermentations with longer residence times occupy space and require energy (e.g. mixing of the fermentation broth and heat to maintain temperature), decreasing residence times is a high priority in the optimization of this technology because it is reasonable to expect that any other performance parameters (e.g. yield) could be optimized by manipulating process conditions (Mark Holtzapple, personal communication to Heather Wilkinson).

Microbes in extreme environments have physiological adaptations that allow them to survive and thrive in normally adverse conditions including high temperatures and high salt concentrations (Mesbah *et al.*, 2007; Meyer-Dombard *et al.*, 2005; Porter *et al.*, 2007). Industrial processes with shorter residence times are most profitable. Thus, optimal production favors higher temperatures running at faster rates (Aitken & Mullennix, 1992). Microbes in industrial processes tend to perform optimally at lower product (i.e. acid or solvent) concentrations (Heipieper *et al.*, 2007; Taylor *et al.*, 2008). However, the higher the product concentration prior to recovery the more efficient the system is. Microbial communities from nature with adaptations for thermal (Rastogi *et al.*, 2010) and high cation (Mesbah & Wiegel, 2009) environments ought to be preadapted for superior performance in industrial processes that involve high temperature conditions and product accumulation in the form of salts.

This dissertation details studies conducted to test the hypothesis that microbial communities from particular environments in nature with features similar to process parameters in industry will be pre-adapted to the industrial process and therefore exhibit

superior performance. Briefly, in Chapter II we screen an extensive collection of sediment samples from putative extreme environments across a wide geographic range to identify microbial communities with superior performances in the CBP. Furthermore, I seek to identify aspects of soil chemistry associated with superior CBP fermentation performance. Chapter III presents an analysis of targeted locus pyrosequence data for communities with superior performances in the fermentations as indicated by performance in particular process parameters (conversion, total acetic acid equivalents, high carbon number volatile fatty acids). Moreover the results provide clear associations between particular bacterial taxa and particular performance parameters. In Chapter IV I compare microbial community compositions across three different process screen technologies employed in research to understand and optimize CBP fermentations. The novel aspects of this experiment were that we used fresh inocula from fresh collected sediment samples and ran the screens simultaneously. The sediments used in this study were from five locations that are ecologically diverse, and also, extensively studied in other CBP experiments (Chapter II), (Hammett, 2011; Hollister *et al.*, 2010b; Hollister *et al.*, 2012). I designed and conducted an analysis of the sequence data to discern the effects of the different screens on community composition. This experiment provides for a direct comparison of the screens that usually are performed sequentially using the same microbial community stored at -20°C in the interim between experiments. Finally, in Chapter V I describe our assembly and characterization of an isolate library generated from a relatively systematic culture approach, which included multiple different media and different oxygen restriction strategies to maximize the library diversity. Based on a

partial 16S rRNA gene segment sequenced for all isolates, I estimated operational taxonomic units (OTUs), and inferred a phylogeny of the OTUs. This isolate library will serve as a tool for future studies of assembled communities and bacterial adaptations useful within the CBP fermentations. In Chapter VI, I discuss the summary and analysis in context of the conclusions from these studies.

CHAPTER II

SCREENING MICROBIAL COMMUNITIES FROM EXTREME ENVIRONMENTS FOR PERFORMANCE IN A CARBOXYLATE BIOFUEL PLATFORM

II.1 Introduction

The carboxylate biofuels platform (CBP), developed at Texas A&M University (Mixalco™) involves the conversion of cellulosic and lignocellulosic biomass into carboxylate salts by fermentation with a mixed microbial community isolated from a sediment environment (Fu & Holtzapple, 2010b; Hollister *et al.*, 2011). In the CBP small chain (C2-C7) volatile fatty acids are produced and provide a substrate for the production of ethanol, gasoline, jet-fuel and many other chemicals (Holtzapple & Granda, 2009). The CBP can use as feedstock many non-food source materials to produce these acids including many landfill targeted wastes (e.g. yard clippings, kitchen waste), and industrial byproducts such as sugarcane bagasse (Fu & Holtzapple, 2011), paper fines and industrial biosludge (Domke *et al.*, 2004). Furthermore the CBP is a non-sterile process in which, under high salt and high temperature conditions, an initial microbial inoculum is sufficient to overtake microbial populations in the feedstock and nutrient sources used in the fermentation, negating the need for sterilization of these inputs.

In the CBP acids are buffered to carboxylate salts (Aiello-Mazzarri *et al.*, 2006). At the termination of the process, all salts are converted to acids and the total acid accumulation is measured. Prior to initiation of this project, we knew little about the

influence of inoculum source on performance of this biofuel platform. However, it should be noted, each of the few informed attempts to improve the community greatly increased yields. The platform was developed originally using terrestrial inocula from environments expected to favor the rapid degradation of biomass (e.g. inocula from a compost pile or a cow gut) (Fu, 2007). Since Dr. Holtzapple recognized that the productivity of the fermentations seemed to fall off at the same time it reached the maximum product concentration (Mark Holtzapple PhD, Texas A&M; personal communication), he reasoned that the communities were sensitive to product concentrations, a well established issue in industry (Taylor *et al.*, 2008). Since the products of these fermentations are salts, he chose to test new inocula from salty environments. Specifically, a switch to a marine community from a Galveston Island, TX, sediment more than doubled the acid yields from the platform relative to the original terrestrial (non-saline) soil community inocula (Thanakoses *et al.*, 2003a; Thanakoses *et al.*, 2003b). Furthermore, a community from an even more saline environment, the Great Salt Lake in Salt Lake City, UT, boosted the performance another 20% relative to the Galveston Island community (Fu, 2007). Thus, it is evident inocula from particular environments have the potential to exhibit superior performances in this process.

Microbes found in extreme environments are known to have physiological adaptations that allow them to live normally in adverse conditions including high temperatures and high salt concentrations (Mesbah *et al.*, 2007; Meyer-Dombard *et al.*, 2005; Porter *et al.*, 2007). Industrial processes that can be achieved in shorter time periods have the best potential to be profitable. Thus, optimal production favors higher

temperatures running at faster rates (Aitken & Mullennix, 1992). Microbes in industrial processes tend to perform optimally at lower product concentrations (Heipieper *et al.*, 2007; Taylor *et al.*, 2008). However, the higher the product concentration prior to recovery the more efficient the system is. Microbial communities from nature with adaptations for thermal (Rastogi *et al.*, 2010) and high cation (Mesbah & Wiegel, 2008) environments ought to be preadapted for superior performance in an industrial process that involves high temperature conditions and product accumulation in the form of salts. Individual cellulose degrading microorganisms have been isolated from soil communities for many years (Skinner, 1960). More recently communities capable of cellulose and lignocellulose degradation have been examined from soils (Deangelis *et al.*, 2012; Deangelis *et al.*, 2010; Haruta *et al.*, 2002); and from compost (Izquierdo *et al.*, 2010; Sizova *et al.*, 2011). Fermentation capable communities are known to be stable after prolonged fermentation (Werner *et al.*, 2011) and still have been shown to be capable of cellulose degradation even after being subjected to heat, cold, and sub-culturing (Haruta *et al.*, 2002).

To address the hypothesis that microbial communities from saline and thermal environments could improve fermentation performance, we collected sediment samples from a wide geographic range of saline and thermal areas and passed them through the 30-day batch fermentation performance screen. We quantified the ability of the sample community to degrade cellulosic substrate and produce acids and then sought to identify those soil environments that favored the most successful communities.

I analyzed soil chemistry and fermentation performance data including the acid

profile and conversion (quantified biomass degradation) outcome using multivariate statistical approaches. I included data for the soil analysis, site details, acid concentrations, and other fermentation performance characteristics in two multivariate statistical analyses to determine if microbial communities from saline and thermal environments produced an improvement of performance in the fermentation and to resolve any relationship between environmental variables and the improvements in performance. I tested the hypothesis that screening diverse salt and thermal community inocula under stressful process conditions will identify candidate communities with superior process performances, and that there will be commonalities between the *in situ* environmental factors among superior performing community samples. Further, by sampling microbial communities from a variety of extreme environments in nature we expected to reveal unique microbial communities, some of which might exhibit superior performances in the process. I predict identification of associations among particular soil characteristics and particular process performance parameters will enhance future efforts to identify and optimize microbial communities used in the carboxylate platform.

II.2 Methods

Study Design and Site Selection

This study was a large-scale effort to examine variation among soil microbial communities collected from nature as inocula in carboxylate platform fermentations. We conducted frequent collection trips from October 2008 to May 2010. In most cases, at a given geographic location (site) we collected multiple samples, we chose to sample

based on variation in physical and ecological features or presumed gradients (e.g. moisture, salt accumulation, temperature). In total we collected 501 samples (Appendix A) from 77 sites (Table 1). We identified sites via literature, database (Boyd, 2002), and internet searches, and by personal communications with site stakeholders.

The study had two stages (hereafter Stage I and Stage II) with distinct site selection criteria and fermentation experimental conditions (detailed below and Table 2). In Stage I of this study, we sought sites within the southern central region of the United States with a history of salt accumulation or commercial salt production and/or sites known to be high in total dissolved solids (TDS) (Figure 1A). Stage I involved evaluation of 102 inoculum samples from 4 collection trips to 17 sites conducted in 2008 (Table 1). In Stage II of this study, we expanded our site selection criteria to include: greater ecological and geographic diversity, and specific addition of sites with thermal features (Table 1). Stage II involved evaluation of 399 inoculum samples from 59 sites and 14 collection trips across the continental United States, Puerto Rico, and Hawaii conducted in 2009-2010 (Table 1; Figure 1A).

Table 1 List of all sites sampled in Chapter II.

Site ID	Site Name - Controlling Agency	State	N	Ecosystem/Biome	Trip Date
GR	Grulla - Grulla National Wildlife Refuge *	NM	12	Lake bed of salt lake, grassland	10/4/2008
MPL	Paul's Lake - Muleshoe National Wildlife Refuge *	TX	9	Three sink-type lakes with no outlets	10/4/2008
MWL	White Lake - Muleshoe National Wildlife Refuge *	TX	4	Three sink-type lakes with no outlets	10/4/2008
MGL	Goose Lake - Muleshoe National Wildlife Refuge *	TX	5	Three sink-type lakes with no outlets	10/4/2008
GSP	Great Salt Plains - Salt Plains National Wildlife Refuge *	OK	11	Saline flats, wetlands	10/9/2008
Bra	Brazoria - Brazoria National Wildlife Refuge *	TX	9	Coastal wetland, prairie, salt and freshwater marshes	10/24/2008
BL	Bitter Lake - Bitter Lake National Wildlife Refuge *	NM	20	Wetland	11/15/2008
LL	Lazy Lagoon - Bottomless Lakes State Park *	NM	4	Fresh water sink hole lakes	11/15/2008
Lea	Lea Lake - Bottomless Lakes State Park *	NM	1	Fresh water sink hole lakes	11/15/2008
BLM	William's Sink - Bureau of Land Management	NM	4	Playa lake, potash slurry	11/16/2008
BLM	Laguna Tuston - Bureau of Land Management	NM	4	Playa lake	11/16/2008
BLM	Laguna Plata - Bureau of Land Management	NM	8	Playa lake	11/16/2008
BLM	Laguna Tonto - Bureau of Land Management	NM	4	Playa lake	11/16/2008
BLM	Laguna Gatuna - Bureau of Land Management	NM	3	Oil brine evaporation pond	11/16/2008
BLM	Laguna Quattro - Bureau of Land Management	NM	2	Potash slurry, oil brine evaporation pond	11/16/2008
BLM	Laguna Walden - Bureau of Land Management	NM	1	Oil brine evaporation pond	11/16/2008
BLM	Laguna Uno - Bureau of Land Management	NM	1	Potash slurry	11/16/2008
SFB	San Francisco Bay - National Wildlife Refuge *	CA	34	Tidal marsh, salt ponds, mud flats, and seasonal wetlands	2/10/2009
OHS	Ogden Hot Springs - Private	UT	4	Natural hot spring	4/28/2009
WHS	Wilson Hot Springs - Dugway Proving Ground US Army	UT	14	Natural hot spring	4/28/2009
FS	Fish Springs - National Wildlife Refuge *	UT	20	Brackish and warm spring	4/28/2009
Topaz	West Topaz - Bureau of Land Management	UT	1	Warm slough spring	4/28/2009
AHS	Abraham Hot Springs - Bureau of Land Management	UT	5	Natural hot spring	4/28/2009
BHS	Baker Hot Springs - Bureau of Land Management	UT	11	Natural hot spring	4/28/2009
AI	Antelope Island - State Park	UT	4	Island in saline lake	5/5/2009
GSL	Great Salt Lake - State Park	UT	3	Saline lake	5/5/2009
SHS	Saratoga Hot Springs - Bureau of Land Management	UT	2	Hot spring	5/5/2009
HIS	Indian Hot Springs - Bureau of Land Management	UT	4	Natural hot spring	5/5/2009
SCW	Salt Creek Waterfowl Preserve - Bureau of Land Management	UT	1	Brackish marsh	5/5/2009
Knoll	Knoll Spring - Bureau of Land Management	UT	1	Warm spring	5/5/2009
LB	Lincoln Beach - City Park	UT	1	Fresh water lake, slightly saline	5/5/2009
UL	Utah Lake - City Park	UT	2	Fresh water lake, slightly saline	5/5/2009

Table 1 Continued.

Site ID	Site Name - Controlling Agency	State	N	Ecosystem/Biome	Trip Date
WS	Warm Springs - City Park	UT	4	Warm spring	5/5/2009
BRR	Bear River Reserve - National Wildlife Refuge *	UT	3	Freshwater marsh	5/5/2009
SWR	Savannah - National Wildlife Refuge *	GA	3	Tidal freshwater marsh	5/19/2009
CR	Cape Romain - National Wildlife Refuge *	SC	25	Salt marsh, fresh and brackish water	5/19/2009
PI	Pinkney Island - National Wildlife Refuge *	SC	14	Salt marsh, freshwater ponds	5/19/2009
SI	Sapelo Island Microbial Observatory - Georgia DNR	GA	21	Salt marsh	5/21/2009
BWR	Laguna Boqueron - National Wildlife Refuge *	PR	7	Brackish ponds	6/2/2009
CAR	Laguna Cartagena - National Wildlife Refuge *	PR	4	Fresh water	6/2/2009
CRR	Cabo Rojo - National Wildlife Refuge *	PR	10	Salt flats, salt terns	6/2/2009
JBR	Jabos Bay Research Reserve - National Wildlife Refuge *	PR	14	Costal mangrove forest	6/2/2009
CIP	Caladesy Island - State Park *	FL	7	Costal sand beach	6/22/2009
HIP	Honeymoon Island - State Park *	FL	4	Costal sand beach	6/22/2009
CHP	Charlotte Harbor - State Park *	FL	6	Costal sand beach	6/22/2009
RBR	Rookery Bay Reserve - National Estuarine Research Reserve System	FL	9	Costal and tidal marshes	6/24/2009
CSP	Collier-Seminole - State Park *	FL	1	Costal and tidal marshes	6/24/2009
TTI	The Thousands Islands - National Wildlife Refuge *	FL	1	Costal and tidal marshes	6/24/2009
JSB	Jemez Spring Baths - Bureau of Land Management	NM	5	Geothermal spring	7/20/2009
NSS	New Mexico Sulfur Springs - Private	NM	7	Geothermal spring	7/20/2009
SLS	Soda Lake, Side - Valles Caldera National Preserve *	NM	1	Geothermal spring	7/20/2009
SAC	San Antonio Cabin - Valles Caldera National Preserve *	NM	1	Hot spring	7/20/2009
CLS	Caribbean Lake Spring - Valles Caldera National Preserve *	NM	1	Warm spring	7/20/2009
NGYS	Norris Geyser - Yellowstone National Park *	WY	19	Geothermal springs	7/28/2009
SMYS	Sentinel Meadows - Yellowstone National Park *	WY	6	Geothermal springs	7/28/2009
HVYS	Hidden Valley - Yellowstone National Park *	WY	12	Geothermal springs	7/28/2009
WFYS	Whisky Flats - Yellowstone National Park *	WY	2	Wetland	7/28/2009
FHYS	Firehole Drive - Yellowstone National Park *	WY	6	Geothermal springs	7/28/2009
STYS	Sulfatara Trail - Yellowstone National Park *	WY	3	Geothermal springs	7/28/2009
SWRN	Stillwater - National Wildlife Refuge *	NV	3	Fresh and brackish water marshes	8/6/2009
GBS	Great Boiling Springs - Private	NV	9	Geothermal spring and pools	8/6/2009
FRN	Fly Ranch - Private	NV	5	Geothermal springs	8/6/2009
CBHS	Buckeye Hot Spring - Inyo National Forest	CA	4	Hot spring	8/6/2009
MLNB	Mono Lake Navy Beach - State Natural Reserve	CA	5	Saline and alkaline lake	8/6/2009
MLIS	Mono Lake Island Hot Springs - BLM, NF, CADWP	CA	7	Hot spring	8/6/2009

Table 1 Continued.

Site ID	Site Name - Controlling Agency	State	N	Ecosystem/Biome	Trip Date
HCMA	Hot Creek at Mammoth - Bureau of Land Management	CA	5	Geothermal spring	8/6/2009
OLCA	Owens Lake - Private	CA	3	Dry salt lake	8/6/2009
HBSP	Hapuna Beach - State Park *	HI	3	Costal sand beach	5/4/2010
APHW	Anchialine Ponds - Private	HI	4	Brackish anchialine pond	5/4/2010
NELH	Natural Energy Lab Hawaii - Private	HI	4	Costal anchialine ponds	5/4/2010
KKHW	Kekaha Kai - State Park *	HI	5	Costal anchialine ponds	5/4/2010
ONHW	Onekahakaha Beach Park - Hawaii County DPR	HI	4	Costal mangrove forest	5/4/2010
WRHW	Wailoa River - State Park *	HI	1	Anchialine ponds	5/4/2010
AFHW	Akaka Falls - State Park *	HI	3	Terrestrial	5/4/2010
CPHW	Carlsmith Beach Park - Hawaii County DPR	HI	4	Costal anchialine ponds	5/4/2010
GAL	Galveston 8 mile Beach Access - Texas General Land Office	TX	1	Costal sand beach	6/1/2008

The following abbreviations appear in the table above: Number of samples (N), Bureau of Land Management (BLM), California Department of Water and Power (CADWP), Department of Natural Resources (DNR), Department of Parks and Recreation (DPR), National Forest (NF). * indicates permit associated with sample(s) see acknowledgements.

Table 2 Non-identical conditions for Stage I and Stage II fermentation experiments conducted with environmental samples collected for this study.

	Stage I	Stage II
Year of experiment	2008	2009-10
Polypropylene bottle volume	1000 mL	250 mL
Mixing method	Constant rolling 2 rpm, horizontal	Shaken upright @100rpm
Temperature	55°C	55°C
Deionized water	400 mL	150 mL
Shredded office paper	36 g	9 g
Yeast extract	4 g	1 g
Calcium acetate	6.4 g	3.2 g
Butyric acid	1.2 g	0.5 g
Calcium propionate	0.4	0.2 g
Sediment	10 g	2.5 g
Calcium carbonate	10-15g	2.2 g
Initial carboxylate salt concentration	20 g L^{-1}	26 g L^{-1}
Volatile loading rate	90 g L^{-1}	60 g L^{-1}

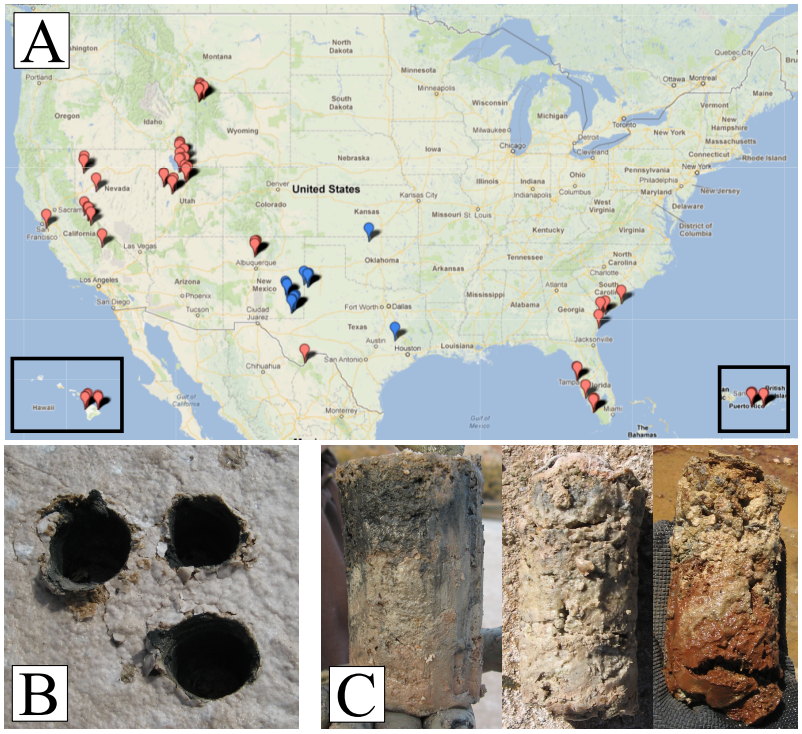


Figure 1 Geographic distribution of sites and sampling method. A) Map exhibits the 77 sites sampled as the basis of these studies (Map data ©2013 Google, INEGI). Blue pins indicate Stage I sites and red pins indicate Stage II sites (Table 2; see Methods). The two insets within the map show Hawaii and Puerto Rico, both sampled in Stage II. B) Three adjacent holes remaining after we collected the 3 cores for the sample at a site. C) Independent soil cores we collected at three different sites in this study, provided to show some of the obvious ecological and physical distinctness across sites.

Sediment Sampling

In most cases, for a single sample we collected three adjacent cores using a standard garden bulb-planting tool to a depth of 10 - 12 cm and with a width of 6.5 cm (Figure 1B and 1C). We took sediment temperature data (Splashproof Thermometer, VWR, PA, USA) at a depth of five centimeters. We sealed each of the three cores in a separate zip-top plastic bag with air removed. As soon after collection as possible, we flash froze one core from each sample with dry ice and subsequently stored this sample

at -80°C for use in future studies. We vacuum-sealed (Sunbeam Products, Inc., FoodSaver Model V2220, DE, USA) the remaining two cores and stored these samples in insulated coolers allowing them to reach ambient temperature during transport to the lab. We used these ambient temperature cores as sources of inocula for fermentation and as the material for sediment characterization.

Sediment Characterization

Upon return to the laboratory and immediately prior to inoculation of the fermentation vessel we homogenized one sample core by hand. We used ~250 cm^3 (or around two tablespoons) of homogenized sediment mixture for measuring volatile solids and moisture content. Following the method of Fu and Holtzapple we measured moisture content of the sediment samples as the difference between the wet and dry weights after drying to constant weight in a 105°C oven (Fu & Holtzapple, 2010b). We measured volatile solids content of the sediment samples as the difference of the constant dry weight and the weight of the sample after ashing at 550°C (Fu & Holtzapple, 2010b). Additionally, we removed sufficient sediment from this homogenized core for soil analysis testing, with the second core being homogenized and used only if additional dry material was necessary. For soil analysis, we dried samples to a constant weight at 40°C in a laboratory oven (Lab-Line Instruments, Inc. Melrose Park, IL, USA). We ground samples and passed them through a 2 mm sieve (USA standard test sieve ASTM E-11 specification) before completely mixing and packaging of each sample for individual tests, and submission to the Soil, Water, and Forage Testing Laboratory at Texas A&M

University (Hollister *et al.*, 2010a). Via these soil analyses we collected data for electrical conductivity of soluble salts (Rhoades, 1982) soil water pH (Schofield & Taylor, 1955) detailed salinity measures of potassium, calcium, magnesium, and sodium (Rhoades & Clark, 1978), plant available phosphorus and sulfur (Mehlich, 1978), analysis of organic carbon, total carbon, total nitrogen (Mcgeehan & Naylor, 1988; Schulte & Hopkins, 1996), and texture, so long as the available soil volume permitted (Day, 1965). We calculated the TDS by multiplying the conductance values (dS m^{-1}) or (mmhos/cm) by the conversion factor 640, resulting in TDS values (mg/L) (Mccauley & Jones, 2005). We stored all sediment remaining after this procedure under vacuum-seal at 4°C for use in further studies.

Fermentation Experiments

Table 2 details the fermentation screens we employed for Stage I and Stage II of this study. Both approaches evaluated inoculum performance in the presence of high product concentrations, as the fermentation broth contained ≥ 2% carboxylate salt concentration. Furthermore, in an attempt to favor shorter residence times we conducted these screens at the high process temperature of 55°C. We conducted the screens in polypropylene centrifuge bottles with screw top caps (Nalgene). Immediately after homogenization of the sample, we used the indicated amount of sediment to inoculate each fermentation. Additionally, we added 100 μL iodoform (20 g L^{-1} in ethanol solution) every other day to inhibit methanogenesis.

Fermentation Characterization

For all fermentations, we harvested 30 days post inoculation and determined the volatile solids remaining and acid products (Golub *et al.*, 2012; Hollister *et al.*, 2010b). Conversion was defined as the grams of volatile solids digested over the grams of volatile solids fed and has been a common indicator of fermentation performance without regard to acid production. In our experiments we used white office paper as the volatile solids or biomass. Acetic acid equivalent concentration (AEQ) is defined as the grams of acetic acid equivalents per liter of fermentation liquid. This variable puts all acid production on the same level so that you can easily compare each sample on the amount of energy contained rather than a suite of acid variables. Selectivity is defined as the grams acetic acid equivalents over the grams volatile solids digested. Yield is defined as the grams total change in acids (end - start) over grams volatile solids fed. We did not study patterns of selectivity or yield data in this study. Both are indicators of the fermentation performance across multiple variables and are useful for quickly assessing how a particular fermentation has degraded biomass into different products (those wanted, and those not).

Data Transformation and Statistical Analyses

I included only those samples with complete data in the statistical analyses (for a complete list of samples see Appendix A). I analyzed the data for Stage I and II fermentations separately.

To better approximate normal distributions and to achieve even scaling, I transformed the sample data; log(1+value) or log(3+value) for discrete quantity data or arcsine [square root (proportion)] for percentage data. I did not transform pH due to the logarithmic character of the variable. Within each analysis, I established two blocks of data: the sediment data block (S-data), and the fermentation data block (F-data). The S-data blocks included the percent volatile and percent moisture content, detailed salinity pH and levels of cation presence, the Mehlich III measured phosphorus and sulfur levels (showing plant available quantities), and the total nitrogen, total carbon, and the organic carbon. F-data blocks included measured individual acid changes (acetic, butyric, iso-propionic, propionic, isovaleric, valeric, caproic, and heptanoic) and the conversion for each sample.

To test for correspondence of fermentation products and soil environmental data I performed multivariate analysis of variance (MANOVA) with JMP 10 Pro (version 10.0 by SAS). The MANOVA provided an overall significance test for the correspondence, and indicated which soil variables were significantly related to fermentation data. I estimated effect strength using eta-squared (Tabachnick & Fidell, 2000). Due to the differences between Stage I and Stage II experiments (Table 2), I analyzed the data separately. However, I compared the similarity of results from the two stages by assessing the correlation among the F statistics.

To explore the relationships among the soil and fermentation variables I performed a two block partial least squares (PLS) regression using Microsoft Office Excel 2007 and the Add-In Poptools version 3.2.5 (Hood, 2010). For this analysis I

standardized all our transformed variables, so that interpretation of eigenvalues would be comparable across different data types. The F-data comprised one block and the S-data formed the second block. I performed PLS on the matrix multiplication product of the transformed, Z-centered blocks of F-data and S-data (Wold *et al.*, 2001). Each row of the F-data contained the results of a single fermentation. I calculated the effect strength as the sum of cross-variance (variance in common between the two data blocks) as a fraction of total variance in fermentation products.

II.3 Results

Variation Among Sediments

As expected, the geographically and ecologically diverse sediments used as inocula in this study (Figure 1A) exhibited extremely wide ranges for characteristics measured (e.g. temperature, texture, conductivity, pH, TDS, etc.) (Appendix A). Across the 501 samples in this study, I analyzed 494 for electrical conductivity used as an indicator of salinity and pH resulting in: 53.5% strongly saline (>16 dS m^{-1}), 11.3% moderately saline (8-15.9 dS m^{-1}), 9.0% slightly saline (4-7.9 dS m^{-1}), 12.1% very slightly saline (2-3.9 dS m^{-1}), 13.9% non-saline (> 2 dS m^{-1}) (Staff, 1993). Based on measurements of pH taken during analysis for detailed salinity: 8.3% of samples were ultra acid (pH< 3.5), 2.4% extremely acid (pH 3.5-4.4), 2.8% very strongly acid (pH 4.5-5.0), 1.4% strongly acid (pH 5.1-5.5), 4.0% moderately acid (pH 5.6-6.0), 7.1% slightly acid (pH 6.1-6.5), 24.8% neutral (pH 6.6-7.3), 28.3% slightly alkaline (pH 7.4-7.8), 16.4% (pH 7.9-8.4), 4.6% strong alkaline (pH 8.5-9.0), 1.8% very strongly alkaline (pH

>9.0) (Staff, 1993). The 459 samples with temperature data available exhibited correspondence to a range of spring categories: 34% to cold springs (<20°C), 54% warm springs (20-50°C), 11.5% hot springs (>50°C). While is helpful to have a standardized nomenclature for reference sake, it is important to recognize that not all samples were from springs. Finally, for the 501 samples with organic carbon data 14 (2.7%) contained greater than 12% organic carbon, the level considered as high organic carbon (Staff, 1993).

Table 3 provides the ranges for the soil characteristics I measured as they distributed across Stage I and Stage II of this study. During Stage I we collected 102 samples isolated from grasslands, fresh and salt water marshes, salt lakes, playas, spring fed lakes and oil brine or potash slurry evaporation ponds sites in TX, NM, and OK (Table 3). Again, we selected these initial sites due to the presumed variation in salinity and total dissolved solids. In fact, these 102 samples span the entire range of conductance (0.01-215.00 dS m^{-1}) and correspondingly the TDS values (5.12-137600.00 mg/L) exhibited across all samples studied. The Stage I sampling trips did not include any sites with thermal features, as such, soil temperatures for these samples spanned a non-thermal range of 6.67 to 30.10°C.

We obtained the 399 samples collected during Stage II from sites representing a much broader geographic and ecological range (Appendix A, Table 1, Figure 1A), selected based on presumed variation in ecology, salinity, total dissolved solids, and/or presence of thermal features. The ecological diversity for these sites included fresh, salt, and alkaline lakes, warm and hot springs, marshes, salt flats, coastal mangroves, and

anchialine ponds (Table 1). The broad ranges for most of the sediment data demonstrated this diversity (Table 3). The broad range of soil temperatures was particularly noteworthy (7.61-92.7°C) but not surprising, in as much as, we collected soils from thermal features.

Fermentation Performance

Appendix A provides all the sediment and fermentation data generated for all samples collected. The inocula were highly variable for fermentation performance for both Stage I and Stage II of this study (Figure 2). In general, if microbial communities exhibited higher conversion of the substrate, then those samples produced more acid (Figure 2). Presumably, in those fermentations with high conversion but low acid production some other unmeasured product was favored (e.g. CO_2 or lactic acid) (M. Holtzapple, personal communication). The wide range in fermentation performances across the diversity of soil samples lends credence to the idea that soil environments harbor different communities and thus it is reasonable to expect we could identify soil parameters associated with community fermentation performance.

There was strong multivariate correspondence between fermentation products and environmental factors of the original soil samples. However, this result did not present as clear univariate responses (e.g. soil factor 1 increases concentration of acid species 1). Instead, we found the strong effects associated with suites of correlated variables in one data block being related to suites of variables in the other block.

Table 3 Distribution of sediment and fermentation variables across the Stage I and Stage II fermentation experiments in this study.

	Stage I		Stage II	
	Min	Max	Min	Max
Sediment data				
Moisture in sediment (%)	1.13	62.89	0.00	100.00
Volatiles in sediment (%)	0.37	47.90	0.00	70.62
Sand (%)	11.00	97.00	5.00	100.00
Silt (%)	0.00	48.00	0.00	79.00
Clay (%)	1.00	86.00	0.00	67.00
pH	5.58	9.50	1.74	10.17
Conductance (dS m^{-1})	0.01	215.00	0.23	202.00
Total Dissolved Solids (mg/L)	5.12	137600.00	0.00	129280.00
Sodium (Na+ mg kg^{-1})	0.32	149756.00	23.69	186457.00
Potassium (K+ mg kg^{-1})	6.24	29952.20	1.04	9973.73
Calcium (Ca++ mg kg^{-1})	11.28	1695.81	1.03	10414.28
Magnesium (Mg++ mg kg^{-1})	5.69	10353.80	0.86	15003.20
Phosphorus (P mg kg^{-1})	1.00	124.00	0.00	384.30
Sulfur (S mg kg^{-1})	15.84	24436.80	18.70	32916.00
Organic carbon (%)	0.07	12.70	0.02	46.43
Total nitrogen (%)	0.06	0.99	0.03	20.78
Total carbon (%)	0.07	13.03	0.07	45.71
Temperature °C	6.67	30.10	7.61	92.70
Fermentation data				
Acetic acid (change in g L^{-1})	-2.71	7.27	-1.06	6.16
Propionic acid (change in g L^{-1})	-0.21	0.86	-0.73	1.91
Isobutyric acid (change in g L^{-1})	0.00	0.35	-0.03	0.14
Butyric acid (change in g L^{-1})	-0.44	3.93	-0.30	1.43
Isovaleric acid (change in g L^{-1})	0.00	0.66	0.00	0.42
Valeric acid (change in g L^{-1})	0.00	0.00	0.00	0.05
Caproic acid (change in g L^{-1})	0.00	0.16	-0.05	0.14
Heptanoic acid (change in g L^{-1})	0.00	0.43	0.00	0.15
Total acids (change in g L^{-1})	-1.27	9.14	0.00	8.08
Conversion (% VS digested/fed in g)	0.15	0.49	0.02	0.57
Selectivity (% g Acid/ g VS dig)	-0.06	0.35	0.00	0.46
Yield (% g Acid/ g VS fed)	-0.01	0.09	0.00	0.13
AEQ (mol/L)	-0.03	0.19	0.00	0.16
AEQ (g L^{-1})	-1.55	11.17	0.00	9.32

Distribution of the sediment and fermentation datasets as separated by stage of experiment. Volatile Solids (Berezovsky & Shakhnovich).

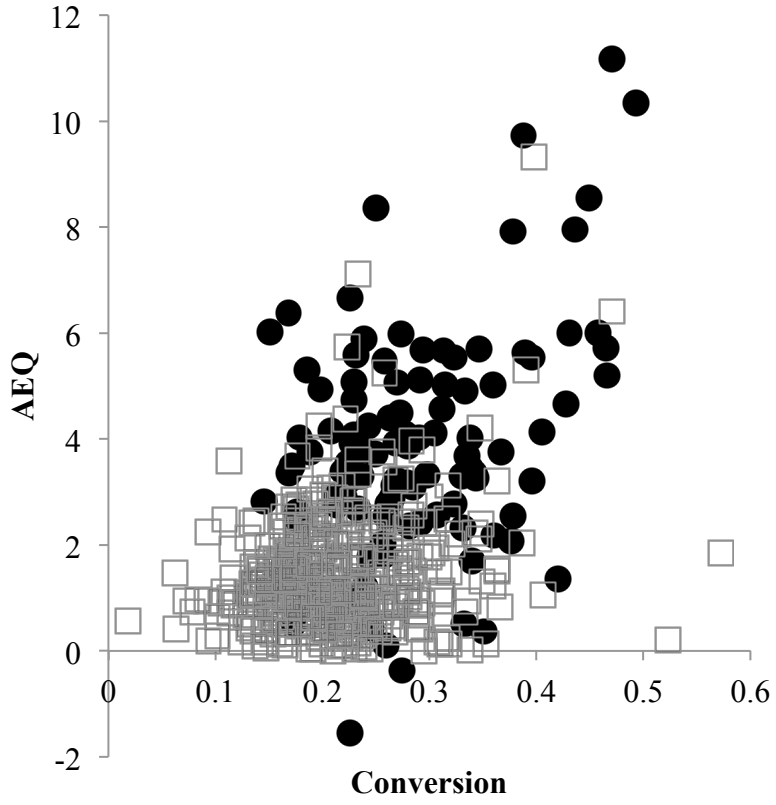


Figure 2 Distribution of acetic acid equivalent concentration (AEQ) by conversion performance for each sample. Since fermentations with different inocula yield different acid profiles, we use AEQ (g L^{-1}) to standardize reporting the acid production across fermentations. Conversion has no units as it is calculated as g volatile solids digested/g volatile solids fed. Black circles represent Stage I, and grey hollow boxes represent Stage II.

We found substantial effect sizes from the MANOVA (Table 4) for both Stage I and Stage II experiments ($\eta^2 = 0.921$ and 0.566, respectively), and the significance structures (F statistics for the array of soil environmental variables) were similar across experiments (r=0.71). Though the effect size was particularly strong for Stage I, note the statistical tests were not powerful, as there were only 6 times the number of samples (N=95) relative to the number of predictor variables. While the only significant effect noted for Stage I was due to temperature (Table 4), some of the cations (K, Ca, S) might also be important determinants or correlates of acid composition, as suggested by the marginally significant *p*-values. It is worth noting that our other analyses confirmed these suggested effects. Stage II results, due to the greater sample size (N=356), revealed more significant effects, despite the fact that the overall effect strength was less than for Stage I results. For Stage II I found significant effects of soil temperature, pH, total N, Mg, Ca, moisture, and P on fermentation variables, with a suggestion (marginal significance) that organic carbon also may have played a causal or correlative role (Table 4). Though these significance tests indicate which independent variables are related to fermentation, the MANOVA framework treats each dependent variable separately, so it is difficult to interpret associations among variables.

Table 4 Multivariate analysis of variance (MANOVA) results for sediment variables across Stage I and Stage II of this study.

	Stage I			Stage II		
	F-stat	Degrees of freedom (Num, Den)[1]	*p*-value	*F*-stat	Degrees of freedom (Num, Den)[1]	*p*-value
Whole Model	2.0226	(112,523.54)	<0.0001	2.3420	(126,2558.1)	<0.0001
Temperature °C	3.2971	(8,73)	0.0029	5.4991	(9,333)	<0.0001
pH	0.9687	(8,73)	0.4672	5.2717	(9,333)	<0.0001
Total nitrogen (%)	0.6035	(8,73)	0.7720	4.8577	(9,333)	<0.0001
Magnesium (Mg++ mg kg^{-1})	0.9346	(8,73)	0.4936	2.9892	(9,333)	0.0019
Calcium (Ca++ mg kg^{-1})	1.8960	(8,73)	0.0735	2.2190	(9,333)	0.0206
Organic carbon (%)	0.7560	(8,73)	0.6422	2.0485	(9,333)	0.0337
Phosphorus (P mg kg^{-1})	1.4973	(8,73)	0.1731	1.9655	(9,333)	0.0426
Moisture in sediment (%)	1.0986	(8,73)	0.3742	1.5265	(9,333)	0.1372
Volatiles in sediment (%)	1.4177	(8,73)	0.2037	1.5265	(9,333)	0.1372
Sodium (Na+ mg kg^{-1})	0.7332	(8,73)	0.6619	1.4944	(9,333)	0.1485
Sulfur (S mg kg^{-1})	1.8564	(8,73)	0.0802	0.8267	(9,333)	0.5919
Conductance (dS m^{-1})	1.2548	(8,73)	0.2806	0.7658	(9,333)	0.6482
Total carbon (%)	0.8502	(8,73)	0.5621	0.6825	(9,333)	0.7248
Potassium (K+ mg kg^{-1})	1.9809	(8,73)	0.0608	0.5897	(9,333)	0.8055
Intercept	4.6127	(8,73)	0.0001	14.9332	(9,333)	<0.0001
Number of samples (N)		N=95			N=356	
Effect Strength (η^2)		η^2=0.921			η^2=0.566	

[1]Numerator (Num), Denominator (Den).

PLS constructs distill axes for both the dependent and independent variables that can be related as multivariate pairs of axes, termed singular axes. The loadings of variables on these axes indicate the relative contribution of variables, as well as, the direction of relationships, and thus, provide a way to interpret two blocks of multivariate data as suites of associated variables. Again, to ensure the loadings could be compared within and between data blocks, I performed this analysis on standardized data. Figure 4 shows the scatter of data relative to the loadings of the first and second Eigen vectors, which represents the majority of the predictable variance. Figure 3 illustrates the loadings of all the variables on that Eigen vector. The Stage I samples exhibited large correlation coefficients for both the first ($R^2 = 0.2682$; Figure 4A) and second ($R^2 = 0.23158$; Figure 4B) axis pairs, which is consistent with the finding of such a strong effect size for Stage I in the MANOVA. The correlation coefficients for the Stage II first ($R^2 = 0.11282$; Figure 4C) and second ($R^2 = 0.06528$; Figure 4D) singular axis pairs were much lower, as was the effect strength in the MANOVA. In fact, the R^2 value for the second singular axis pair for Stage II was so low I found no justification for interpretation of the loadings for this pair. The loadings of the Eigen vectors for Stage I vector 1 and vector 2 were 1.4 and 1.0 respectively. The loadings for Eigen vectors for Stage II vector 1 and vector 2 were 0.8 and 0.5 respectively.

For Stage I, the predominant correspondence across fermentation and soil data was that lower temperatures and higher cations, especially Mg, S, and Na, associated with greater acid production, especially for acetic, isobutyric, isovaleric acids (Figure 3). The second major axis of cross-variance indicated that conversion efficiency particularly

that resulting in propionic and acetic acids, was driven by, or correlated with, variables typically associated with soil fertility (relatively low pH and high N, organic, total and volatile C, and moisture; Figure 3). For Stage II, the soil variables associated with the first Eigen vector (Figure 3) echo those identified for Stage I, namely temperature, pH and nutrients.

The high effect sizes answer our central question about whether soil conditions can be used to predict fermentation products by the soil's microbial community. Yes, they can. However the relationships at hand are complex -- it does not emerge as an isolated few environmental parameters driving an isolated few aspects of organic chemistry. Taken together, these results suggest that in the fermentation conditions we used, and probably to more general conditions (e.g. sampling from other geographic regions) we can maximize desired fermentation products by sampling microbes from soils relatively low (i.e. within the ranges included in our samples) in pH and temperature, but high in cations and soil fertility.

II.4 Discussion

This study shows that CBP productivity is influenced by both fermentation conditions (Stage I and Stage II) and inocula, thus it seems reasonable to expect both can be optimized to target desired outcomes (e.g. particular products and/or productivity levels). Also, we learned that fermentation performance is not as simple as finding one soil parameter that leads to increases in all performance parameters. Rather, there are complex multivariate relationships that are likely indicative of trade-offs associated

within the microbial communities. For example, we found the conductance of soils consistently positively associated with the production of some acids, namely isobutyric and isovaleric but negatively associated with conversion efficiency (Figure 3). Meanwhile, the relationship between temperature and conversion efficiency was consistently positive (Figure 3). The relationship between soil pH and conversion in Stage I was predominantly negative, while pH and conversion were positively associated in the Stage II experiments (Figure 3). The implication of these results is that one might need to choose which performance parameter is of the greatest interest and then optimize that parameter by collecting inocula from those soils with individual parameters that are positively correlated. Alternatively, efforts to identify soils that possess multiple concomitant parameters that tend to influence different aspects of performance (e.g. high conductance and high temperature soil environments might host communities that exhibit both optimal acid production and conversion efficiency) might be justified. Finally, in absence of soils with both conductance and temperature at appropriate levels, it remains to be seen whether one could optimize multiple performance parameters (acid production and conversion simultaneously) by synergistic combination of microbial communities from very different environments (e.g. combining communities from high conductance, low temperature and low conductance, high temperature environments). An important future goal will be to identify the microbial community composition associated with fermentations exhibiting particular performance outcomes.

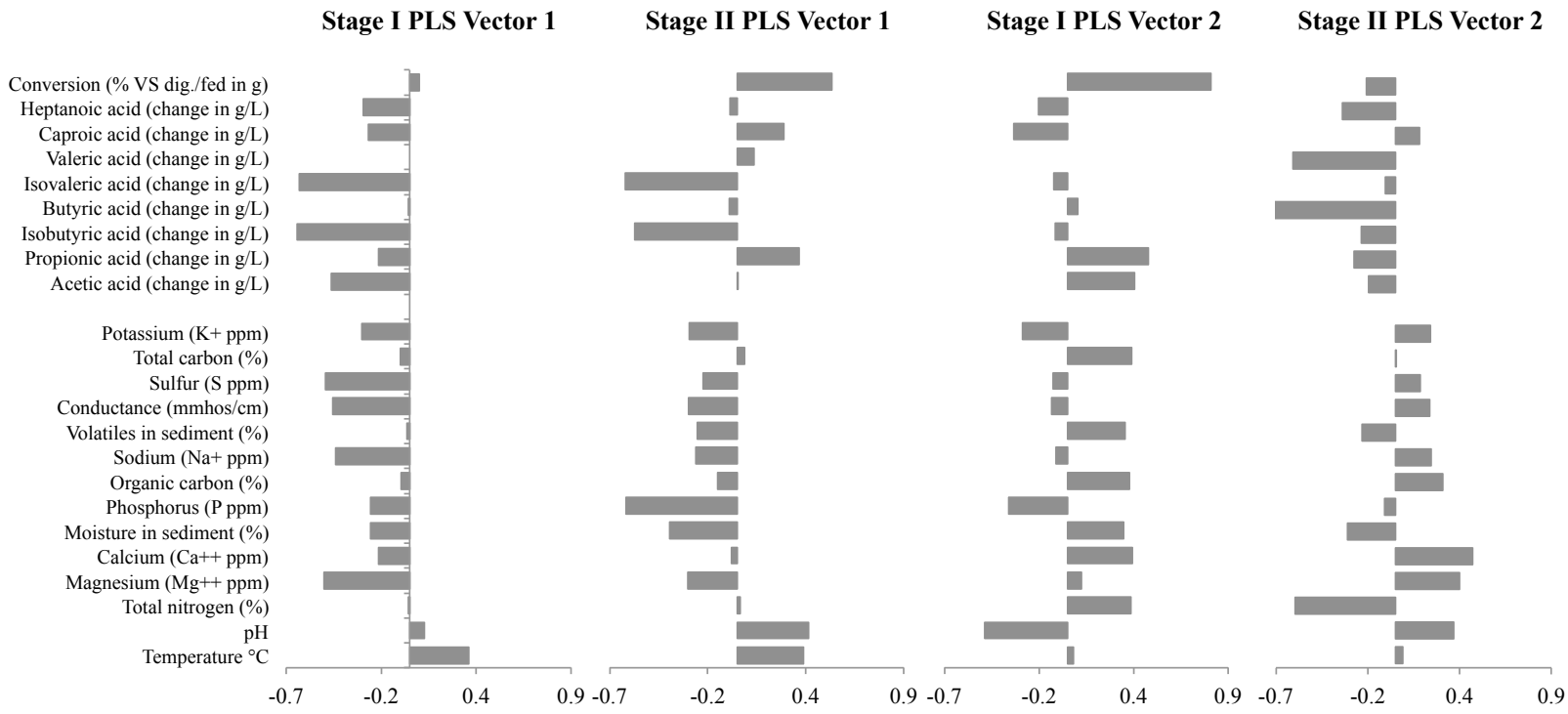


Figure 3 Bar graphs illustrate the loading of the cross-variance onto each Eigen vector across the variables in the 2 block partial least squares (PLS) regressions for Stage I and Stage II data separately.

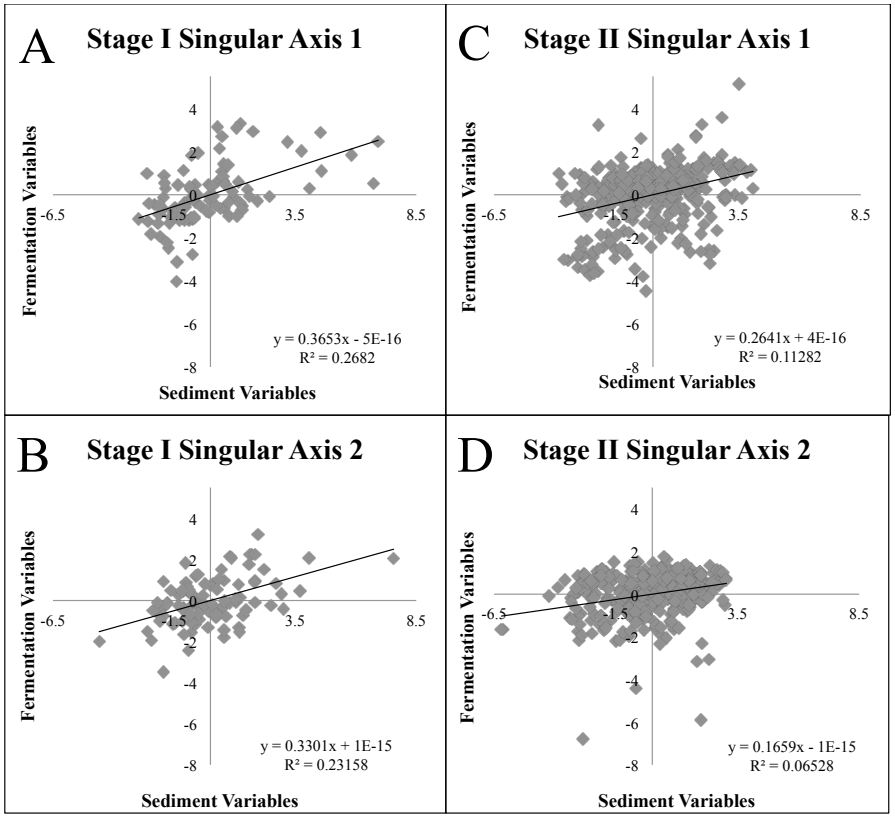


Figure 4 Singular axis pairs for the two-block partial least squares (PLS) regressions conducted separately for the Stage I and Stage II data. A) The Stage I singular axis pair 1 illustrates the majority of the cross-variance between the fermentation performance soil characteristics for Stage I. B) Stage I singular axis pair 2 explains less cross-variance for Stage I. C) Stage II singular axis pair 1 illustrates the majority of the cross-variance between the fermentation performance soil characteristics for Stage II. D) Stage II singular axis pair 2 explains less cross-variance for Stage II.

CHAPTER III

ESTABLISHING THE RELATIONSHIP BETWEEN BACTERIAL COMMUNITY COMPOSITION AND SUPERIOR PERFORMANCE ACROSS PARTICULAR CARBOXYLATE BIOFUEL PLATFORM FERMENTATION PARAMETERS

III.1 Introduction

Previous analyses of carboxylate biofuels platform (CBP) fermentations with a standard inoculum community established clear causal relationships between process parameters and the fermentation product spectrum output, resulting in the predictions that process conditions are the major determinant process performance (Forrest, 2010; Fu, 2007). Subsequently, we established variation among microbial communities for various performance parameters (Chapter II). Furthermore, we recognized that the variation among communities for these performance parameters occurred in both Stage I and Stage II fermentations despite the different process parameters (Chapter II). Thus, we sequenced bacterial communities derived from different natural inocula and associated with a variety of process performances to investigate the role of community composition in the process performance phenotype.

While we measured a variety of performance parameters, a major goal of this project was to identify microbial communities with superior conversion capacities under high salt and temperature conditions. Conversion is defined as the amount of volatile solids digested over the amount of volatile solids fed. Rapid and robust conversion under these stressful conditions can reduce residence time, an important consideration for

implementation of the process at an industrial scale. Interesting acid production parameters for the CBP process include acetic acid equivalence (AEQ), which is a standardized measure for all acid produced in a given fermentation, and the quantity and distribution pattern of individual high molecular weight (HMW) or high carbon number acid outputs (Appendix A, Table 5). High AEQ is desirable since it indicates highly productive communities. Communities producing high levels of HMW products are most desirable for processes associated with production of more complex hydrocarbons (e.g. diesel, jet fuels).

Based on the screen of 501 sources of inocula (Chapter II), ranks of the top performing communities for different performance parameters resulted in overlap, however as one might expect they were not identical. We sequenced 40 bacterial communities that were among the top communities for three performance parameters of greatest importance to industry (conversion, AEQ, and HMW acids) to discern whether particular bacterial taxa were associated with particular performance parameters. I hypothesized that variation in community composition correlates with variation in conversion, AEQ, and select high-molecular weight acids (C5 - valeric, IC5 - isovaleric, C6 - caproic, C7 - heptanoic).

Table 5 Fermentation data for all samples selected for sequence analysis and comparison. We selected these samples from the original 501 microbial communities screened based on ranking among the best for one or more of three performance parameters: High molecular weight acids (IC5-C7), conversion, and acetic acid equivalence (AEQ), each colored on a scale from red (low) to green (high) for their respective values. The horizontal line indicates the change of Stage I to Stage II.

						Acid change in g/L												
Sample identifier	Sediment sample name	Site name	State	Catergory	Stage	C2 Acetic	C3 Propionic	IC4 Isobutyric	C4 Butyric	IC5 Isovaleric	C5 Valeric	C6 Caproic	C7 Heptanoic	Total All Acids	Conversion (g VS dig/g VS fed)	Selectivity (g Acid/ g VS dig)	Yield (g Acid/ g VS fed)	AEQ (g/L)
D04	GR4	Grulla	NM	AH	I	1.95	0.21	0.16	1.95	0.36	0.00	0.00	0.00	4.62	0.32	0.14	0.04	5.54
D05	GR5	Grulla	NM	AH	I	2.39	0.15	0.19	2.37	0.41	0.00	0.00	0.00	5.51	0.23	0.25	0.06	6.66
D18	MPL6R	Muleshoe Paul's Lake	NM	CH	I	0.66	0.32	0.13	0.43	0.27	0.00	0.00	0.00	1.82	0.36	0.05	0.02	2.18
E08	GSP8	Great Salt Plains	OK	3	I	1.95	0.21	0.16	1.95	0.36	0.00	0.00	0.00	4.62	0.40	0.10	0.04	5.54
F01	Bra11	Brazoria NWR	TX	3	I	1.62	0.49	0.22	0.63	0.48	0.00	0.00	0.00	3.44	0.41	0.06	0.02	4.13
F02	Bra12	Brazoria NWR	TX	3	I	2.57	0.16	0.11	1.19	0.30	0.00	0.00	0.00	4.33	0.47	0.07	0.03	5.19
F05	Bra15	Brazoria NWR	TX	CH	I	0.49	-0.04	0.09	0.33	0.27	0.00	0.00	0.00	1.15	0.42	0.02	0.01	1.36
F06	Bra16	Brazoria NWR	TX	3	I	3.15	0.20	0.20	1.01	0.47	0.00	0.00	0.00	5.03	0.43	0.09	0.04	6.01
F09	Bra19	Brazoria NWR	TX	3	I	2.72	0.20	0.12	1.50	0.28	0.00	0.00	0.00	4.83	0.46	0.06	0.03	5.72
G03	BL3	Bitter Lake NWR	NM	AH	I	2.74	0.17	0.30	1.14	0.58	0.00	0.00	0.00	4.92	0.24	0.22	0.05	5.88
G08	BL8	Bitter Lake NWR	NM	3	I	5.02	0.07	0.16	3.53	0.36	0.00	0.00	0.00	9.14	0.47	0.19	0.09	11.17
G09	BL9	Bitter Lake NWR	NM	3	I	7.27	0.26	0.18	0.47	0.32	0.00	0.00	0.00	8.50	0.39	0.22	0.08	9.74
G13	BL13	Bitter Lake NWR	NM	CA	I	2.46	0.12	0.07	2.26	0.07	0.00	0.00	0.00	4.97	0.46	0.12	0.06	5.99
G14	BL14	Bitter Lake NWR	NM	AH	I	2.41	0.28	0.15	1.58	0.36	0.00	0.00	0.00	4.77	0.29	0.18	0.05	5.69
G16	BL16	Bitter Lake NWR	NM	AH	I	4.48	0.36	0.30	1.25	0.60	0.00	0.00	0.00	6.99	0.25	0.30	0.08	8.37
G19	BL19	Bitter Lake NWR	NM	CH	I	1.59	-0.04	0.09	1.99	0.22	0.00	0.00	0.00	3.85	0.43	0.10	0.04	4.66
G22	LL1	Bottomless Lake State Park; Lazy Lagoon	NM	AH	I	3.21	0.86	0.09	2.30	0.19	0.00	0.00	0.00	6.65	0.38	0.19	0.07	7.91
G23	LL2	Bottomless Lake State Park; Lazy Lagoon	NM	AH	I	4.44	-0.04	0.00	3.93	0.13	0.00	0.00	0.00	8.46	0.49	0.18	0.09	10.34
G24	LL3	Bottomless Lake State Park; Lazy Lagoon	NM	3	I	3.24	0.00	0.11	2.92	0.33	0.00	0.00	0.00	6.61	0.44	0.17	0.07	7.96
G34	BLM8	Bureau of Land Management; Laguna Tustin	NM	3	I	2.39	-0.03	0.29	1.45	0.60	0.00	0.00	0.00	4.70	0.35	0.13	0.05	5.70
G38	BLM12	Bureau of Land Management; Laguna Plata	NM	H	I	1.85	0.07	0.25	0.96	0.58	0.00	0.16	0.43	4.31	0.19	0.26	0.05	5.29
G41	BLM15	Bureau of Land Management; Laguna Plata	NM	AH	I	2.77	0.23	0.29	1.40	0.63	0.00	0.00	0.00	5.33	0.17	0.35	0.06	6.37
G45	BLM19	Bureau of Land Management; Laguna Tonto	NM	AH	I	2.16	0.02	0.24	2.02	0.56	0.00	0.00	0.00	5.00	0.15	0.35	0.05	6.02
G46	BLM20	Bureau of Land Management; Laguna Tonto	NM	3	I	4.92	0.05	0.00	1.57	0.11	0.00	0.16	0.33	7.13	0.45	0.17	0.08	8.54
G47	BLM21	Bureau of Land Management; Laguna Gatuna	NM	C	I	1.42	-0.02	0.00	1.22	0.07	0.00	0.00	0.00	2.70	0.40	0.08	0.03	3.20
G48	BLM22	Bureau of Land Management; Laguna Gatuna	NM	AH	I	2.98	0.19	0.19	1.22	0.43	0.00	0.00	0.00	5.00	0.27	0.20	0.05	5.99
H01	SFB1	San Francisco Bay NWR	CA	CA	II	0.07	0.03	0.00	0.48	0.13	0.00	0.00	0.00	0.71	0.37	0.03	0.01	0.83
H20	SFB20	San Francisco Bay NWR	CA	3	II	6.16	0.30	0.12	1.04	0.34	0.00	0.00	0.12	8.08	0.40	0.34	0.13	9.32
J04	Big4	Big Bend NP	TX	C	II	-0.09	-0.04	0.00	1.43	0.00	0.00	0.00	0.00	1.30	0.36	0.06	0.02	1.54
J11	Big11	Big Bend NP	TX	C	II	1.61	0.01	0.00	0.16	0.00	0.00	0.00	0.00	1.78	0.39	0.08	0.03	2.04
J18	Big18	Big Bend NP	TX	C	II	1.07	-0.02	0.00	0.10	0.00	0.00	0.00	0.00	1.15	0.35	0.06	0.02	1.31
J19	Big19	Big Bend NP	TX	C	II	0.20	-0.05	0.00	-0.02	0.00	0.00	0.00	0.00	0.12	0.35	0.01	0.00	0.14
J20	Big20	Big Bend NP	TX	C	II	0.00	-0.02	0.00	0.03	0.00	0.00	0.00	0.00	0.01	0.34	0.00	0.00	0.02
K49	BHS5	Baker Hot Springs	UT	C	II	0.03	-0.08	0.00	0.93	0.00	0.00	0.00	0.00	0.88	0.41	0.04	0.02	1.05
M24	CR22	Cape Romain NWR	SC	C	II	0.62	-0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.62	0.33	0.03	0.01	0.72
N21	SI21	Sapelo Island Microbial Observatory	GA	AH	II	2.99	0.28	0.06	1.30	0.18	0.00	0.00	0.08	4.89	0.22	0.36	0.08	5.73
P01	BWR1	Laguna Boquerón NWR	PR	CH	II	2.08	0.15	0.09	0.24	0.23	0.00	0.00	0.00	2.79	0.36	0.12	0.04	3.20
S44	FHYS5	Firehole drive Yellowstone	WY	C	II	-0.28	-0.02	0.00	0.41	0.05	0.00	0.00	0.00	0.17	0.52	0.01	0.00	0.19
S48	STYS3	Solfatara Trail Yellowstone	WY	C	II	1.41	0.03	0.00	0.13	0.03	0.00	0.00	0.00	1.61	0.57	0.04	0.02	1.84
U22	OLCA1	Owens Lake (dry lake)	CA	CA	II	3.84	0.56	0.00	1.16	0.00	0.00	0.00	0.00	5.56	0.47	0.19	0.09	6.41

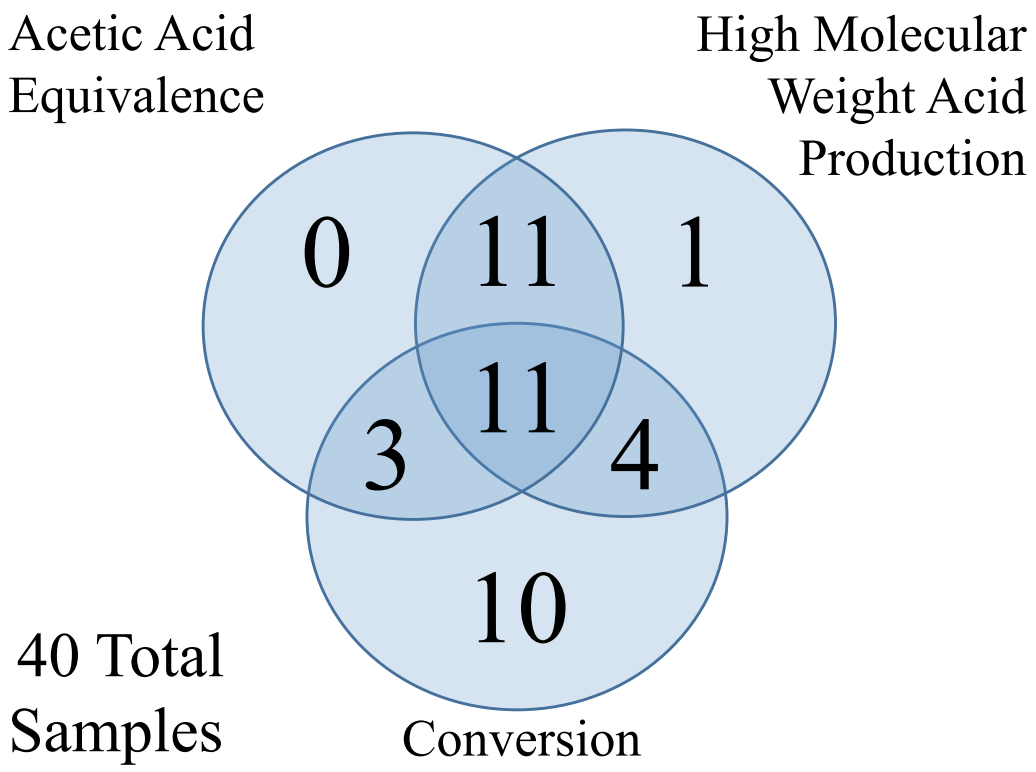


Figure 5 Venn diagram showing distribution of samples across performance parameter categories for samples.

Figure 5 shows the distribution of the 40 sequenced communities across different process performance phenotypes. I conducted community composition analyses, which focused on the repetition or absence of community taxa. The primary focus was on community constituents as measured by presence/absence, diversity indices, and classifications at the order level. The analysis took the form of individual regressions of individual OTUs and multivariate statistical testing of the dominant taxa observed through analysis of the classification. I chose to examine the difference between the communities with 1 to 4 samples, despite the lack of statistical power. I sought to identify community members with known ecological functions that could explain the buildup or breakdown of high molecular weight acids, the acetic acid equivalents, and conversions observed. Discerning the associated taxa distribution pattern for a given performance phenotype should help us define more optimal communities, allowing for testable hypotheses useful for inoculum development and process parameter optimization.

III.2 Materials and Methods

Selection of Samples

We sequenced microbial communities from CBP fermentations conducted as part of a project to screen 501 soil communities from thermal and/or saline environments as inocula (Chapter II). Initially we selected communities with superior conversion performance and compared the microbial community in the soil inoculum to the fermentation community (Hammett, 2011). To expand the comparison, in this study I sequenced additional communities from CBP fermentations with performances superior

for acid production, both quantity (AEQ) and quality (proportion of high molecular weight acids HMW). In total, the comparison includes 40 CBP fermentation communities. Figure 5 shows how the communities distribute across performance categories, it is worth noting that any single community may or may not have ranked among the best for one or more of the parameters: conversion, AEQ, and HMW. Of the 40 samples 28 had high conversion, 25 had high acetic acid equivalents concentration (g L^{-1}), and 16 produced high molecular weight acid concentrations (g L^{-1}) (Figure 5). Again, I hypothesized that comparing the communities will allow me to identify specific taxa that influenced performance parameters.

Fermentation Characterization and Product Storage

We characterized all fermentations as described in Chapter II, (Golub *et al.*, 2012), and (Hollister *et al.*, 2010b). We vortexed the fermentation liquid with the solid portion of the sample to completely re-suspended it. We then placed three aliquots of the mixture in 50 mL Falcon tubes with two aliquots being immediately placed at -80°C and the third sample being dispensed into eight 2 mL cryogenic tubes with the remainder being dried and stored. We filled four of these cryogenic tubes with 1 mL CBP fermentation sample only, and we filled four with 1 mL sample and sufficient 100% sterilized glycerol to bring the mixture to 20% total glycerol. We then vortexed the glycerol and CBP fermentation product tubes until they were homogeneous. We stored eight of these samples at -20°C for further inoculations of the subsequent fermentation screens. We froze the remaining material in the 50mL Falcon tube at -20°C overnight

and lyophilized under vacuum at -40°C (VirTis Sentry Benchtop 3L Company, Inc. Gardener, NY, USA with Welch Chemstar® Vacuum Pump Thomas Industries, Skokie, IL, USA) for possible inoculum for future experiments. As necessary, we maintained the fourth aliquot at 4°C until such time that it could be used as inoculum for the isolate library collection as explained in Chapter V.

Genomic DNA Extraction of Fermentation Materials

We performed DNA extractions from fermentations as described in (Golub *et al.*, 2011b), which involved a modified version of the DNEasy Blood and Tissue Kit Gram Positive Protocol (Qiagen, Venlo, Netherlands, Cat # 69054). We modified the manufacturer's Gram Positive Protocol by adding 40 mg/ml, for a total of 0.04 g lysozyme, to each extraction reaction at the bacterial pellet stage, followed by a 30 minute incubation in a heat block at 37°C. We followed all other steps in the manufacturer's protocol.

454 Pyrosequencing

After extraction and quality checking of community genomic DNA, we normalized samples to 25 ng/µl in a total of 8 µl of 10 mM TRIS HCl. We shipped frozen samples, on dry ice, to the Research and Testing Laboratories (Lubbock, Texas) for bacterial tag-encoded FLX amplicon pyrosequencing (BTEFAP) (Acosta-Martinez *et al.*, 2008) on a Roche Life Sciences 454 FLX sequencer (Roche Applied Science, Werk Penzberg, Germany) with titanium chemistry. We amplified bacterial sequences with the

following primer set: forward and tagged 27f 5'-GAG TTT GAT CNT GGC TCA G-3', and reverse 519r 5'-GTN TTA CNG CGG CKG CTG-3' (Hollister *et al.*, 2011; Lane, 1991), (Table 6). These primers cover the V1 through V3 region of the 16S rRNA gene (Figure 6).

Post Sequencing Processing and Statistical Analysis

I screened all sequenced fasta files and sequences shorter than 350bp, containing ambiguous sequences, and I excluded sequences containing homopolymers larger than 10 bases long from further analysis. I aligned all sequences to the silva.gold.bacteria database (accessed from <http://www.mothur.org/wiki/> on 05/17/11) and also used this database to screen for chimeras using the chimera.slayer function in mothur v1.19.2. I removed all potential chimeras from further analysis. I then compiled the sequences into the larger question set in Bioedit v7.0.9 (Hall, 1999) and submitted them to the Ribosomal Database Project (RDP) pyrosequencing pipeline aligner function (release 10 accessed from <https://rdp.cme.msu.edu/> on 01/01/13) to align all sequences to the known sequence of the bacterial 16S ribosomal gene subunit.

Table 6 All primers used in this dissertation for sequencing.

Name	Citation	Sequence	Use
27f	Lane, 1991	GAGTTTGATCNTGGCTCAG	454 Bacteria
Grey519r	Acosta, 2008	CAGCMGCCGCNGTAANAC	454 Bacteria
27f	Lane, 1991	AGAGTTTGATCCTGGCTCAG	Sanger
1100r	Lane, 1991	AACGAGCGCAACCCT	Sanger
515f	Lane, 1991	GTGCCAGCMGCCGCGGTAA	Sanger
1492r	Lane, 1991	AAGTCGTAACAAGGTAACCG	Sanger

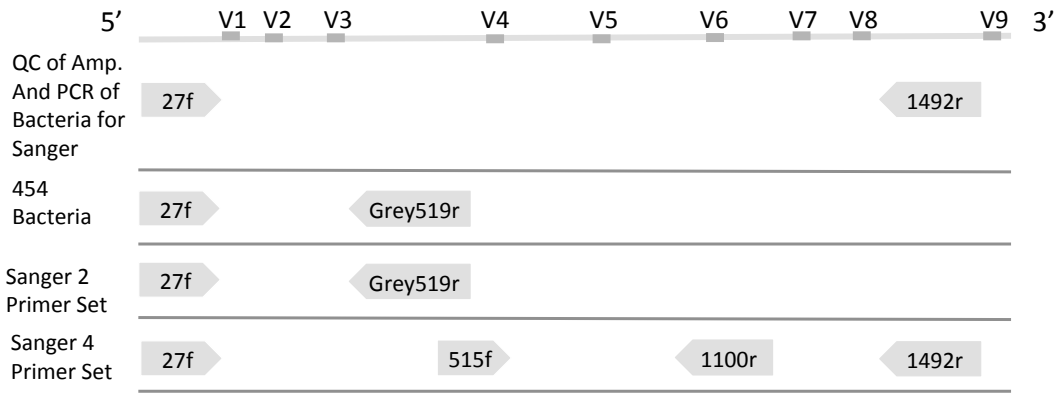


Figure 6 Map of primer placement along the 16S rRNA gene segment for primers used in this dissertation project. Grey arrows indicate primers and the associated direction relative to particular variable regions within the 16S rRNA gene map at the top of the figure. The primers and variable regions are not drawn to scale. Sequences of primers are provided in Table 6.

I utilized mothur (Schloss *et al.*, 2009) for operational taxonomic unit (OTU) classification at 97% similarity or the commonly accepted species level (Stackebrandt & Goebel, 1994). After alignment, I screened the alignments in mothur v1.29.2 for sequences that aligned outside of the 16S ribosomal subunit, were not unique, and for those sequences that started or ended at positions outside of 97% of the samples. Also, I removed positions in the alignment that contained all gaps to speed downstream processing. I removed sequences that didn't align to the 16S and those that started or ended more than 3% away from the majority of aligned sequences. I then used dist.seqs and cluster functions in mothur to cluster the sequence sets at 97% similarity using the average neighbor joining algorithm in cluster. I included all non-unique sequences in further analysis using the names file function of mothur throughout post processing. I used the silva.bacteria database for classification of sequences and OTUs. I made bar graphs based on classifications appearing in >35 of the 40 samples (Appendix C). I determined Chao 1 and Shannon index values with all samples. I also performed regression with JMP v10Pro by SAS with the absolute abundance of OTUs within each sample and the fermentation outcomes of interest conversion, acetic acid equivalence (AEQ) in g L^{-1}, and high molecular weight acid products in g L^{-1}. I examined the top twenty most abundant OTUs and those OTUs present in >20 samples. To examine the relationship between samples I used the normalize.shared function of mothur to give me the same general sample size before calculating the Yue-Clayton theta similarity coefficient and the Bray-Curtis dissimilarity index. We chose the Bray-Curtis dissimilarity index to perform non-metric multidimensional scaling (NMDS) within

mothur to illustrate the maximum differences between the sample OTU classifications. Finally, I performed a partial least squares (PLS) decomposition on the top 14 represented classes with the acid spectrum data and conversion outcomes. I Chi-square normalized the relative abundance of the top 14 classes (Hammett, 2011) before I included them in the PLS.

III.3 Results

Fermentation

Table 5 shows the fermentation results for the 40 selected samples. We never detected valeric acid, a 5 carbon chain acid, in any of the selected samples and thus we discarded this data column before analysis. However, isovaleric acid was present in many samples. Few of the Stage II samples produced HMW acids to any large degree (N21, P01 being the exceptions). Previously, I established there was a Stage I to Stage II change in acid profile (see Chapter II). Yet there still were communities that produce acids in similar quantities (see columns total acids and AEQ in Table 5) even with the change in the volatile solids loading rate (amount of paper) and a change in the initial carboxylate salts challenge condition (salt concentration).

Sequencing and Post Processing

After processing there were 220338 sequences in the library and the number of unique sequences within the library was 50033 with an average length of 300 bases. We deposited all sequences in the National Center for Biotechnology Information (NCBI)

sequence read archive under the project accession number PRJNA208594. There were a total of 1339 OTUs at 3% dissimilarity, 478 of which were found in more than one sample. The number of sequences I obtained for a given community does not predict the number of resulting OTUs (Appendix B). We identified no single OTU shared by all samples. OTU 0006 was the most common, containing 17409 sequences found in all samples except J11 and J19. OTU 0006 was classified as (classification(confidence)) *Bacteria*(100); *Firmicutes*(100); *Bacilli*(100); *Bacillales*(100); *Bacillaceae*(100); *Geobacillus*(100) with all subsequent lower levels unknown. See Table 7 for classifications of the top 30 most abundant OTUs. To further characterize the relative abundance and distribution of OTUs within samples I calculated the Chao 1 richness estimator and the Shannon diversity index for all samples in Appendix B. Appendix C shows the relative abundance of each portion of the order level classified community and the ranking of the sample in each of the three categories of interest (conversion, AEQ, and HMW acid production). As different sample size can affect outcomes (Magurran, 2004), I normalized to examine the similarity of OTUs between samples using the Yue-Clayton theta (θ_{YC}) similarity coefficient. The θ_{YC} coefficient reflects the distribution of OTUs between two samples and the relative abundances of those OTUs.

Table 7 Regression analysis results for the 30 most abundant OTUs. Thirteen of the top 30 OTUs exhibit significant correlations to one or more performance parameters. Signs indicate the direction of the relationship, i.e. + means abundance of the OTU increases the parameter, while - means abundance is correlated with decreases in that performance parameter.

		Conversion		AEQ		HMW Acids		
OTU	Sequences	R^2	*p*-val.	R^2	*p*-val.	R^2	*p*-val.	Taxonomy(Confidence)
0001	42732					+0.137	0.0285	Bacteria(100); Firmicutes(100); Clostridia(100); Clostridiales(100); Family_XI_Incertae_Sedis(100); Tepidimicrobium(100)
0002	30116	+0.117	0.0307	+0.132	0.0215			Bacteria(100); Firmicutes(100); Clostridia(100); Thermoanaerobacterales(100); Family_III_Incertae_Sedis(100); Thermoanaerobacter(100)
0003	18855			+0.151	0.0132			Bacteria(100); Firmicutes(100); Clostridia(100); Clostridiales(100); Ruminococcaceae(100); Incertae_Sedis(99)
0004	18252			-0.211	0.0029	-0.164	0.0096	Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(94); Bacillaceae(94); Bacillus(93)
0005	17756							Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(100); Bacillaceae(100); Ureibacillus(100)
0006	17409					-0.104	0.0275	Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(100); Bacillaceae(100); Geobacillus(100)
0007	11084							Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(100); Thermoactinomycetaceae(97); Thermoactinomyces(97)
0008	8458							Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(100); Bacillaceae(100); Ureibacillus(100)
0009	6121	+0.174	0.0074					Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(100); Bacillaceae(100); Geobacillus(100)
0010	5862	-0.323	0.0001			+0.303	0.0002	Bacteria(100); Thermotogae(100); Thermotogae(100); Thermotogales(100); Thermotogaceae(100); Petrotoga(100)
0011	5347							Bacteria(100); Firmicutes(100); Clostridia(100); Thermoanaerobacterales(100); Family_III_Incertae_Sedis(100); Thermoanaerobacter(100)
0012	5055							Bacteria(100); Proteobacteria(100); Gammaproteobacteria(100); Pseudomonadales(100); Pseudomonadaceae(100); Pseudomonas(100)
0013	4009							Bacteria(100); Firmicutes(99); Clostridia(95); Clostridiales(84); Peptococcaceae(83); Desulfurispora(83)
0014	3315							Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(96); Bacillaceae(74); Bacillus(74)
0015	2044							Bacteria(100); Firmicutes(100); Bacilli(99); Bacillales(99); Paenibacillaceae(94); Thermobacillus(94)
0016	1608			-0.193	0.0059	-0.162	0.0101	Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(84); Bacillaceae(84); Bacillus(78)
0017	1432	-0.125	0.0255			+0.424	0.0001	Bacteria(100); Chloroflexi(94); Chloroflexi(94); Chloroflexales(94); Candidatus_Chlorothrix(94)
0018	1361							Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(100); Bacillaceae(100); Ureibacillus(100)
0019	1313							Bacteria(100); Firmicutes(100); Bacilli(100); Bacillales(100); Planococcaceae(92); Jeotgalibacillus(90)
0020	1182					+0.221	0.0022	Bacteria(100); Firmicutes(100); Clostridia(100); Clostridiales(100); Family_XVIII_Incertae_Sedis(100); Symbiobacterium(100)
0021	1162	-0.235	0.0015			+0.118	0.0303	Bacteria(100);Firmicutes(100);Clostridia(100);Clostridiales(100);Family_XI_Incertae_Sedis(100); Tepidimicrobium(100)
0022	1116							Bacteria(100);Firmicutes(100);Bacilli(100);Bacillales(60);Thermoactinomycetaceae(51);Thermoactinomyces(51)
0023	970							Bacteria(100);Firmicutes(100);Clostridia(100);Thermoanaerobacterales(100);Family_III_Incertae_Sedis(100); Caldanaerobius(100)
0024	912							Bacteria(100);Firmicutes(100);Bacilli(100);Bacillales(100);Thermoactinomycetaceae(75);Thermoactinomyces(75)
0025	724							Bacteria(100);Firmicutes(100);Bacilli(100);Bacillales(100);Thermoactinomycetaceae(82);Thermoactinomyces(82)
0026	691							Bacteria(100);Firmicutes(100);Bacilli(100);Bacillales(69);Bacillaceae(69);Bacillus(57)
0027	628	-0.309	0.0002			+0.138	0.0197	Bacteria(100);Firmicutes(93);Clostridia(79);Clostridiales(79)
0028	591							Bacteria(100);Proteobacteria(100);Betaproteobacteria(100);Burkholderiales(100);Oxalobacteraceae(100); Janthinobacterium(100)
0029	560							Bacteria(100); Firmicutes(100); Clostridia(100); Clostridiales(100); Family_XI_Incertae_Sedis(100); Tepidimicrobium(100)
0030	554			-0.127	0.0261	-0.207	0.0036	Bacteria(100); Firmicutes(100);Bacilli(100);Bacillales(100);Bacillaceae(99);Geobacillus(99)

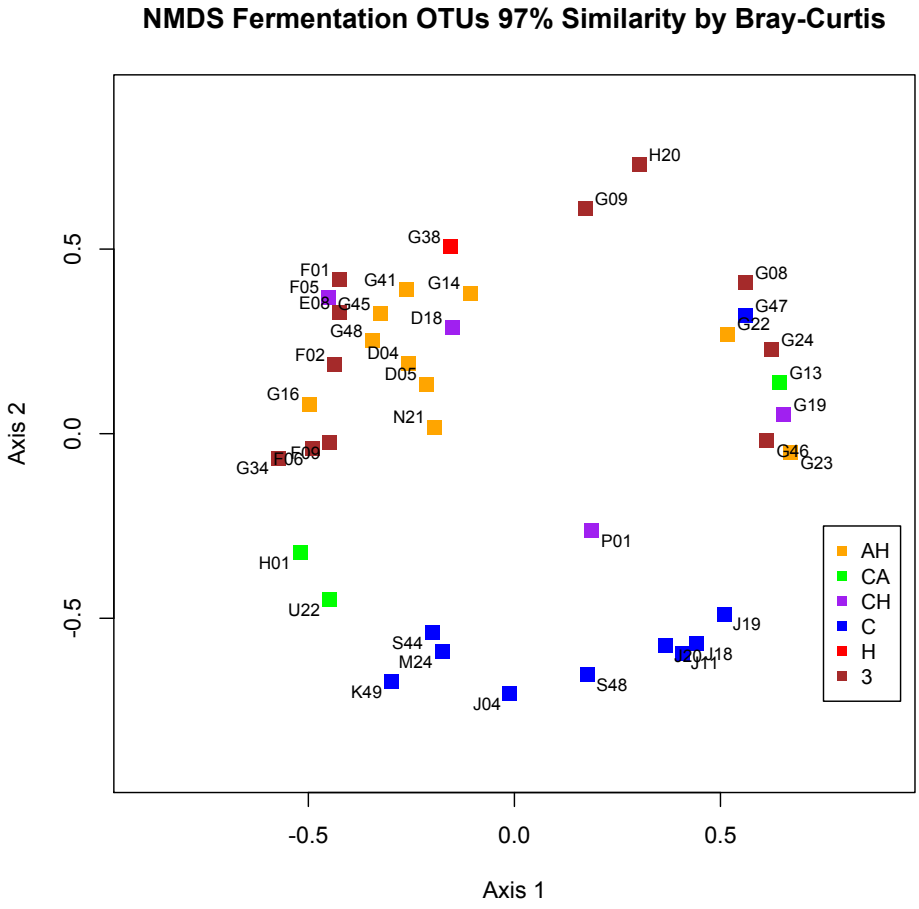


Figure 7 Non-metric multidimensional scaling ordination of Bray-Curtis dissimilarity between operational taxonomic unit (OTU) classifications (at 3% dissimilarity) for 40 selected samples. Sample names and categories of performance parameters both indicated: conversion, C; acetic acid equivalent concentration, A; high molecular weight acid, H; samples high in all three categories are listed as 3.

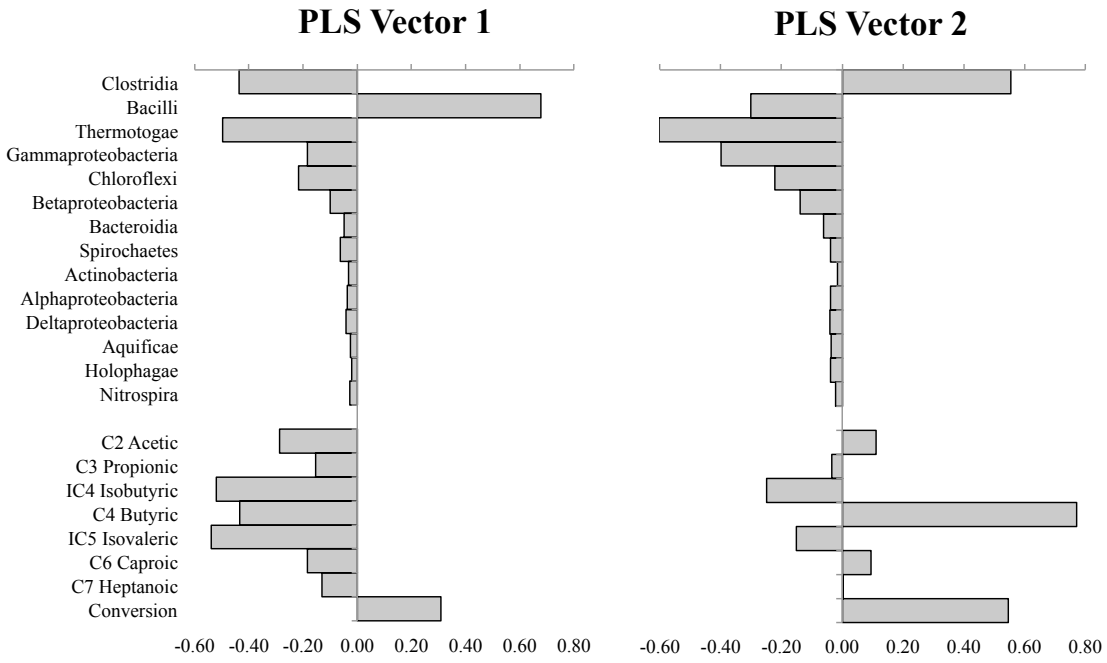


Figure 8 The loadings of the cross-variance onto each axis pair across the variables in the 2-block partial least squares (PLS) regression. The loadings are 0.13 and 0.06 for the first two vectors. A total of 76% of the system wide variance was explained with 52% explained by the first vector and 24% with the second.

Statistical Analyses

The abundance of 13 OTUs in a community showed either a positive or negative regression correlation with one of the categories of interest (Conversion, AEQ, or HMW acids) that was statistically significant at a level of $\alpha=0.05$ (Table 7). In cases where both the conversion and acid production correlated with the OTU sequence number the relationship was opposite in all but one case. In OTU 0002 both conversion and AEQ were positively correlated with sequence numbers reported. OTUs 0010 an 0017 both had high positive R^2 values, 0.30 and 0.42 respectively, with HMW acid production

outcomes. Figure 7 shows the NMDS. The first two Eigen vectors and their loadings are shown in Figure 8.

III.4 Discussion

Process parameters alone do not dictate fermentation product outcome as evidenced by the variety of fermentation outputs observed in Appendix A and Table 5. There was no discernable pattern between sequence numbers and resulting OTU numbers in the sequenced communities (Appendix B). Nor was there a pattern between CBP performance and numbers of OTUs in each sample (data not shown). Community characters, in terms of richness (Chao 1) and diversity (Shannon index), were different across the samples and between geographic locations (Appendix B). The samples from Big Bend (samples with labels beginning with "J") had consistently low numbers of OTUs. Yet in other sample sites (D, G, H, and S; Appendix B) there was no such consistency. The θ_{YC} table (Appendix B) shows that for those communities with very low sequence number, generally the OTUs found were present in other samples. This accounts for the complete similarity numbers between low number samples and those with large numbers of samples. We found no single OTU shared by all samples, OTU 0006 was the most common and was missing from two communities with low numbers of OTUs (J11 and J19 both of which were Big Bend samples). OTU 0006 was classified as genus *Geobacillus* with 100% confidence. *Geobacillus* was also commonly found in the CBP communities with the standard inoculum (Hollister *et al.*, 2012), and is one of the most commonly cultured organisms in Chapter VI of this dissertation.

Regression output can be found in Table 7. In cases where an OTU was correlated with both the conversion and HMW, the relationship was opposite with negative effects on conversion and positive association with HMW. In the one case with both conversion and AEQ associated with an OTU, OTU 0002, both conversion and AEQ were positively correlated with sequence numbers reported. OTU 0002 classifies as the genus *Thermoanaerobacterium* and was one of the most commonly sequenced organisms found in (Hollister *et al.*, 2012), an in depth metagenomic exploration of the communities within the CBP. This genus has been a common target for studies of thermophilic cellulolytic organisms (Taylor *et al.*, 2008). OTUs 0010 and 0017 both had high positive R^2 values, 0.30 and 0.42 respectively, with HMW acid production outcomes. OTU 0010 classifies in phylum *Thermotogae*, known to produce acids (De Vrije *et al.*, 2009) including isovaleric acid (Huber *et al.*, 1990). OTU 0017 was classified in phylum *Chloroflexi*, commonly associated with anaerobic digestion (Riviere *et al.*, 2009), but due to the difficulty of culturing, not well characterized in its role.

The Bray-Curtis dissimilarity index is a measure of the compositional dissimilarity between sites. Figure 7 is a non-metric multidimensional scaling (NMDS) ordination of the 40 fermentation communities based on the Bray-Curtis dissimilarity index. An important organizing factor the ordination of the similarity among the communities is fermentation screen type (Stage I or Stage II). Notice that the bottom third of the ordination space includes only Stage II fermentations. The upper two thirds includes all the Stage I fermentations and two of the Stage II fermentations. Most of the communities sequenced due to superior performance in conversion alone (blue symbols

in Figure 7) are Stage II fermentations clustered near the bottom, except G22 a Stage I fermentation. Likewise, those communities with superior performances due to conversion and AEQ (green symbols in Figure 7) cluster together near the bottom so long as they are Stage II, while G13, a Stage I fermentation, clusters in the upper region. All fermentations with superior performance due to HMW acid production occur in the upper two thirds of Figure 7, including the two Stage II fermentations that cluster in that region of the NMDS ordination (H20 and N21). Both of these samples made heptanoic acids. These two samples are also the only Stage II samples to make heptanoic acids.

There is an obvious and significant (p-value: 0.0005) negative correlation between conversion and HMW acid production (Figure 9). This trade off can also be seen in the first Eigen vector of the PLS that explains 52% of the total variance. The PLS loadings indicate that the resulting community differences were *Bacilli* in opposition to *Clostridia, Thermatogae*, and *Chloroflexi*. The presence of *Bacilli* increased with the increase in conversion. While *Clostridia*, *Thermotogae*, and *Chloroflexi* increased with all measured acids; but acetic, isobutyric, butyric, and isovaleric acids in particular (Figure 8). The second Eigen vector of the PLS, an additional 24% of the variance, showed a strong relationship between conversion and butyric acid with the relative abundance of *Clostridia*. In opposition to this relationship are isobutyric acid and the relative abundance of *Bacilli*, *Thermatogae*, *Gammaproteobacteria*, and *Chloroflexi*. Several *Bacillus* species are known to utilize organic acids as carbon sources (Schleifer, 2009). While *Clostridia* are numerically abundant in the sequenced samples, they have less weight on the PLS than the *Bacilli*.

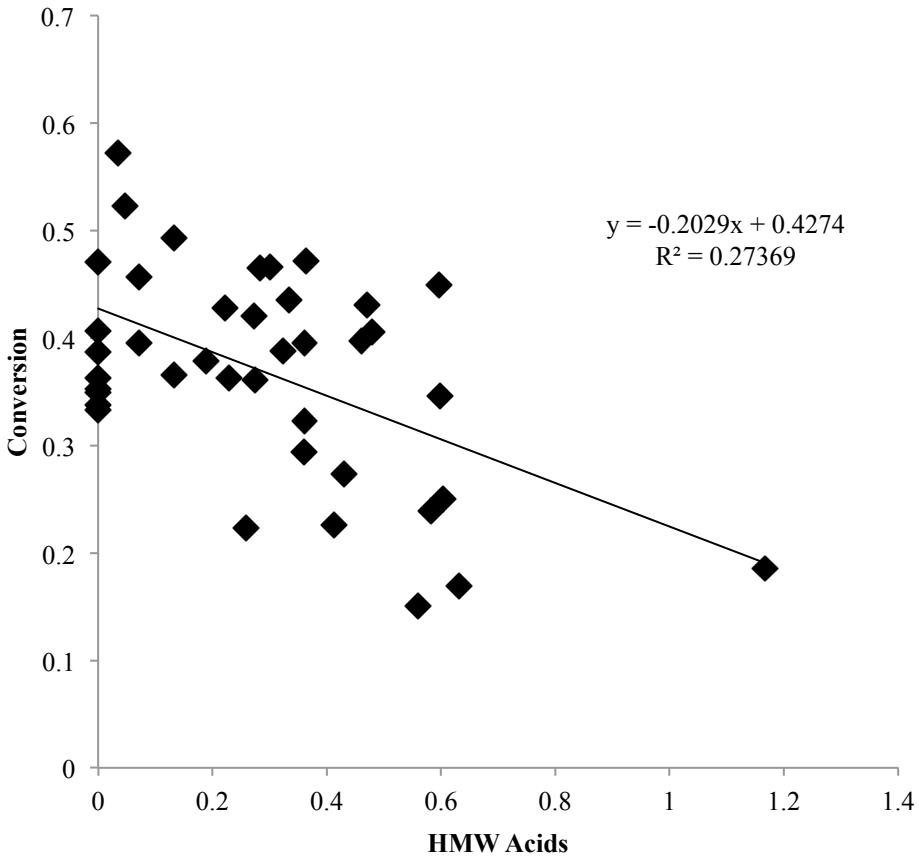


Figure 9 Performance parameter tradeoff: high molecular weight acid concentration (in g L^{-1} IC5-C7) verses conversion. High molecular weight (HMW) acid production is negatively correlated with conversion performance (p-value = 0.0005).

The discovery that the presence of *Clostridia* associates with increases in conversion and acids and that the frequency of *Bacilli* increases conversion is consistent with other work (Hammett, 2011). However, the association of specific taxa (*Thermatoga, Chloroflexi*) with increases in high molecular weight acids and overall diversity of the acid spectrum is a novel finding for 55°C fermentations. In a study of community composition within the CBP fermentations at both 40°C and 55°C the communities had different taxa and different product spectra however the authors proposed that the 40°C process parameter was responsible (Hollister *et al.*, 2011). Thus, here I identified communities capable of HMW acid production at the preferred temperature for the industrial process.

CHAPTER IV

PARALLEL COMPARISON OF FIVE ECOLOGICALLY DIVERSE SOILS AS INOCULA FOR THREE CARBOXYLATE BIOFUEL PLATFORM SCREENING TECHNOLOGIES

IV.1 Introduction

Chemical engineers in the Holtzapple research group at Texas A&M use three fermentation screens to evaluate and model carboxylate biofuel platform (CBP) fermentations under different process conditions: at different scales, times, substrate concentrations and product concentrations. To date, members of the Holzapple research group use these screens have been used sequentially to optimize different features of the process and as a result the inocula for subsequent screens presumably undergoes selection due to the interim storage conditions (e.g. fermentations are frozen while data from the primary screen are analyzed in order to identify the best performing communities, and the frozen material from the target communities are then used to inoculate the next screens). Thus, to date, efficacy of comparisons among the screens for a given community are dubious, in that, starting communities are not equal because it is highly likely they undergo a bottleneck during storage due to repetitious freezing and thawing. To forego this issue we conducted an experiment that compared the different fermentation screens after simultaneous inoculation with some well characterized "high performing" communities (Hammett, 2011). We collected fresh inocula from four sites in the Brazoria National Wildlife Refuge and one site on Galveston Island Public Access

Beach. We then inoculated the three screens simultaneously with sediment from those sites. Post-fermentation we sequenced the communities present in each screen at the point of harvest to compare the communities that establish across the manipulated parameters. I hypothesized that the communities and fermentation performance profiles for the three screens at similar substrate loading ratios and salt concentrations would be most similar. Furthermore, I hypothesized that fermentations with longer time scales, higher salt concentrations, and higher substrate loading ratios would experience stronger selective pressure and thus exhibit major shifts in community composition likely resulting in less diversity. Finally I predicted that plasticity, or the ability to adapt and perform across the substrate levels would be positively related to diversity.

Historically, the chemical engineers developing the CBP have implemented different types of fermentation experiments to model the efficacy of a given community across different process conditions (e.g. substrate or product concentration). In Chapter II of this dissertation I show the primary 30-day batch screen implemented with all samples collected. For communities with the best performances in this primary screen, or for which there were particular targeted questions, additional screening with more complex, time-consuming, resource and labor intensive established screens occurred (see continuum particle dispersal model screen and counter current exchange below). These screen designs mimic salt product or substrate concentrations across stages of product accumulation encountered throughout the process at full scale. Prior to this experiment inoculation of subsequent screens involves frozen material from storage (several days to months). Studying the community compositions and process performances within and

across inoculum sources through the different fermentation screens provides for a more robust understanding of the relationship between communities and the process conditions. These findings can then be related to putative performance in the larger scale production facility. There remained the lingering question, however, of the impact of storage on community viability and fermentation capacity. So, we used fresh inocula for this experiment in each of the three screens and ran them simultaneously.

IV.2 Materials and Methods

Site Selection and Soil Analysis

We chose the five samples in this study because of proximity to Texas A&M; previous characterization (Chapter II), (Golub *et al.*, 2012); extensive sequencing in other experiments in our project (Hammett, 2011); and exhibition of superior performance for several CBP fermentation performance parameters (Forrest *et al.*, 2012). We sampled, stored and characterized the sediments as described in Chapter II of this dissertation (Golub *et al.*, 2012). This experiment represents the sequenced output of the screen evaluations explored in the companion paper (Golub *et al.*, 2013).

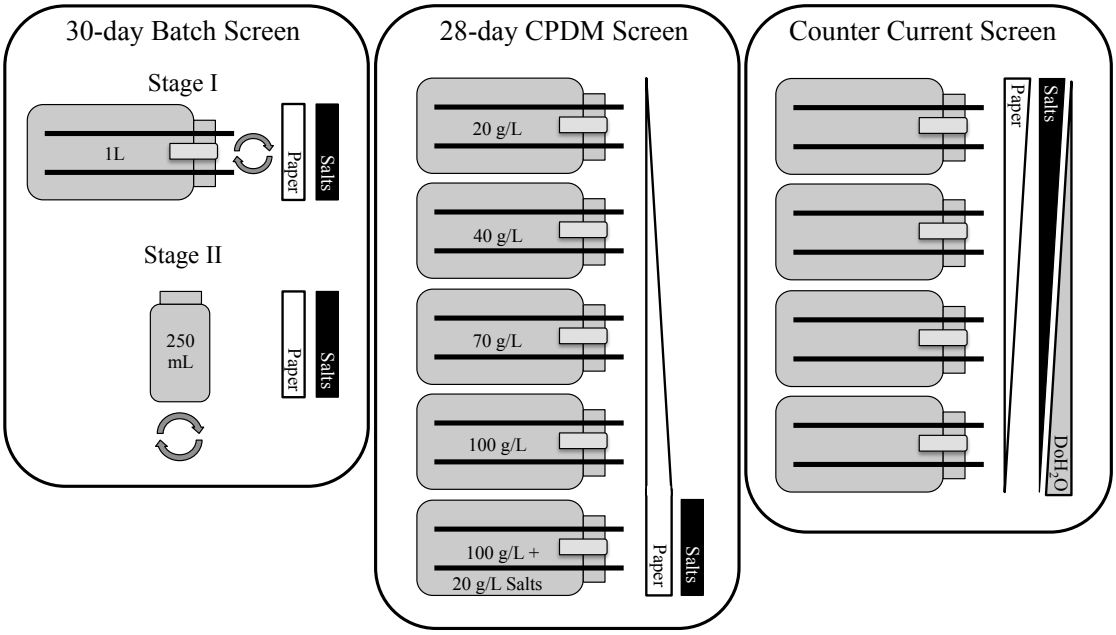


Figure 10 Diagram of screening methods. Historically the screens are run in sequence with the Stage I or Stage II batch screen first, followed by the continuum particle dispersal modeling (CPDM) screen, followed by the counter current screen. Meanwhile the inoculum is derived from stored frozen materials generated in the first screen, thus the resulting differences might well have reflected storage and time associated bottlenecks in microbial inocula. For this experiment all screens (Stage II, CPDM, and Counter Current) were inoculated simultaneously with fresh sediment from five sites, thus allowing direct comparison of results across screens.

Screening Technology

The three screening methods are described in detail in (Golub *et al.*, 2013). We conducted the primary 30-day batch fermentation performance screen as outlined in the description of Stage II fermentations in Chapter II (Figure 10). The continuum particle dispersal model screen (CPDM) design tests samples against a range of substrate and product concentrations (Aiello-Mazzarri *et al.*, 2006). The CPDM analysis in the present study consisted of five fermentations for each inoculum (Figure 10). We ran samples at four substrate levels (20, 40, 70 and 100 g L^{-1} paper/fermentation broth) with no added salts and one sample with the highest substrate level (100 g L^{-1}) with salts added (20 g L^{-1}). We ran the CPDM for twenty-eight days in five 1 L fermentation vessels (Nalgene NNI 3120-1010) with a modified cap that allows for stirring and gas release by way of internal steel tubes and a rubber septum. We took liquid samples (with the same volume of deoxygenated water replaced) every two days to monitor acid levels in this screen. We sequenced the end point community (community at day 30) for each of the CPDM vessels (liquid retention time = 30 days). While the inocula for this experiment was from fresh sediment, we have historically we setup the CPDM fermentation screen from storage products from the 30-day batch screen. Briefly, in the traditional setup we inoculate a 1mL aliquot into a 650 mL fermentation. After thirty days, we homogenize the community from this up-culture, store one portion for sequencing, and inoculate the remainder equally into each of the five conditions. Then these screens are identical in all other ways to the conditions described for the present study. For communities successful in the CPDM screen the counter current exchange screen (CC) is usually the final screen

designed to follow the sample to a steady state of acid production. In this, the third stage, screen researchers load four fermentation vessels with equal ratios of biomass and fermentation broth and over the course of several months biomass and liquid are loaded into and transferred across the vessels to cause opposite gradients of fresh broth and biomass as illustrated in Figure 10. In this study we ran this screen in four 1 L bottles modified for stirring and gas purging with stirring bars as were the CPDM and Stage I 30-day batch fermentations. We took liquid samples every two days with replacement to monitor the acid concentration within this screen. We ran this screen until we detected a steady state of products. Figure 11 is a detailed schematic of the sample names used throughout this chapter. The first half of the sample name indicates the sediment or fermentation from which the sample originated. The sediment sample Bra25 corresponds to fermentation sample W01 and is from a freshwater marsh location. The sediment sample Bra55 corresponds to W02 and is from a saltwater marsh location. Bra65/W03 is a costal prairie soil sample, Bra95/W04 is a salt lake sediment sample, and GAL/W05 is an intertidal sediment sample. Those samples ending in 40C and 55C indicate samples that were run as Stage II screen fermentations at the corresponding temperature (40°C or 55°C) (Figure 10, Table 2). Those samples ending in 20, 40, 70, 100 and 100S are CPDM fermentation screens of the aforementioned concentrations. Samples ending in CC## indicate with those numbers the position in the current, and the time point of collection (Figure 10).

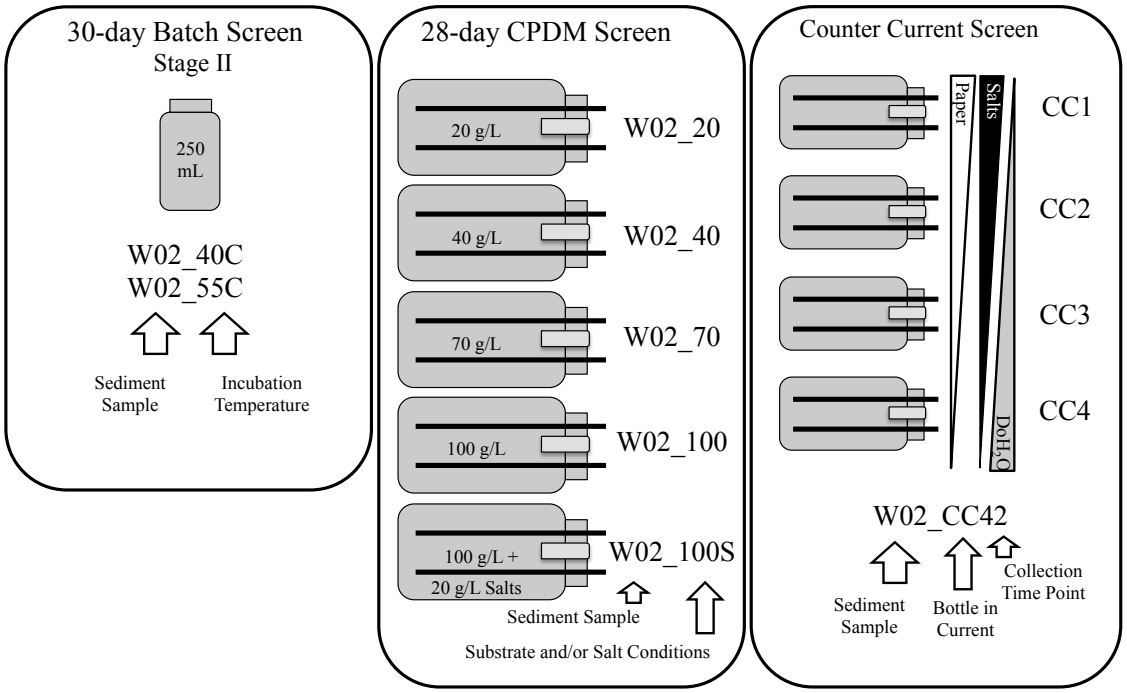


Figure 11 Schematic diagram of samples names for this project.

Fermentation Characterization and Product Storage

We performed fermentation characterizations for the three screens as described in (Golub *et al.*, 2012). We collected and stored products from each fermentation vessel the 30-day batch, CPDM, and CC fermentations separately as described in Chapter III and (Golub *et al.*, 2011a). We stored all samples at -80°C until extraction of genomic DNA for comparison in the community analysis.

Genomic DNA Extraction of Fermentation and Sediment Samples

We extracted genomic DNA from fermentation samples for community analysis as described in (Golub *et al.*, 2011a). We conducted DNA extractions from soil microbial communities as described in (Hollister *et al.*, 2010a). Briefly, we performed the process with PowerMax Soil DNA Isolation Kits (MO BIO Laboratories, Inc., Carlsbad, CA, Cat # 12988-10) modified from manufacturer's protocol by the primary addition of 15 mg of lysozyme in 150 µl + 15 mL of the bead solution per reaction tube. We followed this incubation with a one hour shaking water bath incubation set at 37°C. In absence of a shaking water bath, we vigorously shook the samples every ten minutes during incubation. We followed all other manufacturer's protocols as instructed after this addition.

454 Pyrosequencing

We performed 454 pyrosequencing on soil and fermentation genomic DNA extractions as described in Chapter II of this dissertation. Briefly, after extraction and

quality checking of community genomic DNA, we normalized samples to 25 ng/µl in a total of 8 µl of 10 mM TRIS HCl. We shipped frozen samples, on dry ice, to the Research and Testing Laboratories (Lubbock, Texas) for bacterial tag-encoded FLX amplicon pyrosequencing (BTEFAP) (Acosta-Martinez *et al.*, 2008) on a Roche Life Sciences 454 FLX sequencer (Roche Applied Science, Werk Penzberg, Germany) with titanium chemistry. Bacterial sequences were amplified with the following primer set: forward and tagged 27f 5'-GAG TTT GAT CNT GGC TCA G-3' and reverse 519r 5'-GTN TTA CNG CGG CKG CTG-3' (Hollister *et al.*, 2011; Lane, 1991); (Table 6). These primers cover the V1 through V3 region of the 16S rRNA gene (Figure 6).

Post Sequencing Processing and Data Analysis

All sequencing post processing in this chapter follows the methods outlined in Chapter III of this dissertation. I classified all sequences using the silva.bacteria database (<http://www.mothur.org/wiki/> accessed 5/17/11), and constructed bar graphs at the order level from those classifications. I utilized the mothur Chao 1 richness estimator and the Shannon Index for estimation of diversity. After normalizing the samples sizes as outlined in Chapter III of this dissertation I calculated a dendogram based on the Yue-Clayton theta (Θ_{YC}) similarity coefficient. I also constructed a non-metric multi-dimensional scaling (NMDS) plot of the Bray-Curtis dissimilarity index to display the maximum differences between samples based on OTU binning.

Table 8 Fermentation performance data for 30-day batch screens.

		g L^{-1}			%						g						%		
Sample Name	Fermentation ID	AEQ	Total Acid	Change in Acid	Acetic	Propionic	Butyric	Valeric	Caproic	Heptanoic	C2	C3	C4	C5	C6	C7	Conversion	Selectivity	Yield
Bra25	W01_40C	0.16	19.50	0.14	79.07	3.04	17.89	0.00	0.00	0.00	0.11	0.00	0.03	0.00	0.00	0.00	0.06	0.16	0.01
Bra55	W02_40C	0.20	17.01	0.19	89.42	10.58	0.00	0.00	0.00	0.00	0.17	0.02	0.00	0.00	0.00	0.00	0.10	0.11	0.01
Bra65	W03_40C	1.02	18.30	0.95	89.17	2.51	8.31	0.00	0.00	0.00	0.85	0.02	0.08	0.00	0.00	0.00	0.14	0.05	0.01
Bra95	W04_40C	0.40	20.61	0.34	76.66	3.03	20.30	0.00	0.00	0.00	0.26	0.01	0.07	0.00	0.00	0.00	0.08	0.51	0.04
GAL	W05_40C	2.67	18.94	2.31	75.04	7.11	17.84	0.00	0.00	0.00	1.73	0.16	0.41	0.00	0.00	0.00	0.12	0.06	0.01
Bra25	W01_55C	0.73	19.40	0.64	78.56	0.00	21.44	0.00	0.00	0.00	0.50	0.00	0.14	0.00	0.00	0.00	0.13	0.02	0.00
Bra55	W02_55C	0.68	16.67	0.65	90.93	6.11	2.46	0.50	0.00	0.00	0.59	0.04	0.02	0.00	0.00	0.00	0.04	0.08	0.00
Bra65	W03_55C	0.49	19.68	0.41	70.72	4.39	24.90	0.00	0.00	0.00	0.29	0.02	0.10	0.00	0.00	0.00	0.11	0.15	0.02
Bra95	W04_55C	2.63	18.23	2.43	85.93	4.69	9.37	0.00	0.00	0.00	2.08	0.11	0.23	0.00	0.00	0.00	0.07	0.08	0.01
GAL	W05_55C	0.46	15.73	0.44	92.77	6.51	0.72	0.00	0.00	0.00	0.41	0.03	0.00	0.00	0.00	0.00	0.08	0.46	0.04

Table 9 Soil analysis results for five locations used to simultaneously inoculate three fermentation screens.

							Detailed Salinity						Mehlich III						
Sample Name	State	Fermentation ID	Sand %	Silt %	Clay %	Textural Class Names	pH	Conductance dS m^{-1}	Na mg kg^{-1}	K mg kg^{-1}	Ca mg kg^{-1}	Mg mg kg^{-1}	P mg kg^{-1}	S mg kg^{-1}	Organic C, %	Total N, %	Total C, %	Temp in C	Trip Date
Bra22	TX	W01	32	35	33	Clay Loam	5.2	8.78	2052	81	201	260	33	538	2.96	0.31	2.97	28.7	6/18/10
Bra55	TX	W02	46	39	15	Loam	7.7	48.10	14504	208	473	1843	29	860	0.54	0.08	0.56	34.5	6/18/10
Bra65	TX	W03	42	47	11	Loam	7.8	26.60	7213	126	262	498	4	352	0.44	0.11	0.46	38.4	6/18/10
Bra95	TX	W04	34	29	37	Clay Loam	7.4	32.70	8349	229	351	1034	46	538	2.48	0.24	2.52	32.3	6/18/10
GAL	TX	W05	97	2	1	Sand	7.92	20.60	3999	202	372	410	5	196	0.36	0.06	0.92	27.2	6/18/10

IV.3 Results

Soil Analysis and Fermentation Results

Table 8 shows the 30-day batch screen fermentation results. Table 9 shows the soil analysis. For the primary 30-day batch screen, the inocula Bra25 (W01_55) and Bra65 (W03_55C) were the most optimal screened inocula due to high conversion. The Bra55 inoculated (W02_55C) fermentation made a measurable percentage of longer chain volatile fatty acids, namely valeric acid. The actual amount of acids made in this fermentation was not sufficient to shift the gram weight output of fermentation. Overall, in the 30-Day batch screen the GAL sediment inoculated fermentation W05_55C made the largest amount of acids, though the overall conversion was low for this sample. In the CPDM experiments W01, W04, and W05 fermentations were the preferred inocula for the conversion outputs sought. We condensed all outputs for the CPDM into a single scaled output (W01>W03>W05>W02>W04) and a set of graphs (Figure 12). We combined all substrate concentrations to produce the CPDM maps and did not analyze the data separately (Golub *et al.*, 2013). The counter current exchange screens also resulted in rankings with the order being W02>W01=W04=W05≥W03. The output for these experiments is the ranking as they were not analyzed for fermentation outputs separately (Golub *et al.*, 2013).

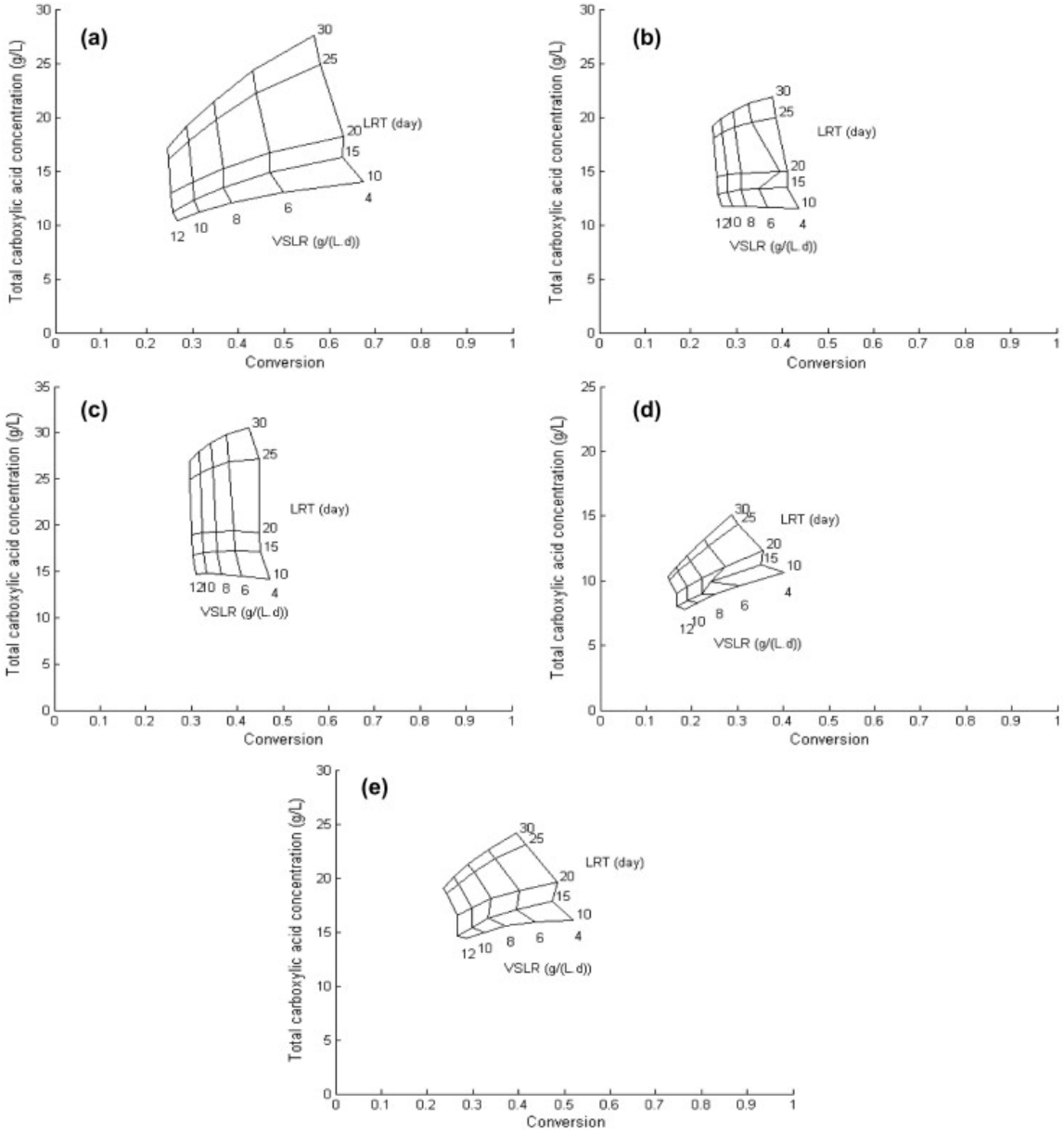


Figure 12 Maps of continuum particle dispersal model (CPDM) screen of five samples. The shift of the plot closer to 1 on the conversion axis and upward on the total carboxylic acid concentration axis implies that the sample will be a better performer in the carboxylate biofuels platform. Reproduced from Golub *et al*. (2013) with permission from Elsiver (originally published in Bioresource Technology an Elsiver publication.)

Sequencing

There were a total of 402076 sequences that were binned into a total of 8928 OTUs after quality checking was complete (Appendix D). All sequences were deposited with the National Center for Biotechnology Information (NCBI) sequence read archive under the project accession number PRJNA196483. There were insufficient numbers of sequences for several samples. These samples are highlighted in grey in Appendix D and were removed from further analysis.

IV.4 Discussion

Appendix D shows the overall diversity by the number of OTUs, and the Shannon index decreased considerably and consistently from the soils to the three screens. There was also a decrease in richness as shown by the Chao 1 richness estimator. I observed a decrease in diversity with the increase in substrate loading of some of the counter current screens though the relationship was not consistent. No such pattern was observed in the CPDM samples. There was a decrease in richness across increasing substrate loading in the counter current exchange. There was also a corresponding decrease in richness as evidenced by the Chao 1 estimator along the subsequent time points (Appendix D). Again, no such pattern appears in the CPDM.

To see the magnitude of diversity change from soil to fermentation compare Figure 13 and Appendix E. Figure 13 shows the large number of order level classifications found in the five inocula. The communities are remarkably similar across screen types from all sediment samples Appendix E. The most striking observation

comes from following concentrations across the counter current exchange experiments where the concentration appears to dictate the community. Further evidence of the similarity can be seen in Figure 14 where several CC samples cluster seemingly on the first number in their exchange identifier. This indicates they were part of the same concentration. The final piece of evidence is observable in Appendix F where the samples are color coded by counter current exchange screen concentration for the counter current samples. Future directions for this dataset should include analysis of these observations with fermentation data, similar to that described in Chapter III.

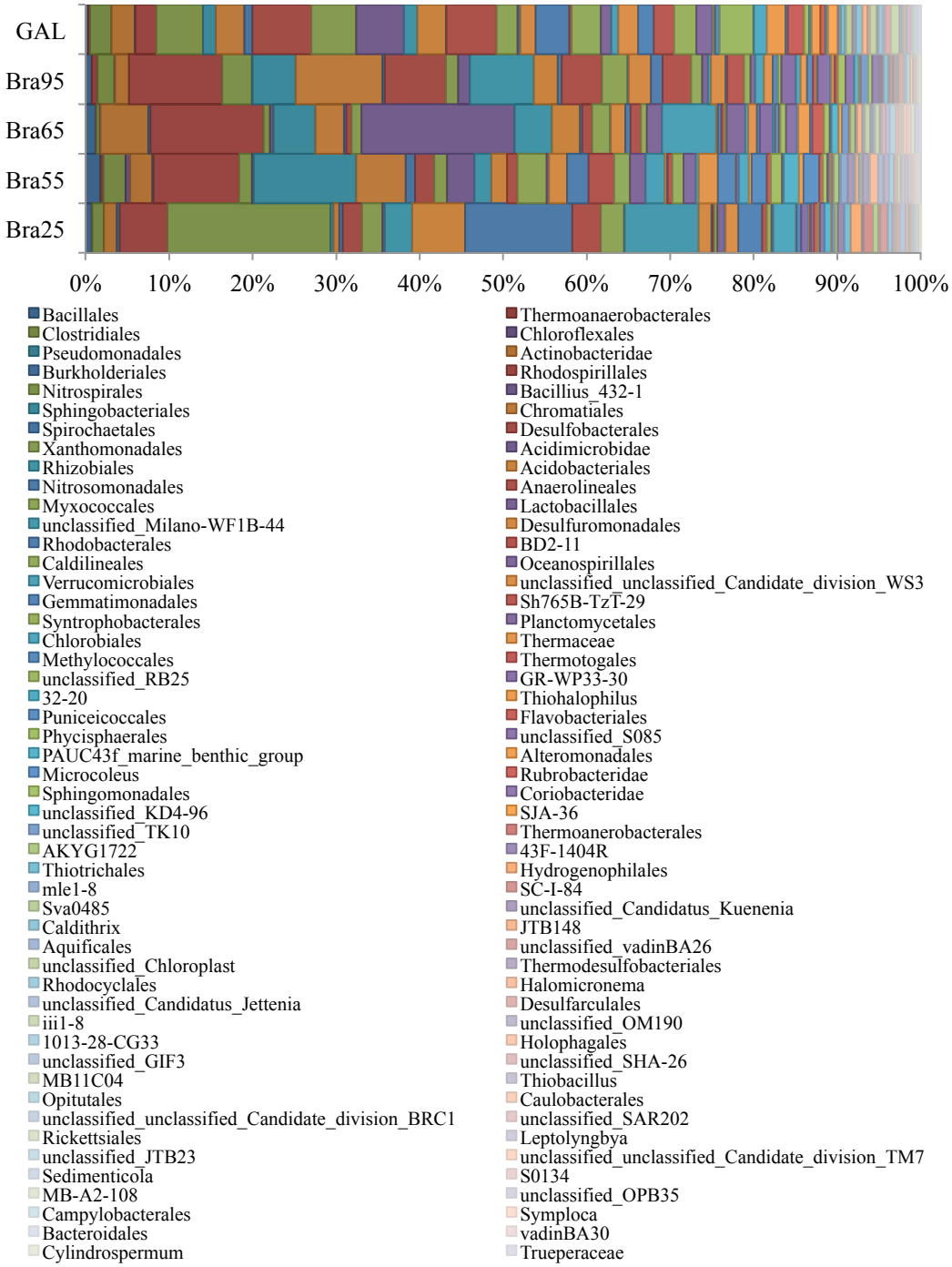


Figure 13 Relative abundance bacterial orders within soil samples from Brazoria National Wildlife Refuge and Galveston Island. Sites were selected for inocula in the screen comparison due to both proximity to the lab and an established understanding of ecological differences among the communities despite proximity to one another.

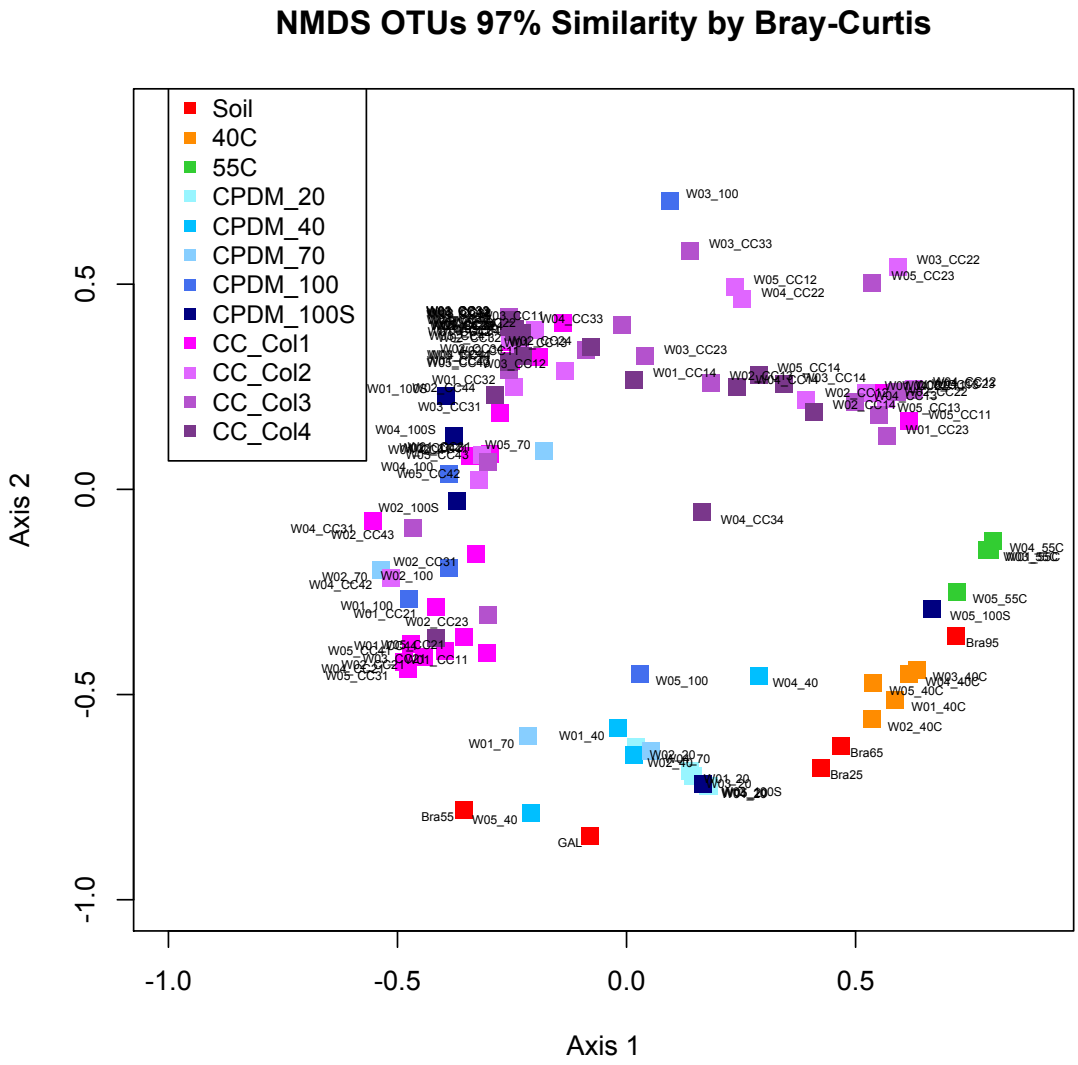


Figure 14 Non-metric multidimensional scaling of three carboxylate biofuel platform screen comparisons. Non-metric multidimensional scaling is an ordination of the Bray-Curtis dissimilarity between OTU classifications (at 3% dissimilarity) for screen comparison experiments. Sample names include indications of the soil inoculum or screen type and manipulated conditions associated with screen design. For a detailed explanation of sample name interpretation see Figure 11.

CHAPTER V

ASSEMBLY AND PHYLOGENETIC ANALYSIS OF AN ISOLATE LIBRARY DERIVED FROM CARBOXYLATE BIOFUEL PLATFORM FERMENTATIONS INOCULATED WITH ECOLOGICALLY AND GEOGRAPHICALLY DIVERSE MICROBIAL COMMUNITIES

V.1 Introduction

Metagenomic analyses of microbial communities provide for identification of organisms and putative insights into the metabolic abilities of the organisms. Ultimately to confirm the functional ecology and specific phenotypes, pure cultures of the organisms are necessary. We undertook a large-scale isolation project to capture a diverse collection of individual isolates from 30-day batch screen fermentation communities for a carboxylate biofuel platform (CBP). This affords us the opportunity to conduct more traditional direct studies of the microbes and microbial processes selected for by the fermentation screens. Bacteria from fermentation communities potentially possess relevant traits to improve industrial processes (Sizova *et al.*, 2011). By culturing microbes from fermentation communities ranked best for conversion we provide for future studies of: 1) individual strains; 2) genes and gene networks important for fermentation or survival in an extreme environments; 3) industrially relevant traits; 4) assembled community dynamics and; 5) booster inocula for directed fermentations. Thus we chose to characterize the library inasmuch as knowing the species of the isolates serves to prioritize the candidates for particular traits.

In the CBP fermentations, intermittent oxygen exposure is common. Possible future uses of these isolated microorganisms may be in a production plant that will not be kept completely anaerobic. Thus, the microbes used to inoculate the CBP must tolerate bursts of atmospheric levels of oxygen. Each culture approach in this study is designed to select for a diversity of facultative anaerobes. Our objective was to assemble and phylogenetically characterize a library of isolates from: CBP fermentations with particular performance outcomes, fermentations from sites that expanded the geographic diversity of the samples targeted for isolates, and from samples from other studies of traits relevant to industry (Hollister *et al.*, 2012; Hou *et al.*, 2012). We used a variety of culturing techniques to maximize library diversity including multiple media types and two oxygenation conditions. We sequenced a partial 16S subunit of the ribosome gene (16S rRNA gene) segment for all isolates and inferred a phylogeny for the library. We also compared the partial 16S isolate sequencing with partial 16S 454 pyro-sequencing from other experiments conducted in our lab (Chapter III), (Hammett, 2011), to determine whether any of these isolates were also detected in either the soil communities or the fermentation communities.

Table 10 Summary of isolate and sequence distribution across soil sites, fermentation experiments and culture conditions imposed. This library of isolates originated from fermentation experiments inoculated with soils from different sites. Focus for isolate library construction efforts was mainly fermentations with conversion performances that were among the top 10%, other fermentations included were based on maximizing geographic diversity. Isolation targets with fermentation identification number, corresponding soil identification number, location of soil sample, site name of soil sample, conversion percentage of fermentation, isolated (Iso.), and sequenced (Seq.) numbers of strains, across media and oxygenation regime, growth priority, rationale for targeting, and totals within sites of isolates and sequenced isolates.

Fermentation ID	Soil Sample Name	State	Site Name	Conversion	Facultative Aerobic Acetogenic		Facultative Aerobic Halophilic		Facultative Aerobic Cellulose		Anaerobic Acetogenic		Anaerobic Cellulose		Total Isolates Per Site	Total Seq. Per Site
					Iso.	Seq.	Iso.	Seq.	Iso.	Seq.	Iso.	Seq.	Iso.	Seq.		
S48	STYS3	WY	Yellowstone	0.57	30	30	20	19	30	27	0	0	0	0	80	76
S44	FHYS5	WY	Yellowstone	0.52	34	34	21	21	30	24	13	13	12	12	110	104
F02	Bra2	TX	Brazoria	0.47	32	30	26	26	27	27	0	0	13	13	98	96
G23	LL2	NM	Lazy Lagoon (Roswell)	0.49	44	39	20	20	24	27	0	0	12	12	100	98
H01	SFB1	CA	San Francisco Bay	0.37	28	23	27	25	12	11	0	0	0	0	67	59
G13	BL13	NM	Bitter Lake (Roswell)	0.46	45	43	20	20	30	24	0	0	0	0	95	87
J04	Big4	TX	Big Bend	0.36	32	30	15	13	25	27	0	0	23	23	95	93
G08	BL8	NM	Bitter Lake (Roswell)	0.47	26	23	22	21	25	26	0	0	0	0	73	70
F05	Bra5	TX	Brazoria	0.42	30	30	21	16	0	0	0	0	0	0	51	46
U22	OLCA1	CA	Owens Lake	0.47	30	30	27	27	30	28	9	9	7	7	103	101
J19	Big19	TX	Big Bend	0.35	30	28	19	19	30	20	0	0	0	0	79	67
H20	SFB20	CA	San Fran Bay	0.40	8	8	12	12	25	23	0	0	0	0	45	43
G19	BL19	NM	Bitter Lake (Roswell)	0.43	30	27	19	19	25	26	0	0	0	0	74	72
J18	Big18	TX	Big Bend	0.35	30	29	0	0	27	25	0	0	0	0	57	54
G24	LL3	NM	Lazy Lagoon (Roswell)	0.44	30	24	26	26	30	23	0	0	0	0	86	73
K49	BHS5	UT	Baker Hot Spring	0.41	30	16	23	23	30	26	0	0	0	0	83	65
P01	BWR1	PR	Puerto Rico	0.36	1	1	18	18	24	18	0	0	0	0	43	37
E08	GSP8	OK	Great Salt Plains	0.40	31	30	23	22	29	31	0	0	13	13	96	96

Table 10 Continued.

Fermentation ID	Soil Sample Name	State	Site Name	Conversion	Facultative Aerobic Acetogenic		Facultative Aerobic Halophilic		Facultative Aerobic Cellulose		Anaerobic Acetogenic		Anaerobic Cellulose		Total Isolates Per Site	Total Seq. Per Site
					Iso.	Seq.	Iso.	Seq.	Iso.	Seq.	Iso.	Seq.	Iso.	Seq.		
M24	CR22	SC	Cape Romain	0.33	30	27	25	25	30	22	0	0	0	0	85	74
N09	SI9	GA	Sapelo Island	0.27	30	24	22	21	30	23	0	0	0	0	82	68
R08	NSS3	NM	Vellas Caldera	0.22	31	23	16	16	30	21	0	0	0	0	77	60
T02	SWRN2	NV	Still Water	0.21	30	20	19	19	30	25	0	0	0	0	79	64
A07	LSDR T1-325	TX	La Sal Del Ray	N/A	0	0	24	24	0	0	0	0	0	0	24	24
J11	Big11	TX	Big Bend	0.39	2	2	13	11	0	0	0	0	0	0	15	13
J20	Big20	TX	Big Bend	0.34	0	0	19	17	0	0	0	0	0	0	19	17
G09	BL9	NM	Bitter Lake	0.39	30	26	0	0	0	0	0	0	0	0	30	26
G21	BL21	NM	Bitter Lake	0.38	0	0	0	0	0	0	0	0	0	0	0	0
F09	Bra91	TX	Brazoria	0.46	30	30	9	9	0	0	0	0	0	0	39	39
F06	Bra61	TX	Brazoria	0.43	0	0	23	21	0	0	0	0	0	0	23	21
F22	Bra12	TX	Brazoria	N/A	0	0	7	7	0	0	0	0	0	0	7	7
F08	Bra81	TX	Brazoria	0.35	30	28	0	0	0	0	0	0	0	0	30	28
F01	Bra11	TX	Brazoria	0.41	0	0	17	17	0	0	0	0	0	0	17	17
E07	GSP7	OK	Great Salt Plains	0.36	0	0	14	13	16	16	0	0	0	0	30	29
G47	BLM21	NM	BLM; Laguna Gatuna	0.40	0	0	5	5	0	0	0	0	0	0	5	5
N/A	JGI	N/A	Day15 - EBH Experiment	N/A	0	0	0	0	0	0	0	0	30	30	N/A	30
N/A	EB7	N/A	Electric Bacteria	N/A	0	0	0	0	0	0	0	0	0	0	N/A	7
				Total:	734	655	572	552	589	520	22	22	110	110	1997	1866

V.2 Materials and Methods

Selection of Samples Cultured

The selection of communities targeted for isolation (Table 10) was based on including communities with superior conversion performance (g volatile solids digested/ g volatile solids fed), selected from the top ~5% of communities in the 30-day batch screen (Chapter II). Also, we included communities from sites chosen to maximize possible geographic diversity in the isolates. In total, we cultured from at total of 36 samples, 34 were from fermentation screens and two additional samples came from collaborative projects (Hollister *et al.*, 2012; Hou *et al.*, 2012).

Since we were interested in selecting for diversity using different media and oxygen conditions, we attempted to isolate similar numbers from each sample on each medium, ≥25 per sample. Table 10 indicates the actual numbers of isolates derived from each culture approach.

Isolate Culture Methods

We collected a combination of wet and solid materials from the 30-day batch screen fermentation and stored it at 4°C until ready to culture on solid agar medium. For this study we employed three different defined media: Drake's Thermophilic Acetogen Medium (Drake, 1994) (DTAM), modified by aerobic handling (Non-Hungate), and SL9 trace elements (Tschech & Pfennig, 1984); Cellulose Agar for Thermophiles (Atlas, 1997) (CAT), modified by aerobic handling; and Modified Growth Medium (Nuttall & Dyall-Smith, 1993) (MGM) a halophile medium. We grew cultures on these three media

at 55°C. For six of the samples we used two different approaches to limit oxygen during culturing: growth in a nitrogen triple flushed and sealed food grade plastic bag (Sunbeam Products, Inc., Boca Raton, FL, USA, FoodSaver™ Model V2220,) or an anaerobic chamber (Oxoid Limited, Hampshire, England, 2.5L AnaeroJar Assembly) with chemical reduction of oxygen (Oxoid Limited, Hampshire, England, AneroGen 2.5L Cat. #AN0025A), both performed after transferring within a biosafety cabinet with exposure to ambient oxygen levels. Based on the color of the oxygen indicator (resazurin) in the CAT plates we estimate that oxygen was more limited in the chambers (light grey) than in the sealed plastic bags (pink). For the other thirty of the thirty-six samples we used only the nitrogen triple flushed and sealed food grade plastic bag method.

We isolated cultures to single colonies on solid medium and re-streaked 2 to 4 times to ensure a pure culture. Then we grew the isolates in the liquid broth version of the given defined medium under the same conditions as the primary culture, with either nitrogen flushing in sealed bags or chemical oxygen purging in the anaerobic chambers. We allowed the liquid cultures to grow until we saw turbidity or high colony numbers after transfer back onto solid media. We established stock cultures containing glycerol (20% of total volume) for storage at both -20°C and -80°C.

Genomic DNA Extraction

To collect DNA from an isolate we transferred a single colony into a 250µL polymerase chain reaction (PCR) tube containing 20µL of Alkaline Lysis Buffer (0.25%

SDS, 0.05 N NaOH) (Storms *et al.*, 2004), then incubated at 95°C for 30 min. in a stable heat block with water, frozen at -20°C for 30 min., and when removed from the freezer we combined 80µL of room temperature Promega nuclease free water (Madison, WI, USA) with the lysed cell suspension. We vortexed or vigorously shook the mixture to both thaw and homogenize the contents. We used 5µL of this preparation for PCR amplification of the 16S rRNA gene. In instances when this process failed to provide a sufficient quality or yield of genomic DNA we used Wizard® SV Genomic DNA Purification System (Promega Corp., Madison, WI, USA) extraction kits. We froze all lysis products at -20^{o}C until we accumulated enough samples for a 96 well PCR amplification or for a maximum of one week. Controls to monitor for uncontaminated DNA extractions included placing stock Alkaline Lysis Buffer alone through the extraction protocol and PCR to test the sterility of the solutions and using *Escherichia coli* for the positive control throughout this procedure.

PCR Amplification of Bacterial 16S rDNA

To prepare isolate genomic DNA extracts for Sanger sequencing and to verify the amplification of community DNA extracts, we amplified a ~1450bp region of the 16S rRNA gene spanning variable region V1 to variable region V8 as shown in Figure 6**.** Primers for PCR amplification were forward primer 27f: 5'-AGA GTT TGA TCC TGG CTC AG-3' and reverse primer 1492r: 5'-CGG TTA CCT TGT TAC GAC TT-3' (Lane, 1991); (Table 6), supplied by Integrated DNA Technologies, Inc. (Coralville, IA, USA). PCR amplification primers were rehydrated to a concentration of 100 µM, and diluted to

a working stock concentration of 20 µM with Promega nuclease free water. PCR reaction constituents for a 50 µL reaction include: 5 µL 10x buffer (500 mM KCl, 300 mM TRIS pH 8.3, 15 mM $MgCl_2$), 5 µL bovine serum albumin (BSA) (1.0mg/ml), 2 µL $MgCl_2$ (25 µM), 4 µL GeneAmp dNTPs (0.1 mM, Life Technologies, Carlsbad, CA, USA, formerly Applied Biosystems), 1 µL 27f, 1 µL 1492r, 0.5 µL *Taq* DNA Polymerase (2.5 Units New England Biolabs; Ipswich, MA, USA), 26.5 µL H2O, and 5 µL template DNA. Bovine serum albumin was used in the PCR reaction to alleviate inhibition of replication (Kreader, 1996) Amplification consisted of initial denaturation for 1 min. at 95°C; followed by 35 cycles of 1 min. at 95°C (denaturation), 1 min. at 55°C (annealing), 1.5 min. at 72°C (extension); followed by one cycle of final extension for 10 min. at 72°C; and then held at 4°C until separation by gel electrophoresis. Each group of PCR reactions included a positive and negative control (*E. coli* DNA and PCR grade water used as template for each, respectively). All isolates were amplified by PCR with the nearly complete 16S rRNA gene primers listed even when smaller portions of the 16S rRNA gene were the sequencing target, as the PCR product may be used for further sequencing should the need arise. Large 50 µL PCR reactions were required for this project, as all samples were normalized to 15-25 ng/µL for submission to Sanger sequencing. For isolate sequencing we normalized the DNA concentration of the full 96-well plates, which often required the full volume of the PCR reaction remaining after visualization. For agarose gel electrophoresis and visualization of PCR amplifications we used ethidium bromide (2 µL at 20 mg/100 mL) mixed per 100 mL 0.75% agarose (SeaKem® LE Agarose, Lonza, Walkersville, MD, USA). We performed gel

electrophoresis with 6 volts per cm of gel length between electrical contacts (Cambrex, East Rutherford, NJ, USA). We ran a 5 µl aliquot of each PCR product and 2 µL 1 kb Ladder (New England Biolabs Inc., Ipswich, MA, USA) on the gel to compare to the band sizes with the predicted product measuring 1450 bp. To track the migration of the PCR product in the gel we added 1 µl 0.015% bromophenol blue in each sample. We quantified the PCR products using an AlphaImager 2000 v5.5 with spot density software (Alpha Innotech Corp., San Leandro, CA, USA) and picture capture and visualization occurred on a UV Illuminator (Alpha Innotech Corp., San Leandro, CA, USA) with the same software. To estimate the concentration of the PCR projects we compared the intensity of each band to the intensity of a known concentration of DNA in the ladder bands. We used PCR grade nuclease-free water to dilute PCR products when necessary.

Sanger Sequencing of the 16S rRNA Gene

Beckman Coulter Genomics, formerly Agencourt Bioscience Corporation (Danvers, MA, USA) sequenced the 16S rRNA gene of all isolates using the Sanger single pass method (Sanger *et al.*, 1977). For the majority of samples sequenced for this study we supplied the following primers at a concentration of 100 µM: forward 27f: 5'-AGA GTT TGA TCC TGG CTC AG-3' (Lane, 1991) and reverse 519r-Gray: 5'-GTN TTA CNG CGG CKG CTG-3' (Acosta-Martinez *et al.*, 2008); Table 6). For a more complete look at the 16S rRNA gene (Figure 6) and a more reliable determination of classification, we generated a larger sequence with the following forward primers 27f: 5'-AGA GTT TGA TCC TGG CTC AG-3' and 515f: 5'-GTG CCA GCM GCC GCG

GTA A-3', and reverse primers 1100r: 5'-AGG GTT GCG CTC GTT-3', and 1492r: 5'-CGG TTA CCT TGT TAC GAC TT-3' (Lane, 1991); Table 6), (Integrated DNA Technologies, Inc., Coralville, IA, USA).

Post Sequencing Processing

We assembled the reads for each sample into contigs using Sequencher 4.10.1 (Gene Codes Corporation, Ann Arbor, MI, USA). We used the Ribosomal Database Project (RDP) classifier function (<https://rdp.cme.msu.edu/> accessed 11/11/11) (Cole *et al.*, 2009) to reference against the contigs for putative identities of the isolates. I used the RDP pyrosequencing pipeline aligner function (accessed 11/11/11) to align all sequences. I trimmed the alignment to a constant length using Bioedit v7.0.9 (Hall, 1999) To assess phylogenetic diversity, I analyzed all samples that formed a contig with a 97% identity cut off for assignment of OTU using the cluster command (average neighbor default) of mothur v1.22 (Schloss *et al.*, 2009). I deposited sequences into National Center for Biotechnology Information (NCBI) GenBank accession numbers KC847997 - KC848049 and KC849718 - KC851724. I used a representative library of different isolates as determined by OTU assignment and the get.oturep function in mothur v1.22 in the RDP Tree Builder program (accessed 11/11/11) with the inclusion of the highest similarity sequence in the RDP Hierarchy Browser (accessed 11/11/11) feature as nearest neighbors for classification. I obtained strains for comparison from the SeqMatch feature of the RDP website (accessed 11/11/11). I aligned all nearest neighbor sequences using the RDP aligner function (accessed 11/11/11) with the corresponding

database of representative sequences and I trimmed the sequences to the same length as the project isolate sequences using Bioedit. I uploaded the alignment into the myRDP database and selected for inclusion in the RDP Tree Builder program (accessed 11/11/11.) The RDP Tree Builder used Jukes-Cantor corrected distance matrix modeling, and Weighbor, a weighted neighbor-joining tree algorithm (Bruno *et al.*, 2000) or tree creation with bootstrapping at 100 iterations.

Analysis

I compared the outcome of isolation by sample, by media, and by media and oxygenation condition through OTU clustering and classification. I also compared the whole isolate library (Sanger *et al.* & 1977) composition to the post fermentation 454 pyrosequenced community libraries and to the original soil community composition (Chapter III). Since the regions of sequencing for that study and this on are the same, the two collections (454 pyro-sequencing and Sanger sequencing) were compatible for this purpose. This resulted, however, in a shortened segment of comparison, as the 454 pyrosequencing reads were shorter in length.

Table 11 Distribution of operational taxonomic units by original fermentation sample. Rows correspond to each Operational Taxonomic Unit (OTU) numbers based on a phylogenetic analysis. Columns correspond to the fermentation community or other source for the isolates (Elec. indicates electrical fuel cell project (Hou *et al.*, 2012); JGI indicates Joint Genome Institute project isolates (Hollister *et al.*, 2012); and E. coli indicates internal control isolates). The final column and final row correspond to the sum of all isolates within each OTU and each fermentation sample respectively. Coloring indicates red (no or few samples) to green (many samples) and is calculated by category.

OTU	A07	E07	E08	F01	F02	F05	F06	F08	F09	F22	G08	G09	G13	G19	G21	G23	G24	G47	H01	H20	J04	J11	J18	J19	J20	K49	M24	N09	P01	R08	S44	S48	T02	U22	Elec.	JGI	E. coli	Sum
1	4	22	51	14	35	40	20	28	33	5	62	27	86	51	89	53	0	5	47	7	39	5	29	65	7	27	58	61	23	41	73	46	37	70	0	0	0	1260
2	0	17	8	0	41	1	0	1	2	0	4	0	11	21	0	10	13	0	4	11	18	0	2	12	3	0	6	6	7	16	18	10	0	9	0	0	0	251
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
4	20	0	2	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	6	12	4	0	0	0	0	0	0	4	4	2	22	0	24	0	0	0	0	102
5	0	0	1	4	23	6	0	0	4	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	3	0	0	0	44
6	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	2	26	0	22	0	0	0	57
7	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
8	0	0	17	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	2	40	0	16	1	0	23	0	0	0	0	1	0	0	6	0	0	0	110
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	2	0	0	6	0	0	0	3	0	0	0	1	0	0	19
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
11	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	5	0	0	5	4	0	0	0	0	0	0	0	0	0	0	0	17
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	4	0	3	3	0	0	0	3	2	0	0	0	16
13	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
14	0	0	0	0	2	0	0	1	0	0	0	0	0	11	0	0	11	0	0	0	0	0	0	0	0	4	0	0	0	0	0	6	0	0	0	0	0	35
15	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
16	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
19	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2

Table 11 Continued.

OTU	A07	E07	E08	F01	F02	F05	F06	F08	F09	F22	G08	G09	G13	G19	G21	G23	G24	G47	H01	H20	J04	J11	J18	J19	J20	K49	M24	N09	P01	R08	S44	S48	T02	U22	Elec.	JGI	E. coli	Sum
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
24	0	0	5	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0	7	0	0	0	9	0	0	0	5	15	0	73
25	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	21
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
27	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	16
28	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
30	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	3	0	7
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
33	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
37	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2
39	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
43	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
44	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2
46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

V.3 Results

We collected a total of 1997 independent isolates across 34 fermentation samples inoculated with naturally occurring microbial communities (Chapter II), as well as, two additional projects related to examining the microbial communities in a CBP fermentation (Hollister *et al.*, 2012; Hou *et al.*, 2012); (Table 10). We sequenced a total of 1866 of those isolates. We generated a total of 2078 sequences due to repeated efforts for some sequences, and included 2060 sequences in the analysis after quality control (NCBI Genbank accessions: KC847997-KC848049, KC849718-KC851724). The length of the alignment was 482 bases while the average length of sequences was 445 base pairs. I found a total of 230 unique sequencing reads. I define unique as identical at the DNA sequence level. Binned at 97% similarity (3% dissimilarity), the commonly proposed level for species grouping, there were a total of 46 groups or operational taxonomic units (OTUs), hereafter referred to as OTU 1 through OTU 46. OTU 31 included an *Escherichia coli* control group (one OTU and two unique sequences) that we sequenced as a positive indicator of quality during extraction, PCR, and sequencing reactions. By examining the columns of Table 11 you can observe the OTU overlap between each of the sites.

Appendix G shows a phylogeny of the distinct bacterial lineages (OTUs). Of the 230 unique sequences, 223 or 97% were in the phylum *Firmicutes*. Of the *Firmicutes* I identified class *Bacilli* as 85% (196/230) of unique isolates. Class *Clostridia* made up the remainder of those sequences within *Firmicutes* with 11.7% (27/230) of unique isolate sequences. Additionally I found phylum *"Actinobacteria"* class *Actinobacteridae*

at 0.9% (2/230) unique level sequences, and phylum *Proteobacteria* class *Gammaproteobacteria* with 1.3% (3/230) of the library at the unique sequence level.

V.4 Discussion

One facet of our ongoing effort to characterize the microbial communities successful at fermentation in the carboxylate platform was to establish a library of individual isolates. We sought to maximize diversity within the library in order to increase its usefulness. To this end, we attempted to collect at least 25 independent single colony isolates with each of three distinct defined media. Furthermore, for a subset of the fermentation communities we imposed more stringent oxygen restriction in an attempt to capture greater diversity. Our efforts were successful, we collected distinct bacterial lineages (OTUs). In the very well represented OTUs we observed overlap across the media types (Table 11, Figure 15). Specifically 4% were captured on all three media, 6% on both CAT and MGM, 9% on CAT and DTAM, 4% on both DTAM and MGM, and 63% of the OTUs resulted from one media or one media and oxygenation condition. Perhaps a most noteworthy example is that by employing the anaerobic chamber in combination with the cellulose agar (CAT) we captured 4 OTUs (OTUs 23, 29, 34, and 38) not otherwise represented in the library. It seems reasonable to conclude our approach of using multiple media types and oxygenation conditions successfully expanded the diversity of OTUs represented in the library.

Table 12 Comparison of the operational taxonomic unit (OTU) (based on 3% dissimilarity) sequence matches across all possible pairings of three sequence collections in this dissertation. For each site with sequences available for the soil, the fermentation screen (Ferm.) and our isolate library (Iso.) (N=19) the numbers of OTU sequences for each collection and the number that overlap between and among them at the 3% dissimilarity level are listed.

Sample	Soil Only	Soil-Ferm.	Ferm. Only	Ferm.-Iso.	Iso. Only	Iso.-Soil	All
E08	651	1	59	4	7	0	0
F01	1266	4	55	1	1	0	0
F02	2068	2	68	2	5	0	0
F05	1655	0	37	0	3	0	0
F06	1010	3	39	1	0	0	0
F09	2158	1	49	0	3	0	0
G08	1059	2	38	1	4	0	0
H01	748	0	60	2	2	0	0
H20	1675	4	16	1	3	0	0
J04	699	2	28	3	2	0	0
J11	574	0	17	0	3	0	0
J18	569	1	16	2	3	0	0
J19	411	0	24	1	3	0	0
J20	556	1	26	0	4	0	0
K49	522	0	31	2	4	1	0
P01	2804	0	66	2	2	0	0
S44	130	2	39	5	2	0	0
S48	317	0	15	2	1	0	0
U22	391	1	27	2	3	0	0

Table 13 Comparison of the unique sequence matches across all possible pairings of three sequence collections in this dissertation. For each site with sequences available for the soil, the fermentation screen (Ferm.) and our isolate library (Iso.) (N=19) the numbers of unique sequences for each collection and the number that overlap between and among them at the unique level are listed.

Sample	Soil Only	Soil-Ferm.	Ferm. Only	Ferm.-Iso.	Iso. Only	Iso.-Soil	All
E08	1293	0	1828	4	17	0	0
F01	2782	1	1330	1	2	0	0
F02	4380	1	1530	3	11	0	0
F05	3557	0	1812	0	3	0	0
F06	2961	0	1313	1	0	0	0
F09	3843	0	1637	0	3	0	0
G08	2831	1	1751	0	6	0	0
H01	1459	0	1692	2	5	0	0
H20	3961	1	1660	1	6	0	0
J04	1128	0	1681	3	6	0	0
J11	980	0	1697	0	6	0	0
J18	1050	1	1920	1	8	0	0
J19	1205	0	1929	1	9	0	0
J20	1157	0	2122	0	4	0	0
K49	1526	0	1398	2	5	0	0
P01	5270	0	988	2	2	0	0
S44	1003	0	1926	6	4	0	0
S48	1601	0	920	2	3	0	0
U22	1827	0	1327	2	6	0	0

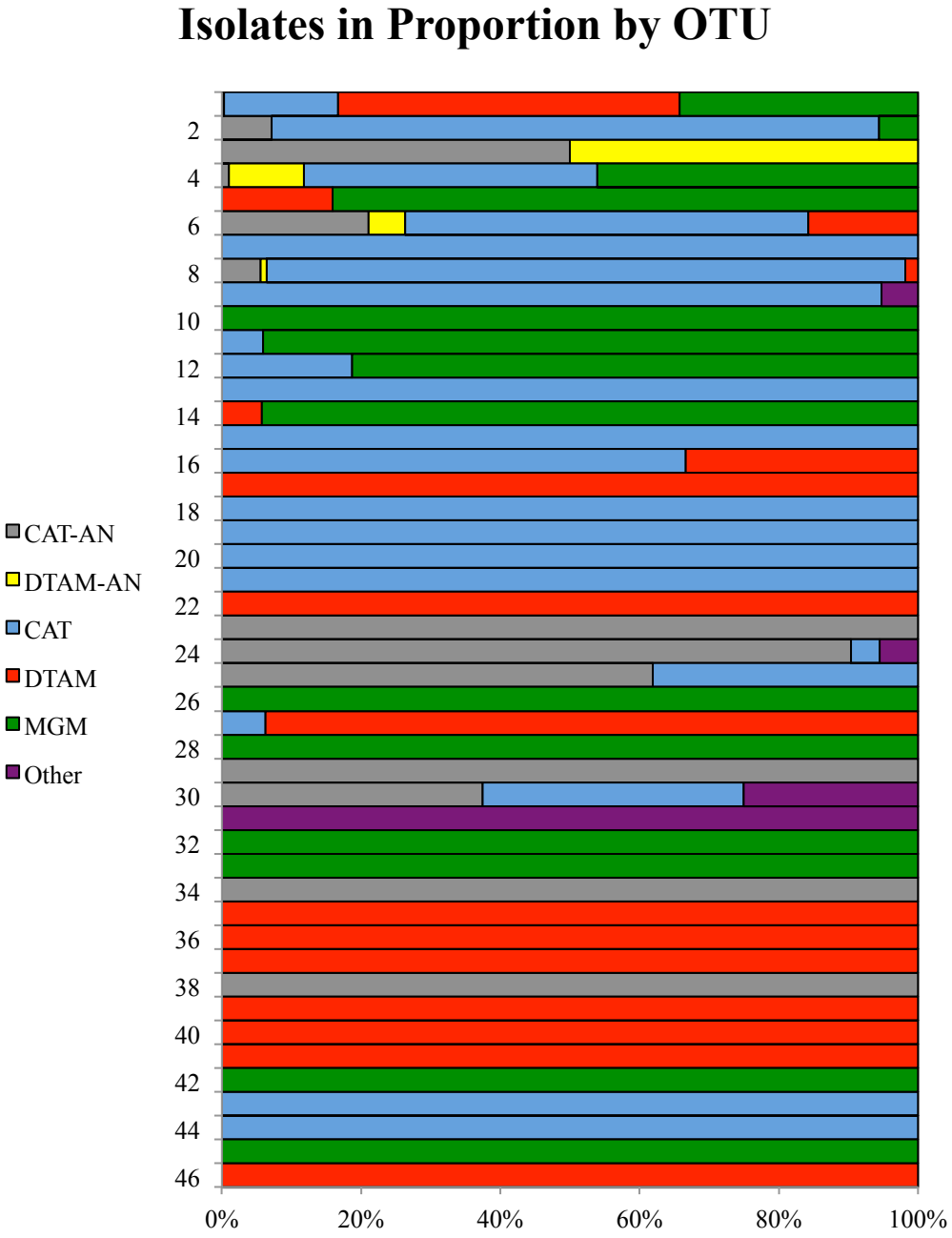


Figure 15 Media and oxygen incubation regime associated with each operational taxonomic unit (OTU) identified within the library. Bars indicate the relative proportion of each culture condition for each OTU (labelled by number on the left axis). Column to the right indicates the number of isolates within each OTU. Abbreviations: cellulose (CAT), acetogen (DTAM), halophile (MGM), and anerobic chamber incubations (-AN).

Most of the isolates we cultured fell into an OTU that was similar to a known sequence described and named within the 16S databases. In fact, of the 46 resulting OTUs 45 were identified to class and order at 100% confidence. Only 1 of 46, OTU 33, resulted in 92% confidence for class *Bacilli*, order *Bacillales*. At the family level, 7 of 46 OTUs were below 95% confidence. All 7 of these were within the family *Bacillaceae*; a diverse group and known to be difficult to classify by 16S sequence alone (Schleifer, 2009).

During our efforts to determine whether natural microbial communities from saline and thermal environments might exhibit superior performance in CBP fermentations (Chapter II), we conducted targeted locus sequencing with several 30-day batch screen fermentation communities (Chapter III) inclusive of many of the samples from which we cultured for this study. Furthermore, for a subset of the 30-day batch screen fermentation communities we conducted a targeted locus community survey of the original soils used as inocula for the fermentations (Chapter III). Thus, I wanted to see whether, and to what extent, the sequences from the isolates matched sequences in the community analyses of those same fermentation communities and soils. Table 12 shows the overlap of the sequences across these three datasets at the OTU level and that 74% of sites had at least one sequence overlap between the isolates and fermentation community sequencing. Table 13 shows the overlap of these sequences across these three datasets at the unique level and that and 79% of sites had at least one sequence overlap between the isolates and the fermentation community sequencing. No sequences in this subset shared identity between all three categories (soil, fermentation, and isolate)

at either level. Only one sample, K49, shared a sequence at the OTU level between the isolates and the soil community.

Also of interest was the idea that some strains could overlap from soil to 30-day batch screen to isolation. To examine this question I compared the sequence libraries of all isolates, all 30-day batch screen sequences, and all soil sequences. Figure 16 and Figure 17 show the overlap in these complete sets of sequencing without regard to sample site. Even at the unique level of comparison (every base over the examined portion), 23 isolate sequences matched sequences in the fermentation screen community sequences (Figure 17). One isolate sequence matched a soil community sequence. At the 97% similarity level, there were 10 operational taxonomic units (OTUs) in common only between the isolate library and the fermentation sequencing, and 2 OTUs in common only between the isolate library and the soil sequences, and 5 OTUs in common among the isolate library, the fermentation sequences, and the soil sequences (See Figure 16).

This diverse library of bacteria isolated from communities successful at CBP fermentations will serve as a resource for our ongoing efforts to optimize the use of this approach for associated biofuel and bioproduct industries. For example, we may be able to assemble an optimized community for use as a booster inoculum for carboxylate platform applications. Furthermore, with isolates in hand and characterized, any such communities assembled from these resources can be studied in depth via a variety of approaches that require generation of baseline genome sequence(s). In addition the library can serve as a source for strains for use in other industrial processes that use

individual isolates and require growth under conditions similar to those involved in these fermentations (e.g. 2% salt, 55°C).

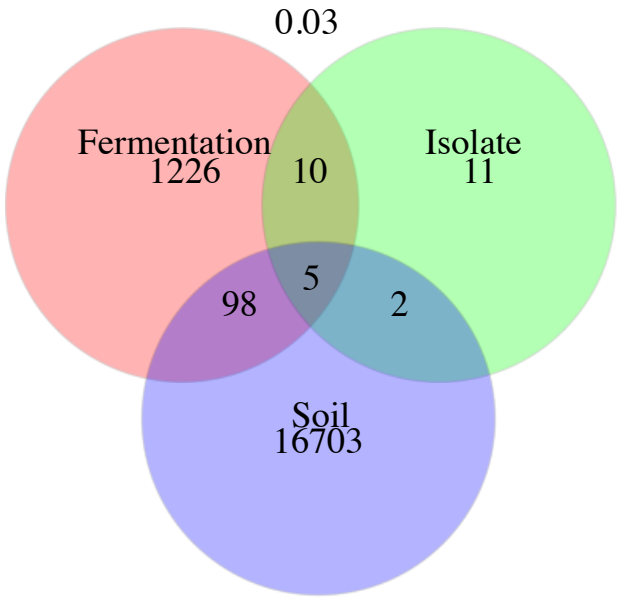


Figure 16 Venn diagram exhibiting the numbers of operational taxonomic units (OTUs) (at 97% similarity) and operational taxonomic unit (OTU) matches among sequence collections in this dissertation. For a list of sequenced communities see Table 9.

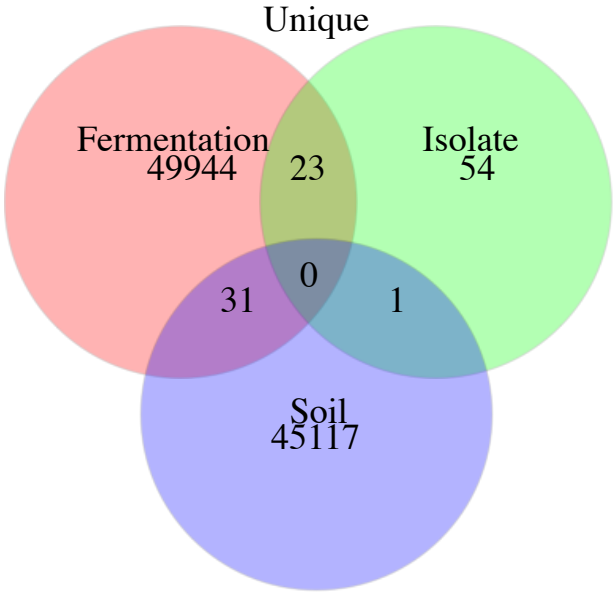


Figure 17 Venn diagram exhibiting the numbers of unique sequences and unique level sequence matches among sequence collections in this dissertation. For a list of sequenced communities see Table 9.

CHAPTER VI

CONCLUSION

In Chapter II we extensively sampled natural communities to explore the hypothesis that saline and thermal environments would select for microbes that perform in similar conditions. Variation in geographically and ecologically diverse sediments resulted in wide ranges of sediment chemistry. There were three times as many samples in the Stage II style fermentations than were collected in Stage I. This resulted in more power in the analysis of the relationships between the sediments and the fermentation performance. When effects were observed, they bore out in both the Stage I models and the Stage II models. The most notable of these was the influence of temperature of the original in situ soil. The effect of temperature was negative on acid production while positive on conversion. With the increase in sample size, other effects were also observed such as pH, total nitrogen, magnesium and calcium as measured through detailed salinity, and the (plant available) phosphorus. We also observed the evidence that process conditions have an influence on fermentation outcomes (Stage I verses Stage II), yet do not dictate the performance of the fermentation. Evidence of this can be seen in the variability of fermentation outcomes and the overlap of several outputs across fermentation conditions. I predict that the resulting variability in fermentation we observed was due to the variability of communities used as inocula.

In Chapter III I explored the resultant bacterial community structure from many of the top performing communities across several important fermentation characteristics

in (conversion, acetic acid equivalents (AEQ), and high molecular weight (C5-C7) acids). I saw through regression and Partial Least Squares decomposition two major and several minor relationships that were consistent. The major conclusions are that the presence and frequency of *Clostridia* sequences was associated with increased conversion and AEQ and that the presence and sequence frequency of *Bacilli* increased with a corresponding increase in conversion. There are several other classes of organisms (*Thermatoga, Chloroflexi*) that when present increase the diversity of the fermentation (a common observation for fermentation improvement) and the subsequent acid spectrum as per the major conclusions in (Hammett, 2011). I was able, through increased sampling, to establish these conclusions at 55°C the common operating temperature of the CBP.

In Chapter IV we explore the restrictions to community diversity imposed by the screening technology and the effect of fresh inoculum and concurrent inoculation of three different fermentation screens. We knew that diversity decreases dramatically from sediment through fermentation, but were able to see that no one screen has a more severe impact on diversity than any other.

We focused on bacterial communities even though the extreme and variable nature of our sampling sites could predispose our inocula to contain other forms of life (protists, archaea, eukaryotes). A prior in depth study (Hollister *et al.*, 2011) showed that bacteria were numerically abundant at the time point of interest (30 days). Throughout the experiments in this dissertation and the project as a whole I sought parallels of communities for performance that would allow me to design and test

community assemblages that could lend themselves toward an inocula for the industrial scale fermentation. In Chapter V we cultured with three media under two oxygenation conditions from top performing communities to try to assemble a diverse library of organisms. We sought to capture sufficient diversity to reassemble communities I observed influencing performance to directly test the hypotheses about patterns identified in this dissertation. I was able to culture representative of *Clostridia* and a diverse range of *Bacillus*. Again, I demonstrated (Chapter III) that both of these classes were important for superior conversion.

REFERENCES

Acosta-Martinez, V, Dowd, S, Sun, Y and Allen, V. (2008). Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. *Soil Biology and Biochemistry* 40:2762-2770.

Aiello-Mazzarri, C, Agbogbo, FK and Holtzapple, MT. (2006). Conversion of municipal solid waste to carboxylic acids using a mixed culture of mesophilic microorganisms. *Bioresour Technol* 97:47-56.

Aitken, MD and Mullennix, RW. (1992). Another look at thermophilic anaerobic digestion of wastewater sludge. *Water Environment Research* 64:915-919.

Atlas, RM. (1997). Handbook of Microbiological Media. 2 ed. CRC Press, Boca Raton, FL.

Berezovsky, IN and Shakhnovich, EI. (2005). Physics and evolution of thermophilic adaptation. *Proc Natl Acad Sci U S A* 102:12742-12747.

Boyd, TL. (2002). Western States Geothermal Databases CD. *Geo-Heat Center Quarterly Bulliten* 23:1-5.

Bruno, WJ, Socci, ND and Halpern, AL. (2000). Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Molecular Biology and Evolution* 17:189-197.

Cole, JR, Wang, Q, Cardenas, E, Fish, J, Chai, B, Farris, RJ *et al.* (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37:D141-145.

Day, PR. (1965). Particle fractionation and particle-size analysis In Black, A S (ed.), Agronomy, a series of monographs. American Society of Agronomy, Soil Science Society of America, and American Society for Testing Materials, Madison, WI, Vol. 1, pp. 545-567.

de Vrije, T, Bakker, RR, Budde, MA, Lai, MH, Mars, AE and Claassen, PA. (2009). Efficient hydrogen production from the lignocellulosic energy crop *Miscanthus* by the extreme thermophilic bacteria *Caldicellulosiruptor saccharolyticus* and *Thermotoga neapolitana*. *Biotechnol Biofuels* 2:12.

DeAngelis, KM, Gladden, JM, Allgaier, M, D'haeseleer, P, Fortney, JL, Reddy, A *et al.* (2010). Strategies for enhancing the effectiveness of metagenomic-based enzyme discovery in lignocellulolytic microbial communities. *BioEnergy Research* 3:146-158.

DeAngelis, KM, Fortney, JL, Borglin, S, Silver, WL, Simmons, BA and Hazen, TC. (2012). Anaerobic decomposition of switchgrass by tropical soil-derived feedstock-adapted consortia. *MBio* 3:e00249-00211.

Domke, SB, Aiello-Mazzarri, C and Holtzapple, MT. (2004). Mixed acid fermentation of paper fines and industrial biosludge. *Bioresour Technol* 91:41-51.

Drake, HL. (1994). Acetogenesis. Chapman & Hall, New York, NY. 223-224.

Forrest, AK. (2010). Doctoral dissertation, Texas A&M University Department of Chemical Engineering, College Station, TX.

Forrest, AK, Hernandez, J and Holtzapple, MT. (2010). Effects of temperature and pretreatment conditions on mixed-acid fermentation of water hyacinths using a mixed culture of thermophilic microorganisms. *Bioresour Technol* 101:7510-7515.

Forrest, AK, Hollister, EB, Gentry, TJ, Wilkinson, HH and Holtzapple, MT. (2012). Comparison of mixed-acid fermentations inoculated with six different mixed cultures. *Bioresour Technol* 118:343-349.

Fu, ZH. (2007). Doctoral dissertation, Texas A&M University Department of Chemical Engineering, College Station, TX.

Fu, ZH and Holtzapple, MT. (2010a). Consolidated bioprocessing of sugarcane bagasse and chicken manure to ammonium carboxylates by a mixed culture of marine microorganisms. *Bioresour Technol* 101:2825-2836.

Fu, ZH and Holtzapple, MT. (2010b). Fermentation of sugarcane bagasse and chicken manure to calcium carboxylates under thermophilic conditions. *Appl Biochem Biotechnol* 162:561-578.

Fu, ZH and Holtzapple, MT. (2011). Anaerobic thermophilic fermentation for carboxylic acid production from in-storage air-lime-treated sugarcane bagasse. *Appl Microbiol Biotechnol* 90:1669-1679.

Garlock, RJ, Balan, V, Dale, BE, Pallapolu, VR, Lee, YY, Kim, Y *et al.* (2011). Comparative material balances around pretreatment technologies for the conversion of switchgrass to soluble sugars. *Bioresour Technol* 102:11063-11071.

Golub, KW, Forrest, AK, Mercy, KL and Holtzapple, MT. (2011a). Propagated fixed-bed mixed-acid fermentation: Part I: Effect of volatile solid loading rate and agitation at high pH. *Bioresour Technol* 102:10592-10601.

Golub, KW, Smith, AD, Hollister, EB, Gentry, TJ and Holtzapple, MT. (2011b). Investigation of intermittent air exposure on four-stage and one-stage anaerobic semi-continuous mixed-acid fermentations. *Bioresour Technol* 102:5066-5075.

Golub, KW. (2012). Doctoral Dissertation, Texas A&M University Department of Chemical Engineering, College Station, TX.

Golub, KW, Golub, SR, Meysing, DM and Holtzapple, MT. (2012). Propagated fixed-bed mixed-acid fermentation: effect of volatile solid loading rate and agitation at near-neutral pH. *Bioresour Technol* 124:146-156.

Golub, KW, Forrest, AK, Wales, ME, Hammett, AJ, Cope, JL, Wilkinson, HH *et al.* (2013). Comparison of three screening methods to select mixed-microbial inoculum for mixed-acid fermentations. *Bioresour Technol* 130:739-749.

Hall, TA. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41:95-98.

Hammett, AJM. (2011). Doctoral dissertation, Texas A&M University Department of Plant Pathology and Microbiology, College Station, TX.

Haruta, S, Cui, Z, Huang, Z, Li, M, Ishii, M and Igarashi, Y. (2002). Construction of a stable microbial community with high cellulose-degradation ability. *Appl Microbiol Biotechnol* 59:529-534.

Heipieper, H, Neumann, G, Cornelissen, S and Meinhardt, F. (2007). Solvent-tolerant bacteria for biotransformations in two-phase fermentation systems. *Appl Microbiol Biotechnol* 74:961-973.

Hollister, EB, Engledow, AS, Hammett, AJ, Provin, TL, Wilkinson, HH and Gentry, TJ. (2010a). Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. *ISME J* 4:829-838.

Hollister, EB, Forrest, AK, Wilkinson, HH, Ebbole, DJ, Malfatti, SA, Tringe, SG *et al.* (2010b). Structure and dynamics of the microbial communities underlying the carboxylate platform for biofuel production. *Appl Microbiol Biotechnol* 88:389-399.

Hollister, EB, Hammett, AM, Holtzapple, MT, Gentry, TJ and Wilkinson, HH. (2011). Microbial community composition and dynamics in a semi-industrial-scale facility operating under the MixAlco™ bioconversion platform. *Journal of Applied Microbiology* 110:587-596.

Hollister, EB, Forrest, AK, Wilkinson, HH, Ebbole, DJ, Tringe, SG, Malfatti, SA *et al.* (2012). Mesophilic and thermophilic conditions select for unique but highly parallel

microbial communities to perform carboxylate platform biomass conversion. *PLoS One* 7:e39689.

Holtzapple, MT and Granda, CB. (2009). Carboxylate platform: the MixAlco process part 1: comparison of three biomass conversion platforms. *Appl Biochem Biotechnol* 156:95-106.

Hood, GM. (2010). *Poptools version 3.2.5.* Available at: http://www.poptools.org

Hou, H, Li, L, Ceylan, CU, Haynes, A, Cope, J, Wilkinson, HH *et al.* (2012). A microfluidic microbial fuel cell array that supports long-term multiplexed analyses of electricigens. *Lab Chip* 12:4151-4159.

Huber, R, Woese, C, Langworthy, T, Kristjansson, J and Stetter, K. (1990). *Fervidobacterium islandicum* sp. nov., a new extremely thermophilic eubacterium belonging to the "*Thermotogales*". *Archives of Microbiology* 154:105-111.

Izquierdo, JA, Sizova, MV and Lynd, LR. (2010). Diversity of bacteria and glycosyl hydrolase family 48 genes in cellulolytic consortia enriched from thermophilic biocompost. *Appl Environ Microbiol* 76:3545-3553.

Kreader, CA. (1996). Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Appl Environ Microbiol* 62:1102-1106.

Lane, DJ. (1991). 16S/23S rRNA sequencing In Stackebrandt, E, and Goodfellow, M (eds.), Nucleic acid techniques in bacterial systematics - Modern microbiological methods. Wiley & Sons Ltd., London, UK, pp. 329.

Magurran, AE. (2004). Measuring Biological Diversity. Blackwell Pub., Malden, MA.

McCauley, A and Jones, C. (2005). Soil and Water Management Modules. Montana State University Extension Service, Bozeman, MT, Vol. 2, pp. 1-16.

McGeehan, SL and Naylor, DV. (1988). Automated instrumental analysis of carbon and nitrogen in plant and soil samples 1. *Communications in Soil Science and Plant Analysis* 19:493-505.

Mehlich, A. (1978). New extractant for soil test evaluation of phosphorus, potassium, magnesium, calcium, sodium, manganese, and zinc. *Communications in Soil Science and Plant Analysis* 9:477-492.

Mesbah, NM, Abou-El-Ela, SH and Wiegel, J. (2007). Novel and unexpected prokaryotic diversity in water and sediments of the alkaline, hypersaline lakes of the Wadi An Natrun, Egypt. *Microb Ecol* 54:598-617.

Mesbah, NM and Wiegel, J. (2008). Life at extreme limits: the anaerobic halophilic alkalithermophiles. *Ann N Y Acad Sci* 1125:44-57.

Mesbah, NM and Wiegel, J. (2009). *Natronovirga wadinatrunensis* gen. nov., sp. nov. and *Natranaerobius trueperi* sp. nov., halophilic, alkalithermophilic micro-organisms from soda lakes of the Wadi An Natrun, Egypt. *Int J Syst Evol Microbiol* 59:2042-2048.

Meyer-Dombard, DR, Shock, EL and Amend, JP. (2005). Archaeal and bacterial communities in geochemically diverse hot springs of Yellowstone National Park, USA. *Geobiology* 3:211-227.

Nuttall, SD and Dyall-Smith, ML. (1993). Ch2, a novel halophilic archaeon from an Australian solar saltern. *Int J Syst Bacteriol* 43:729-734.

Porter, D, Roychoudhury, AN and Cowan, D. (2007). Dissimilatory sulfate reduction in hypersaline coastal pans: Activity across a salinity gradient. *Geochimica et Cosmochimica Acta* 71:5102-5116.

Rastogi, G, Bhalla, A, Adhikari, A, Bischoff, KM, Hughes, SR, Christopher, LP *et al.* (2010). Characterization of thermostable cellulases produced by *Bacillus* and *Geobacillus* strains. *Bioresour Technol* 101:8798-8806.

Rhoades, JD and Clark, M. (1978). In U.S. Salinity Laboratory USDA, R C (ed.), Washington, DC, pp. 11-12.

Rhoades, JD. (1982). Soluble salts. American Society of Agronomy, Soil Science Society of America, Madison, WI, pp. 167-179.

Riviere, D, Desvignes, V, Pelletier, E, Chaussonnerie, S, Guermazi, S, Weissenbach, J *et al.* (2009). Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *ISME J* 3:700-714.

Sanger, F, Nicklen, S and Coulson, AR. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463-5467.

Schleifer, K-H. (2009). Phylum XIII.*Firmicutes* Gibbons and Murray 1978, 5 (*Firmacutes* [sic] Gibbons and Murray 1978, 5) In Vos, P, Garrity, G, Jones, D, Krieg, N, Ludwig, W, Rainey, F *et al.* (eds.), Bergey's Manual of Systematic Bacteriology. Springer, NY, pp. 19-1317.

Schloss, PD, Westcott, SL, Ryabin, T, Hall, JR, Hartmann, M, Hollister, EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported

software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537-7541.

Schofield, RK and Taylor, AW. (1955). Measurement of soil pH. *Proc of the Soil Sci Soc America* 19:164-167.

Schulte, EE and Hopkins, BG. (1996). Estimation of Soil Organic Matter by Weight Loss-On-Ignition. *Soil Organic Matter: Analysis and Interpretation* Soil Science Society of America Special Publication:21-31.

Sizova, MV, Izquierdo, JA, Panikov, NS and Lynd, LR. (2011). Cellulose- and xylan-degrading thermophilic anaerobic bacteria from biocompost. *Appl Environ Microbiol* 77:2282-2291.

Skinner, FA. (1960). The isolation of anaerobic cellulose-decomposing bacteria from soil. *J Gen Microbiol* 22:539-554.

Stackebrandt, E and 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.

Staff, SSD. (1993). Soil Survey Manual. Soil Conservation Service, U.S. Department of Agriculture Handbook, Washington, DC, Vol. 18.

Storms, V, Van Den Vreken, N, Coenye, T, Mahenthiralingam, E, LiPuma, JJ, Gillis, M *et al.* (2004). Polyphasic characterisation of burkholderia cepacia-like isolates leading to the emended description of *Burkholderia pyrrocinia*. *Systematic and Applied Microbiology* 27:517-526.

Tabachnick, BG and Fidell, LS. (2000). Computer-assisted research design and analysis. Allyn and Bacon, Boston, MA.

Taylor, M, Tuffin, M, Burton, S, Eley, K and Cowan, D. (2008). Microbial responses to solvent and alcohol stress. *J Biotechnol* 3:1388-1397.

Thanakoses, P, Black, AS and Holtzapple, MT. (2003a). Fermentation of corn stover to carboxylic acids. *Biotechnol Bioeng* 83:191-200.

Thanakoses, P, Mostafa, NA and Holtzapple, MT. (2003b). Conversion of sugarcane bagasse to carboxylic acids using a mixed culture of mesophilic microorganisms. *Appl Biochem Biotechnol* 105 -108:523-546.

Tschech, A and Pfennig, N. (1984). Growth yield increase linked to caffeate reduction in *Acetobacterium woodii*. *Archives of Microbiology* 137:163-167.

Werner, JJ, Knights, D, Garcia, ML, Scalfone, NB, Smith, S, Yarasheski, K *et al.* (2011). Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proc Natl Acad Sci U S A* 108:4158-4163.

Wold, S, Sjöström, M and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58:109-130.

APPENDIX A

FERMENTATION AND SOIL ANALYSIS DATA FOR ALL SAMPLES IN EXPERIMENT

Appendix A is provided as a supplemental Microsoft Excel file.

APPENDIX B

OTU DIVERSITY, RICHNESS AND OVERLAP AMONG 40 SAMPLES

# seqs	# OTUs	Chao	Shannon	Θyc	D04	D05	D18	E08	F01	F02	F05	F06	F09	G03	G08	G09	G13	G14	G16	G19	G22	G23	G24	G34	G38	G41	G45	G46	G47	G48	H01	H20	J04	J11	J18	J19	J20	K49	M24	N21	P01	S44	S48	U22
3572	79	247	1.87	D04																																									
4244	97	180	1.93	D05	0.098																																									
5229	138	243	2.14	D18	0.317	0.274																																							
5247	64	130	1.23	E08	0.164	0.194	0.346																																						
5766	60	106	0.86	F01	0.231	0.276	0.455	0.038																																					
4269	72	129	1.75	F02	0.163	0.170	0.265	0.174	0.289																																				
6863	37	59	0.73	F05	0.237	0.267	0.451	0.031	0.006	0.273																																			
4221	43	54	1.62	F06	0.557	0.515	0.562	0.561	0.668	0.366	0.634																																		
4738	50	96	1.51	F09	0.553	0.485	0.465	0.503	0.622	0.330	0.586	0.073																																	
2410	72	151	1.89	G03	0.073	0.162	0.276	0.160	0.253	0.121	0.253	0.479	0.484																																
7168	41	68	1.08	G08	0.997	0.985	0.996	0.996	0.998	0.998	0.998	0.995	0.995	0.992																															
2747	149	346	2.90	G09	0.947	0.891	0.884	0.973	0.979	0.968	0.982	0.970	0.965	0.943	0.967																														
10002	36	43	1.07	G13	0.999	0.997	0.998	0.999	0.999	1.000	0.999	0.998	0.998	0.999	0.601	0.993																													
3817	104	392	2.20	G14	0.328	0.295	0.174	0.387	0.480	0.339	0.481	0.655	0.587	0.312	0.990	0.853	0.995																												
4994	107	213	1.99	G16	0.466	0.500	0.496	0.551	0.654	0.341	0.634	0.150	0.243	0.337	0.989	0.945	0.997	0.585																											
7274	37	54	1.21	G19	0.997	0.985	0.998	0.996	0.997	0.999	0.998	0.996	0.996	0.993	0.901	0.970	0.329	0.988	0.990																										
6478	85	148	1.80	G22	0.989	0.961	0.980	0.991	0.993	0.993	0.994	0.988	0.987	0.981	0.294	0.904	0.378	0.971	0.974	0.596																									
8818	29	47	0.64	G23	0.999	0.997	0.999	1.000	0.998	1.000	1.000	0.999	0.999	0.999	0.977	0.996	0.341	0.995	0.999	0.159	0.759																								
8851	25	29	1.27	G24	0.996	0.988	0.998	0.997	0.997	0.999	0.999	0.996	0.996	0.994	0.317	0.972	0.146	0.991	0.992	0.509	0.121	0.616																							
3755	394	1411	2.46	G34	0.709	0.643	0.619	0.700	0.793	0.514	0.764	0.102	0.118	0.629	0.999	0.901	1.000	0.712	0.240	0.999	0.985	1.000	0.999																						
4598	107	310	2.58	G38	0.577	0.597	0.520	0.679	0.747	0.555	0.745	0.559	0.598	0.530	0.992	0.816	0.998	0.398	0.447	0.994	0.976	0.999	0.995	0.590																					
5085	84	140	1.76	G41	0.278	0.414	0.513	0.445	0.508	0.353	0.510	0.511	0.600	0.161	0.994	0.930	0.999	0.525	0.256	0.994	0.980	0.999	0.995	0.646	0.511																			
6006	104	175	1.60	G45	0.079	0.123	0.322	0.066	0.118	0.158	0.120	0.564	0.541	0.067	0.996	0.921	0.999	0.342	0.491	0.996	0.982	1.000	0.996	0.692	0.621	0.300																	
9282	43	113	0.89	G46	0.997	0.995	0.996	0.999	0.998	0.999	0.999	0.997	0.997	0.999	0.907	0.996	0.212	0.992	0.998	0.151	0.677	0.032	0.502	1.000	0.997	0.999	1.000																	
6721	30	36	1.57	G47	0.992	0.962	0.990	0.991	0.994	0.996	0.996	0.990	0.989	0.982	0.303	0.925	0.617	0.979	0.973	0.693	0.132	0.877	0.343	0.998	0.984	0.986	0.990	0.828															
4207	112	190	2.14	G48	0.106	0.110	0.175	0.133	0.253	0.072	0.242	0.425	0.378	0.071	0.995	0.931	0.999	0.218	0.378	0.995	0.983	0.999	0.996	0.557	0.500	0.341	0.094	0.999	0.987													
4732	62	145	1.22	H01	0.923	0.878	0.862	0.883	0.950	0.764	0.922	0.330	0.325	0.873	0.999	0.985	0.997	0.929	0.540	0.996	0.995	0.997	0.998	0.263	0.819	0.870	0.913	0.997	0.998	0.820												
7475	21	28	0.41	H20	0.993	0.994	0.993	0.985	0.992	0.994	0.991	0.995	0.995	0.992	0.995	0.999	0.997	0.839	0.995	0.999	0.998	0.999	0.997	0.998	0.833	0.985	0.993	0.999	0.998	0.974	0.996										
5279	33	119	1.33	J04	0.949	0.961	0.993	0.999	0.993	0.992	0.997	0.981	0.987	0.997	0.967	0.975	0.998	0.987	0.997	0.964	0.968	0.986	0.961	0.998	0.996	0.999	1.000	0.990	0.950	0.999	0.998	1.000									
4599	17	83	0.56	J11	0.992	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.901	0.979	0.996								
5152	19	97	1.03	J18	0.992	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.902	0.980	0.983	0.072							
4439	25	85	1.45	J19	0.994	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	0.932	0.987	0.855	0.531	0.451						
6405	27	57	0.42	J20	0.991	0.996	0.999	0.999	1.000	0.999	1.000	0.998	0.999	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.903	0.978	0.993	0.018	0.106	0.559					
7710	33	62	0.54	K49	0.933	0.874	0.980	0.998	0.996	0.981	0.990	0.948	0.968	0.997	0.997	0.997	0.973	0.977	0.991	0.968	0.982	0.965	0.982	0.998	0.989	0.996	0.999	0.950	0.982	0.996	0.994	0.999	0.845	1.000	0.999	0.992	0.984				
2737	34	139	1.48	M24	0.933	0.835	0.940	0.999	0.997	0.986	0.994	0.961	0.976	0.980	0.998	0.739	1.000	0.923	0.982	0.999	0.963	0.999	1.000	0.946	0.940	0.962	0.953	0.990	0.991	0.975	0.996	1.000	0.896	1.000	1.000	1.000	0.991	0.580			
4316	62	140	1.36	N21	0.125	0.089	0.379	0.148	0.175	0.251	0.173	0.635	0.602	0.228	0.999	0.956	0.996	0.389	0.629	0.995	0.989	0.996	0.997	0.776	0.702	0.485	0.140	0.991	0.994	0.208	0.945	0.991	0.952	0.949	0.948	0.963	0.945	0.821	0.855		
3506	68	123	1.98	P01	0.915	0.917	0.926	0.920	0.924	0.919	0.925	0.952	0.950	0.918	1.000	0.995	1.000	0.927	0.949	1.000	0.999	1.000	1.000	0.967	0.951	0.937	0.917	1.000	1.000	0.913	0.988	0.998	0.998	0.974	0.972	0.979	0.975	0.997	0.998	0.916	
6710	46	88	1.75	S44	0.889	0.838	0.936	0.949	0.949	0.934	0.947	0.927	0.942	0.946	0.998	0.993	0.999	0.933	0.960	0.999	0.994	0.999	1.000	0.978	0.960	0.958	0.948	0.990	0.990	0.945	0.984	0.998	0.887	0.968	0.966	0.961	0.959	0.562	0.406	0.807	0.986
5299	17	26	0.80	S48	0.994	0.995	0.999	1.000	1.000	0.999	1.000	0.998	0.999	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.961	0.991	0.994	0.746	0.749	0.841	0.748	0.980	0.748	0.972	0.988	0.625
5517	30	50	1.06	U22	0.923	0.838	0.915	0.955	0.994	0.862	0.972	0.560	0.593	0.944	0.997	0.988	1.000	0.956	0.728	0.999	0.993	0.999	1.000	0.591	0.881	0.933	0.972	0.988	0.990	0.911	0.496	0.999	0.885	0.989	0.989	0.992	0.978	0.457	0.672	0.865	0.997	0.647	0.970
				Θyc	D04	D05	D18	E08	F01	F02	F05	F06	F09	G03	G08	G09	G13	G14	G16	G19	G22	G23	G24	G34	G38	G41	G45	G46	G47	G48	H01	H20	J04	J11	J18	J19	J20	K49	M24	N21	P01	S44	S48	U22

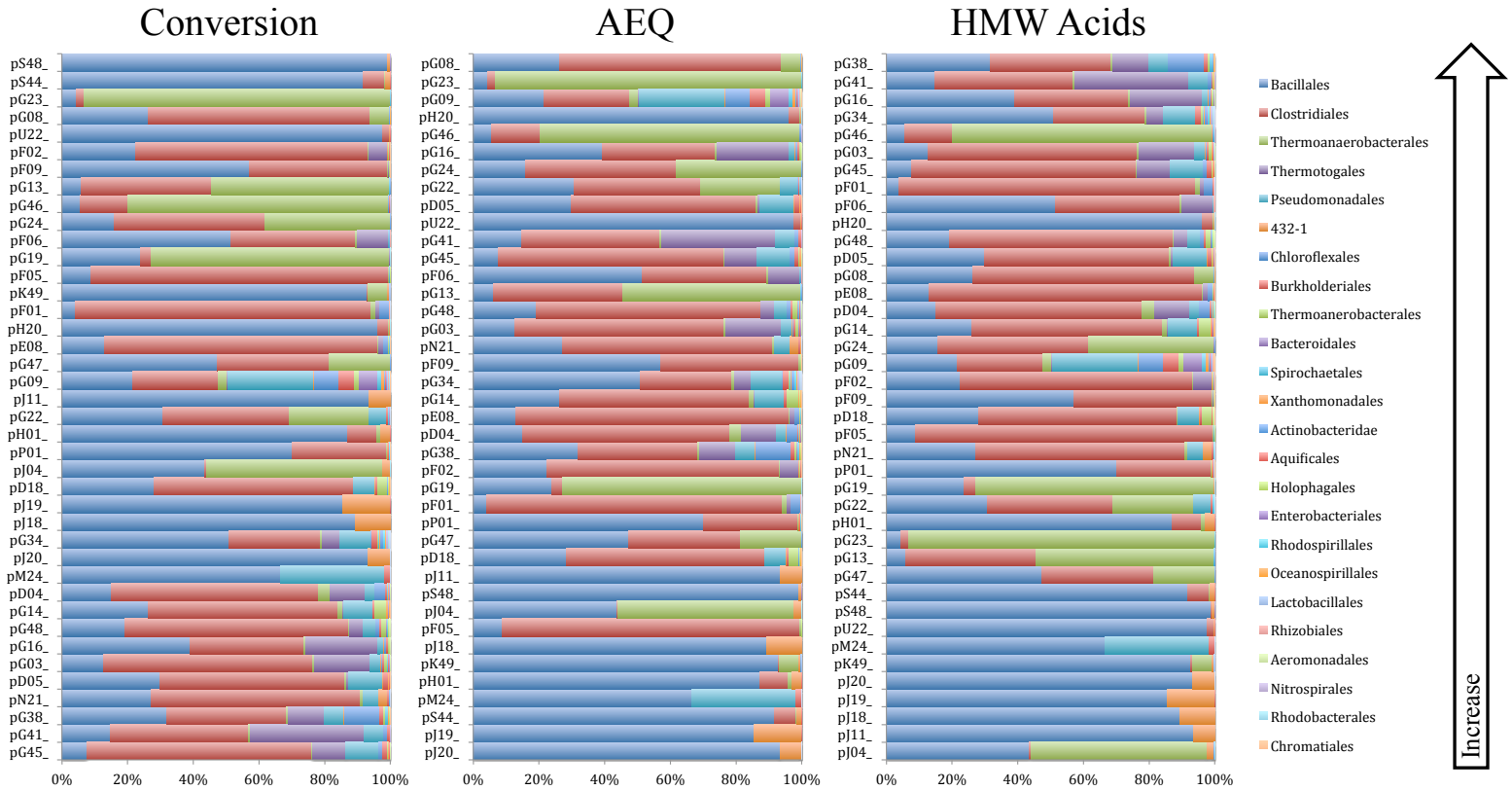
These are not normalized populations.

These are normalized populations.

Appendix B OTU diversity, richness and overlap among 40 samples. Data within the first block of the figure are not normalized. The second block were normalized. The Chao index is a measure of OTU richness. The Shannon index is a measure of OTU diversity. Yue-Clayton theta is a measure of the overlap of OTU composition across samples (ranges from 0 to 1, with 1 = complete similarity in OTU membership and 0 = complete dissimilarity). Based on Yue-Clayton theta similarity from normalized populations at 0.03 dissimilarity.

APPENDIX C

BAR GRAPHS DEPICTING RELATIVE FREQUENCY OF BACTERIAL ORDERS FOR THE 40 SEQUENCED SAMPLES ARRANGED IN THREE WAYS BASED ON RANKING FOR EACH PERFORMANCE PARAMETER OF INTEREST



Appendix C Bar graphs depicting relative frequency of bacterial orders for the 40 sequenced samples arranged in three ways based on ranking for each performance parameter of interest. Identical bar graphs for each community are arranged by rank (top = best performance) for conversion (left), AEQ (middle), and HMW (right). Abbreviations: Acetic acid equivalence concentration (AEQ), high molecular weight (HMW) acid production.

APPENDIX D

OTU DIVERSITY AND RICHNESS FOR THREE CARBOXYLATE BIOFUEL PLATFORM SCREEN COMPARISONS EXPERIMENT

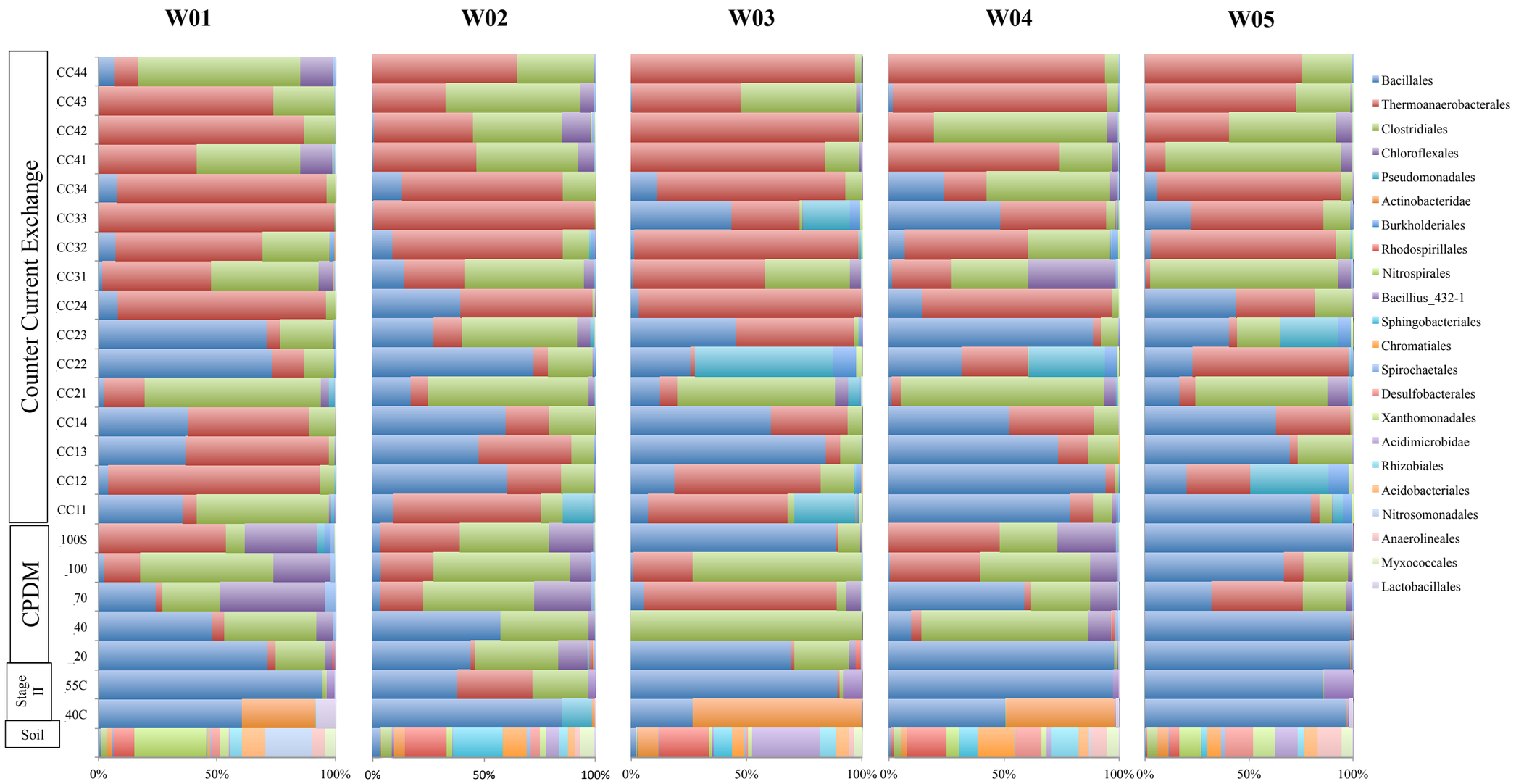
Planned Brazoria	Sample	# Seqs	# OTUs	Chao 1	Shannon Index	Sample	# Seqs	# OTUs	Chao 1	Shannon Index	Sample	# Seqs	# OTUs	Chao 1	Shannon Index	Sample	# Seqs	# OTUs	Chao 1	Shannon Index	Sample	# Seqs	# OTUs	Chao 1	Shannon Index
Soil	Bra25	5707	1281	3721.47	5.21	Bra55	6097	2267	5348.55	6.92	Bra65	1952	855	2105.57	5.92	Bra95	4270	1629	5112.95	6.46	GAL	4666	2179	6310.04	7.06
Stage II 55C	W01_55C	3683	42	96.17	1.07	W02_55C	1016	30	56.25	1.92	W03_55C	7716	58	128.13	1.17	W04_55C	6114	18	40.50	0.45	W05_55C	1919	26	56.33	1.51
Stage II 40C	W01_40C	6728	89	159.91	1.83	W02_40C	4349	44	68.43	1.01	W03_40C	1767	34	76.00	0.98	W04_40C	3225	49	88.43	1.67	W05_40C	2584	83	147.50	2.13
CPDM	W01_20	1606	32	72.00	1.24	W02_20	2243	50	76.25	1.93	W03_20	3034	43	127.33	1.15	W04_20	2638	25	38.20	0.26	W05_20	4319	16	23.00	0.12
CPDM	W01_40	2679	41	62.86	1.65	W02_40	2009	32	59.20	1.42	W03_40	3	2	2.00	0.64	W04_40	3073	62	140.75	2.14	W05_40	4232	18	23.00	0.29
CPDM	W01_70	1896	32	83.00	1.38	W02_70	2997	37	62.50	1.51	W03_70	7272	35	46.00	0.80	W04_70	3101	36	70.20	1.37	W05_70	3618	36	58.75	1.78
CPDM	W01_100	3197	54	100.20	1.95	W02_100	3626	36	58.75	1.68	W03_100	6432	45	87.00	1.29	W04_100	5869	46	67.11	1.76	W05_100	2819	33	63.00	1.95
CPDM	W01_100S	1402	29	44.17	1.33	W02_100S	2219	28	37.00	1.59	W03_100S	4212	34	68.00	0.72	W04_100S	2302	22	25.00	1.27	W05_100S	4271	16	34.00	0.28
Counter Current	W01_CC11	1765	38	74.14	1.89	W02_CC11	2306	38	76.25	1.38	W03_CC11	1537	25	43.33	1.23	W04_CC11	2657	41	69.88	1.15	W05_CC11	1410	45	95.00	1.77
Counter Current	W01_CC12	1913	22	33.25	0.57	W02_CC12	3356	51	76.67	1.42	W03_CC12	1739	36	99.33	1.31	W04_CC12	2562	21	30.00	0.69	W05_CC12	4955	42	57.00	1.55
Counter Current	W01_CC13	2248	33	63.00	1.30	W02_CC13	2079	38	47.75	1.43	W03_CC13	1975	29	45.50	1.03	W04_CC13	12345	108	187.50	1.56	W05_CC13	9327	111	188.00	1.55
Counter Current	W01_CC14	2580	40	70.60	1.49	W02_CC14	1619	37	54.14	1.54	W03_CC14	2457	27	46.50	1.08	W04_CC14	2047	33	53.00	1.40	W05_CC14	12727	78	126.46	1.30
Counter Current	W01_CC21	2591	39	123.33	1.50	W02_CC21	4067	37	71.00	1.54	W03_CC21	2047	58	113.11	2.21	W04_CC21	1604	33	138.00	1.35	W05_CC21	2271	61	141.14	2.17
Counter Current	W01_CC22	2962	54	69.00	1.47	W02_CC22	1511	38	58.00	1.43	W03_CC22	4008	49	91.86	1.37	W04_CC22	5273	55	80.67	1.66	W05_CC22	2183	23	50.50	0.81
Counter Current	W01_CC23	1540	33	37.00	1.71	W02_CC23	6178	73	147.00	2.12	W03_CC23	1473	31	61.00	1.39	W04_CC23	11471	78	160.50	0.88	W05_CC23	3893	64	100.91	2.23
Counter Current	W01_CC24	1229	21	33.00	0.68	W02_CC24	1874	35	60.50	1.30	W03_CC24	2679	14	59.00	0.20	W04_CC24	2156	25	44.50	0.71	W05_CC24	1270	31	65.00	1.82
Counter Current	W01_CC31	4235	54	135.20	1.62	W02_CC31	3123	48	186.00	1.90	W03_CC31	4862	48	98.00	1.37	W04_CC31	1843	48	85.50	1.75	W05_CC31	2649	43	93.60	1.54
Counter Current	W01_CC32	1850	47	89.86	1.53	W02_CC32	1656	49	92.88	1.16	W03_CC32	1979	12	15.33	0.20	W04_CC32	1189	39	91.50	1.62	W05_CC32	2501	34	79.33	0.64
Counter Current	W01_CC33	3415	15	22.00	0.10	W02_CC33	2907	8	14.00	0.07	W03_CC33	5726	75	133.58	1.75	W04_CC33	5269	81	193.88	1.57	W05_CC33	1301	27	53.00	1.27
Counter Current	W01_CC34	2109	30	49.43	0.57	W02_CC34	1641	28	54.25	1.13	W03_CC34	3965	38	65.14	0.83	W04_CC34	1565	60	111.67	2.21	W05_CC34	2766	21	60.00	0.57
Counter Current	W01_CC41	4063	50	87.50	1.72	W02_CC41	5443	51	90.43	1.50	W03_CC41	6015	37	72.00	0.76	W04_CC41	3931	41	68.14	1.11	W05_CC41	2169	46	85.43	1.57
Counter Current	W01_CC42	2332	20	29.00	0.58	W02_CC42	1967	40	59.00	1.65	W03_CC42	2163	15	33.00	0.13	W04_CC42	2933	57	119.14	1.62	W05_CC42	2374	49	99.14	1.70
Counter Current	W01_CC43	2033	23	53.33	0.95	W02_CC43	2458	30	52.75	1.23	W03_CC43	8909	58	116.13	1.38	W04_CC43	6015	40	74.20	0.42	W05_CC43	1615	25	44.50	0.99
Counter Current	W01_CC44	1462	49	85.14	2.24	W02_CC44	2299	19	26.00	0.85	W03_CC44	3362	16	26.50	0.22	W04_CC44	3307	20	42.50	0.32	W05_CC44	1929	16	21.00	0.85

All 402076

Appendix D OTU diversity and richness for three carboxylate biofuel platform screen comparisons experiment. The Chao 1 richness estimator is an estimation of richness. The Shannon Index is a measure of OTU diversity. Each category is separately scaled from high values (green) to low values (red). Those samples highlighted in grey were removed from further analysis due to low sequencing returns after quality checking. For a detailed explanation of sample name interpretation see Figure 11.

APPENDIX E

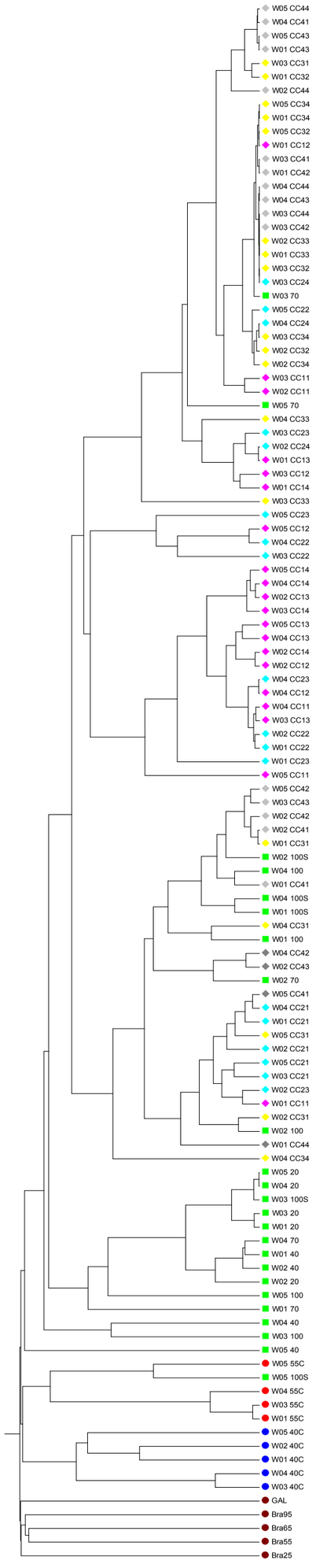
BAR GRAPHS SHOWING THE RELATIVE FREQUENCY OF THE TOP 22 MOST COMMON ORDER LEVEL CLASSIFICATIONS ACROSS ALL SAMPLES IN THE CBP SCREEN COMPARISON EXPERIMENT



Appendix E Bar graphs showing the relative frequency of the top 22 most common order level classifications across all samples sequenced in the carboxylate biofuel platform screen comparison experiment. W01 is a freshwater marsh sediment sample, W02 is a saltwater marsh sediment sample, W03 is a costal prairie soil sample, W04 is a salt lake sediment sample, and W05 is an intertidal sediment sample.

APPENDIX F

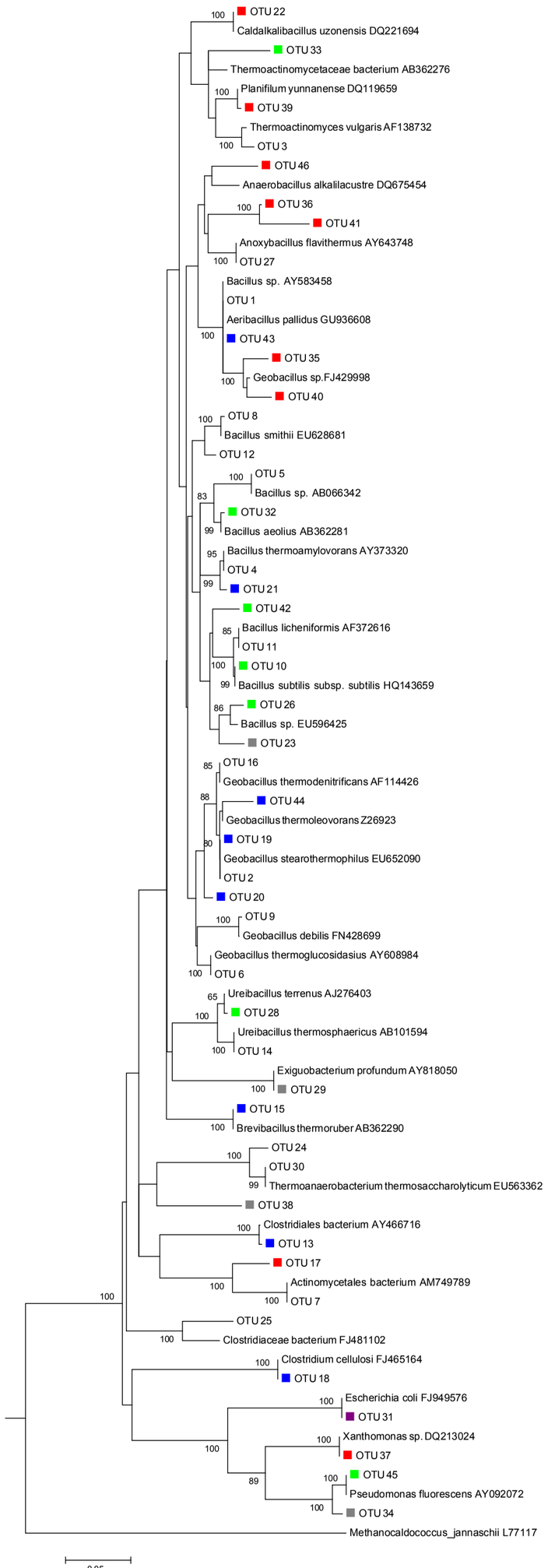
DENDOGRAM OF ALL SAMPLES BASED ON YUE-CLAYTON THETA DIVERSITY SIMILARITY MEASURE



Appendix F Dendogram of all samples based on Yue-Clayton theta diversity similarity measure. Samples are colored separately soils brown; 30-day screens 40C dark blue, 55C red; CPDMs green; counter current trains by collection point 1st magenta, 2nd light blue, 3rd yellow, and 4th grey. Bra25 or W01 is a freshwater marsh sediment sample, Bra55 and W02 is a saltwater marsh sediment sample, Bra65 and W03 is a costal prairie soil sample, Bra95 and W04 is a salt lake sediment sample, and GAL and W05 is an intertidal sediment sample.

APPENDIX G

PHYLOGENETIC TREE OF ALL REPRESENTATIVE OPERATIONAL TAXONOMIC UNITS AND CLOSEST SEQUENCES WITHIN RIBOSOMAL DATABASE PROJECT SEQMATCH AT 97% SIMILARITY



Appendix G Phylogenetic tree of all representative operational taxonomic units (OTUs) and closest sequences within Ribosomal Database Project (RDP) SeqMatch (accessed 11/11/11) at 97% similarity. Tree constructed using Jukes-Cantor corrected distance matrix modeling and weighted neighbor-joining with bootstrapping at 100 iterations. Colored squares indicate those branches resulting from one media/oxygen incubation type: cellulose/ anaerobic chamber (grey), cellulose (blue), acetogen (red), halophile (green), other (purple).