Western Graduate&PostdoctoralStudies

Western University

Scholarship@Western

Electronic Thesis and Dissertation Repository

5-26-2015 12:00 AM

Genomic analyses of Paenibacillus polymyxa CR1, a bacterium with potential applications in biomass degradation and biofuel production

Alexander W. Eastman
The University of Western Ontario

Supervisor
Dr. Ze-Chun Yuan
The University of Western Ontario

Graduate Program in Microbiology and Immunology
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science
© Alexander W. Eastman 2015

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Other Microbiology Commons

Recommended Citation

Eastman, Alexander W., "Genomic analyses of Paenibacillus polymyxa CR1, a bacterium with potential applications in biomass degradation and biofuel production" (2015). *Electronic Thesis and Dissertation Repository*. 2873.
https://ir.lib.uwo.ca/etd/2873

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Genomic analyses of *Paenibacillus polymyxa* CR1, a bacterium with potential applications in biomass degradation and biofuel production (Thesis format: Monograph)

by

Alexander William Eastman

Graduate Program in Microbiology and Immunology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Masters of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Alexander William Eastman 2015

Abstract

Lignin is a polyphenolic heteropolymer constituting between 18 to 35% of lignocellulose and is recognized as preventative of cellulosic biofuel commercialization. *Paenibacillus polymyxa* CR1 was isolated from naturally degrading corn stover and shown to produce alcohols using lignin as a sole carbon source. Genome sequencing and comparative genomics of *P. polymyxa* CR1 identified two homologs, a Dyp-type peroxidase and a laccase, which have previously been implicated in lignin metabolism in other bacteria. Knockout mutants of the identified genes displayed no growth deficiency and *P. polymyxa* CR1 is incapable of metabolizing common aromatic intermediates of lignin, suggesting the bacterium employs a novel catabolic pathway. To identify genes involved in lignin metabolism, a transposon library was generated and screened for abnormal lignin growth phenotypes. The results contained within will help elucidate the genetic basis of known functions helping delineate regulatory pathways and metabolic versatility in *P. polymyxa* relevant to lignin metabolism.

Keywords

Sequencing, genomics, genetics, comparative genomics, lignocellulose, lignin, metabolism, biofuels, renewable energy, finishing, transposon mutagenesis, ribosomal RNA

Co-Authorship Statements

Complete Genome Sequence of *Paenibacillus polymyxa* CR1, a Plant Growth-Promoting Bacterium Isolated from the Corn Rhizosphere Exhibiting Potential for Biocontrol, Biomass Degradation, and Biofuel Production.

Alexander W. Eastman [1,2], Brian Weselowski [1], Naeem Nathoo [1,3], and Ze-Chun Yuan[1,2]

[1] Southern Crop Protection and Food Research Centre, Agriculture and Agri-Food Canada, Government of Canada, London, Ontario, Canada N5V 4T3
[2] Western University, Department of Microbiology and Immunology, Schulich School of Medicine and Dentistry, London, Ontario, Canada N6A 5C1
[3]Western University, Department of Biology, London, Ontario, Canada N6A 5B7

Portions of this work have been published in *Genome Announcements* 2014, **2**(1)**:** e01218-13

Brian Weselowski aided in performing experiments. Naeem Nathoo aided in analysis of scaffolding and *in silico* alignments. Dr. Yuan contributed to editing of the manuscript.

Development and validation of an rDNA operon based primer walking strategy applicable to *de novo* bacterial genome finishing.

Alexander W. Eastman [1,2], and Ze-Chun Yuan [1,2]

[1] Southern Crop Protection and Food Research Centre, Agriculture and Agri-Food Canada, Government of Canada, London, Ontario, Canada N5V 4T3
[2] Western University, Department of Microbiology and Immunology, Schulich School of Medicine and Dentistry, London, Ontario, Canada N6A 5C1

Portions of this work has been published in *Frontiers in Microbiology* 2015, **7**:769

Dr. Yuan contributed to the design of the experiments and editing of the manuscript.

Comparative and genetic analysis of the four sequenced *Paenibacillus polymyxa* genomes reveals a diverse metabolism and conservation of genes relevant to plant-growth promotion and competitiveness.

Alexander W. Eastman [1,2], David E Heinrichs [2], and Ze-Chun Yuan [1,2]

[1] Southern Crop Protection and Food Research Centre, Agriculture and Agri-Food Canada, Government of Canada, London, Ontario, Canada N5V 4T3
[2] Western University, Department of Microbiology and Immunology, Schulich School of Medicine and Dentistry, London, Ontario, Canada N6A 5C1

Portions of this work have been published in *BMC Genomics* 2014, **15**:851

Dr. Yuan contributed to the conceptualization of the research and the writing and editing of the manuscript.
Dr. Heinrichs supervised and edited the writing of the manuscript.

Acknowledgments

I would like to express my gratitude to my co-supervisors Dr. Ze-Chun Yuan and Dr. David Heinrichs who have always been available for guidance and support over the course of my research.

I would also like to thank the members of my committee Dr. Martin McGavin and Dr. Yuhai Cui for their invaluable insights into limitations and areas in need of clarification and focus in my work, as well their time and efforts put into reading and editing my thesis. A special thanks goes out to Alex Molnar for his work on figures.

I would like to thank members of my immediate lab group, Naeem Nathoo and Brian Weselowski who have helped maintain my sanity and dedication at various times and made the long hours at the lab seem like mere minutes.

Finally I would like to thank my friends and family, with expressive thanks to Shilpa Goel, for enduring my endless ranting and incessant science both when things we running smoothly and when experiments were inevitably failing.

List of Abbreviations

ABC transporter- ATP binding cassette transporter

ATP- adenosine triphosphate

BLAST- basic local alignment search tool

Bp- base pairs

CDS- coding sequences

CO_2- carbon dioxide

COG- cluster of orthologous groups

Contig – contiguous DNA sequence

DNA – deoxyribonucleic acid

DyP- dye-decolourizing peroxidase

GC- gas chromatography

GH- glycoside hydrolase

GHG- greenhouse gas

Kan- kanamycin sulphate

KO- KEGG orthology

LB- lysogeny broth

LCB- local collinear block

Lig- Kraft lignin

LiP- lignin peroxidase

MFS- major facilitator superfamily

MM- minimal media

Mn- manganese

NCBI- National Centre for Bioinformatics

PCR- polymerase chain reaction

PTS- phosphor transferase system

rDNA- DNA region encoding ribosomal subunit genes

RNA- ribonucleic acid

rRNA- mature ribosomal subunit RNA

TAE- tris base, acetic acid ethylenediaminetetraacetic acid

Tet- tetracycline

Tn- transposon

tRNA- transfer RNA

US- United States of America

VP-versatile peroxidase

G- giga (10^9)

M- mega (10^6)

k-kilo (10^3)

m- milli (10^{-3})

n- nano (10^{-6})

p- pico (10^{-9})

Pa- Pascal

M- molar (mol/L)

g- gram

L- litre

Table of Contents

Abstract i

Co-Authorship Statements ii

Acknowledgments iv

Table of Contents vii

List of Tables 2

List of Figures 3

List of Appendices 5

Chapter 1 – Introduction 6

- **Biofuels History** 6
- **First-Generation Biofuels** 7
- **Cellulosic Biofuels** 9
- **Lignocellulose** 12
  - Lignin Structure and Synthesis 16
  - Current Limitations of Lignin on Cellulosic Biofuel 19
- **Chemical and Industrial Delignification Processes** 20
  - Acid or Alkali Hydrolysis 20
  - Ionic Liquids 21
  - Steam-Explosion 22
  - Kraft Process 22
  - Organosolv Lignin 24
- **Biological Delignification Processes** 26
  - Fungal Metabolism 26
  - Bacterial Lignin Metabolism 29
- **Features of Paenibacillus polymyxa** 33

Chapter 2 - Materials and Methods 38

- **Characterization of *Paenibacillus polymyxa* CR1** 38
  - Isolation and Plant-Derived Carbon Metabolism 38

16S Ribosomal Subunit Sequencing 39

Genome Sequencing of *Paenibacillus polymyxa* CR1 .... **40**

Phylogenetic Analyses 41

Genome Assembly of *Paenibacillus polymyxa* CR1 .... **42**

Draft Genome Assembly 42

Contig reduction and Draft Genome Scaffolding 42

Targeted Sequencing of Ambiguous Bases 43

Contig Gap Closure 44

Investigation of Rearrangements 45

Genome Annotation 46

Genome Annotation and Analyses 46

Genomic Island Identification 47

General Comparative Genomics 47

Insertional Gene Knockouts 48

Transposon Mutagenesis 51

Phenotypic Characterization 52

Identification of Disrupted Genes 53

Nucleotide Accession Numbers 54

Chapter 3 – Results **55**

Isolation and Characterization of *P. polymyxa* CR1 .... **55**

Genome Sequencing of *P. polymyxa* CR1 57

Novel Gap Closure Method 59

Genome Annotation 64

Comparative Genomics of *P. polymyxa* Strains **65**

General Features 65

Phylogeny **72**

Plant-Derived Carbon Metabolism **75**

Functional Genomics **82**

Targeted Knockouts 83

Transposon Mutagenesis 85

Chapter 4 – Discussion **89**

Significance **100**

References **103**

Appendices **123**

Curriculum Vitae **146**

List of Tables

Table 1. United States of America Environmental Protection Agency renewable fuel classifications. 6

Table 2. Comparisons of greenhouse gas emission reduction of biofuels from various feedstock. 12

Table 3. Ribosomal subunit 16S and 23S primers. 39

Table 4. Primers for confirming insertions in pUCP30T. 51

Table 5. General genome features of completely sequenced *Paenibacillus polymyxa* strains 66

Table 6. Plant-growth promoting traits of sequenced *P. polymyxa* strains. 76

Table 7. Putative ligninolytic enzymes of sequenced *Paenibacillus polymyxa* strains. 81

List of Figures

Figure 1. Schematic depicting lignocellulose structure and components. 13

Figure 2. Most Common Monolignol Monomers in Nature. 17

Figure 3. Comparison of Lignocellulose Component Streams from Kraft Process Versus Organosolv Process. 24

Figure 4. Common Fungal Delignification Pathways.. 28

Figure 5. Lignin Metabolism Network in *Sphingomonas paucimobilis* SYK-6.............. 30

Figure 6. Plasmid Map of pUCP30T. 49

Figure 7. *P. polymyxa* CR1 is capable of metabolizing a wide range of plant-derived carbon sources... 56

Figure 8. Gas-chromatography of lignin fermenting *P. polymyxa* CR1........................... 57

Figure 9. Phred-like quality score for *P. polymyxa* genome sequencing......................... 58

Figure 10. Comparison of scaffolding against closest related strain versus species. 61

Figure 11. Confirmation of rRNA operons within contig gaps. 62

Figure 12. Identification and resolution of rearrangement in draft contig assembly........ 64

Figure 13. Global alignment of chromosomes of completely sequenced *P. polymyxa* strains. 68

Figure 14. COG functional categorization of sequenced *Paenibacillus polymyxa* genomes. 69

Figure 15. Circular representation of the *P. polymyxa* CR1 genome. 71

Figure 16. Phylogenetic tree of completely sequenced *Paenibacillus polymyxa* strains.. 73

Figure 17. Comparison of *Paenibacillus polymyxa* CR1 *nif* cluster to other free-living diazotrophic bacteria. 74

Figure 19. *P. polymyxa* CR1 cannot utilize protocatechuate as a sole carbon source. 82

Figure 20. Growth curve of *P. polymyxa* CR1 deletion mutants. 85

Figure 21. Representative growth curves of fast growing *P. polymyxa* CR1 Tn5 mutants. 87

Figure 22. Representative growth curves of slow-growing *P. polymyxa* CR1 Tn5 mutants. 88

List of Appendices

Appendix 1. Locations of genomic islands and encoded genes... 123

Appendix 2 Insertion sequences identified in *P. polymyxa* CR1 133

Appendix 3. Identified antimicrobial compounds encoded by *P. polymyxa* CR1 133

Appendix 4. Transporter classification for sequenced *P. polymyxa* strains. 138

Appendix 5. CaZY profile of sequenced *P. polymyxa* strains. 141

Chapter 1 – Introduction

Biofuels History

As nations continue to invest in sustainable and renewable energy production, the interest in biofuels continues to rise, in part a result of their feasibility using current technologies. Both the European Union and the United States of America have put forward ambitious timelines to increase domestic biofuel usage, thereby facilitating energy independence and reducing greenhouse gas production (Sorda et al., 2010). The expansion of the biofuel sector is predicted to continue as China and other rapidly developing countries continue to set ambitious targets for renewable energy consumption to accommodate their rapidly expanding energy needs (Zhou and Thomson, 2015). Expansion of the North American biofuel industry has been steady, punctuated by dramatic jumps as individual regions mandate a minimum ethanol blend in consumer gasoline.

Table 1. United States of America Environmental Protection Agency renewable fuel classifications. Required greenhouse gas emission reductions for biofuels to be classified into various categories. Bioethanol from sugarcane is considered to be a first-generation biofuel despite its inclusion in the "Advanced biofuel" category.

Fuel Standard	GHG Reduction	Feedstock
Bioethanol	20%	Corn starch
Biobutanol	20%	Corn starch
Biodiesel	50%	Soy oil, fats, grease, algal oils
Advanced biofuel	50%	Sugarcane
Cellulosic biofuel	60%	Crop/forestry residues, cover crops, perennial grasses, consumer waste

First-Generation Biofuels

The increased demand for ethanol resulting from its use in fuels, has necessitated the construction of a large number of first-generation ethanol plants in both North America and Brazil. Although the feedstock for fermentation differs depending on regional availabilities, the underlying theory and basis of first-generation biofuels is the same regardless of the substrate used (Naik et al., 2010; Pimentel and Patzek, 2005). In first-generation ethanol manufactories, simple sugars and starches are broken down into constituent mono and di-saccharides (Bai et al., 2008). Yeasts, most commonly *Saccharomyces cerevisae*, ferment these sugars into ethanol and CO_2. The breakdown of starches or sugars for biofuel production is most commonly performed by either acid-hydrolysis or enzymatic pretreatment (Sanchez and Cardona, 2008). In North America, market scale first-generation ethanol plants use a valuable commodity, corn (*Zea mays*), as a feedstock (Naik et al., 2010; Sanchez and Cardona, 2008). One of the macro-economic effects of the expansion of the biofuel industry is the dramatic increase in corn commodity prices, roughly corresponding to the development and expansion of the North American ethanol market (Fortenbery and Park, 2008). The diversion of maize from livestock feed to ethanol production was the major factor contributing to an increase in the raw commodity price of maize from 2005 to 2008, contrasting with a continuous reduction in corn prices over the preceding three decades.

Industry experts, economists and government organizations agree that first-generation ethanol production methods do not have the potential capacity to replace petroleum fuels with our current energy demand. Research has determined that if first-generation biofuels were to expand to the scale necessary to replace gasoline in motor vehicles, food-producing agricultural land would need to be transitioned to maize production for biofuel feedstock (Searchinger et al., 2008; Fargione et al., 2008). This challenge has fueled the most major area of discourse against the adoption of biofuels, commonly known as the "food versus fuel" debate. The argument follows that if fuel production uses the same substrate as common food goods, the competition between the two markets will inevitably lead to increases in food pricing, thereby driving up the cost of living. However, mounting research demonstrates that maize based ethanol production does not necessitate competition with food production. There exists major disparity in the public's understanding of the percentages of food costs that derive from production costs versus those that stem from downstream consumer costs such as packing, transport and retail costs (Canning, 2011). The fear of increased food prices is a consequence of the large proportion of processed food consumed in the Western diet that contains high-fructose corn syrup or other corn-derivatives. Consumer groups fear that increased ethanol production would cause a concomitant price increase in a wide variety of food products. However, research by the US Department of Agriculture demonstrates that the production costs of common food crops (including corn) only account for approximately 10% of consumer food purchasing price (Canning, 2011). Therefore, even in an extreme situation where the price of corn doubles or triples, there is not a mirrored price increase in corn-derived products. Even with the corn commodity price increases between 2005 and 2008 resulting from first-generation ethanol production, a corresponding increase in corn-derived products has not been seen. Furthermore, recent increases in food prices are better correlated to increased gasoline and transportation costs rather than decreased available agricultural land or increased corn prices (Baffes and Dennis, 2013).

However, addressing concerns with first generation biofuels and the concerns of consumer groups stands to increase sustainability and supply issues of biofuels by instead producing ethanol from bulk plant biomass. The production of ethanol from plant biomass (lignocellulose) has been termed the “second-generation” of biofuels or more appropriately, cellulosic biofuels (Pimentel and Patzek, 2005; Naik et al., 2010; Bensah and Mensah, 2013; Li et al., 2010).

Cellulosic Biofuels

Cellulosic biofuels aim to hydrolyze bulk plant biomass to generate usable alcohols, in contrast to monosaccharide-dense feedstock used in first-generation biofuels (Pimentel and Patzek, 2005; Naik et al., 2010). Plant biomass is abundant in current North American agricultural practices in the form of otherwise low-value wastes that are underutilized or completely unused by manufacturing, agricultural or industrial streams. Plant biomass can be obtained in the form of wheat straw, corn stover, grass clippings, forestry residues and specially grown biofuel crops (Sanchez and Cardona, 2008; Pimentel and Patzek, 2005; Naik et al., 2010). Specialty biofuel crops can be cultivated on lands currently unsuitable for growth of current agricultural products, which have been shown to sustainably produce high yields of biomass-specific crops such as *Miscanthus x giganteus* or *Panicum virgatum* (switch grass) (Bensah and Mensah, 2013; Naik et al., 2010). Additionally, as corporations and the public at large continue trends towards a paperless economy, poplar grown on former forestry land have been shown to be able to provide large quantities of rapidly growing sustainable biomass (Littlewood et al., 2014).

A major force driving the development of cellulosic technologies are greenhouse gas (GHG) emissions concerns of first-generation biofuels (Slade et al., 2009). As governments plan long-term solutions to meet national energy demands, major challenges concerning the viability of alternative energy will continue to be a divisive issue. Currently, North American agriculture uses the majority of arable land for either food or livestock feed production. Ideally, for maximal sustainability, there needs to be a minimal effect on the global food supply caused by mandated adoption of biofuels. A major advantage of cellulosic biofuels is their high-energy yield/acre when compared to first-generation fuels, low carbon footprint in comparison to other alternative fuels and the ability to use low-value agricultural waste residues as a feedstock (Sanchez and Cardona, 2008).

Many cellulosic biofuel pathways aim to utilize otherwise waste products of agriculture, thereby reducing the potential negative effects of widespread direct land-use changes. However more complex land use considerations become a concern when such large-scale infrastructure developments are proposed. Indirect-land-use changes refer to the clearing of new lands for agriculture to replace those lands which products are diverted to biofuel production (Searchinger et al., 2008; Fargione et al., 2008). Many cellulosic biofuels are produced using agricultural residues from commodity crops that are otherwise discarded, such as corn stover or wheat straw, and as such have no indirect land-use change associated with them. In North America, the quantity of farmed agricultural lands has remained static since the late 1980s (Lambin et al., 2001). As such, the majority of estimates for indirect-land-use changes are concerned with Brazil and other developing nations. In rapidly expanding agricultural systems, large areas of rain forest or densely vegetated grasslands and marshes are cleared for use as farmlands via burning of biomass, resulting in the immediate release of carbon sequestered in dense vegetation.

Estimates of biofuel impacts by the US government use what is called a 30-year payback period that incorporates concerns about indirect-land-use changes (Kim and Dale, 2005). The 30-year payback period accounts for the estimated 25% release of all carbon contained in the converted soils, immediately following the use of the crop cover for biofuel, and the calculation incorporates the subsequent reduction in current greenhouse gas (GHG) production over a 30-year span by the continued use of these lands for biofuel feedstock production (U.S. Enviornmental Protection Agency, 2010). These most recent estimates suggest only a modest decrease in GHG emissions for first-generation biofuels over petroleum fuels, at 7-32% and 52-72% for corn and sugarcane, respectively (Table 2). Meanwhile, cellulosic ethanol from switchgrass and corn stover results in a much more pronounced GHG emission savings versus gasoline, demonstrating the amicability of further development of these technologies (Table 2).

Table 2. Comparisons of greenhouse gas emission reduction of biofuels from various feedstock. Greenhouse gas (GHG) emission reductions based off of the US Department of Energy estimations for indirect-land-use changes (ILUC). Data obtained from the Environmental Protection Agency Renewable Fuels Standards 2 Report (U.S. Enviornmental Protection Agency, 2010).

Fuel type	GHG Reduction	GHG + ILUC	Assumptions
Corn ethanol	21%	7-32%	Using currently available technology
Sugarcane ethanol	61%	52-72%	Including import emissions
Cellulosic switchgrass	110%	102-117%	Unproductive land use changes
Cellulosic corn stover	129%	129%	No ILUC, exclusively waste feedstock

Lignocellulose

The physical attractiveness of cellulosic biofuels stems from the structure of bulk plant biomass, containing approximately 65-82% hexose and pentose sugars by weight (Sanchez and Cardona, 2008; Naik et al., 2010). Lignocellulose is the major component of the secondary cell wall of plants and account for the majority of mature plant mass. However much of the available hexose and pentose sugars contained in plant biomass is sequestered in long chain polymers. Hydrolyzing and liberating constituent mono- and di-saccharides for fermentation is both technically difficult and expensive, when compared to the hydrolysis of simple starches or sugars (Bensah and Mensah, 2013; Naik et al., 2010; Bai et al., 2008; Sanchez and Cardona, 2008). The pretreatment of plant biomass to release sugars using current methods adds significant costs and accordingly, cellulosic ethanol is substantially more expensive than first-generation biofuels. However, the pretreatment bottleneck affords a potential for immediate savings by addressing limitations and reducing the cost of the expensive pretreatment process. Limited research has investigated the metabolic pathways and chemical processes by which complex plant biomass is degraded and metabolized in the natural environment and researchers are only now beginning to appreciate its complexity (Littlewood et al., 2014; Naik et al., 2010).

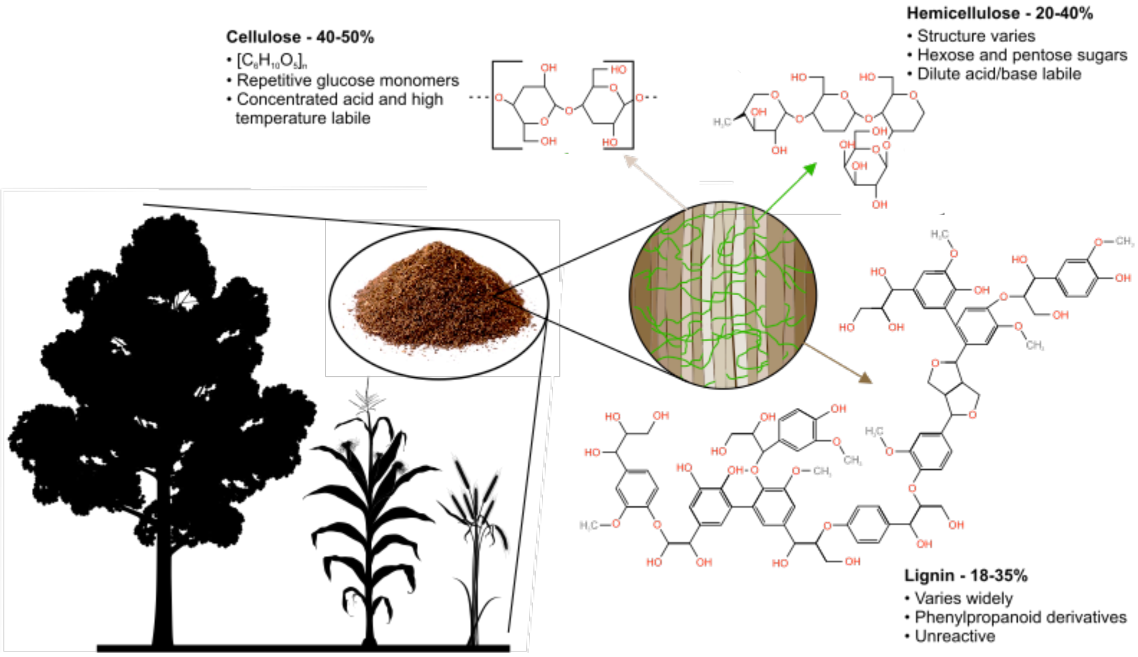

Cellulose - 40-50%
• [C6H10O5]n
• Repetitive glucose monomers
• Concentrated acid and high temperature labile
Hemicellulose - 20-40%
• Structure varies
• Hexose and pentose sugars
• Dilute acid/base labile
Lignin - 18-35%
• Varies widely
• Phenylpropanoid derivatives
• Unreactive

Figure 1. Schematic depicting lignocellulose structure and components. The precise stoichiometry of plant biomass varies between species, especially in terms of hemicellulose and lignin composition. Cellulose is unique in being the only plant component comprised entirely of a single sugar monomer, glucose, contrasting with the variable sugar content of hemicellulose and the variable phenolic units of lignin.

The majority of bulk plant biomass is comprised of lignocellulose, a carbon dense matrix incorporating three different polymers, cellulose, hemi-cellulose, and lignin. Cellulose has a general $(C_6H_{10}O_5)_n$ chemical formula and is a long-chain of repetitive β-1, 4-D-glucose glycosidic linkages (Sullivan, 1997). The repetitive β-1, 4 glycosidic linkages results in long, straight chains of glucose molecules that can form extensive hydrogen bonds with adjacent, cellulose fibrils, which is known as crystalline cellulose. These structures are extremely stable and generate much of the strength of terrestrial plants. Chains of cellulose vary widely in length, dependent on the organism, and can reach upwards of 15 000 repeating glucose units with a molecular weight of 100 000g/mol (Sullivan, 1997). Cellulose is not limited to the plant kingdom and is also a component of various fungal and algal cell walls as well as its role as a common component of bacterial biofilms (Ross et al., 1991; Ude et al., 2006). For chemical cleavage of cellulose, a very high temperature and pressure is necessary to catalyze the crystalline cellulose transition into amorphous, water soluble cellulose.

In contrast, hemi-cellulose is a combination of both hexose and pentose sugars. Hemi-celluloses are branched heteropolymers that incorporate any number of hexose and pentose sugar combinations, with a few types predominating (Scheller and Ulvskov, 2010). Hemicelluloses have been isolated containing almost all of the known D-pentose sugars, with intermittent usage of L-pentose sugars and various D-hexose sugars. The most predominant monosaccharides in plant hemicelluloses are glucose, xylose, mannose, rhamnose, arabinose and galactose. As a result of branching, variable linkages and a random structure, hemi-celluloses are amorphous and easily hydrolyzed by either dilute acids or bases (Yang et al., 2007). In plant biomass, hemicellulose acts to link cellulose fibrils to pectin, found in the primary cell wall, as well as to lignin (Scheller and Ulvskov, 2010) (discussed further in the following section).

Until recently, research on cellulosic biofuels has been directed towards increasing the fermentation efficiency of cellulose and hemi-cellulose by increasing the enzymatic efficiency and engineering better bioreactors (Balachandrababu Malini et al., 2012; Brunecky et al., 2014; Song et al., 2014; Bandounas et al., 2011). In current systems, cellulose is broken down into glucose either chemically, using acid treatments, or enzymatically by cellulases, a diverse category of glycosidic bond cleavage enzymes belonging to a wide variety of glycoside hydrolase (GH) families (Henrissat and Daviest, 1997). Cellulases are either secreted into the extracellular space, or in complexes called cellulosomes (Zhang and Lynd, 2004; Schwarz, 2001). Cellulosomes are extremely variable both between and within organisms and are comprised of any combination of a multitude of uniquely encoded GH family enzymes including endocellulases, exocellulases, mannases, xylanases among many others. Cellulosomes are commonly encoded by anaerobes and facultatively aerobic bacteria where they are either physically attached to the outer membrane or secreted into the extracellular space as a complex. The wide variation in cellulosomes and their constituent subunits has been an area of investigation for structure-directed enzyme engineering of cellulases and hemi-cellulases degradation for *in vitro* use in biomass pretreatment (Papoutsakis, 2008; Mazzoli et al., 2012). Cellulases are commonly encoded by soil dwelling fungi and bacteria and are especially prevalent in organisms that survive on degrading plant tissues, known as saprotrophs (López-Guerrero et al., 2013). Saprotrophs are widespread in soils where decaying plant matter acts as an oasis of readily available carbon, energy and nutrients in otherwise nutrient limited soil.

Pentose sugars derived from hemi-celluloses, such as xylose and arabinose, are fermentation inhibitors that are poorly metabolized by *Saccharomyces cerevisiae,* the most commonly used yeast in biofuel alcohol fermentation (Klinke et al., 2004). This has lead to genetic engineering of the pentose phosphate pathway in *S. cerevisiae,* increasing expression of xylose isomerase, thus facilitating co-fermentation of xylose and glucose, as well as development and optimization of other yeast species for combinatorial hemi-cellulose/cellulose metabolism (Hahn-Hägerdal et al., 2007; Kricka et al., 2014).

Researchers are continually developing ameliorated methods for the deconstruction and subsequent fermentation of celluloses and hemicellulose, through pre-treatments of plant biomass, cellulosome/enzyme development, and genetic engineering of pentose fermenting yeast. However research has continually neglected the potential uses of the lignin component of lignocellulose, as both an energy source and as an intermediate in the development of other value-added products.

Lignin Structure and Synthesis

Despite its central role in lignocellulose, the structure, production and catabolism of lignin is poorly understood (Martínez et al., 2009; Boerjan et al., 2003). Lignin is a polyphenolic heteropolymer comprised of cinnamic alcohol derivatives, called monolignols, that coat and covalently link hemi-cellulose and cellulose molecules together (Boerjan et al., 2003). The phenolic, and by extension hydrophobic, nature of lignin fills in the spaces between cellulose fibrils, excluding access of solvents and water from the lignocellulose matrix. Lignin provides structural support to terrestrial plants and allows xylem cells to conduct water from plant roots to distal plant tissues. Although comprising between 18-35% of plant biomass, researchers have only recently begun to appreciate its limiting role in cellulosic ethanol production (Zeng et al., 2014; Sannigrahi et al., 2010; Doherty et al., 2011). Lignin is extremely recalcitrant to chemical, biochemical and physical degradation, and accounts for a significant proportion of costs associated with cellulosic biofuels (Zeng et al., 2014; Sannigrahi et al., 2010). Due to its pervasiveness throughout the secondary cell wall, in combination with its hydrophobicity and lack of reactivity, lignin acts as a shield for reactive chains of cellulose and hemicellulose occluding the access of enzymes and acids (Boerjan et al., 2003).

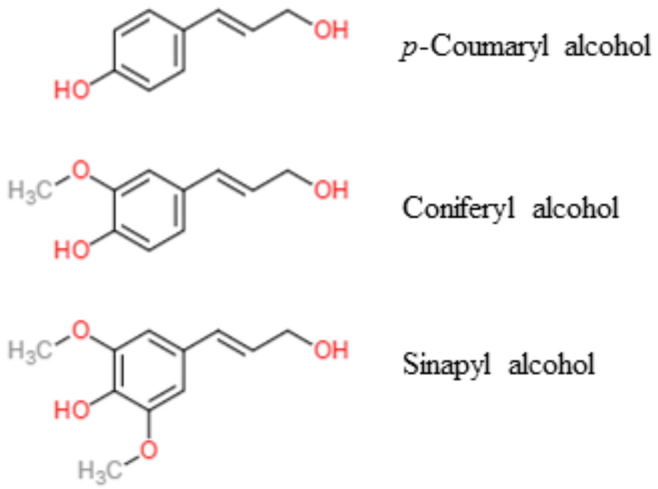


Figure 2. Most Common Monolignol Monomers in Nature. Although mature lignin is a heteropolymer with no defined secondary structure, cells synthesize monolignols as glycosides which are exported prior to polymerization

The prevailing theory for the heterogeneity of the lignin matrix was that lignin deposition is not under tight control in the plant secondary cell wall and deposition is controlled only at the subunit synthesis and export levels. This is seemingly supported by the observation that plant lignin recycling is rare and many plants do not encode the enzymes necessary to recycle mature lignin once monolignols are exported and polymerized. The first published mechanisms detailing lignin synthesis postulated polymerization occurs through a resonance-stabilized monolignol radical, with experiments demonstrating that similar structures can be generated *in vitro* using synthetic monolignols (Freudenburg and Neish, 1968). However, various linkages and bonds not found in nature were present in the synthetic lignin, and ratios of the types of linkages found in nature could not be obtained, suggesting at least nominal control over polymerization. However, further research has challenged the uncontrolled polymerization model demonstrating that plants actually exert control over lignin matrix deposition, structure and recycling through a wide range of regulatory mechanisms (Boerjan et al., 2003). Research has shown that monolignols are exported as glucosides to increase their solubility and once past the plant cell wall the sugar is cleaved and monolignols are polymerized through a unique and complex series of redox reactions that are yet to be fully elucidated (Boerjan et al., 2003). Regardless of the levels and types of control of lignin depositions in the secondary cell wall, researchers agree that the structure is not as tightly controlled or homogenous as cellulose or hemi-cellulose, and lignin offers a diverse set of side-chains and subunits available for interaction with solvents and enzymes.

Various applications for lignin are currently being investigated including use as an adhesive, an industrial heat insulator or as a precursor for other valuable chemicals (Doherty et al., 2011; Sannigrahi et al., 2010). Although potential applications are promising in terms of their efficacy, no existing process or product has the potential to utilize the entire lignin stream that would be generated by industrial scale cellulosic biofuel production. For a potential application of waste lignin to increase the overall economic viability of cellulosic biofuels, a large market of lignin-derived products is needed. Use of lignin as a feedstock for further fuel or chemical production is ideal since cellulosic biofuel production would not be dependent on co-expansion of two diverse markets to ensure economic feasibility.

Current Limitations of Lignin on Cellulosic Biofuel

The heterogeneity, pervasiveness and recalcitrance of the lignin component of lignocellulose is currently a major limiting factor preventing economical development of cellulosic biofuels (U.S. Department of Energy, 2012). Lignin coats and occupies the secondary cell wall forming an unreactive shell around valuable cellulose and hemi-cellulose fractions of lignocellulose thereby sterically inhibiting chemical and biochemical degradation of these fractions. As a result of its low reactivity, current treatment options targeting lignin are either economically or environmentally prohibitive for the scale needed for commodity fuel production. Current technologies that approach economic feasibility do not fully exploit the inherent value of the lignin constituent of lignocellulose (Doherty et al., 2011; Sannigrahi et al., 2010). Therefore, new technologies and processes should either reduce costs associated with lignin removal or provide downstream value to the separated lignin fractions. These ameliorations have the potential to significantly increase the viability of cellulosic biofuels without serious modifications to existing infrastructure and processes (Doherty et al., 2011).

Attempts have been made at simply reducing the lignin content of plants used as a feedstock to reduce its deleterious effects on biofuel production (Chen and Dixon, 2007).

However, success has been limited, as a negative relationship between lignin content and plant biomass prevents a significant reduction of lignin content (Chapple et al., 2007). Researchers are investigating potential mechanisms to modify lignin *in planta,* in hopes of making more chemically labile lignin (Mansfield, 2009). Since lignin plays such a vital role in plant physiology, the xylem/parenchyma to sclerenchyma development, mechanical support of woody terrestrial plants, and responses to plant stresses, perturbations in lignin synthesis pathways are inundated with reductions in plant biomass quantity and poor plant health.

Chemical and Industrial Delignification Processes

Despite the chemical recalcitrance of lignin polymers, various chemical processes have been developed to either pretreat lignocellulose or to the reduce lignin content of wood chips/pulp. Only those pre-treatments or processes either in development for or being currently used in demonstration scale cellulosic biofuel production are discussed below.

Acid or Alkali Hydrolysis

There are two differing types of acid hydrolysis employed for pre-treatment of biomass in regards to cellulosic alcohol, dilute acid and concentrated acid (Bensah and Mensah, 2013). In a dilute acid biomass pre-treatment (0.2-2.5% w/w), biomass is degraded under high pressure/temperature extremely quickly with a low input chemical cost. Although drawing a large amount of attention from researchers, dilute acids have a major drawback of poor lignin removal, expensive reaction conditions, reduced sugar yields and enzyme inhibiting by-products resulting from the breakdown of lignin and hemi-cellulose.

Conversely, concentrated acid (~70% w/v) biomass pre-treatment involves highly concentrated acid incubation with lignocellulose for a long duration (24 hours) at atmospheric pressures and low temperatures. Concentrated acid pretreatment has higher sugar yields with the drawbacks of longer reaction times, higher input chemical costs, expensive corrosive resistant systems and large volumes of wastewater from subsequent neutralization reactions.

Conversely, alkali pre-treatments have the distinct advantage of better delignification of biomass, lower reaction temperature, less corrosion concerns and less inhibitory by-product release (Bensah and Mensah, 2013). However, alkali pre-treatments have lower hemicellulose yields and high levels of salt formation that presents major environmental challenges for disposal and subsequent sugar purification and fermentation.

Ionic Liquids

Recent research into ionic liquids has developed extremely promising alternatives to current biomass pre-treatment options, however many of the processes are still in the experimental stages (Bensah and Mensah, 2013; Li et al., 2010). Ionic liquids are non-flammable salts comprised of large organic cations and small inorganic anions that remain in liquid phase at temperatures below 100°C. When used in pre-treatment of biomass, ionic liquids attack the hydrogen bonds formed between cellulose and lignin, allowing cellulose to dissolve into the solvent (Bensah and Mensah, 2013). The solubilized cellulose is regenerated upon addition of an antisolvent, typically water, resulting in remarkably pure amorphous cellulose and hemicellulose and a separated lignin fraction. However, since the identification of ionic liquids is relatively new, the deleterious effects of ionic liquids on the environment are unknown and the associated production and disposal costs are prohibitive for industrial-scale cellulosic ethanol production.

Steam-Explosion

Theoretically and mechanistically simple, steam explosion has gained considerable attention in the pre-treatment of biomass (Bensah and Mensah, 2013). Steam-explosion pre-treatment is a physiochemical process where milled biomass is elevated to extremely high pressures and temperatures followed by an explosive decompression to atmospheric conditions. This explosive decompression physically tears apart the lignocellulose and increases accessibility of cellulose and hemicellulose for enzymatic pre-treatment. Remarkably steam-explosion has been shown to result in an increase from 15%-90% cellulose hydrolysis efficiency over untreated biomass. Steam-explosion reactors are costly and do not separate lignocellulose components necessitating an increased number of subsequent sugar purification steps and a large waste stream. Despite its limitations, steam-explosion is currently the most favoured process in cellulosic biofuel production. However, lignin is treated as waste and is not recovered in this process.

Kraft Process

The most common method used in the paper making industry for producing high-quality paper is the Kraft process, otherwise known as sulphate pulping. The Kraft process treats wood chips (or any cellulose dense biomass) with concentrated sodium hydroxide and sodium sulfate (white liquor) to break the bonds between lignin and cellulose (Chakar and Ragauskas, 2004). The reaction causing the delignification of wood pulp takes several hours at highly elevated temperatures and results in production of black liquor, a viscous liquid that contains approximately 50% of the energy content of the original wood pulp and is comprised of between 35-45% sulfonated lignin residues (known as Kraft lignin). The cellulose content of plant biomass is not solubilized into the alkaline black liquor and remains in the wood chip particulate during the reaction (Figure 3). Black liquor is separated and combusted in recovery boilers or is gasified to produce syngas (a biogas comprised mainly of H_2). The insoluble fraction, containing cellulose and hemi-cellulose, forms the feedstock for production of ethanol.

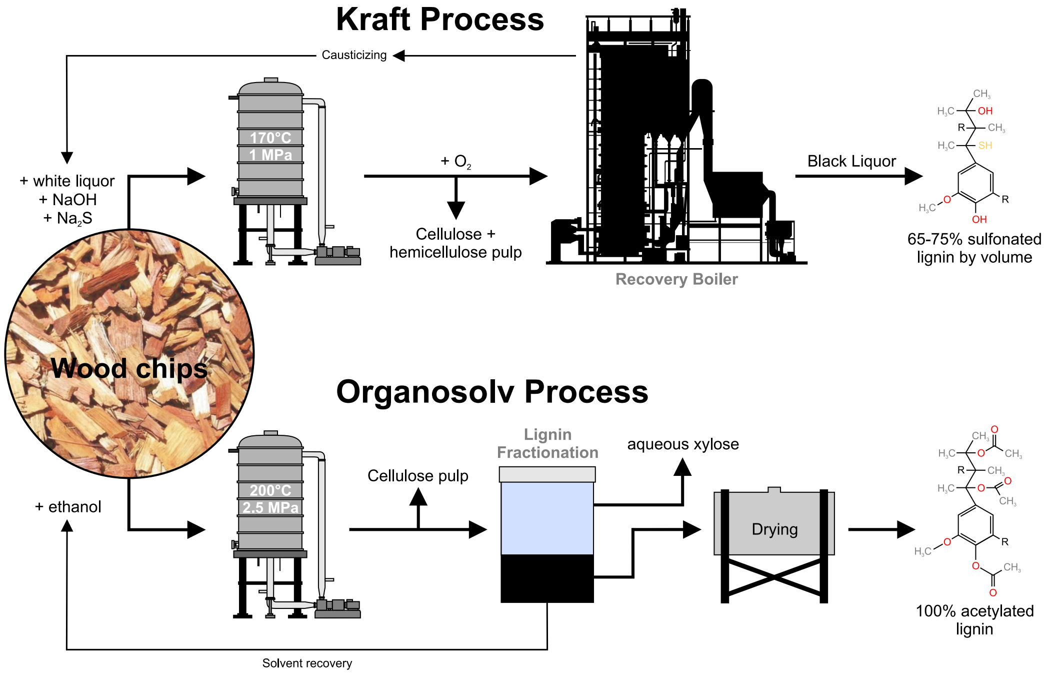


Figure 3. Comparison of Lignocellulose Component Streams from Kraft Process Versus Organosolv Process. The two most common processes differ in the resulting component purity, reaction chemistry and solvents used during processing. While the Kraft process is cheaper and more widely used, obtained fractions have higher levels of impurities, highly sulfonated lignin derivatives and the process employs caustic chemicals. Conversely, Organosolv is more expensive in both input chemicals and reaction costs, however the resulting components are highly pure, relatively native state and required solvents are easily recycled.

Organosolv Lignin

Although other methods exist for removing lignin from bulk plant biomass in regards to papermaking only one other method has gained considerable interest for its potential in cellulosic biofuels (Zhao et al., 2009). The Organosolv process was invented as an environmentally friendly alternative to the Kraft process, although its high unit cost prevented wide spread adoption in papermaking. However, the Organosolv process has the distinct advantage of a high purity separation of cellulose, hemicellulose and lignin, as well as a more renewable and recoverable set of solvents, both desirable features for sustainable biofuel production (Zhao et al., 2009). Organosolv pulping involves submerging wood chips into an aqueous organic solvent at elevated temperatures of 140-220°C at 2.5 MPa. This causes a breakdown of the lignin by hydrolytic cleavage of α aryl-ether linkages, common in the lignin matrix, resulting in water-soluble fragments that are free to dissolve into the aqueous phase. The aqueous phase is separated and distilled to remove solvents and hemicellulose resulting in a remarkably pure lignin, cellulose and hemi-cellulose product. The complete separation of streams allows for easy fermentation of sugars without the troublesome lignin moieties, as well as a pure lignin substrate for further downstream applications. Furthermore, various organic solvents can be used as the solvent including ethanol, 1-butanol and acetone, all common fermentation products, potentially allowing for development of a self-sufficient biofuel pretreatment/fermentation factory (Zhao et al., 2009).

Regardless of the pre-treatment process used, lignin presents a major barrier to cellulosic ethanol. Even in pretreatment processes where lignin is effectively removed, no revenue generating process currently exists to utilize these lignin-containing fractions. The development of value-added lignin-based products has been identified as the next logical step to increase the feasibility of cellulosic biofuels (Doherty et al., 2011).

Biological Delignification Processes

The US Department of Energy and various industry groups have identified lignin as a major technical barrier preventing economically viable development of biofuels (U.S. Department of Energy, 2012). Previous insights into the metabolism, structure and construction of the lignin matrix are limited since the compound has historically played only a minor role in processes where it is typically treated as a low-value waste product. Organisms must possess processes to naturally degrade lignin in the environment or it would be expected that massive deposits of lignin would exist worldwide. Various identified fungal species and an ever-increasing number of bacterial species have been identified as capable lignin metabolizers, presumably evolving to use lignin as an energy source due to its wide-spread availability and reduced competition from other organisms (Blanchette, 1991; Sharma et al., 2006; Paliwal et al., 2012; Brown and Chang, 2014; Pollegioni et al., 2015).

Fungal Metabolism

The most commonly researched lignin metabolizing organisms are the white-rot fungi, whose common name derives from their ability to metabolize lignin, leaving behind light white/yellow rotted wood comprised of mainly cellulose and hemicellulose (Ryu et al., 2013; Hofrichter, 2002; Sun and Cheng, 2002). Fungi typically employ peroxidases and laccase enzymes to metabolize lignin. Unique to *Basidomycota*, manganese peroxidases are widely employed during the depolymerization of lignin, the oxidation of sulfuric compounds and the unsaturation of fatty acids (Hofrichter, 2002; Janusz et al., 2013). Manganese peroxidases catalyze the oxidation of Mn(II) to Mn(III) through an irreversible set of redox reactions generating two oxidized Mn(III) that are typically secreted as Mn(III)-oxalate (although other carboxylic acid chelators exist). The Mn(III)-carboxylic acid then randomly oxidizes phenolic moieties in lignin, most commonly causing alkyl-aryl cleavages and α-carbon oxidation (Paliwal et al., 2012).

Various fungi employ specialized peroxidase enzymes known as lignin peroxidases (LiP) as well as versatile peroxidases (VP) (Janusz et al., 2013; Paliwal et al., 2012; Brown and Chang, 2014). LiP interact directly with the lignin phenolic side chains, in contrast to the Mn(III)-mediated mechanism seen with manganese peroxidases. Crystal structures exist for a variety of lignin peroxidases, which appear to co-metabolize lignin with other carbon sources (Camarero et al., 1999; Martinez, 2002). The direct cleavage mechanism of lignin by LiP makes them undesirable for pretreatment of bulk biomass where lignin is often variably accessible, but has potential use if lignin components are separated prior to treatment. Lignin metabolizing fungi that lack LiP instead appear to express a combination of VP, manganese peroxidases and laccases to degrade lignin (Fernández-Fueyo et al., 2014).

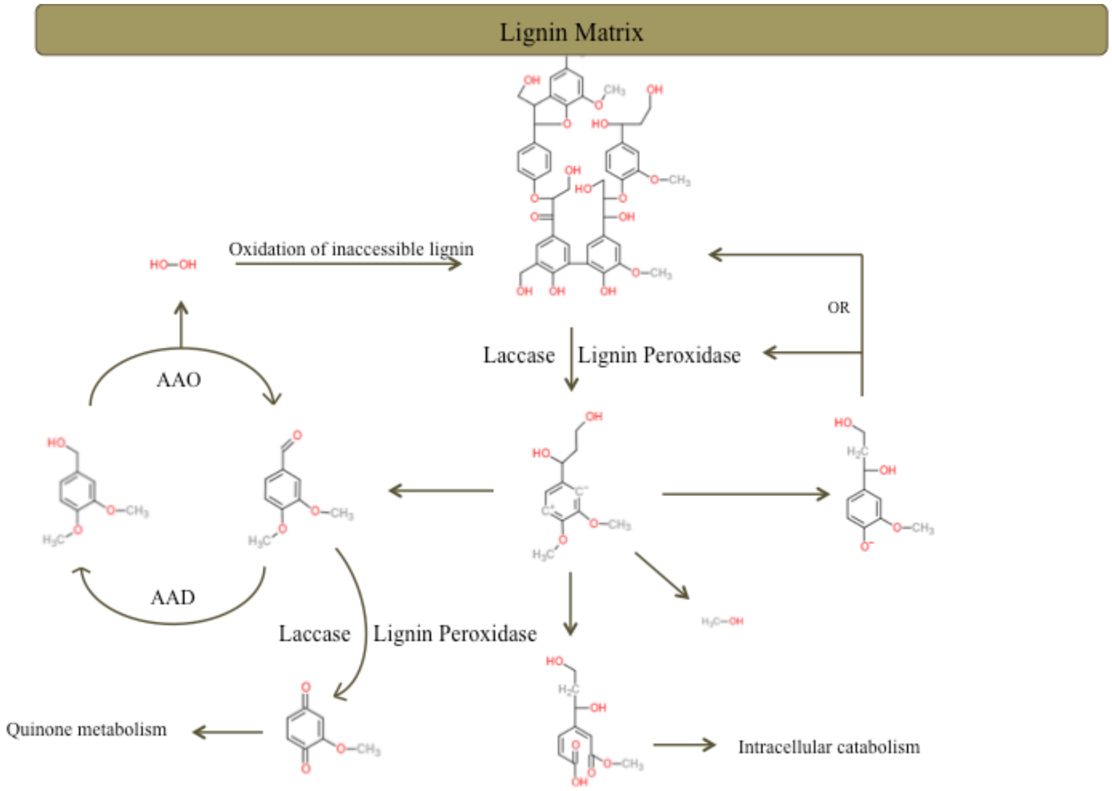


Figure 4. Common Fungal Delignification Pathways. White-rot fungi are known to employ a variety of mechanisms for delignification and subsequent metabolism of plant lignin. Various enzymes involved in the downstream lignin derivative catabolism are not shown as they are highly species specific. Various pathways generated free radicals and peroxide which are then able to further mediate lignin breakdown. AAO – aryl alcohol oxidase, AAD – aryl alcohol dehydrogenase.

The most heavily studied family of lignin metabolism genes in fungi are laccases; multi copper oxidases secreted by a variety of fungi and bacteria, that are capable of catalyzing the ring opening of various phenolic compounds (Ryu et al., 2013; Janusz et al., 2013; Dwivedi et al., 2011; Sun and Cheng, 2002; Martínez et al., 2009; Blanchette, 1991; Paliwal et al., 2012). An interesting avenue for biomass delignification, laccases can function through both direct and indirect mechanisms. The most promising application for laccases in biomass pretreatment is the decoupling of cellulose and hemi-cellulose residues from the lignin matrix as opposed to complete lignin breakdown. This approach still does not result in a potential application for lignin, instead opting to reduce costs associated with the delignification of biomass. A more comprehensive review of laccase function and potential is discussed in the following section.

Bacterial Lignin Metabolism

Research into fungal lignin metabolism genes has been limited by the difficulties inherent to fungal genomics and genetics. Compared to bacterial genetics and genomics, fungal studies are time-consuming, complex and problematic, exacerbated by the eukaryotic nature of fungi. This has lead researchers to bacterial lignin metabolizers to identify potential pathways, genes and mechanisms to address current limitations of cellulosic fuels (Bandounas et al., 2011; Bugg et al., 2011).

Various bacterial species are known to metabolize lignin, especially Gram-positive species, however, little research has investigated the genetic basis of these phenotypes. Despite the growing interest in lignin metabolism in bacteria, the only complete characterization of the lignin metabolism network in a bacterium was performed in *Sphingomonas paucimobilis* SYK-6, representing one of the few complete characterizations of lignin metabolism in any organism. Genes involved in the lignin metabolism network in *S. paucimobilis* SYK-6 were extensively characterized detailing the function, substrate specificity and reaction rates of each intermediate and its cognate enzyme(s) (Hara et al., 2003; Masai et al., 1999; Peng et al., 2005; Sonoki et al., 2000; Masai et al., 2007, 2000; Abe et al., 2005; Hara et al., 2000; Peng et al., 2002, 1999, 1998). *S. paucimobilis* SYK-6 is unique in that it possesses multiple operons involved in the lignin metabolism network, which specifically recognize a substrate and perform a single reaction, in contrast to fungal pathways mediated mainly by non-specific oxidases and peroxidases. Previously, researchers believed only higher fungi were capable of degrading lignin by oxidative processes, of which the majority employs non-specific mechanisms.

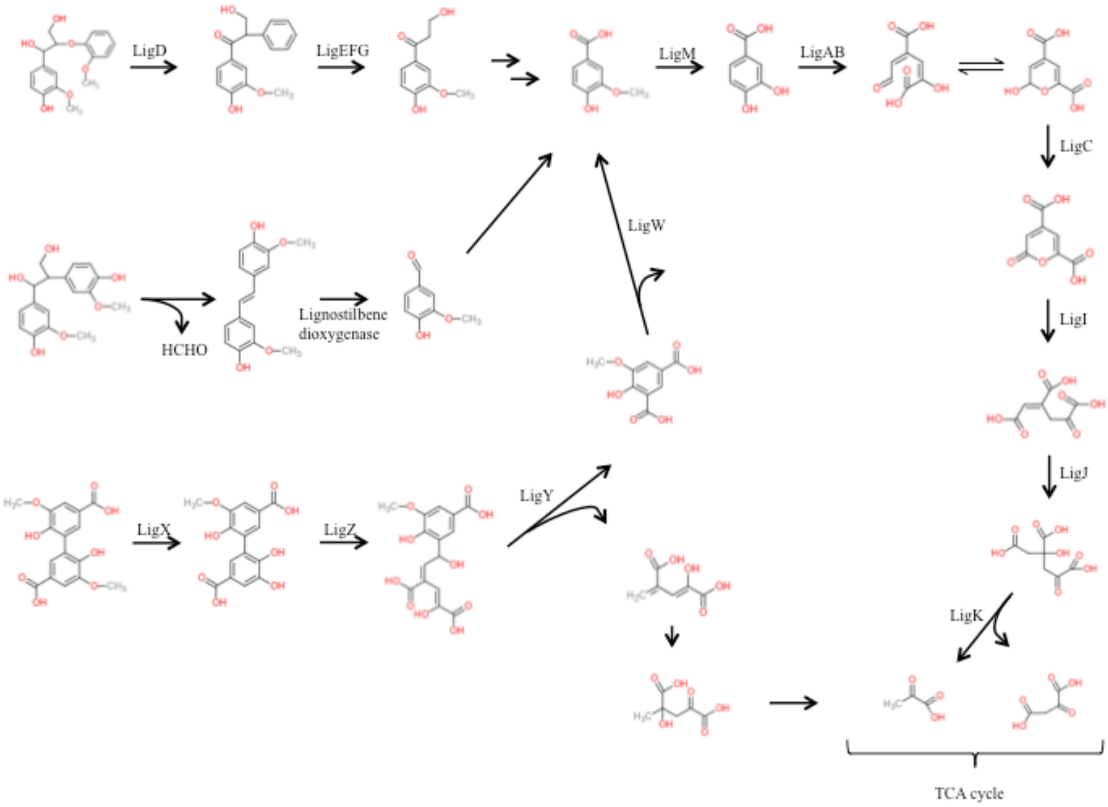


Figure 5. Lignin Metabolism Network in *Sphingomonas paucimobilis* SYK-6. Despite encoding for an entire catabolism pathway of a compound widely distributed in nature, *lig* genes are seemingly limited to a small number of lignin metabolizing Gram-negative bacterial isolates.

The lignin metabolism pathway of *S. paucimobilis* SYK-6 contrasts with other known lignin metabolism pathways, where lignin de-polymerization proceeds mainly through non-specific oxidation or radical generation (Pollegioni et al., 2015). One of the key findings of the *S. paucimobilis* SYK-6 characterization was the identification of specific genes mediating deconstruction of lignin into vanillin followed by genes mediating the conversion of vanillin into protocatechuate, a key nodal intermediate in previously characterized bacterial phenolic metabolism pathways. Furthermore, *S. paucimobilis* SYK-6 appears to metabolize lignin derived protocatechuate through an uncommon aromatic ring cleavage pathway. The discovery of a lignin metabolism network that involves a series of substrate-specific redox reactions, with enzyme complexes possessing tight substrate-specificities, was in stark contrast with previously described mechanisms that were not structure specific.

Much like in fungi, lignin matrix deconstruction in bacteria typically occurs via non-specific redox reactions or free radical generation. For example various *Streptomyces* sp. are capable lignin metabolizers employing single subunit laccases from a unique family of enzymes (Bugg et al., 2011). The *Streptomyces* sp. laccases are structurally different than their fungal counterparts despite the catalytic centre of both families showing similar catalytic pockets (Lu et al., 2014; Machczynski et al., 2004; Majumdar et al., 2014). Laccase enzymes are multi-copper oxidase enzymes that depolymerize lignin by directly oxidizing the phenolic backbone, catalyzing aromatic ring opening through either an intradiol or extradiol substitution. Laccases found in *Streptomyces* sp. are small in size compared to fungal laccases and have been shown to have an abnormally wide substrate specificity *in vitro* (Majumdar et al., 2014). Recent research investigating the small two-domain laccase encoded by *Streptomyces coelicolor* suggests these enzymes mediate delignification through two separate processes, the catabolism of phenolic β–O-4 lignin without mediators and degradation of non-phenolic moieties in the presence of specific mediators (Lu et al., 2014; Majumdar et al., 2014). Fungal lignin degraders that employ laccases typically encode heterologous enzymes that act to prevent repolymerization of the radical lignin species by coproducing a reducing agent, yet no similar processes have been identified as of yet in bacterial degraders (Martínez et al., 2009). However, even laccase deletion mutants of *S. coelicolor* retain a portion of their lignin metabolism capacity, suggesting that other pathways and genes are at least nominally involved in these processes (Majumdar et al., 2014).

Recent research in bacteria has identified a novel family of peroxidases known as Dye-decolourizing peroxidases (DyP). DyPs were discovered as a novel family of heme peroxidase enzymes and were first noted for their ability to degrade industrial dyes found in textile factory wastewater, which have a similar structure to lignin moieties (Kim and Shoda, 1999). DyP-type peroxidases are grouped into one of two subclasses, DyP type-A and DyP type-B (Paliwal et al., 2012). Further research into DyP type-A peroxidases subsequently demonstrated their cleavage of heme, and DyP type A peroxidases are currently believed to be a mechanism of scavenging iron. Conversely, the DyP type-B peroxidases have been shown to degrade lignin and lignin mimetic dyes (Roberts et al., 2011). Although first identified as a lignin metabolizing gene in *Rhodococcus jostii* RHA1, DyPB genes have subsequently been identified in other soil associated Gram-positive bacteria capable of metabolizing lignin (Bugg et al., 2011; Eastman et al., 2014a; Brown and Chang, 2014; Majumdar et al., 2014; Ahmad et al., 2011). Recently, researchers have begun identifying delignification by various Gram-negative bacterial species utilizing Dyp1B homologs that were previously annotated as peroxidases with no known function (Rahmanpour and Bugg, 2015). Various crystal structures of DyPB have been determined, which point to a mechanism similar to Manganese-peroxidases of fungi, which employ Mn^{2+} as a redox shuttle (Roberts et al., 2011). However, the DyP type B peroxidases contain a novel fold and active site confirmation.

The recent identification of a variety of novel mechanisms employed by bacterial species during delignification is promising for identification and characterization of the mechanistic details of bacterial lignin metabolism. Furthermore, many of the identified enzymes identified in bacteria are fully functional within heterologous organisms, allowing for more rapid detailed studies of their function and intricacies .

Features of Paenibacillus polymyxa

Paenibacillus polymyxa are Gram-positive, facultatively anaerobic motile, sporulating, free-living, soil-dwelling bacteria that are distributed worldwide in soils and aquatic environments (Anand et al., 2013; Lal and Tabacchioni, 2009). *P. polymyxa* strains are typically found free-living in the rhizosphere, however strains have been found in association with plant hosts in a diazotrophic mutualistic relationship (Anand et al., 2013; McSpadden Gardener, 2004; Haggag and Timmusk, 2008; Holl and Chanway, 1992; Timmusk et al., 2005; von der Weid et al., 2000). The most fervent area of research on *P. polymyxa* involves nitrogen fixation, the process by which microorganisms convert atmospheric nitrogen into ammonia, catalyzed by the enzyme nitrogenase (Wang et al., 2013). In this way, the bacterium provides the plant with biologically active ammonia or urea in exchange for fixed carbon provided through plant-derived compounds (Bohlool et al., 1992; Udvardi and Poole, 2013). This important topic of *P. polymyxa* research currently garners a lot of attention since agricultural practices in the developed and developing world involve heavy addition of nitrogen fertilizers to fields to increase crop yields (Bohlool et al., 1992). These fertilizers result in a wide array of unintended and devastating side effects including eutrophication of water systems, large-scale changes in the ecology of the surrounding ecosystem and methemoglobinemia when nitrite contaminated waters are ingested.

P. polymyxa strains are well known for their production of wide spectrum antimicrobial compounds including polymyxins, bacitracins, gramicidins and fuscaricidins, among a multitude of strain specific non-ribosomal peptides and polyketides (Eastman et al., 2014a; Beatty and Jensen, 2002; Niu et al., 2013; Haggag and Timmusk, 2008; Choi et al., 2009; Khan et al., 2008; Shaheen et al., 2011; Timmusk et al., 2009; Ryu et al., 2006). Various non-ribosomal antimicrobial peptides produced by *P. polymyxa* are used both clinically and over the counter in topical solutions. For example, polymyxin B and E (trade name colistin) are employed as a last resort antibiotic in the treatment of multi-drug resistant Gram-negative bacteria (Petrosillo et al., 2014; Velkov et al., 2010). However due to its relatively high nephro-toxicity, neuro-toxicity and allergenic potential polymyxin B intravenous usage remains limited to extreme cases. Additionally, polymyxins, gramicidins and bacitracins produced by *P. polymyxa* strains, are also used in combinations in a variety of both over-the-counter and prescription antibacterial ointments.

Although focus has historically been on the plant association and antagonistic features of *P. polymyxa*, recent investigations have begun appreciating the potential industrial applications of the species. Various strains of *P. polymyxa* are currently under investigation for stereoselective fermentation of 2,3-butane-diol, a precursor used in industrial chemical production (Tong et al., 2013; Yu et al., 2011). Optimization of feedstock and culture conditions have resulted in a very high production of (R,R)-2,3-butane-diol and the species is currently under development for alcohol production for biofuels (Li et al., 2013).

Previous research has established *P. polymyxa* strains as plant-growth promoting rhizobacteria, found in association with both dicotylous and monocotylous plant hosts. As expected, many strains of *P. polymyxa* are known to encode a variety of plant cell wall metabolism proteins, which are necessary for survival in the soil environment and for gaining entry to host plant tissues (Shin et al., 2012; Eastman et al., 2014a; Lal and Tabacchioni, 2009). The genus *Paenibacillus* contains a wide variety of species that have been extensively characterized for their abilities to produce glycoside hydrolase family enzymes, of which various enzymes are currently under investigation for the potential in cellulose and hemi-cellulose degradation in *in vitro* bioreactors (Song et al., 2014; Balachandrababu Malini et al., 2012).

Our lab isolated a previously unreported strain of *P. polymyxa* from degrading corn stover from a field at the Southern Crop Protection and Food Research Centre of Agriculture and Agri-Food Canada in London, Ontario (Eastman et al., 2014b). Isolated based on its ability to metabolize lignin, hemi-cellulose and cellulose as sole carbon sources, the strain was further noted for its ability to fix nitrogen and antagonize growth of fungal, bacterial and oocyte pathogens. The 16S ribosomal gene sequence identified the isolate as a novel strain of *Paenibacillus polymyxa,* which we named *Paenibacillus polymyxa* CR1.

Our isolation and characterization of *P. polymyxa* CR1 is the first report of a *P. polymyxa* strain capable of metabolizing lignin, hemi-cellulose and cellulose. The ability to metabolize otherwise resilient plant constituents in combination with the species' known affinity for alcohol fermentation and hardiness lends to the applicability of the strain to the biofuel sector. This study aims to identify the potential mechanisms and pathways employed by *P. polymyxa* CR1 during the deconstruction of lignin matrices to allow for future investigations into their desirability for development of novel biofuel production strategies.

The post-genomics era of biology is characterized by the wide availability of genomic data and the speed at which genomes can be sequenced, annotated and compared. Whereas a decade ago when complete bacterial genome sequencing was prohibitively expensive and exceedingly arduous, current technologies and approaches allow for rapid sequencing and assembly of a genome at a fraction of the cost of previous endeavors (Bentley, 2010; Delseny et al., 2010). These technologies have lead to an explosion of bacterial genomic data and have resulted in genome sequencing of a novel and interesting isolate becoming a routine stage in their characterization (Medini et al., 2008; MacLean et al., 2009). Furthermore, the availability of a wide variety of previous sequenced organisms allows for a much faster identification of potential genetic elements underpinning desirable or interesting phenotypes.

The investigation into genes involved in the lignin metabolism network of *P. polymyxa* CR1 presented within allows for future targeted engineering of metabolic flux and optimization of culture and feedstock conditions for the development of valuable lignin pretreatment and/or lignin bio-product pathways.

Chapter 2 - Materials and Methods

Characterization of *Paenibacillus polymyxa* CR1

Isolation and Plant-Derived Carbon Metabolism

Corn residues remaining on the research farm fields at the Southern Crop Protection and Food Research Centre of Agriculture and Agri-Food Canada in London, Ontario, were buried during routine post-harvest tilling. In the following spring, April 2013, degrading corn stalks were excavated and transported to the lab in sterile plastic bags. Soil was washed from degrading corn biomass in a sterilized 0.85% saline solution overnight with agitation. Corn tissues were macerated in a sterile bench top blender prior to inoculation into a modified Minimal Media Davis (minimal media or MM) [7 grams/L (g/L) K_2HPO_4; 2 g/L KH_2PO_4; 0.1 g/L $MgSO_4$; 1 g/L $(NH_4)_2SO_4$] liquid culture containing 0.1% weight/volume (w/v) alkali Kraft lignin from Sigma-Aldrich (catalog number 370959) as a sole carbon source. After 2 weeks of enrichment culture growth with shaking, samples were diluted and plated onto minimal media + 0.1% Kraft lignin agar plates (15 g/L agar) to obtain individual isolates for further characterization.

Isolates were screened based on their ability to grow rapidly utilizing lignin as a sole carbon source. Isolates were further assayed for their ability to metabolize cellulose and hemi-cellulose as a sole carbon source by supplementation of minimal media agar plates with either 1% crystalline or carboxymethylcellulose (for cellulose metabolism) or 0.1% xylan (for hemicellulose metabolism) from Sigma-Aldrich (catalog numbers 22182, 419311 and X4252). Plating isolates on minimal media agar plates supplemented with the lignin mimetic dyes toluidine blue or methylene blue at a concentration of 25mM identified secretion of lignolytic enzymes as described previously (Bandounas et al., 2011).

16S Ribosomal Subunit Sequencing

16S ribosomal DNA was amplified by end-point polymerase chain reaction (PCR) using purified genomic DNA as a template from isolates using the universal primers 8F and 1492R (sequences provided in Table 3). Genomic DNA was purified using a GenElute™ Bacterial Genomic DNA Kit from Sigma-Aldrich (catalog number NA2120) according to the manufacturers protocol with the exception of elution into UltraPure™ DNase/RNase-Free Distilled Water from Life Technologies (catalog number 10977-023).

Table 3. Ribosomal subunit 16S and 23S primers.

Primer	Sequence	Gene	Direction	Reference
8F	AGAGTTTGATCCTGGCTCAG	16S	F	Anzai *et al*, 2000
1492R	CGTTACCTTGTTACGACTT	16S	R	Anzai *et al*, 2000
127F	CYGAATGGGRVAACC	23S	F	Hunt *et al*, 2006
2241R	ACCGCCCCAGTHAAACT	23S	R	Hunt *et al*, 2006
U1	TGGGATACCACCCTGATCGT	16S	F	Eastman *et al*, 2015
U2	GTTTGGGCTAATCCGCGTTC	16S	R	Eastman *et al*, 2015
U3	CCGTCACACCACGAGAGTT	23S	F	Eastman *et al*, 2015
U4	GTCCGCCGCTAGGTTGATTA	23S	R	Eastman *et al*, 2015

Primers are listed 5'-3'. Nucleotides follow IUPAC conventions where; Y = C or T, R = A or G, V = A, C, or G, H = A, C, or T.

PCR reactions for 16S ribosomal subunit amplification used *Taq* DNA Polymerase from Qiagen (catalog number 201203) with 1.5 mM Mg^{2+} and 10 nM forward and reverse primers (8F and 1492R respectively, Table 3 (Anzai et al., 2000)), 2.5 mM dNTP Mix, PCR Grade from Qiagen (catalog number 201900). End-point PCR conditions for routine 16S rDNA amplification were as follows, 95°C for 5 minutes followed by 35 cycles of; 95°C for 30 seconds, 57°C for 45 seconds, 72°C for 1 minute, followed a final extension at 72°C for 15 minutes. PCR products were visualized on a 40 mM Tris, 20 mM acetic acid, 1mM ethylenediaminetetraacetic acid (TAE) buffered, 1% agarose gel. PCR products of expected length were purified using a Qiagen PCR Purification Kit (catalog number 28106) according to the manufacturers protocol with the exception of elution into UltraPure™ DNase/RNase-Free Distilled Water from Life Technologies (catalog number 10977-023).

Sequencing of 16S ribosomal subunit PCR products were performed on an Applied Biosystems 3730xl DNA Analyzer using the universal 16S ribosomal subunit DNA primers 8F and 1492R (Anzai et al., 2000). Obtained sequences were aligned and manually refined using the SeqMan Pro application from DNAStar. Isolate CR1 was tentatively identified as a novel strain of *Paenibacillus polymyxa* based upon a BLASTn alignment (99.5% identity, E-value = 0) of the sequenced 16S gene against the GenBank database and was assigned the strain identifier CR1 (corn rhizobacterium 1).

Genome Sequencing of *Paenibacillus polymyxa* CR1

Whole genome sequencing of *Paenibacillus polymyxa* CR1 was performed using both a mate-pair and short-insert read library on the Illumina MiSeq Desktop Sequencer at

AGCT Inc. (Chicago, Illinois). The short-insert read library was prepared with a target insert size of 400 base pairs using the NexteraXT DNA sample preparation kit (catalog number FC-131-1024) from Illumina. A mate-pair read library was also generated with an average insert size of 1.25 kbp. The two libraries were loaded onto an Illumina MiSeq desktop genome analyzer at 9pmol and 2 x 150 base pair chemistry sequencing was performed using the MiSeq Reagent Kit v2. Adaptor sequences were removed and mate-pair reads with insert sizes shorter than 400 base pairs were removed. The resulting mate-pair read and short-insert read libraries were merged and prepared for contig assembly.

Phylogenetic Analyses

16S sequences were obtained for publically available *Paenibacillacae* from the NCBI Nucleotide database and aligned using Clustal (Sievers et al., 2011) and manually refined within MEGA6 (Tamura et al., 2013). 16S rRNA sequences were obtained from publically available whole genome sequences for strains where individual 16S sequences were not available. From these aligned sequences a phylogenetic was generated using the Maximum-likelihood method (Felsenstein, 1981) with default parameters using *Agrobacterium fabrum* C58 as an out-group. Support for the produced phylogenetic tree was determined by performing 1000 bootstrap replications and branches with less than 60% support were collapsed to polytomies. The neighbour joining whole genome phylogeny was generated using the dnadist and neighbour packages in PHYLIP and visualized using phyloXML, in addition to tools publically available on the Joint Genome Institute Integrated Microbial Genomes Database (Felsenstein, 1989; Han and Zmasek, 2009; Markowitz et al., 2014).

Genome Assembly of *Paenibacillus polymyxa* CR1

Draft Genome Assembly

The merged short-insert read and mate-pair read library from the whole genome sequencing of *P. polymyxa* CR1 were assembled *de novo* by three separate contig assembly programs. SOAPdenovo, ABySS and Velvet were run on default settings with k-mer lengths of 55 base pairs, 67 base pairs and 31 base pairs, respectively (Luo et al., 2012; Zerbino and Birney, 2008; Birol et al., 2009). The three separately generated draft genome assemblies were integrated using CISA on default settings (Lin and Liao, 2013).

Contig reduction and Draft Genome Scaffolding

PCR products corresponding to *Paenibacillus polymyxa* CR1 16S and 23S ribosomal subunit DNA genes were generated from purified genomic DNA using the universal primers 8F/1492R for 16S and 127F/2442R for 23S, respectively (Anzai et al., 2000; Hunt et al., 2006). PCR conditions were as follows; 95°C for 7 minutes followed by 40 cycles of; 95°C for 40 seconds, 57°C for 45 seconds, 72°C for 1.5 minutes, followed a final extension at 72°C for 20 minutes. 16S and 23S PCR products were visualized on a 40 mM TAE buffered, 1% agarose gel, excised from the gel and purified with an UltraClean® 15 DNA Purification Kit from MoBio (catalog number 12100-300) according to the manufacturers protocol with the modification of elution into UltraPure™ DNase/RNase-Free Distilled Water from Life Technologies (catalog number 10977-023). Sequencing of 16S and 23S ribosomal subunit PCR products were performed on an Applied Biosystems 3730xl DNA Analyzer using the same primers used for amplification of 16S and 23S. Sequences were aligned and a consensus sequence for 16S and 23S ribosomal subunit DNA was generated using the SeqMan Pro program from DNAStar.

Contigs from the *P. polymyxa* draft genome assembly were individually scaffolded against the *Paenibacillus terrae* HPL-003 (Shin et al., 2012), *Paenibacillus polymyxa* E681 (Kim et al., 2010), *Paenibacillus polymyxa* M1 (Niu et al., 2011), and *Paenibacillus polymyxa* SC2 (Ma et al., 2011) genomes using the contig reorder tool in progressiveMauve, using each completed genome sequence as a reference individually (Darling et al., 2010; Rissman et al., 2009).

Targeted Sequencing of Ambiguous Bases

Ambiguous bases contained within the draft *P. polymyxa* CR1 genome were identified and annotated in Artemis (Rutherford et al., 2000). Specific primer sets were designed using Primer3 to prime flanking protein-coding sequences and amplified each region corresponding to ambiguous bases (Untergasser et al., 2012). PCR reactions for generating products corresponding to the ambiguous base stretches used *Taq* DNA Polymerase from Qiagen (catalog number 201203) with 1.5 mM Mg^{2+} and 10 nM forward and reverse primers, 2.5 mM dNTP Mix, PCR Grade from Qiagen (catalog number 201900). End-point PCR conditions were as follows, 95°C for 5 minutes followed by 35 cycles of; 95°C for 30 seconds, Annealing Temperature for 45 seconds, 72°C for 1 minute, followed a final extension at 72°C for 15 minutes. Annealing temperatures were dependent on each specific primer set and varied between 54°C and 64°C. PCR products were visualized on a TAE-buffered 1% agarose gel. PCR products that contained the expected product length were purified using a Qiagen PCR Purification Kit (catalog number 28106) according to the manufacturers protocol with the exception of elution into UltraPure™ DNase/RNase-Free Distilled Water from Life Technologies (catalog number 10977-023). Sequencing of ambiguous base PCR products was performed on an Applied Biosystems 3730xl DNA Analyzer and sequences were aligned to the draft genome using Blastn with a word size of 28 and an E-value cut-off of 10^{-10}. Ambiguous base sequencing results were integrated into the draft genome using Artemis (Rutherford et al., 2000).

Contig Gap Closure

Putative gaps identified by scaffolding were targeted for Long and Accurate PCR from genomic DNA using primers designed to specifically amplify each contig gap (average length 7.5 kb). Primers were designed to hybridize between 250-300 bp from the 3' and 5' ends of each contig gap pair. PCR conditions for amplification of contig gaps were as follows; 95°C for 1 minute followed by 40 cycles of 95°C for 30 seconds, annealing temperature for 45s, 72°C for 10 minutes and a final extension of 72°C for 20 minutes. All PCR amplifications of contig gaps were performed using Phusion® Taq polymerase from New England Biolabs (catalog number M0530L). Resulting PCR products were run on an TAE-buffered 1% agarose gel, excised and gel-purified using a UltraClean® 15 DNA Purification Kit from MoBio (catalog number 12100-300) according to the manufacturers protocol with the modification of elution into UltraPure™ DNase/RNase-Free Distilled Water from Life Technologies to facilitate downstream sequencing (catalog number 10977-023).

Purified contig gaps were tested for the presence of ribosomal subunit gene operons using the universal primers 8F/1492R and 127F/2442R and for contaminating genomic DNA using primers corresponding to distant genomic regions not contained in the purified contig gap.

Investigation of Rearrangements

Rearrangements relative to other sequenced *P. polymyxa* strains were identified during contig reordering by localizing perturbations in the local collinear block composition, compared to each of the previously sequenced strains (Darling et al., 2011). Potential rearrangements were further scrutinized by local BLASTn alignments of identified regions using a sliding window of 100 nucleotides to identify precise locations of rearrangements. Locations where the draft genome of *P. polymyxa* CR1 jumped more than 10kb relative to other completely sequenced *P. polymyxa* strains were flagged for assembly confirmation. Multiple primers were designed to amplify across a region of unexpected sequence length versus *P. polymyxa* E681 and resulting PCR products of expected length corresponding to these gaps assessed assembly fidelity. Regions where assembly fidelity could not be confirmed were considered positive for the presence of a contig misassembly.

Potential resolutions of contig misassemblies were generated by alignment of the identified region against the genome of *P. polymyxa* E681 and primers were designed to generate products corresponding to the structure in *P. polymyxa* E681. These primers were then used to assess assembly fidelity of the newly identified structure as described above.

Genome Annotation

Annotation of the *P. polymyxa* draft genome was performed by the Rapid Annotation using Subsystem Technology (RAST) server from the National Microbial Pathogen Data Resource (rast.nmpdr.org) to identify protein coding sequences for downstream primer design (Aziz et al., 2008). These annotations do not appear in the publically available sequence on the NCBI database and were used only to identify priming locations in the draft genome.

The completely sequenced *P. polymyxa* CR1 genome was annotated using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) without consideration of the previous RAST annotations (Angiuoli et al., 2008).

Genome Annotation and Analyses

Genome annotations of each *P. polymyxa* strain were performed as described previously. Annotations were obtained from Genebank on January 1st, 2014. Genomic features were annotated in Artemis and visualized using DNAplotter (Rutherford et al., 2000; Carver et al., 2009). General genome features (rRNA, tRNA, and CDS) were identified using the provided annotations from Genebank and the genomic sequences were reanalyzed using tRNAscan and RNAmmer (Schattner et al., 2005; Lagesen et al., 2007). Information regarding clusters of orthologous groups (COGs), KEGG orthology (KO), protein localization, and gene ontology were obtained from the Joint Genome Institute Integrated Microbial Genomes database (Markowitz et al., 2014; Tatusov et al., 2001, 2003; Kanehisa et al., 2012; Kanehisa and Goto, 2000; Kanehisa et al., 2014). Tandem repeats were determined using TandemFinder (Benson, 1999). Prophage elements and features were identified using PHAST and visualized in Artemis (Zhou et al., 2011). Insertion sequences were predicted by the IS Finder database (Siguier et al., 2006).

Genomic Island Identification

Putative horizontally transferred genes were identified using IslandViewer 2.0, which scans the genome and identifies putative genomic islands by regional differences in GC-content and skew (Langille and Brinkman, 2009). Genomic islands identified by this method containing greater than 5 genes or larger than 4 kb in size were considered for analysis. Phage related genes contained within putative genomic islands were identified by manual curation of genomic island encoded genes.

General Comparative Genomics

The genome of each *P. polymyxa* strain was aligned against other sequenced *P. polymyxa* genomes accessible on Genebank on January 1st, 2014 by determining local collinear blocks (LCBs) using the progressiveMauve algorithm in Mauve (Darling et al., 2011, 2010). Dot-plots were created by iteratively comparing homologous protein coding sequences using the available tools on the JGI IMG database (Markowitz et al., 2014). Conserved and strain-specific genes were identified using mGenomeSubtractor on default parameters with H- value cut-offs of <0.41 and >0.8 for strain-specific and conserved proteins respectively (Shao et al., 2010).

Genes putatively responsible for plant-growth promotion, bio-mass degradation and solventogenesis were identified by using KO and homology searches using tBLASTx to previously characterized homologs. Metabolic and signaling pathways were constructed using the KEGG database. Homologs within these pathways were identified using a cut-off threshold of >50% positive amino acid identity against the closest related available homologue. Encoded transport proteins were identified by a BLAST search against the Transporter Classification Database and KO classification (Saier et al., 2014).

Glycoside hydrolase, pectin lyase, carbohydrate esterase and carbohydrate binding motifs were identified using the CAzY database (Lombard et al., 2014).

Protein homology was determined by performing a BLASTp search against identified homologs. Proteins that met an E-value and positive amino acid identity cut- off of $\leq 10^{-25}$ and ≥60%, respectively, were considered homologous.

Insertional Gene Knockouts

Insertional gene knockouts of DypB and a multicopper oxidase enzyme (laccase) in *P. polymyxa* CR1 were generated using pUCP30T as a vector (Figure 5). Primers DyPBF and DyPBR were designed using Primer3 to generate a product of ~300bp localized to the start of the DyPB gene. PCR conditions for amplification of DypB and laccase were as follows; 94°C for 1 minute followed by 35 cycles of 94°C for 30 seconds, annealing temperature for 45s, 72°C for 45 seconds and a final extension of 72°C for 7 minutes. PCR products were run on an TAE-buffered 1% agarose gel, excised and gel-purified using a UltraClean® 15 DNA Purification Kit from MoBio (catalog number 12100-300) according to the manufacturers protocol with the exception of elution into UltraPure™ DNase/RNase-Free Distilled Water.

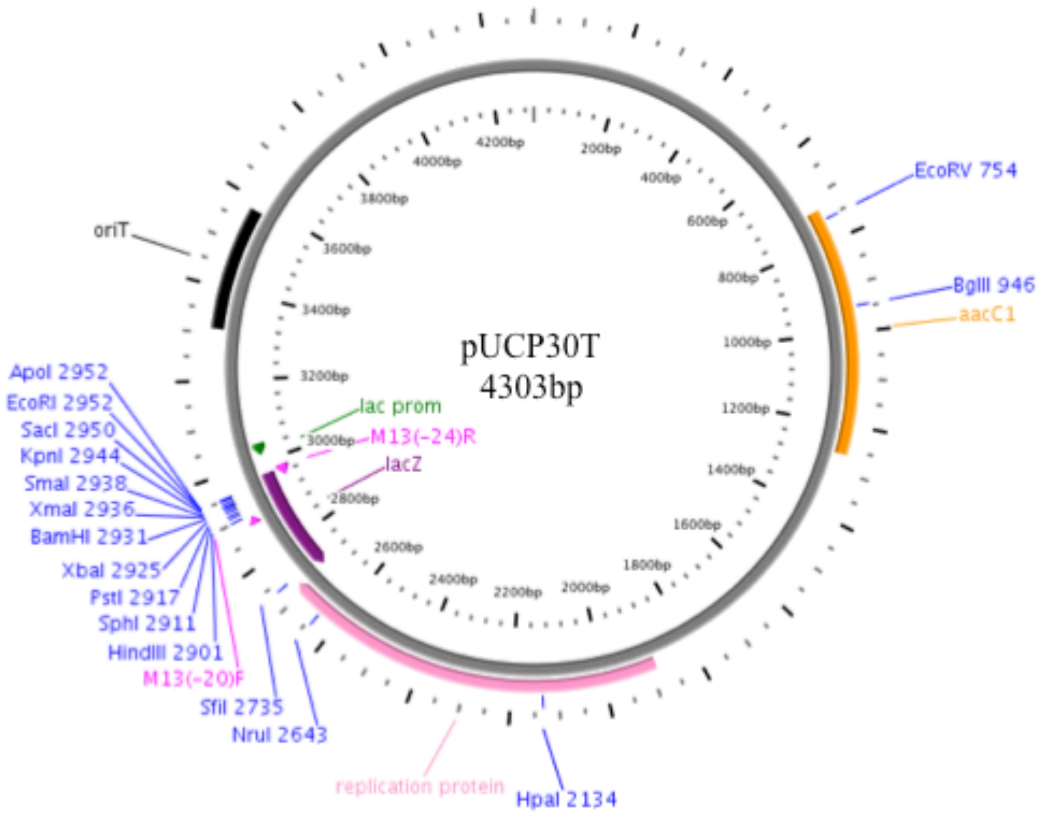


Figure 6. Plasmid Map of pUCP30T. The pUCP30T vector acts as a suicide plasmid in *Paenibacillus polymyxa*, and also contains aacC1, a gentamycin resistance marker, and a multiple cloning site within *lacZ*.

DypB or laccase fragment PCR products and p30T were digested using EcoRI for 1 hour at 37°C minutes. Following digestion reaction mixtures were ethanol precipitated using the following steps; 100 µL of ice cold 100% anhydrous ethanol mix by inverting followed by incubation on ice for 20 minutes. Solutions were centrifuged at 13000 rpm on an Eppendorf 5424 tabletop centrifuge for 10 minutes followed by removal of supernatant. The resulting pellet was washed with 500 µL ice-cold 100% anhydrous ethanol and vortexed briefly followed by centrifugation at 13000 rpm for 10 minutes.

Supernatant was removed completely and the pellet was allowed to air dry in a continuous flow fume hood.

Digested and purified p30T and DypB/Laccase PCR products were combined and re-suspended in 17 µL of UltraPure™ DNase/RNase-Free Distilled Water, 2 µL 10X T4 DNA ligase buffer and 1 µL of T4 ligase from New England Biolabs (catalog number M0202S). Reaction mixtures were incubated at room temperature for 2 hours to allow for ligation. Ligated plasmids were transformed into One Shot® TOP10 Chemically Competent *Escherichia coli* DH5α cells (catalog number C4040-03). 1.5 µL of ligation reaction was incubated on ice for 30 minutes with *E. coli* DH5α followed by submersion in a 42°C water bath for 2 minutes and returned to ice for 2 minutes. Following incubation on ice, 250 µL of SOC Broth was added and cells were incubated at 37°C for 1 hour in a vertical rotary wheel. Cells were plated on LB-Miller agar supplemented with 10 mg/µL of gentamycin and 40 mg/µL X-Gal. White colonies were picked and streaked onto LB-Miller agar containing 10mg/µL gentamycin to prepare for plasmid extraction.

Plasmids were extracted using a QiaPrep Spin Miniprep Kit from Qiagen (catalog number 27106). Aliquots of plasmids were re-digested as described previously and digestion products were run on a 1X TAE buffered 5% agarose gel at 100V for 30 minutes to confirm correct size of insertion. Plasmids with positive insertion sizes were sent for sequencing using primers M13F and M13R (Table 4, priming location shown in Figure 5) on an Applied Biosystems 3730xl DNA Analyzer confirm correct incorporation of the desired gene fragments.

Table 4. Primers for confirming insertions in pUCP30T. Primers are listed 5'-3' using standard nucleotide convention.

Primer	Sequence	Direction
M13(-20)F	GTAAAACGACGGCCAGT	Forward
M13(-24)R	AACAGCTATGACCATG	Reverse

1 µL of plasmids pDyp and pLac were electroporated into electrocompetent *P. polymyxa* CR1 using conditions as previously described (Kim and Timmusk, 2013). Insertional mutants were screened based on their ability to grow on LB agar plates supplemented with 25ng/µL gentamycin. Clones that grew were confirmed as knock-outs by PCR amplification of the insertion location using primers designed to hybridize immediately flanking the insertion, followed by sequencing of the region on an Applied Biosystems 3730xl DNA Analyzer.

For testing growth phenotypes of knockout mutants 250mL culture flasks of 0.2% lignin liquid M9 media [3g/L Na_2HPO_4, 1.5g/L KH_2PO_4, 0.5g/L NH_4Cl, 0.25g/L NaCl, 0.2% lignin (w/v)] were inoculated with 10^5 cells/mL and incubated at 37°C for 60 hours with shaking. Aliquots were taken, serially diluted and plated in triplicate to determine CFU/mL at various time points.

Transposon Mutagenesis

Electrocompetent *P. polymyxa* CR1 cells were prepared by washing cells grown to an OD_{600} of 0.6 twice with ice-cold SG buffer (0.5M sucrose, 1mM $MgCl_2$, 10% glycerol) prior to snap freezing in SG buffer liquid nitrogen and storage at -80°C. Tn5-Tet transposomes were prepared from the EZ-TN5-TET Transposon (catalog number EZI921T) by Epicentre (Illumina) by incubation of transposons with transposase in the absence of Mg^{2+}. Electrocompetent *P. polymyxa* CR1 were thawed on ice and mixed with 1 μL of arrested transposomes and the tubes were inverted and incubated on ice for 5 minutes. Cells were subsequently transferred into a 0.1cm cuvette and pulsed at 625kV for 15ms with 15 mΩ of resistance. Cells were recovered in 350 μL of Super Optimal Broth with catabolite repression (SOC) media [20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCL, 10 mM $MgCl_2$, 20 mM glucose] at 37°C with shaking for 3 hours. After recovery, cells were serially diluted and plated on LB-Miller agar plates [10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar] containing 10 ng/μL tetracycline hydrochloride. Cells which grew on LB-Miller + 10 mg/L tetracycline were streaked and single colony purified twice prior to maintenance of the transposon library by patching onto LB-Miller + 10 mg/L tetracycline and storage at 4°C.

Phenotypic Characterization

The TN5-Tet CR1 transposon library containing 5678 individually isolated and purified clones was streaked onto Lignin Minimal Media agar plates [7 g/L K_2HPO_4; 2 g/L KH_2PO_4; 0.1 g/L $MgSO_4$; 1 g/L $(NH_4)_2SO_4$, 0.1% weight/volume (w/v) Kraft lignin (Sigma), 15 g/L agar]. Identified clones that either grew faster, did not grow or grew at a reduced rate compared to wild-type *P. polymyxa* CR1 were confirmed for their growth phenotype and labeled with a unique identifier. Those clones that showed the desired phenotypes were flagged for identification of the disrupted gene by direct genomic DNA primer walking or reverse PCR in situations where direct sequencing failed to yield usable sequence.

Identification of Disrupted Genes

Sequencing of the regions flanking the transposon insertion were identified by primer walking genomic DNA using primers designed to specifically anneal to unique regions within the TN5 transposon with 3' ends directed out of the transposon (Figure 6).

Those clones that were unable to be identified by a genomic DNA primer walking strategy were flagged for identified of the disrupted genes by inverse PCR (Ochman et al., 1988). Isolated genomic DNA from clones was digested using EcoRI from New England Bio-labs (catalog number) at 37°C for 2 hours followed by heat inactivation of the enzyme at 80°C for 20 minutes. Sheared genomic DNA was isolated by ethanol precipitation (described protocol). Following DNA purification, sheared genomic DNA was diluted to a concentration that promotes intramolecular over intermolecular interactions. The reaction mixture was circularized using T4 ligase from New England Bio-labs by incubation at 37°C for 1 hour. The circularized fragments were used as a template for end-point PCR using primers designed to specifically anneal to unique regions within the TN5 transposon whose 3' ends direct out of the transposon. PCR conditions were as follows, 95°C for 5 minutes followed by 35 cycles of; 95°C for 30 seconds, 57°C for 45 seconds, 72°C for 1 minute, followed a final extension at 72°C for 15 minutes. Amplified PCR products corresponding to Tn5 genomic flanking regions were purified using the Qiagen PCR Purification Kit (catalog number 28106) according to the manufacturers protocol with the exception of elution into UltraPure™ DNase/RNase-Free Distilled Water from Life Technologies (catalog number 10977-023).

Nucleotide Accession Numbers

All complete genomic sequences referenced in the text are publically available on the NCBI Nucleotide database with the following accession numbers; *P. polymyxa* CR1 – NC_023037.2, *P. polymyxa* E681 – NC_014483.1, *P. polymyxa* M1 – NC_01752.1, *P. polymyxa* SC2 – NC_014622.1, *P. terrae* HPL-003 – NC_016641.1. Plasmid sequenced referenced in the text were obtained from the NCBI Nucleotide database with the following accession numbers; pSC2 – NC_014628.1, pM1 – NC_017542.1 16S ribosomal subunit gene sequences used in the construction of phylograms were obtained from the NCBI Nucleotide database or from the completed genome where available.

Chapter 3 – Results

Isolation and Characterization of *P. polymyxa* CR1

Bacterial strains showing wide spectrum antibiotic production and plant-derived carbon metabolism were isolated from degrading corn stalks recovered from the Southern Crop Protection and Food Research Centre of Agriculture and Agri-Food Canada in London, Ontario. Partially degraded corn tissues were macerated in a sterile tabletop blender and inoculated into parallel liquid minimal media containing 0.2% lignin. Replicates of cultures were left to grow for 2 weeks at 22°C, 37°C and 56°C with shaking. After 2 weeks of growth, enrichment cultures were serially diluted and plated onto 0.1% (w/v) Kraft lignin minimal media (MM) agar. Isolates that showed rapid growth were streak purified and assigned a unique identifier. The isolate that demonstrated the most rapid growth on lignin, cellulose and hemi-cellulose medias, CR1, was also able to metabolize lignin mimetic industrial dyes, suggesting the secretion of lignolytic enzymes (Figure 7) (Ahmad et al., 2010).

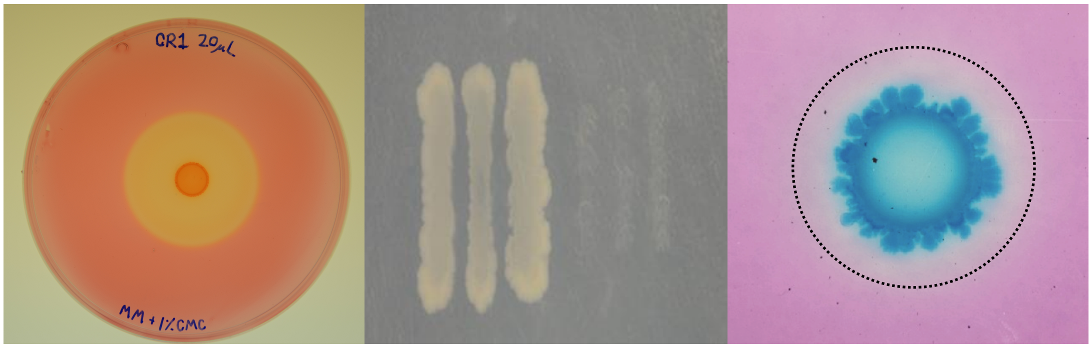


Figure 7. *P. polymyxa* CR1 is capable of metabolizing a wide range of plant-derived carbon sources. Cellulose, hemi-cellulose and lignin degradation capacities were qualitatively assayed as follows (left to right). Carboxymethylcellulose Congo Red stain assay, zone of clearing surrounding inoculated red colony of *P. polymyxa* CR1 represents secretion of endocellulases capable of cleaving reducing end of CMC. Hemi-cellulose hydrolysis was assayed by positive growth on minimal salts media containing hemi-cellulose as a sole carbon source. Methylene Blue dye clearance assay, zone of clearance assays secretion of dye-decolourizing peroxidase (DyP).

Sequencing of the 16S ribosomal subunit tentatively placed the isolate as a novel strain of *Paenibacillus polymyxa,* which we assigned the strain identifier CR1 (corn rhizosphere 1). Inoculation of *P. polymyxa* CR1 into 0.1% lignin MM agar media at low oxygen concentrations resulted in the production of solvents and evolution of gas. Preliminarily analysis of produced solvents using LC-GS identified ethanol, 2,3-methylbutanol, propene and butane production; solvents widely used in industrial production of a variety of precursor chemicals and as fuels in the biofuel sector (Figure 8).

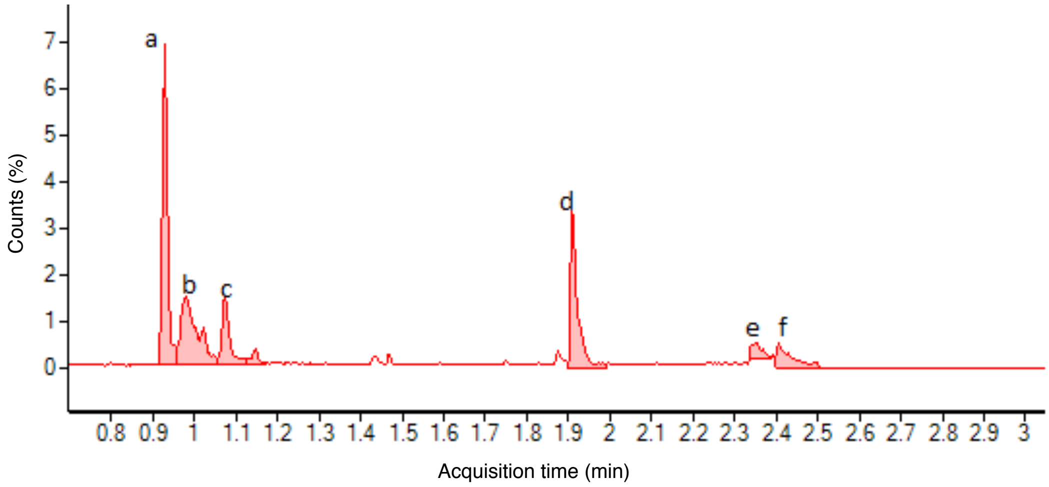


Figure 8. Gas-chromatography of lignin fermenting *P. polymyxa* CR1. *P. polymyxa* CR1 was grown in low oxygen concentration in minimal salts agar supplemented with 1% lignin. Peaks were identified as follows; a) CO_2, b) unidentified, c) ethanol, d) 2-methylbutanol, e) propene, f) 2-butene.

Identification of a bacterium capable of degrading lignin substrates directly into usable alcohols would represent a significant development for biofuel production since lignin rich biomass fractions are currently treated as waste products (Zeng et al., 2014; Bugg et al., 2011; Doherty et al., 2011). Further development of the isolate as an industrial strain requires a robust knowledge of the genetic basis of the lignin metabolism and solventogenesis.

Genome Sequencing of *P. polymyxa* CR1

Second-generation sequencing of the *P. polymyxa* CR1 genome was performed on an Illumina MiSeq Sequencer. 2x150 bp chemistry was performed with both a mate-pair read library and a short-insert read library, generating 2.9 million reads and 2.2 million reads for the short-insert read and mate-pair read libraries respectively. The resulting unmerged short-insert read library had 140x genome coverage. Conversely, the unmerged mate-pair read library had 107x genome coverage. Adaptor sequences were removed, mate-pair reads with inserts shorter than 400 bp were filtered and the two libraries merged. The merged read library had 40x genome coverage with a mean mate-pair insert size of 1.25 kb with 86% Q_{30} bases with a total genome size of 6.0 Mb was determined (Figure 9) (Ewing and Green, 1998; Ewing et al., 1998).

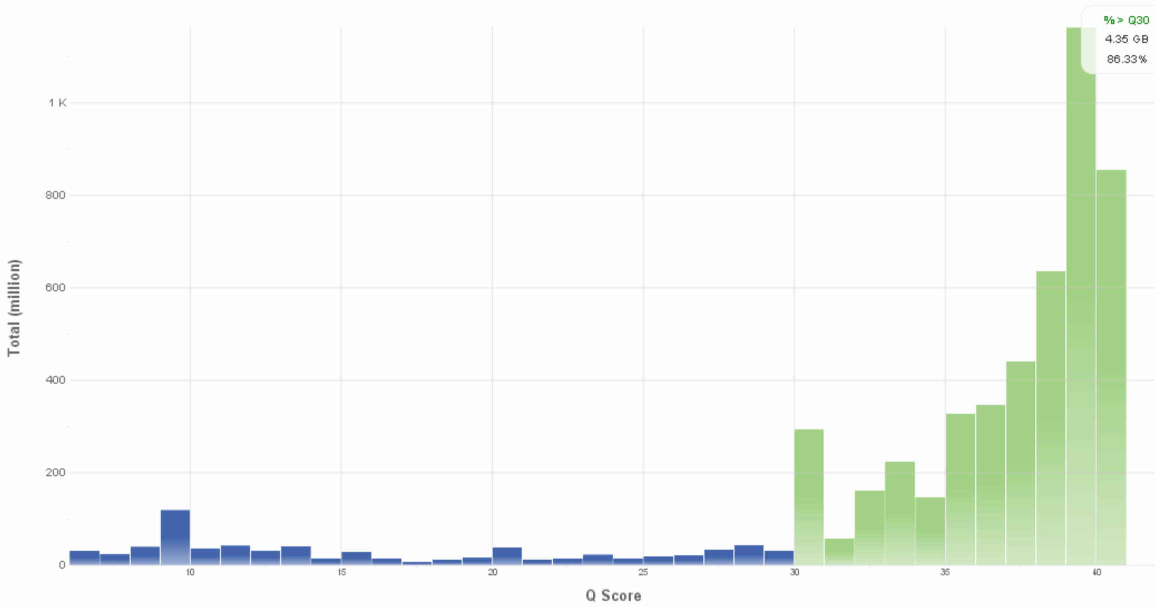


Figure 9. Phred-like quality score for *P. polymyxa* genome sequencing. The Phred-like quality score (Q-score) is a probability measure of an incorrect base call for a given base, where $Q = -10 \log_{10} P$, and is considered the gold standard for assessing the accuracy of Illumina sequencing runs. The green portion of the histogram corresponds to bases with a Q-score equal or greater than 30, corresponding to 99.9% accuracy for any given base.

Continual development of second-generation sequencing platforms has yielded ever-larger data sets of shotgun genomic sequence, requiring constant amelioration of computational read alignment and assembly algorithms to ensure high fidelity assemblies. Assembly of sequencing reads into contiguous sequences (contigs) relies on iterative overlapping of millions of reads using mathematical models (typically de Bruijn graphs) (Delseny et al., 2010). Different programs vary on the specifics of the algorithm, cut-offs and parameters used to ensure assembly fidelity, which can have a substantial impact on the assembly output. In addition, variability in the k-mer value of each program has a significant impact on the resulting assembly. Programs with a short k-mer value are highly contiguous at the expense of assembly fidelity. Conversely, programs with long k-mer values are comparatively fragmented, albeit with higher assembly fidelity (Gibbons et al., 2009). Recently, progress has focused on the development of contig integration programs, designed to amalgamate advantages of multiple different contig assembly programs while minimizing shortcomings inherent to each program.

The *P. polymyxa* CR1 genome was assembled *de novo* using three separate contig assembly programs SOAPdenovo (k-mer = 55), ABySS (k-mer = 67) and Velvet (k-mer= 37). The three resulting contig assemblies were integrated into a single assembly using CISA on standard parameters. The CISA integrated draft assembly of the *P. polymyxa* CR1 genome contained 38 contigs with an N_{50} contig size of 1.5 Mbp (N_{50} refers to the size of the contig, ordered in descending order by size where 50% of the total bases in the genome are represented).

Novel Gap Closure Method

A genome is defined as finished when all ambiguous regions have been sequenced, rearrangements investigated, and gaps between contigs determined, resulting in an accurate contiguous genetic element (or elements in the case of multiple chromosomes or plasmids) (Tsai et al., 2010; Wetzel et al., 2011). This contrasts with draft genomes where the presence of sequencing/assembly mistakes, chimeric regions and assembly artifacts may not have been addressed (Mardis et al., 2002). Despite the rate of sequencing developments outpacing Moore's Law of Computing, concomitant advances in bacterial genome finishing techniques and methods have not maintained pace. The short read lengths of second-generation technologies limits the ability of current platforms to cross multi-copy features of genomes longer than the insert length. First manifesting as a limitation in the assembly of telomeres of mammalian cells, limitations created as a consequence of read length have become apparent for the assembly of finished prokaryotic genomes as well.

Prokaryotic genes for ribosomal subunit RNA are encoded in a linear operon consisting of 16S, 23S and 5S subunits (here forth referred to as rDNA when discussing chromosomally encoded genes/operons and rRNA when referring to mature ribosomal RNA). These operons are multi-copy in any given prokaryotic cell and copy number is correlated to the environmental response rate of the bacterium (Klappenbach et al., 2000). Assembly algorithms are unable to assemble short-insert reads into rDNA-spanning contigs since short-insert read chemistry (on the Illumina platform) is currently limited to ~2x250bp, meanwhile, rDNA reaches upwards of 5kb in length.

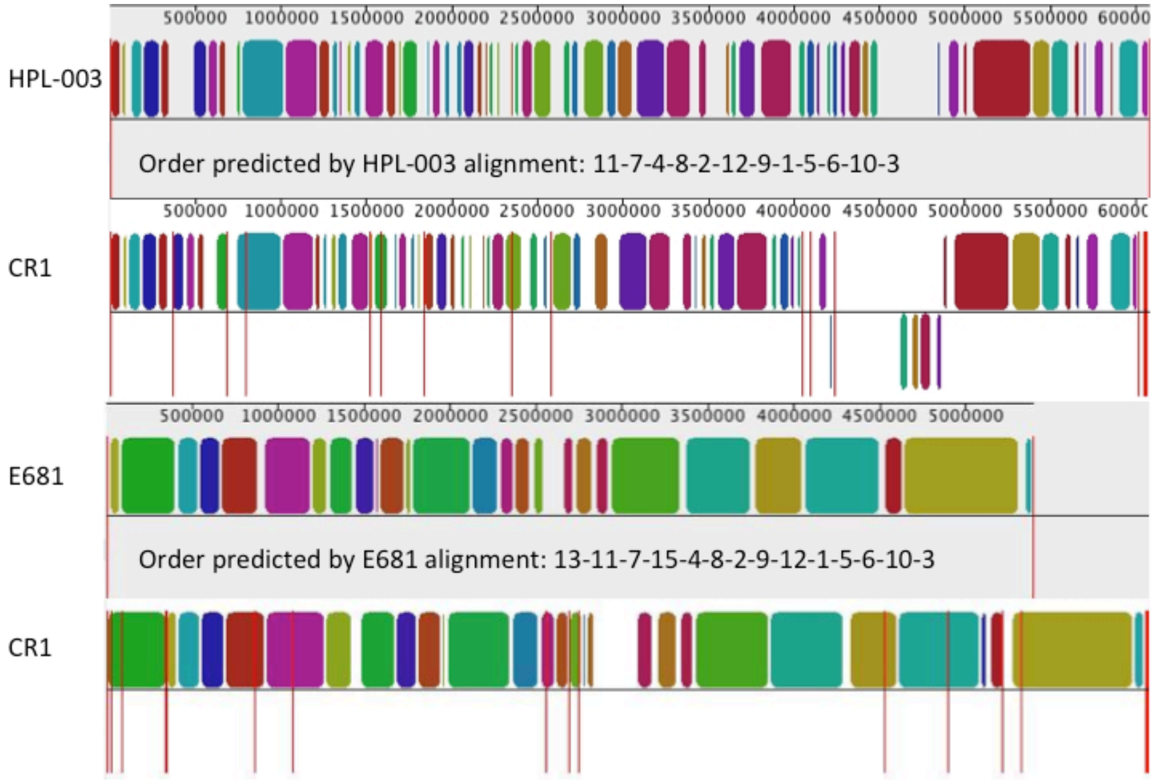


Figure 10. Comparison of scaffolding against closest related strain versus species. The local collinear block (LCB) plot was generated with the contig reordering tool within Mauve using default parameters. The name of the strain represented is listed to the left of each LCB plot. HPL-003 represents *Paenibacillus terrae* strain HPL-003, a closely related species to *P. polymyxa*. Conversely, E681 represents *P. polymyxa* strain E681, the closest related *P. polymyxa* strain to CR1. The contigs contained within the *P. polymyxa* strain CR1 draft genome are reordered to approximate the LCB plot of the above completely finished genome. Global alignments are visualized as LCBs, which represent regions with high levels of nucleotide similarity between genomes. LCBs are colored according to homology to LCBs of the compared genome. LCBs drawn below the horizontal line correspond to inversions relative to the reference genome.

Scaffolding of the *P. polymyxa* CR1 draft genome against the closely related *P. polymyxa* E681 and *Paenibacillus terrae* HPL-003 allowed for identification of putative contig gaps for experimental confirmation (Figure 10).

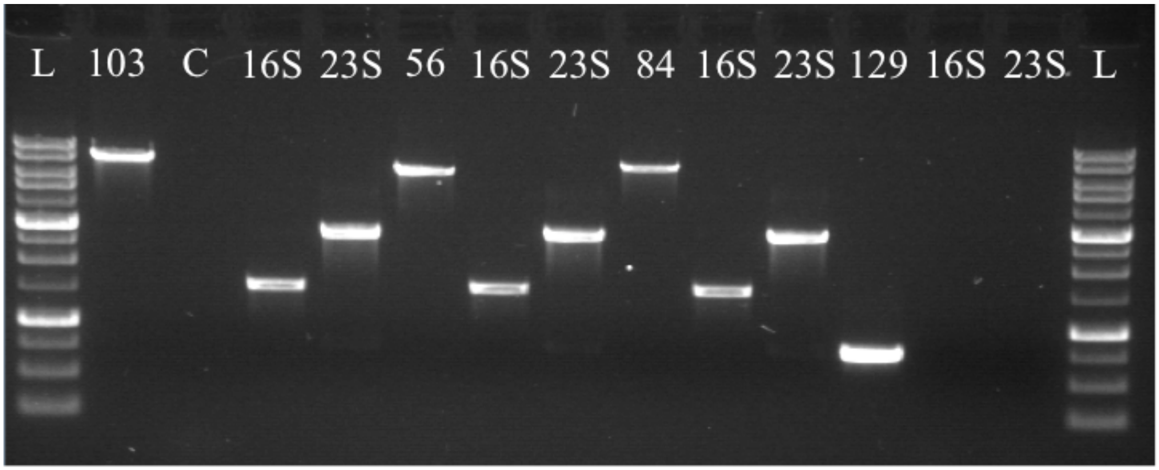


Figure 11. Confirmation of rRNA operons within contig gaps. PCR products joining the ends of two contigs were used as templates for PCR to determine presence of rRNA genes within the gaps. Gaps that tested positive for rRNA operons were closed using our innovative rRNA gap sequencing procedure as illustrated in Figure 12. Gaps that did not contain rRNA operons required traditional primer walking to determine the missing sequence. PCR products labeled as 16S and 23S rRNA genes were amplified from the respective preceding PCR product corresponding to the following gaps; 10/3 – gap between contig 10 and contig 3, 5/6 – gap between contig 5 and contig 6, 8/4 – gap between contig 8 and contig 4, 12/9 – gap between contig 12 and contig 9. L represents a 1 kb molecular weight ladder and C is a representative template control utilizing distant primers not expected to be contained within the 10/3 gap to confirm template purity.

For sequencing of gaps identified by the Mauve alignments of the draft genome against the *P. polymyxa* reference genome, primers were designed to the ends of identified contig gaps. Primers were designed to specifically hybridize to protein coding sequences, identified by our RAST annotation, to minimize off-target products. Long and Accurate PCR allowed for targeted amplification of a desired contig gap, which was then gel purified prior to use as a template for a modified primer walking parallel strategy. In our method, a candidate contig gap was used as a template in end-point PCR with universal 16S and 23S primers (Table 3 and Figure 11). If the candidate PCR amplified contig gap was positive for presence of 16S and/or 23S rDNA, the entire sequence of the rDNA contained within the contig gap could be determined in parallel Sanger sequencing reactions using the primers listed in Table 3 (schematic of priming locations shown in Figure 12). These primers were designed to hybridize to highly conserved regions of 23S and 16S rDNA, working in conjunction with previously established and widely used universal primers to generate overlapping, oppositely oriented sequences to bridge rDNA (Figure 12).

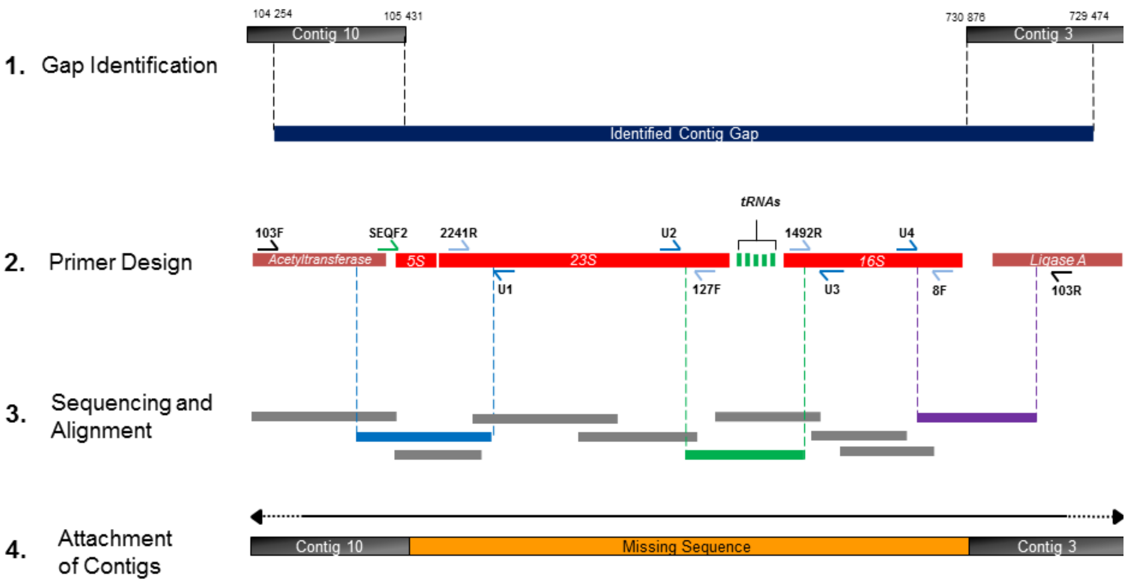


Figure 12. Identification and resolution of rearrangement in draft contig assembly. The combined contigs of 13, 11, 7, and 15 was predicted to be located within an ambiguous base stretch based on Mauve alignments, which was confirmed by PCR amplification of the adjoining ends of contigs 4(5′N) to contig 13 and contig 15 to contig 4 (3′N) as shown by lanes 4/13 and 4/15, respectively. L represents a 1 kb molecular weight ladder, 16S and 23S genes were amplified from the preceding PCR product correspond to the following contig gaps; 3/4 – gap between contig 3 and contig 4, 4/15 – PCR product corresponding to insertion of contig 15 within an ambiguous base stretch of contig 4, 4/13 – PCR product corresponding to insertion of contig 13 within an ambiguous base stretch of contig 4.

Genome Annotation

The completed genome of *P. polymyxa* CR1 was 6.02Mb and was annotated using the NCBI Prokaryotic Genome Automatic Annotation Pipeline and annotations were used irrespective of genes previously annotated using the RAST server.

Comparative Genomics of *P. polymyxa* Strains

The sequencing of the *P. polymyxa* CR1 genome represented the fourth completely sequenced strain of *P. polymyxa* and other researchers had generated a range of useful information pertaining to the previously sequenced strains. Taking advantage of the wealth of information contained within a complete genome sequence is a daunting task. Comparative genomics identifies genes relevant to desired phenotypes, mechanisms, and pathways conserved between organisms that have been previously identified in related bacteria, allowing researchers to place their organism within the framework established previously.

General Features

General features of the four completely sequenced *P. polymyxa* genomes are presented in Table 5. Interestingly, the genome size and structure is quite variable between strains, with a size difference of approximately 700kb between the chromosomes of *P. polymyxa* E681 versus *P. polymyxa* CR1, as well as the presence of two seemingly unrelated plasmids in *P. polymyxa* SC2 and *P. polymyxa* M1 (here forth E681, CR1, SC2 and M1 respectively). Plasmids encoded by SC2 and M1 have lower G + C mol % content compared to the genome in accordance with previously reported research (Nishida, 2012). The species has a mean G + C mol% content of 45.4% with between 12 (E681/CR1) and 14 (SC2/M1) ribosomal DNA operons encoded by each strain. The variation in protein coding sequences is striking with SC2 encoding 337 more CDS than M1, despite a 164kb difference in genome size between the strains (Table 5). The genome of CR1 contains 5306 coding sequences and a relatively large genome size of 6 Mb despite the absence of a plasmid.

Table 5. General genome features of completely sequenced *Paenibacillus polymyxa* strains

	CR1	**E681**	**SC2**	**M1**
Accession Numbers	NC_023037	NC_014483	NC_014622	NC_017542
Location of Isolation	Rhizosphere	Rhizosphere	Rhizosphere	Roots
Genome Size (base pairs)	6 024 666	5 394 884	5 731 816	5 864 546
GC content (%)	45.6	45.8	45.2	44.8
Coding Sequences	5 306	4 805	5 406	5 069
Plasmid Size (base pairs)	NA	NA	510 115	366 576
Pseudogenes	217	1	52	
rRNA genes	36	36	42	42
tRNA genes	87	91	110	110
Other RNA genes	1			
Conserved CDS	3463	3457	3505	3338
Strain Specific CDS	955	443	11	121

NA, not applicable, refers to strains in which no plasmids are naturally present. Accession numbers refer to the genome sequence entry in NCBI Nucleotide database. Coding sequences, pseudogenes, and RNA genes were identified from available annotations in the NCBI Genebank database. tRNA and rRNA genes were re-identified using tRNAscan and RNAmmer respectively. Conserved and strain-specific sequences were determined using mGenomeSubtractor, with an H-value cut-off 0.81 and 0.41 respectively.

Of the total 5306 genes encoded by CR1, 955 are strain-specific and 3463 are conserved amongst strains (Table 5); representing 18.1% of the total genes in the CR1 genome as strain-specific, significantly higher than other *P. polymyxa* strains (9.2%, 0.2% and 2.4% for E681, SC2 and M1 respectively).

Despite differences in genome size, CDS and strain-specific genes, genetic structure amongst *P. polymyxa* strains is highly conserved (Figure 13) with SC2 and M1 showing almost identical structure (compared local collinear block composition between SC2 and M1). However, readily apparent is the relative dissimilarity of CR1, as opposed to any other grouping of strains.

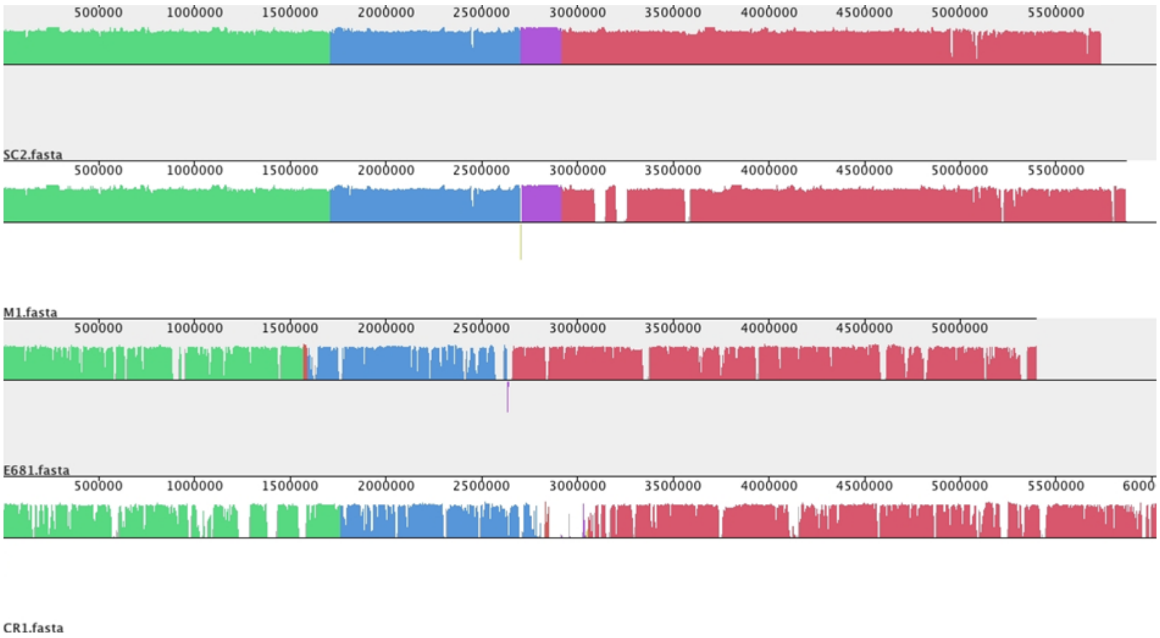
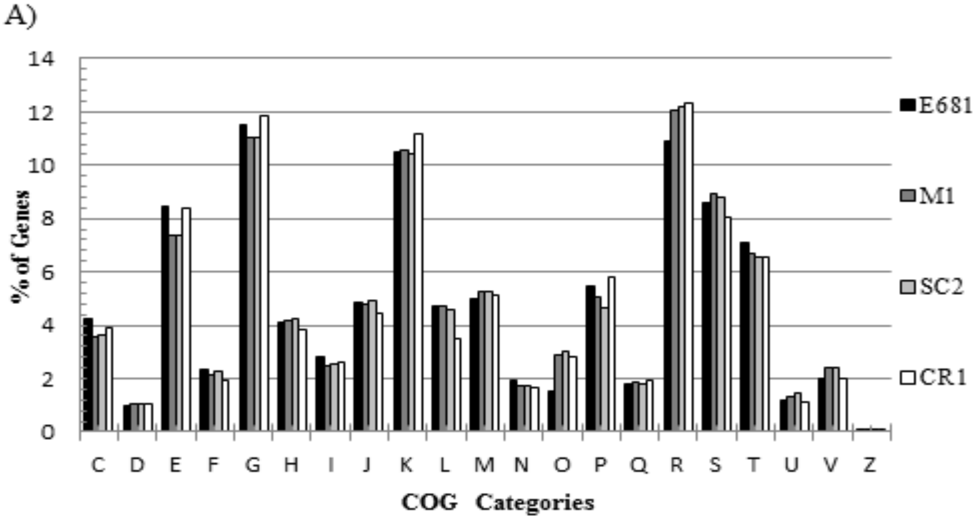


Figure 13. Global alignment of chromosomes of completely sequenced *P. polymyxa* strains. Local collinear block plot was generated using the progressiveMauve algorithm using default parameters. The name of each strain is listed below block plot, which represent regions of chromosomal similarity amongst strains. Regions without colour represent the presence of strain-specific sequence. Regions drawn below the horizontal correspond to inversions.

A comparison of COG category composition of the *P. polymyxa* CR1 genome versus other sequenced *P. polymyxa* strains identified both a larger absolute number and a higher proportion of genes dedicated to energy metabolism, inorganic transport and metabolism (Figure 14).



B)

COG	Description	COG	Description
C	Energy Conversion	N	Cell Motility
D	Cell Cycle and Division	O	Protein Modification/Folding
E	Amino Acid Metabolism	P	Inorganic Ion Metabolism
F	Nucleotide Metabolism	Q	Secondary Metabolites
G	Carbohydrate Metabolism	R	General Function
H	Coenzyme Metabolism	S	Function Unknown
I	Lipid Metabolism	T	Signal Transduction
J	Translation	U	Trafficking and Secretion
K	Transcription	V	Defense Mechanisms
L	DNA Replication and Repair	Z	Cytoskeleton
M	Cell Wall/Membrane Biogenesis		

Figure 14. COG functional categorization of sequenced *Paenibacillus polymyxa* genomes. Functional categorization was performed using available tools on the JGI IMG database. A) Proportion of total CDS versus COG categories, categories A, B, W, and Y correspond to eukaryotic functions and are thus omitted. B) List of COG categories and their respective functions.

To identify potential horizontally transferred genes for later corroboration of lignin metabolism genes, we identified genomic islands using IslandViewer2.0, phage-related genes using PHAST, tandem repeats using TandemFinder, and insertion sequences using the ISFinder database. To visualize the general structure of the *P. polymyxa* CR1 genome, plus and minus strand CDS, RNA genes, transposons, phage-related genes, insertion elements, putatively horizontally transferred genes, and strain-specific genes were annotated in Artemis and plotted using DNAPlotter (Figure 15). Genes contained within identified genomic islands of CR1 are mainly hypothetical proteins without an annotated function, however interesting genes encoded within genomic islands include antimicrobial compound synthesis clusters as well as a minimal *nif* cluster. Furthermore, a large number of insertion sequences, prophages and tandem repeats were localized to genomic islands suggesting CR1 is a common phage target, with genomic islands encoding multiple phage-related genes (Appendix 1).

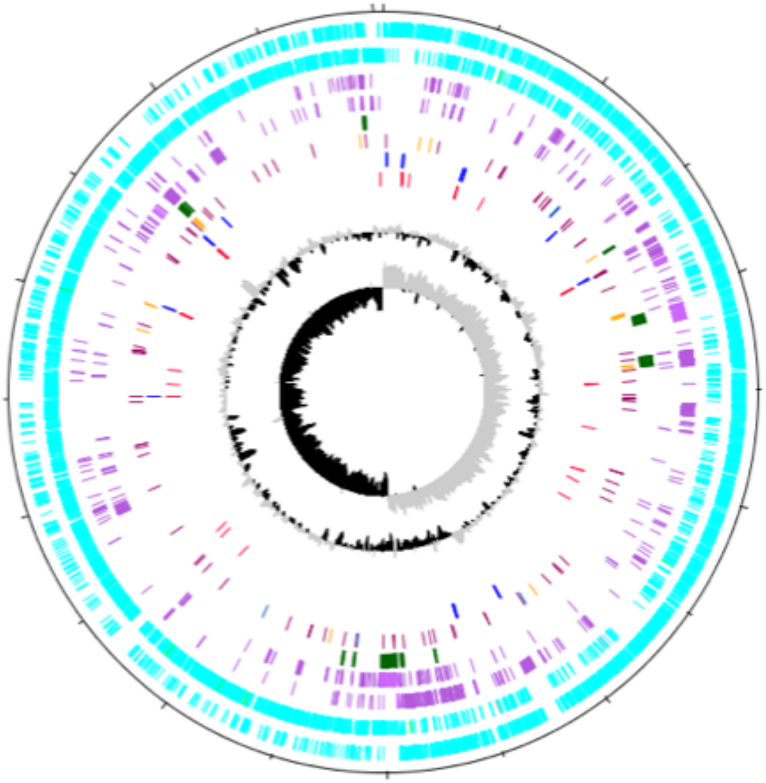


Figure 15. Circular representation of the *P. polymyxa* CR1 genome. Rings represent the following features labeled from outside to centre, where the outermost circle represents the scale in Mbps where each tick mark represents 250Mbps. 1st ring; plus-strand CDS (cyan), 2nd ring; minus-strand CDS (cyan), 3rd ring; plus-strand strain specific CDS (purple), 4th ring; minus-strand strain specific CDS (purple), 5th ring; putative horizontally transferred genes (dark green), 6th ring; phage-related genes (orange), tandem repeats (brick red), transposons (dark blue), 7th ring; ribosomal rRNA genes (bright blue), 8th ring; tRNA genes (red), 9th ring; GC-plot where black and grey correspond to above and below average GC content respectively, 10th ring; GC-skew where black and grey correspond to above and below average GC-skew respectively. Strain-specific genes were identified using mGenomeSubtractor with an H-value cut off of ≤0.41. Putative horizontally transferred genes were identified using IslandViewer 2.0. Annotation was obtained from the NCBI GeneBank database. Phage genes, tandem repeats and transposons were identified using PHAST and IS Finder, respectively. rRNA and tRNA genes were obtained from available annotations.

Phylogeny

A phylogenetic tree was generated using the Maximum-likelihood method based on 16S RNA sequences of available completely sequenced species within the *Paenibacillus* genus (Figure 16 panel A). A whole-genome neighbour-joining phylogenetic tree was generated using the dnadist and neighbour packages in Phylip, and visualized using PhyloXML (Figure 16 panel B).

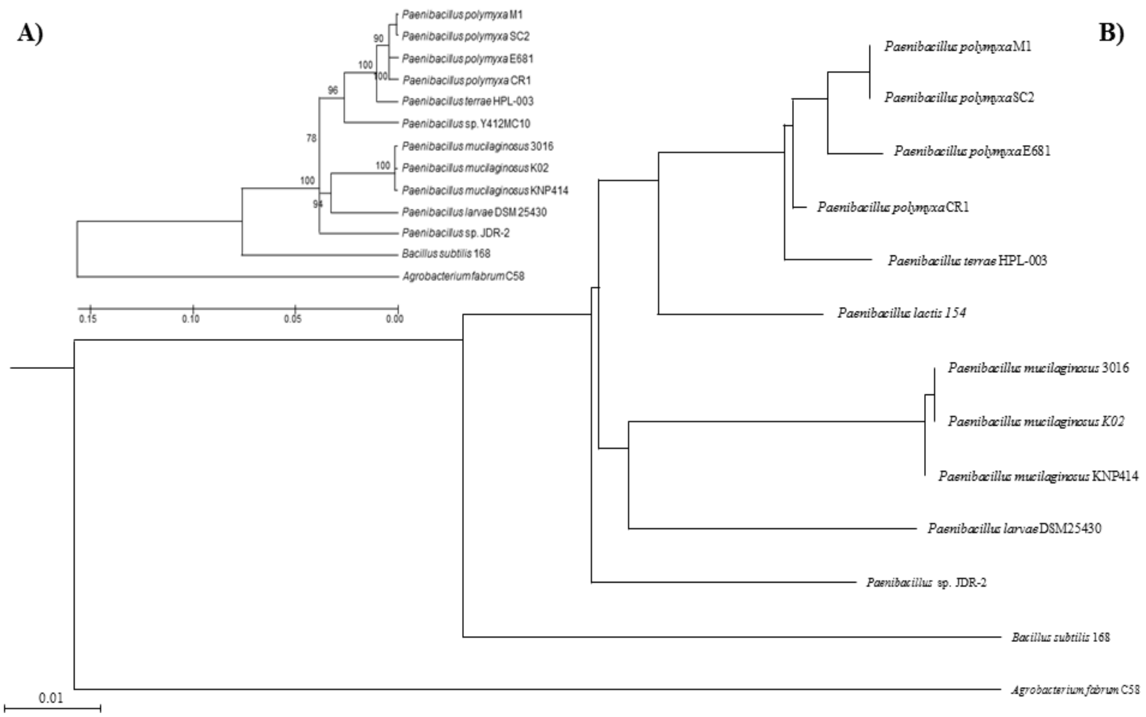

A)
Paenibacillus polymyxa M1
Paenibacillus polymyxa SC2
Paenibacillus polymyxa E681
Paenibacillus polymyxa CR1
Paenibacillus terrae HPL-003
Paenibacillus sp. Y412MC10
Paenibacillus mucilaginosus 3016
Paenibacillus mucilaginosus K02
Paenibacillus mucilaginosus KNP414
Paenibacillus larvae DSM 25430
Paenibacillus sp. JDR-2
Bacillus subtilis 168
Agrobacterium fabrum C58
90
100
100
96
78
100
100
94
0.15
0.10
0.05
0.00
B)
Paenibacillus polymyxa M1
Paenibacillus polymyxa SC2
Paenibacillus polymyxa E681
Paenibacillus polymyxa CR1
Paenibacillus terrae HPL-003
Paenibacillus lactis 154
Paenibacillus mucilaginosus 3016
Paenibacillus mucilaginosus K02
Paenibacillus mucilaginosus KNP414
Paenibacillus larvae DSM25430
Paenibacillus sp. JDR-2
Bacillus subtilis 168
Agrobacterium fabrum C58
0.01

Figure 16. Phylogenetic tree of completely sequenced *Paenibacillus polymyxa* strains. A) Sequences of complete genomes were obtained from the NCBI Nucleotide database. The phylogeny was generated in MEGA6 using the maximum-likelihood method with 1000 bootstrap replications. Numbers at each branch point correspond to the proportion of positive results from bootstrapping. B) Neighbour-joining whole-genome phylogram generated using the dnadist and neighbor packages in PHYLIP visualized using phyloXML. Branch lengths are representative of the number of nucleotide substitutions per site. *Agrobacterium fabrum* C58 was used as an out-group.

As expected, *P. polymyxa* strains form a monophyletic group when 16S sequences are taken alone, however when the whole genome is used to compute phylogeny, CR1 appears to form its own subclade. The close relationship between *P. polymyxa* and *P. terrae* in our phylogram is not surprising considering the species show remarkable genomic structural homology (Figure 10) and harbor a unique nitrogen-fixation cluster found only in specific strains of *Paenibacillus* sp (Figure 17).

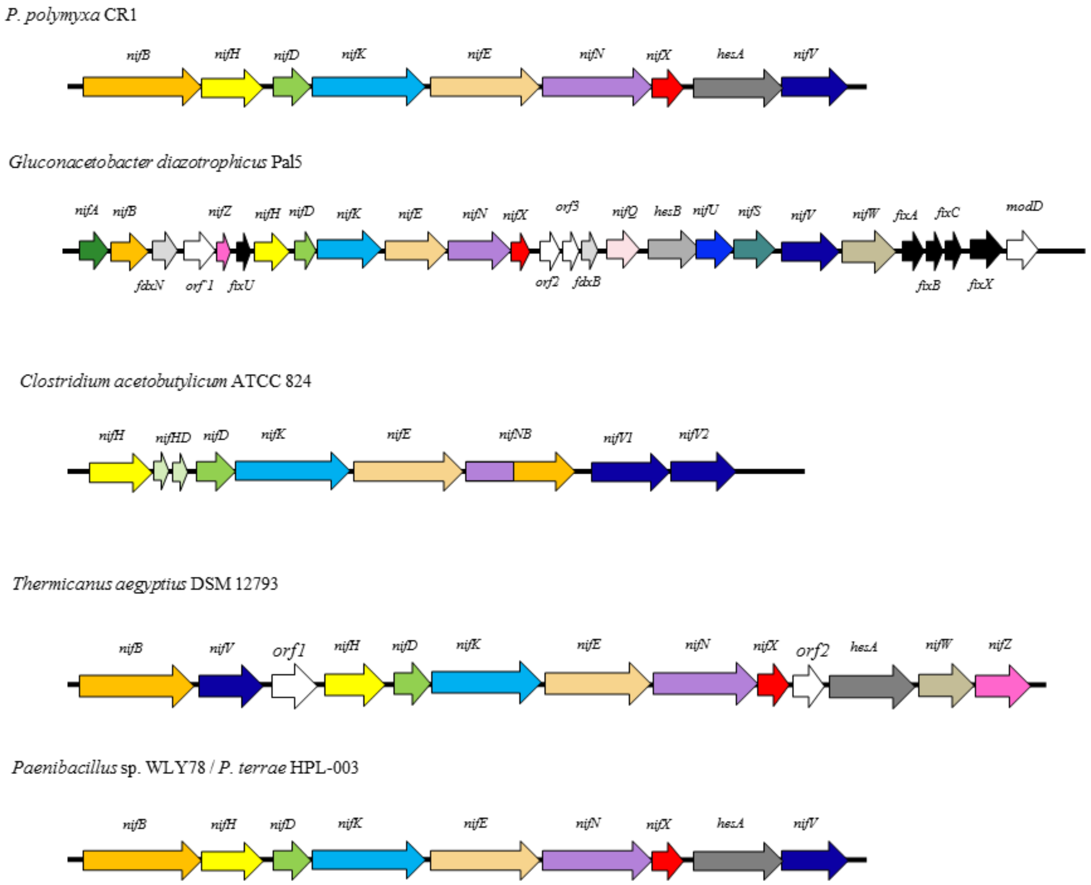


Figure 17. Comparison of *Paenibacillus polymyxa* CR1 *nif* cluster to other free-living diazotrophic bacteria. Genes indicated by the same colour represent functional or structural homologs. Cluster homology is based off of gene clustering using available tools on the JGI Integrated Microbial Genomics Database. Representative *nif* clusters encoded by other free-living diazotrophic bacteria are included for comparison. Bacteria from *Rhizobia* are excluded due to the relative complexity of their nitrogen fixation clusters gene organization, as well as their requirement for nodulation, a trait not observed in *Paenibacillus* sp.

Plant-Derived Carbon Metabolism

P. polymyxa strains encode a wide variety of genes involved in plant association and plant-derived compound metabolism, which are highly conserved amongst strains (Table 6). As previously mentioned, CR1 is the only *P. polymyxa* strain that encodes a functional nitrogenase (Figure 17). In natural systems, the majority of bacteria are not capable of fixing diatmospheric nitrogen to ammonia, which requires absence of oxygen to prevent irreversible inhibition of nitrogenase, the enzyme responsible for nitrogen fixation (Bohlool et al., 1992). Bacteria that are capable of fixing nitrogen typically do so in association with plant hosts (Oldroyd and Dixon, 2014). For example, Rhizobia are well known for their interaction with legumous plants where they form nodules, specialized organs that provide the bacteria with nutrients, prevent oxygenation of nitrogenase, and allow for a plant-mediated differentiation into specialized nitrogen-producing bacterioids (Hirsch, 2015). As of yet, no species of *Paenibacillus* have been found that encodes *nod* genes, which are responsible for controlling nodulation, suggesting these complex relationships are not formed by *Paenibacillus* sp. (Wang et al., 2013). Additionally, *nifA*, the gene responsible for activating expression of plant-derived compound transporters and the nitrogenase structural gene operon *nifHDK*, is yet to be identified in any *Paenibacillus* sp.. In Rhizobia, *nifA* is dispensable since NtrC (nitrogen regulatory protein C) is able to activate expression of the *nif* cluster to 80% of wild-type expression in Δ*nifA* strains (Labes et al., 1993; Labes and Finan, 1993). Furthermore in Rhizobia, DctD is a C4- dicarboxylate response regulator involved in C4 carbon uptake and complex plant-microbe signaling pathways (Yurgel and Kahn, 2004). Interestingly, CR1 appears to encode a hybrid DctD/NtrC homolog (YP_008912290) that may function in plant signaling and plant-derived carbon source perception. Upon perception of plant derived carbon compounds *P. polymxa* CR1 induces expression of plant-dervied compound transporters and nitrogenase.

Table 6. Plant-growth promoting traits of sequenced *P. polymyxa* strains.

Trait	*Gene Name*	CR1	E681	M1	SC2
Indole-3-acetic acid production					
	ipdC	YP_008911027	YP_003869749	YP_005958986	YP_003945692
	auxin efflux carriers	YP_008912813 YP_008912323 YP_008911849	YP_003871347 YP_003870860 YP_003870585	YP_005960839 YP_005960204 YP_005959850	YP_003947565 YP_003947053 YP_003946661
Phosphate solubilization					
	gcd	YP_008912273	YP_003870830	YP_005960174	YP_005960174
Phosphonate cluster (*phn*)					
	phnA	YP_008914717	YP_003873107	YP_008050528	YP_003949521
	phnB	YP_008910326	YP_003869234	YP_003945144	YP_005958491
	phnC	YP_008913947	YP_003872434	YP_008049904	YP_003948836
	phnD	YP_008913946	YP_003872433	YP_008049903	YP_003948835
	phnE	YP_008913948	YP_003872435	YP_008049905	YP_003948837
	phnW	YP_008914692	YP_003873086	-	-
	phnX	YP_008909947	YP_003868868	YP_005958118	YP_003944734
	ppd	YP_008914693	YP_003873087	-	-
	pepM	YP_008914694	YP_003873088	-	-
Phosphate transporter (*pst*)					
	pstS	YP_008911198	YP_003869955	YP_005959210	YP_003945962
	pstA	YP_008911200	YP_003869957	YP_005959212	YP_003945964

	pstB	YP_008911201	YP_003869958	YP_005959213	YP_003945965
	pstC	YP_008911199	YP_003869956	YP_005959211	YP_003945963
	phoP	YP_008911212	YP_003869969	YP_005959224	YP_003945977
	phoR	YP_008911211	YP_003869968	YP_005959223	YP_003945976
Nitrogen Fixation					
	nifB	YP_008910495	-	-	-
	nifH	YP_008910496	-	-	-
	nifD	YP_008910497	-	-	-
	nifK	YP_008910498	-	-	-
	nifE	YP_008910499	-	-	-
	nifN	YP_008910500	-	-	-
	nifX	YP_008910501	-	-	-
	hesA	YP_008910502	-	-	-
	nifV	YP_008910503	-	-	-

"-" corresponds to no genes with homology to the gene listed on the left. Accession numbers listed refers to protein sequences in the NCBI Protein database. Genes were identified using annotations provided in Genebank followed by BLASTx searches of the genomes using previously characterized homologs. Auxin efflux carrier proteins were identified using Transporter Classification on the JGI IMG database.

The *Bacillus subtilis* DegS/DegU two-component system is involved in the regulation of post-exponential phase processes. During the transition from exponential to stationary phase, *B. subtilis* secretes a variety of hydrolytic and proteolytic enzymes, which are controlled at the transcriptional level by the DegS/DegU two-component system (Msadek et al., 1991; Murray et al., 2009). All sequenced strains of *P. polymyxa* encode a DegS/DegU two-component system, suggesting biomass metabolism may be coupled to cell cycle stages in *P. polymyxa* strains. Further research is needed to confirm the link between the DegS/DegU two-component system and biomass metabolism in *Paenibacillus sp*.

We compared encoded transporter families and their specificities amongst sequenced *P. polymyxa* genomes to gain insights into metabolism intermediates transported and by extension, potential pathways used during lignin metabolism (Appendix 4, Figure 18). CR1 encodes both a larger absolute number and higher proportion relative to genome size of ATP-binding Cassette transporters. Additionally, transporters corresponding to cellobiose, arabinose, chitobiose and various other plant-derived compounds are encoded by CR1, corroborating the wide metabolism of plant-derived carbon sources.

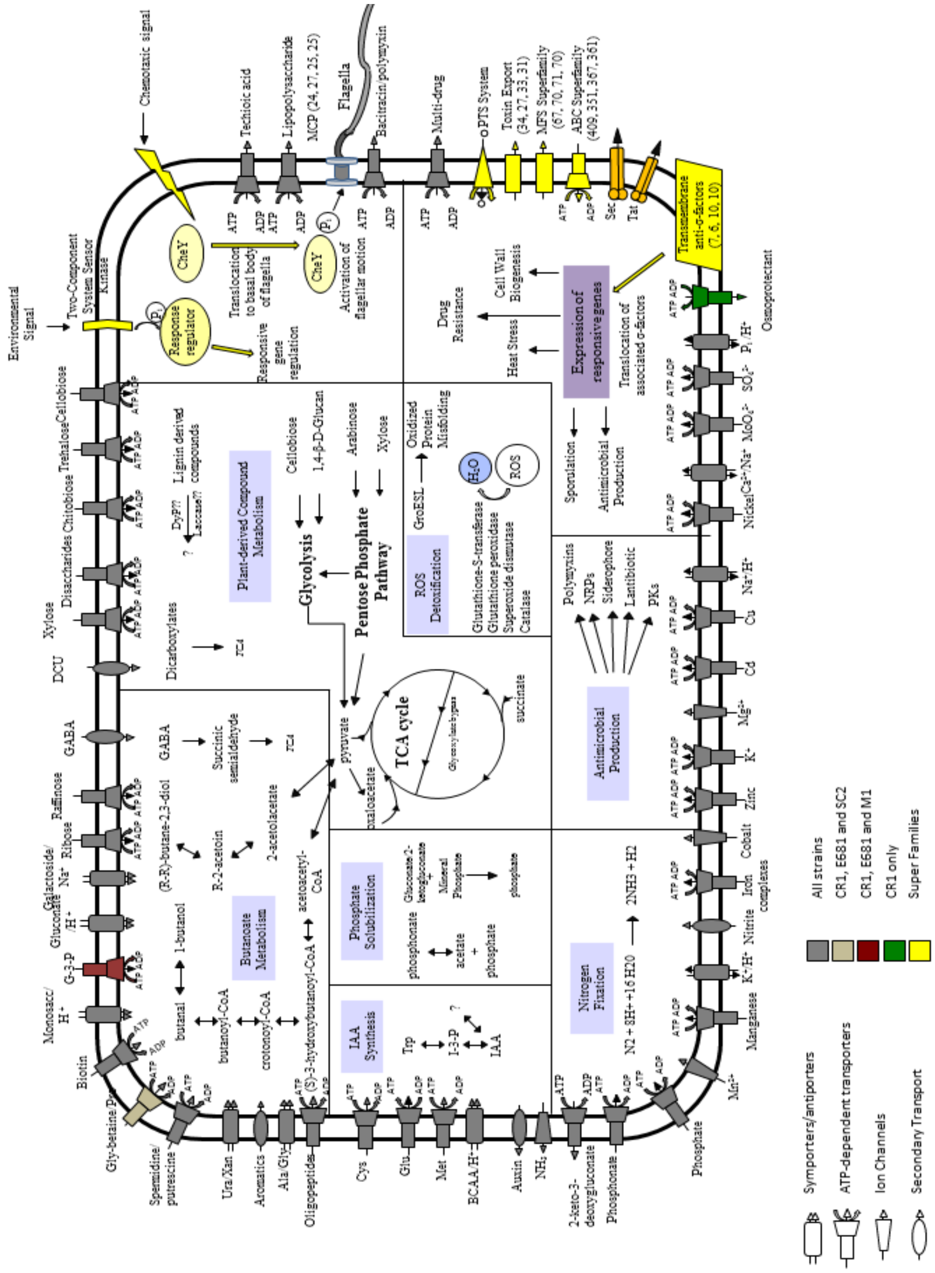


Figure 18. Schematic summary of *Paenibacillus polymyxa* metabolism. Listed beside each superfamily is the number of CDS found in the following order; CR1, E681, M1, SC2. Metabolic and regulatory pathways involved in survival in the rhizosphere niche and plant-growth promoting traits are included in the interior of the cell diagram.

Glycoside hydrolase (GH) enzymes are responsible for the hydrolytic cleavage of a wide variety of plant-derived compounds (Henrissat and Daviest, 1997). GHs are divided into families based on their structural homology and function (Lombard et al., 2014). As expected by our COG categorization showing a diverse metabolic profile, CR1 encodes a wide variety of GH family enzymes and the largest repertoire of out of all completely sequenced *P. polymyxa* strains (Appendix 5). Interestingly, CR1 encodes a large number of GH family 1,2,3, and 42, while encoding a fewer amount of GH family 5 proteins. GH family 1,2,3 and 42 correspond to enzymes responsible for cleaving a variety of glycosidic bonds, including all the types of bonds commonly found in cellulose and hemicellulose. No GH family enzyme has previously been linked to lignin metabolism since glycosidic bonds are not found in lignin polymers, however lignin fractions from cellulosic ethanol production unanimously contain cellulose and hemi-cellulose at low concentrations.

As previously reported for other *P. polymyxa* strains under investigation for application in industrial fermentation, *P. polymyxa* CR1 contains the entire repertoire of genes necessary for 1-butanol, ethanol, acetone and 2,3-butane-diol production (Figure 18), corroborating our preliminary results showing production of various alcohol compounds (Figure 8).

No strain of *P. polymyxa* was found to encode genes involved in carbon fixation, ruling out the possibility that strains were fixing carbon dioxide to obtain necessary carbon. All strains of *P. polymyxa* encode Dyp-type peroxidases and laccase-like oxidase homologs

(Table 7), supporting our findings that CR1 is capable of metabolizing extracellular methylene and toluidine blue dyes.

Table 7. Putative lignolytic enzymes of sequenced *Paenibacillus polymyxa* strains.

Gene Homology	**Accession Number**			
	CR1	**E681**	**M1**	**SC2**
DyP peroxidase	YP_008913589	YP_003872107	YP_008049549	YP_003948444
Laccase	YP_008912908	YP_003871475	YP_008048909	YP_003947734

Location refers to the chromosome location on the genome of each respective strain. Accession numbers refer to the protein sequence contained in the NCBI Protein database. Genes encoding putative Dyp type peroxidases were identified based off of BLASTx homology with DyPB from *Rhodococcus jostii* RHA1. Laccase enzymes were based off of annotations provided by the NCBI Genebank database and homology to fungal encoded laccase enzymes.

Unexpectedly, our comparative analyses failed to identify homologs to genes involved in aromatic compound metabolism in bacteria. Bacterial lignin metabolism is believed to occur in one of two fashions: either extracellular catabolism into short chain carbon molecules and subsequent uptake, or extracellular catabolism into aromatic compounds followed by internalization where the compounds act as intermediates in aromatic metabolism pathways (Pollegioni et al., 2015). *P. polymyxa* CR1 does not appear to encode any genes involved in the metabolism of protocatechuate (3,4-dihydroxybenzoic acid), a key nodal intermediate in bacterial aromatic metabolism pathways . As expected by our *in silico* analyses, *P. polymyxa* CR1 is unable to metabolize protocatechuate, as well as a multitude of other similar aromatic compounds as a carbon source (Figure 19). Absence of an identifiable aromatic metabolic network suggests *P. polymyxa* CR1 is able to gain necessary carbon to support growth, without degrading aromatic components of lignin or by metabolizing lignin-derived aromatic compounds through an unconventional pathway. Dyp peroxidases and laccases are known to oxidize and degrade aromatic compounds to tricarboxylate cycle intermediates supporting the possibility that *P. polymyxa* utilizes non-specific mechanisms for lignin metabolism.

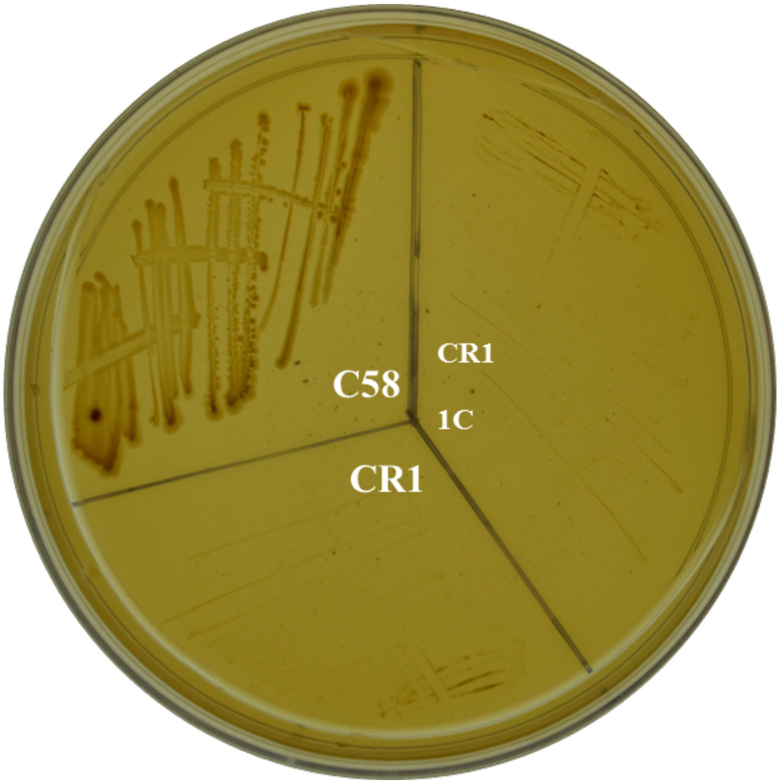


Figure 18. *P. polymyxa* CR1 cannot utilize protocatechuate as a sole carbon source. *P. polymyxa* CR1 and its derivative strain 1C were grown on minimal media with protocatechuate as the sole carbon source for 4 days at 37°C. *Agrobacterium fabrum* C58 protocatechuate degradation has been established previously and acts as a positive control.

Functional Genomics

Bacterial DyP-type peroxidases and laccases have been extensively linked to lignin metabolism in various bacterial species (Thevenot et al., 2010; Brown and Chang, 2014;

Bugg et al., 2011; Pollegioni et al., 2015; Ahmad et al., 2011; Paliwal et al., 2012; Sharma et al., 2006). Therefore, to investigate the possibility that lignin metabolism in *P. polymyxa* CR1 is mediated by the chromosomally encoded Dyp-type peroxidase and/or laccase homologs, insertion knockouts of each respective gene were generated. Phenotypes of the mutants were assessed for defects in lignin metabolism and growth phenotypes on lignin minimal media.

Targeted Knockouts

The locus tag of both DypB and laccase are listed in Table 7. To generate insertional knockout mutants of both genes in *P. polymyxa* CR1, primers were designed to generate PCR products corresponding to a 350 bp region within 50bp of the 5' end of each gene, while incorporating an EcoRI restriction site. These products were cloned into the vector pUCP30T, a pUC18 derivative encoding *aacC1,* and a multiple cloning site within *lacZ*. Previous research has established that pUC18 derivative plasmids do not replicate within *P. polymyxa* (Kim and Timmusk, 2013). The PCR products were cloned into recipient plasmids by restriction digest using EcoRI, prior to heat-shock transformation into chemically competent *Escherichia coli* DH5α cells. After blue-white screening for insert confirmation, plasmids were extracted and desired insertion was confirmed by sequencing.

Plasmid p30T-Dyp and p30T-Lyc were electroporated into electrocompetent *P. polymyxa* CR1, and mutants were isolated based on their ability to grow on LB-Miller +10 ng/mL gentamycin agar plates. Sequencing confirmed gene knockouts were as intended, and gentamycin resistance was not a consequence of cytosolic propagation of plasmids.

CR1 ΔDypB and CR1 ΔLyc were assayed for growth deficiencies on 0.2% w/v lignin minimal media agarose plates. In this experiment, agarose was used in place of agar as a media-solidifying agent to prevent false negatives resulting from metabolism of sugar impurities commonly found in commercial agar.

Unexpectedly, neither CR1 ΔDypB nor CR1 ΔLyc showed any growth defect on 0.2% lignin MM agarose compared to wild-type, suggesting these genes are not absolutely necessary for lignin metabolism. To further confirm a wild-type lignin metabolism phenotype, cell counts from cultures grown in modified M9 media containing lignin, showed no differences in either growth rates or total growth between wild type and knockout mutants (Figure 20).

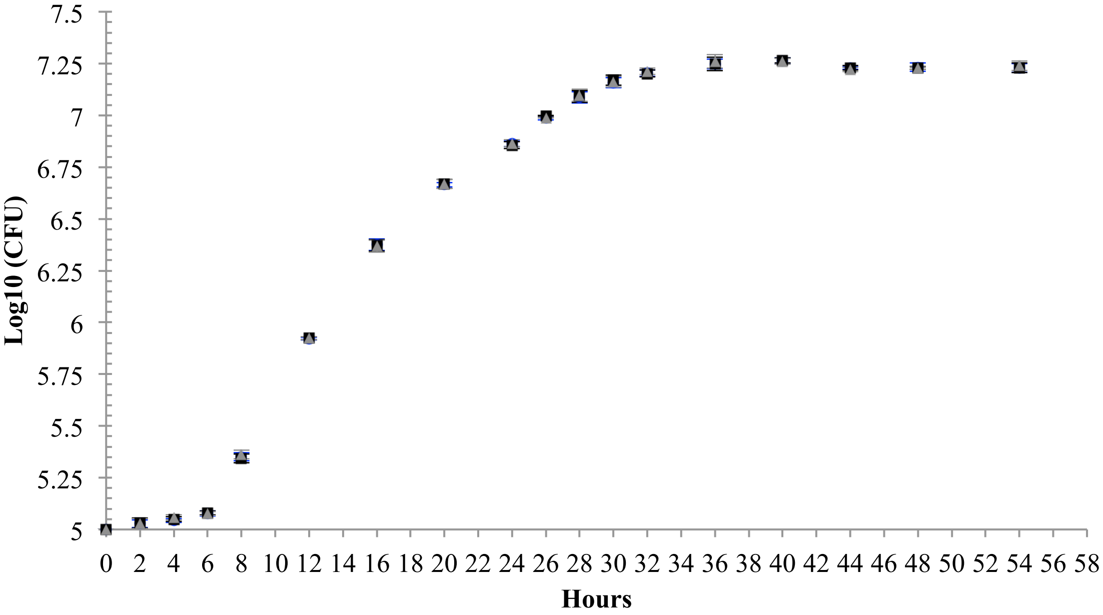


Figure 19. Growth curve of *P. polymyxa* CR1 deletion mutants. Insertional knockout mutants for DyP-type peroxidase and laccase were generated as described in methods. Cells (1x10^5 CFU) were inoculated into 0.2% lignin liquid media and incubated at 37°C with shaking. Growth curves represent three independent experiments repeated in triplicate. CFU measurements were taken by dilution plating onto LB-Miller plates incubated at 37°C.

Transposon Mutagenesis

The combination of bioinformatics and targeted knockout approaches failed to identify genes involved in the lignin metabolism network of *P. polymyxa* CR1, suggesting the bacterium may employ novel, yet uncharacterized delignification mechanisms. Transposon mutagenesis is an established and widely used technique for generating libraries of thousands of random gene knockouts that can be screened for desired or interesting phenotypes (Barquist et al., 2013). Transposons are DNA sequences flanked by two inverted repeats, otherwise known as insertion sequences, enable the fragment to move (or transpose) around a genome. Modified transposons where the encoded transposase has been removed and various selectable markers incorporated are widely used for identifying novel pathways and generating knockout mutants.

Two transposon mutagenized libraries of *P. polymyxa* CR1 clones were generated to screen for lignin growth phenotypes. The first library consisted of a Tn5-Kan transposon and library preparation yielded approximately 5000 mutants. However, attempts to identify insertion locations of the transposon in mutants were inconsistent and further analysis determined many of the clones did not contain the kanamycin resistance gene encoded by the transposon, suggesting *P. polymyxa* readily develops kanamycin resistance via other mechanisms. To avoid potential false-positives, a second transposon library was generated using Tn5-Tet transposons, an antibiotic with a low minimal inhibitory concentration (MIC) in *P. polymyxa* CR1. After electroporation of Tn5-Tet transposons into *P. polymyxa* CR1, cells were recovered for 3 hours before plating on LB-Miller + [10 µg/mL tetracycline]. Libraries were single colony purified on LB_{Tet} media twice prior to screening on 0.2% lignin MM.

Individual isolates from the Tn5-Tet library were streaked onto 0.2% lignin MM agar plates at a density of four unique isolates per plate. The same inoculant stick was used to replicate inoculation onto LB_{Tet} plates to ensure bacterial viability and transfer. Preliminary phenotypic assessment identified clones with increased (Tet_R, Lig^{++}), reduced (Tet_R, Lig^{-}) or no growth (Tet_R, Lig^{o}) using lignin as a sole carbon source. The phenotypic characterization of the 6215 Tn5-Tet CR1 isolates yielded 9 isolates with increased growth (Tet_R, Lig^{++}), 12 isolates with reduced (Tet_R, Lig^{-}) growth and 4 isolates that showed no growth (Tet_R, Lig^{o}). The lignin growth phenotype of Tn5 mutants were confirmed in the same manner as described for gene knockouts prior to preparation for identification of the Tn5 insertion location (Figures 21 and 22).

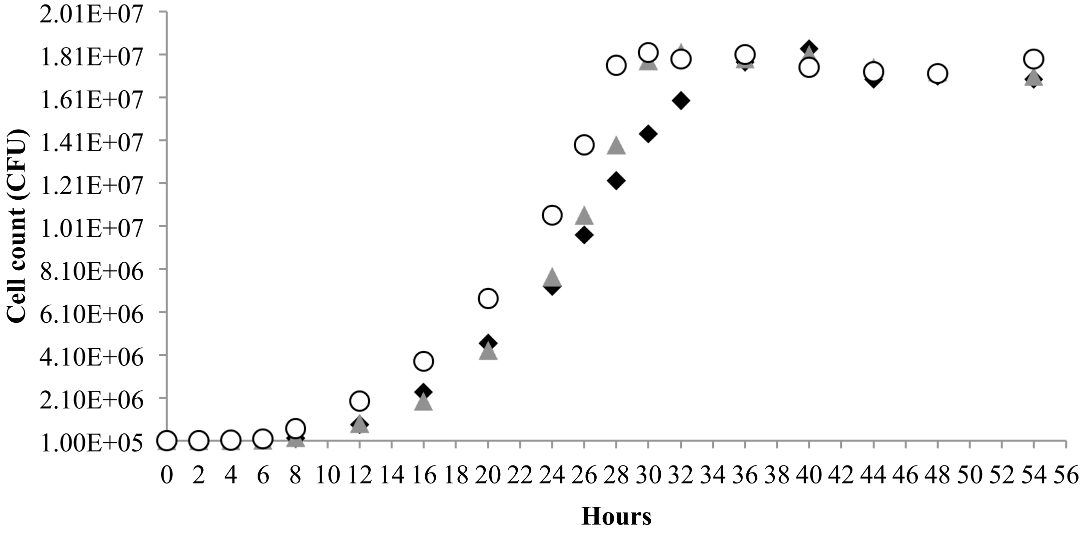


Figure 20. Representative growth curves of fast growing *P. polymyxa* CR1 Tn5 mutants. Black diamond - Wt, grey triangle – Isolate 3G2, white circle – Isolate 2H2. Fast-growing Tn5 mutants followed one of two trends; a shorten lag time or a more rapid expansion of cell numbers during late-exponential phase. Growth curves depict one experiment repeated in triplicate. In all situations cultures reached an approximately same colony forming unit density at steady state. Growth curves for all obtained Tn5 mutants are shown in Appendix 6.

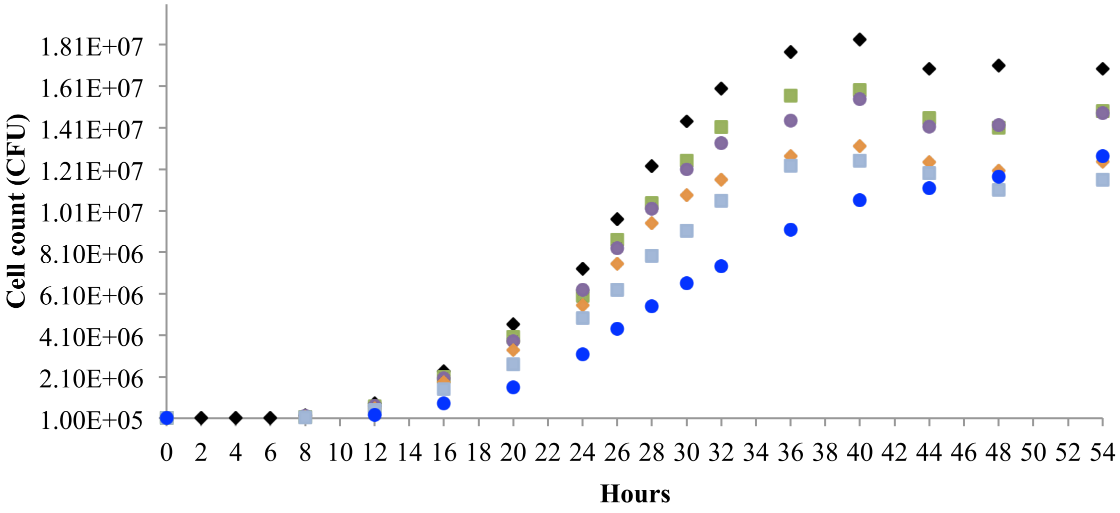


Figure 21. Representative growth curves of slow-growing *P. polymyxa* CR1 Tn5 mutants. Black - Wt, blue – mutant 1D5, light green – mutant 30B5, grey – mutant 30B6, orange – 21B5, light blue – 5E4. Slow-growing Tn5 mutants showed diverse trends in growth. Growth curves depict one experiment repeated in triplicate. Growth curves for all obtained Tn5 mutants are shown in Appendix 7.

Chapter 4 – Discussion

The development of second-generation sequencing has allowed for unparalleled insights into the genetic features and structures of both model and non-model organisms alike (Bentley, 2010). Continual developments in sequencing technologies, chemistry and bioinformatics tools has drastically reduced costs while concurrently increasing the speed of *de novo* whole genome sequencing and allowed for its routine use in the characterization of novel bacterial strains (Medini et al., 2008; MacLean et al., 2009; Delseny et al., 2010). The first bacterial genomes sequenced were published with fanfare and high readership, however, announcements for the availability of new bacterial genomic sequences has become commonplace.

The publicity associated with developments in genome sequencing techniques has occluded the reality of the limitations of our understanding and slow progression to automated complete bacterial genomes. Despite recent progress in genome sequencing speed, the manual finishing stages of the majority of *de novo* bacterial genome sequencing projects remains essentially identical to the approaches used during sequencing of the first bacterial genomes (Hurt et al., 2012). Fortunately the continual publication of new genomes permits the identification of the most probable organization of contigs based off of the results of previous genome sequencing efforts. This subtle change to the sequencing method facilitates a different approach than what is typically performed during bacterial genome sequencing. Researchers commonly use a long-read sequencing platform in conjunction with a draft genome assembled from short-reads. Long-read platforms allow for sequencing of repetitive, misassembly prone regions, while still utilizing the data generating potential of short-read technologies. While robust, this approach does not necessarily result in a finished contiguous genome and has the potential to complicate assemblies by increasing the number of sequencing artifacts, rearrangements and contigs (Hurt et al., 2012).

Our modification to genome finishing takes advantage of the multi-copy nature of rDNA in bacterial genomes, targeting these regions to allow the determination of rDNA containing contig gaps in parallel. Stemming from its central role in biology, rDNA is highly conserved within all bacteria, with genes possessing well-defined regions of variability. Highly conserved, multi-copy DNA elements in conjunction with short read length technologies employed by the majority of bacterial genome sequencing projects, results in rDNA being highly prone to chimerism during *in silico* assembly (Ricker et al., 2012). Assembly algorithms are designed to account for the potential deleterious effects of chimerism, by intentionally ending contigs at repetitive regions, such as rDNA, despite overlapping reads. This results in an enrichment of rDNA and other repetitive features at contig-contig boundaries.

The method developed over the course of sequencing the *P. polymyxa* CR1 genome first identifies likely contig gaps, utilizing the work of previous genome sequencing projects to limit the amount of arduous blind primer walking (Eastman and Yuan, 2015). Gaps identified via contig scaffolding software are amplified by priming flanking protein coding sequences and the presence of rDNA operons within the contig gap is assessed by PCR (Figure 11). The highly conserved nature of rDNA lends to robust universal primers that are rDNA-specific in a range of bacterial species. However the rDNA primers presented within would need to be optimized for diverse species to be universally applicable in bacterial genome sequencing by introducing degenerate bases in a similar fashion to the development of universal 23S primers. This method resolves approximately 10kb of sequence contained within a gap in parallel sequencing reactions (assuming a mean Sanger sequencing read length of 1kb), which in all cases of rDNA mediated gaps in the *P. polymyxa* CR1 draft genome, allowed for resolution of gaps in parallel (Figure 12).

Although our method is robust for sequencing of gaps between contigs containing rDNA, many gaps in bacterial draft genome sequences are caused by a multitude of other repetitive features and isolate-specific variability. Further research and protocol development will be necessary to identify and design strategies to address other common limitations to ensure that published genomic data maintains a high standard.

Comparative genomic analyses and whole genome phylogenetics firmly established our lignin metabolizing isolate as a novel strain of *P. polymyxa* with significant departures from previously characterized strains (Figures 15 and 16) (Eastman et al., 2014a). When the whole genome is used to compute phylogeny, *P. polymyxa* strains M1 and SC2 form a subclade within the *P. polymyxa* species. Interestingly despite *P. polymyxa* CR1 being most closely related to *P. polymyxa* E681, there are significant differences in gene content and genome size, with the CR1 genome being nearly 700Mb larger than its E681 counterpart (Table 5).

Comparative analyses of metabolic networks within non-model families, such as *Paenibacillacae,* are hindered by the availability of in-depth experimental data of homologs from closely related bacteria. The clusters of orthologous groups (COG) categories R and S represent general function prediction and unknown function, respectively. These categories offer minimal evidence of the potential substrate specificity, protein function or metabolic pathways an annotated protein may contribute to. Protein function in *P. polymyxa* CR1 is not well characterized, as evidenced by the large proportion of genes assigned to the COG categories R and S (Figure 14). Future investigations into the potential use of *P. polymyxa* in industry could benefit greatly from an increased knowledge of genetic regulation and metabolic pathways employed by the species. Regardless of the limitations of available data and tools, some important inferences into the metabolic capacities of *P. polymyxa* CR1 are still possible. The CR1 genome has a higher proportion of encoded genes dedicated to energy metabolism, inorganic transport and metabolism (categories C, E, G and P), compared to M1 and SC2. This suggests *P. polymyxa* CR1 either possesses a wider metabolic capacity or a higher level of functional redundancy of its encoded proteins.

Perhaps the most striking characteristic of *P. polymyxa* CR1 is the high proportion of strain-specific genes, which account for approximately 18% of encoded open reading frames. Many of the strain-specific genes of *P. polymyxa* CR1 are localized to genomic islands, defined as putative regions of horizontal gene transfer. Genomic islands are typically identified via perturbations in the G+C mol % content of the bacterium, however this method misses potential horizontal transfer events from bacteria with close G+C mol % content (Dobrindt et al., 2004). This results in an underestimation of the extent of horizontal gene transfer since an individual bacterium is more likely to incorporate DNA with a similar G+C mol % content.

Genomic islands (GIs) are thought to be genetic elements acquired during evolution from distantly related organisms and as such, horizontally transferred genes contribute to genome flux and variation. Initially identified and established as important mediators of virulence in pathogenic bacteria (pathogenicity islands), GIs were subsequently identified in non-pathogenic bacteria from dense populated niches such as the rhizosphere (Dobrindt et al., 2004). Many GIs encode traits that enhance bacterial fitness including iron-uptake systems, polyketide synthesis clusters, resistance cassettes, symbiosis genes, xenobiotic compound degradation and primary metabolism pathways (Dobrindt et al., 2004). The locations of each identified genomic islands and their encoded genes are provided in Appendix 1. The majority of CDSs in the correspondent genomic island are annotated as hypothetical genes although many genomic islands include antibiotic synthesis and resistance cassettes.

Most interestingly, the majority of completely sequenced *P. polymyxa* strains do not encode a minimal 9 open reading frame *nif* cluster found exclusively in *Paenibacillus* sp. Anaerobic bacterial species with limited auxotrophic requirements and high survivability/adaptability are desirable for biofuel production, where feedstock nutrient composition is variable. Of the studied strains, only *P. polymyxa* CR1 encodes a functional nitrogenase, which we confirmed via growth on nitrogen-free media. Interestingly, *P. polymyxa* CR1 is capable of fixing nitrogen aerobically, which contrasts with oxygen sensitive nitrogenases of *Rhizobia* that are irreversibly inhibited by molecular oxygen. This suggests *P. polymyxa* CR1 either has some mechanism of preventing the interaction between nitrogenase and oxygen, or that the nitrogenase encoded is not irreversibly inhibited by oxygen. Furthermore, *P. polymyxa* CR1 does not appear to encode a canonical *ntrA* homolog. NtrA plays a vital role in sensing nitrogen limited conditions and activating the expression of nitrogenase genes encoded in multiple operons in *Rhizobia* (Yurgel and Kahn, 2004). Conversely DctD, a dicarboxylate response regulator, senses the presence of dicarboxylates produced by plants and activates the expression of plant-association genes and DctA, a dicarboxylate transporter (Yurgel and Kahn, 2004). Instead, *P. polymyxa* CR1 appears to encode a hybrid DctD/NtrA fusion (YP_008912290) that may act to sense plant-derived compounds and activate nitrogen-fixation genes, and may aid in the formation of the non-symbiotic mutualistic relationship with a plant host identified by other researchers.

P. polymyxa strains contained within our analyses show large differences in the number and nature of encoded glycoside hydrolase enzymes. *P. polymyxa* CR1 encodes both the largest absolute number and largest number relative to its genome size of GH family enzymes, suggesting a wide metabolism and diverse substrate capacity for sugar glycosides. Bacteria commonly employ diverse combinations of glycoside hydrolase family enzymes in what is known as the cellulosome complex (Schwarz, 2001; Henrissat and Daviest, 1997). In bacteria cellulosomes typically contain a cellulose binding protein in conjunction with endocellulase, exocellulase, cellulobiase, oxidative cellulases and cellulose phosphorylases. Cellulosome complexes have also been identified that contain non-canonical subunit compositions, incorporating arabinoglucanases, xylase among a multitude of other subunits (Schwarz, 2001). With *in silico* analyses it is exceedingly difficult to predict the substrate specificity of any given GH family enzyme since families are characterized by structure, and individual members may not be reflective of the overall families' substrate specificity (Henrissat and Daviest, 1997). Members of an individual GH family may hydrolyze diverse substrates, exemplified by GH families 1-5, which encode glucosidases, mannosidases, chitosanases, cellulases, and cellobiase among others. Despite the heterogeneity and modular nature of GH enzymes important inferences can be gleaned from GH family composition. *P. polymyxa* CR1 contains a large number of GH family 1, 2, 3 and 43 enzymes all of which are enriched in cellulobiose and cellulose hydrolase enzymes, suggesting strong cellulolytic potential. Our results demonstrate that *P. polymyxa* CR1 secretes cellulases capable of metabolizing cellulose, since the zone of reducing end cellulose cleavage greatly exceeds the size of the bacterial colony (Figure 7). Identification of the major GH enzymes expressed during growth on lignocellulose will aid in future development of fermentation pathways, where residual cellulose and hemicellulose components may act as easily accessible energy source for establishing robust expression of necessary genes thereby facilitating lignin metabolism.

Current biofuel production strategies are heavily reliant on the hexose and pentose rich cellulose and hemi-cellulose portions of plant biomass (Sannigrahi et al., 2010; Doherty et al., 2011; Paliwal et al., 2012). Accounting for between 18 and 35% of the total available carbon in biomass, lignin is still heavily underutilized by industry and affords an exciting potential avenue for developing value added processes for the bio-based energy economy. The isolation and characterization of the lignin metabolizing bacterium *Paenibacillus polymyxa* CR1, is the first report in the literature of an isolate capable of metabolizing lignin into valuable solvents/alcohols.

Highlighting our gap in knowledge of bacterial lignin metabolism is the absence of a complete identifiable lignin metabolism network in *P. polymyxa* CR1. Although various factors, including a DyP-like peroxidase and a laccase-like enzyme, previously shown to be involved in lignin metabolism were identified in the genome of *P. polymyxa* CR1

(Table 7), neither of these genes were shown to be necessary for growth on lignin containing media (Figure 20). This suggests that there are either compensatory mechanisms to accommodate the loss of these genes, or a novel, as of yet uncharacterized, pathway is used during lignin metabolism.

Although DyP-type peroxidases were first identified as industrial dye degrading peroxidases, recent research has shown the family of enzymes appears to be further subdivided into two classes (Brown and Chang, 2014). The DyP-type A peroxidases have been shown to catalyze porphyrin ring cleavage, and are potentially regulated by iron levels. Despite showing higher structural homology to the DyP-type B enzymes, it is possible that these enzymes are used as an iron scavenging mechanism by *P. polymyxa*. This scenario is supported by the absence of traditional siderophore synthesis clusters in *P. polymyxa* SC2 and *P. polymyxa* E681, these strains may conceivably use DyP-type enzymes to supply necessary iron in nutrient limited conditions.

Laccase-like enzymes are polyphenol oxidase enzymes that employ histidine coordinated copper for reduction-oxidation reactions on phenolic compounds. Multi-subunit laccases are the major factor responsible for lignin metabolism in *Basidomyctes* fungi, and unique two-domain laccases are the primary mechanism by which *Streptomyces coeilcor* degrades lignin (Majumdar et al., 2014; Sharma et al., 2006; Blanchette, 1991; Pollegioni et al., 2015). Laccase-like enzymes encoded by *P. polymyxa* strains appear to be two-domain oxidase enzymes and are encoded in an operon containing a pyridoxal-5-phosphate dependent enzyme with homology to racemase. This lead to the hypothesis that the encoded laccase oxidizes the lignin backbone, causing the release of moieties of which a subset can be modified into intermediates in any number of other pathways.

However, our data suggests that the laccase encoded by *P. polymyxa* CR1 is dispensable to lignin metabolism (Figure 20). Laccases are implicated in a wide range of aromatic compound metabolism and likely is involved in other processes within the cell.

Surprisingly, our comparative analyses of the *P. polymyxa* CR1 genome did not identify common aromatic tolerance and metabolism genes. Current knowledge suggests lignin-metabolizing bacteria employ one of two strategies for lignin metabolism: either assimilatory or dissimilatory lignin metabolism (Pollegioni et al., 2015; Brown and Chang, 2014). In assimilatory metabolism, specific enzymes with narrow substrate specificities degrade lignin substrates into aromatic compounds, typically protocatechuate or vanillin, that are imported into the cell where they act as substrates in common aromatic metabolism pathways. On the other hand, dissimilatory metabolism degrades lignin extracellularly into short chain carbon compounds via non-specific redox or free-radical generating mechanisms, which then act a substrate for short chain carbon transport systems. Assimilatory lignin metabolism has only been previously reported in a small subset of Gram-negative bacteria (Pollegioni et al., 2015). *P. polymyxa* appears to use a dissimilatory lignin metabolism approach, as it is incapable of metabolizing assimilatory pathway intermediates, yet is capable of metabolizing lignin and lignin mimetic dyes (Figures 7 and 19). This suggests that the enzymes responsible do not have tight substrate specificity and that aromatic compounds are not an intermediate in lignin metabolism pathways. Typically bacteria utilizing a dissimilatory approach require oxygenation for lignin metabolism. However, the ability of *P. polymyxa* CR1 to ferment lignin into alcohols and solvents suggests that oxygen is not necessary for lignin metabolism. This suggests *P. polymyxa* CR1 employs an oxygen-independent, non-specific mechanism to degrade lignin into short-chain carbon compounds, which would be the first identification of such pathway reported in the literature. Our results demonstrate that *P. polymyxa* is capable of both aerobic and anaerobic lignin metabolism using previously unidentified mechanisms. Since *P. polymyxa* CR1 is capable of metabolizing industrial lignin mimetic dyes, it is likely that oxidases and peroxidases play at least a nominal role in aerobic metabolism. However, it is unlikely these reactions form the basis of anaerobic lignin metabolism since oxygen is not available to act as a cofactor.

Future identification of disrupted genes in the *P. polymyxa* CR1 Tn5-mutants generated in this work has the potential for identifying novel genes implicated in lignin metabolism in bacteria. Development of *P. polymyxa* CR1 as a bioreactor for lignin necessitates in depth analyses into lignin metabolic genes to either modify or mobilize existing pathways and necessary regulatory genes. The comparative genomics work presented within did not identify a previously characterized lignin metabolism network, suggesting Tn5-mutants displaying abnormal growth phenotypes may represent genes not previously linked to lignin metabolism in bacteria. Unexpectedly, Tn5 mutants were identified that grew faster than Wt, implying *P. polymyxa* CR1 negatively regulates lignin metabolism (Figure 21). However, the possibility remains that mutants with increased growth have insertions in an unrelated, metabolically intensive pathway, such as antimicrobial compound synthesis, where the presented growth deficiency is below detection in nutrient rich media and only offers an advantage in nutrient-limited conditions. Interestingly, Tn5-mutants with slowed growth display a variety of phenotypes (Figure 22). Various mutants were identified which grow at a similar rate to Wt, albeit with a reduced final CFU. This supports *P. polymyxa* CR1 utilizing a multi-step pathway in the metabolism of lignin, where Tn5 mutants with insertions in genes late in the pathway still retain the ability to metabolize and grow using lignin but are unable to metabolize intermediates fully and thus do not use the full energy potential of the media. However, other mutants were obtained with a marked decrease in rate of growth, where the population appears to expand geometrically, and is still dividing when wild-type cells has already reached steady state. The most promising information detailing the lignin metabolism network of *P. polymyxa* CR1 will be obtained from the identification of disrupted genes in mutants that are incapable of growth on lignin. These mutants show normal growth when grown in rich media, however are incapable of sustaining growth on lignin media and accordingly represent genes required for lignin metabolism or tolerance of compounds produced by lignin catabolism.

Significance

Our research aims to address one of the fundamental problems plaguing the cellulosic biofuel industry: the high biomass pretreatment cost and low return of the lignin component of lignocellulose. Knowledge generated from this study will aid future genetic and metabolic engineering efforts in *P. polymyxa* to enhance performance in renewable energy programs and sustainable industrial bioreactors.

As the global economy surges towards sustainable and renewable resources in the face of dwindling petroleum supplies, continued development of novel, innovative solutions will be necessary to meet increasing global energy needs. Other renewable energy platforms such as wind, solar and hydroelectric have distinct disadvantages of low portability, transmission/infrastructure costs and storage. Conversely, renewable chemical fuels are disadvantaged by a higher associated production cost, and concerns over reliable sustained production compared to their conventional petroleum counterparts. Recent attempts to address the limitations of biofuels have yielded promising developments in the field of cellulosic alcohols. However with current technologies, cellulosic biofuels are neither economically viable, nor wholly renewable and sustainable with a high unit cost and an unacceptably large waste stream.

Canada has vastly abundant and inexpensive non-food biomass resources in the form of forestry and agricultural residues, creating both business opportunities and waste management challenges. Developing cost effective lignin degradation techniques can revitalize the biofuel industry and allow for economically viable production of cellulosic biofuels. Amelioration of the pitfalls of cellulosic biofuels has the potential to address multiple challenges facing Canada, including reduced reliance on petroleum based products and better management of cellulose-dense organic wastes.

Given the current lignin depolymerization bottleneck in cellulosic biofuel production, it is very important to understand the metabolic pathways and regulatory mechanisms underpinning lignin degradation and biofuel production in bacteria. To the best of our knowledge, lignin degradation and direct fermentative biofuel production has never been characterized in any *Paenibacillus* sp.. In fact, complete lignin metabolic pathways are not described for any Gram-positive bacterium. These results will help elucidate the genetic basis of known functions and delineate putative regulatory pathways/metabolic versatility in *P. polymyxa* relevant to lignin metabolism. Identification of the insertion locations in Tn5 mutagenized *P. polymyxa* CR1 will facilitate deeper insights into the novel pathways employed during lignin metabolism. Future studies into gene functions and implicated pathways will aid in genetic engineering of metabolic flux in *P. polymyxa* CR1 to optimize lignin metabolism and bioproduct synthesis.

Genomics of *P. polymyxa* CR1 is a prerequisite and integral part of developing an efficient delignification bioreactor. In particular, this research will lay the necessary ground work for future work into CR1 including transcriptome, secretome and metabolome based approaches to identify CR1 metabolic pathways and regulatory circuitry required for lignin degradation and biofuel production.

Development of a high-yield bioreactor based on *P. polymyxa* CR1 lignin metabolism necessitates a deep understanding of regulatory and metabolic circuits that underpin relevant phenotypes. The identification of *P. polymyxa* as a previously unreported lignin metabolizing bacteria employing a novel metabolic pathway demonstrates how our understanding of non-pathogenic soil bacterial metabolism is still in its infancy.

References

Abe, T., Masai, E., Miyauchi, K., Katayama, Y., and Fukuda, M. (2005). A Tetrahydrofolate-Dependent O-Demethylase, LigM, Is Crucial for Catabolism of Vanillate and Syringate in *Sphingomonas paucimobilis* SYK-6. *J. Bacteriol.* 187, 2030–2037. doi:10.1128/JB.187.6.2030.

Ahmad, M., Roberts, J. N., Hardiman, E. M., Singh, R., Eltis, L. D., and Bugg, T. D. H. (2011). Identification of DypB from *Rhodococcus jostii* RHA1 as a lignin peroxidase. *Biochemistry* 50, 5096–5107.

Ahmad, M., Taylor, C. R., Pink, D., Burton, K., Eastwood, D., Bending, G. D., and Bugg, T. D. H. (2010). Development of novel assays for lignin degradation: comparative analysis of bacterial and fungal lignin degraders. *Mol. Biosyst.* 6, 815–21. doi:10.1039/b908966g.

Anand, R., Grayston, S., and Chanway, C. (2013). N_2-Fixation and Seedling Growth Promotion of Lodgepole Pine by Endophytic *Paenibacillus polymyxa*. *Microb. Ecol.* 66, 369–374. doi:10.1007/s00248-013-0196-1.

Angiuoli, S. V, Gussman, A., Klimke, W., Cochrane, G., Field, D., Garrity, G., Kodira, C. D., Kyrpides, N., Madupu, R., Markowitz, V., et al. (2008). Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. *OMICS* 12, 137–41. doi:10.1089/omi.2008.0017.

Anzai, Y., Kim, H., Park, J., Wakabayashi, H., and Oyaizu, H. (2000). Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *Int. J. Syst. Evol. Microbiol.* 50, 1563–1589.

Aziz, R. K., Bartels, D., Best, A. a, DeJongh, M., Disz, T., Edwards, R. a, Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75. doi:10.1186/1471-2164-9-75.

Baffes, J., and Dennis, A. (2013). Long-Term Drivers of Food Prices.

Bai, F. W., Anderson, W. A., and Moo-young, M. (2008). Ethanol fermentation technologies from sugar and starch feedstocks. *Biotechnol. Adv.* 26, 89–105. doi:10.1016/j.biotechadv.2007.09.002.

Balachandrababu Malini, A., Revathi, M., Yadav, A., and Sakthivel, N. (2012). Purification and Characterization of a Thermophilic Cellulase from a Novel Cellulolytic Strain, *Paenibacillus barcinonensis*. *J. Microbiol. Biotechnol.* 22, 1501–1509.

Bandounas, L., Wierckx, N. J., de Winde, J. H., and Ruijssenaars, H. J. (2011). Isolation and characterization of novel bacterial strains exhibiting ligninolytic potential. *BMC Biotechnol.* 11, 94. doi:10.1186/1472-6750-11-94.

Barquist, L., Boinett, C. J., and Cain, A. K. (2013). Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol.* 10, 1161–9. doi:10.4161/rna.24765.

Beatty, P. H., and Jensen, S. E. (2002). *Paenibacillus polymyxa* produces fusaricidin-type antifungal antibiotics active against *Leptosphaeria maculans*, the causative agent of blackleg disease of canola. *Can. J. Microbiol.* 48, 159–169. doi:10.1139/W02-002.

Bensah, E. C., and Mensah, M. (2013). Chemical Pretreatment Methods for the Production of Cellulosic Ethanol: Technologies and Innovations. *Int. J. Chem. Eng.* 2013, 719607.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580.

Bentley, S. (2010). Taming the next-gen beast. *Nat. Rev. Microbiol.* 8, 161. doi:10.1038/nrmicro2322.

Birol, I., Jackman, S. D., Nielsen, C. B., Qian, J. Q., Varhol, R., Stazyk, G., Morin, R. D., Zhao, Y., Hirst, M., Schein, J. E., et al. (2009). *De novo* transcriptome assembly with ABySS. *Bioinformatics* 25, 2872–2877. doi:10.1093/bioinformatics/btp367.

Blanchette, R. A. (1991). Delignification by wood-decay fungi. *Annu. Rev. Phytopathol.* 29, 381–398.

Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., and Weber, T. (2013). antiSMASH 2.0— a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* 41, 204–212. doi:10.1093/nar/gkt449.

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54, 519–46. doi:10.1146/annurev.arplant.54.031902.134938.

Bohlool, B. B., Ladha, J. K., Garrity, D. P., and George, T. (1992). Biological nitrogen fixation for sustainable agriculture : A perspective. *Plant Soil* 141, 1–11.

Brown, M. E., and Chang, M. C. Y. (2014). Exploring bacterial lignin degradation. *Curr. Opin. Chem. Biol.* 19, 1–7. doi:10.1016/j.cbpa.2013.11.015.

Brunecky, R., Alahuhta, M., Xu, Q., Donohoe, B. S., Crowley, M. F., Kataeva, I. A., Yang, S., Resch, M. G., Adams, M. W. W., Lunin, V. V, et al. (2014). Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science (80-.).* 342, 1513–1516. doi:10.1126/science.1244273.

Bugg, T. D. H., Ahmad, M., Hardiman, E. M., and Singh, R. (2011). The emerging role for bacteria in lignin degradation and bio-product formation. *Curr. Opin. Biotechnol.* 22, 394–400. doi:10.1016/j.copbio.2010.10.009.

Camarero, S., Sarkar, S., Ruiz-duen, F. J., Martinez, M. J., and Martinez, A. (1999). Description of a Versatile Peroxidase Involved in the Natural Degradation of Lignin That Has Both Manganese Peroxidase and Lignin Peroxidase Substrate Interaction Sites. *J. Biol. Chem.* 274, 10324–10330.

Canning, P. (2011). A Revised and Expanded Food Dollar Series A Better Understanding of Our Food Costs.

Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J. (2009). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25, 119–20. doi:10.1093/bioinformatics/btn578.

Chakar, F. S., and Ragauskas, A. J. (2004). Review of current and future softwood kraft lignin process chemistry. *Ind. Crops Prod.* 20, 131–141. doi:10.1016/j.indcrop.2004.04.016.

Chapple, C., Ladisch, M., and Meilan, R. (2007). Loosening lignin' s grip on biofuel production. *Nat. Biotechnol.* 25, 746–748.

Chen, F., and Dixon, R. a (2007). Lignin modification improves fermentable sugar yields for biofuel production. *Nat. Biotechnol.* 25, 759–61. doi:10.1038/nbt1316.

Choi, S., Park, S., Kim, R., Kim, S., Lee, C., Kim, J. F., and Park, S. (2009). Identification of a Polymyxin Synthetase Gene Cluster of *Paenibacillus polymyxa* and Heterologous Expression of the Gene in *Bacillus subtilis*. *J. Bacteriol.* 191, 3350–3358. doi:10.1128/JB.01728-08.

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5, e11147. doi:10.1371/journal.pone.0011147.

Darling, A. E., Tritt, A., Eisen, J. A., and Facciotti, M. T. (2011). Mauve assembly metrics. *Bioinformatics* 27, 2756–2757. doi:10.1093/bioinformatics/btr451.

Delseny, M., Han, B., and Hsing, Y. (2010). High throughput DNA sequencing: The new sequencing revolution. *Plant Sci.* 179, 407–422. doi:10.1016/j.plantsci.2010.07.019.

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* 2, 414–424. doi:10.1038/nrmicro884.

Doherty, W. O. S., Mousavioun, P., and Fellows, C. M. (2011). Value-adding to cellulosic ethanol: Lignin polymers. *Ind. Crops Prod.* 33, 259–276. doi:10.1016/j.indcrop.2010.10.022.

Dwivedi, U. N., Singh, P., Pandey, V. P., and Kumar, A. (2011). Structure–function relationship among bacterial, fungal and plant laccases. *J. Mol. Catal. B Enzym.* 68, 117–128. doi:10.1016/j.molcatb.2010.11.002.

Eastman, A. W., Heinrichs, D. E., and Yuan, Z.-C. (2014a). Comparative and genetic analysis of the four sequenced *Paenibacillus polymyxa* genomes reveals a diverse metabolism and conservation of genes relevant to plant-growth promotion and competitiveness. *BMC Genomics* 15, 851. doi:10.1186/1471-2164-15-851.

Eastman, A. W., Weselowski, B., Nathoo, N., and Yuan, Z.-C. (2014b). Complete Genome Sequence of *Paenibacillus polymyxa* CR1, a Plant Growth-Promoting Bacterium Isolated from the Corn Rhizosphere Exhibiting Potential for Biocontrol, Biomass Degradation, and Biofuel. *Genome Announc.* 2, e01218–13. doi:10.1128/genomeA.01218-13.

Eastman, A. W., and Yuan, Z. (2015). Development and validation of an rDNA operon based primer walking strategy applicable to de novo bacterial genome finishing. *Front. Microbiol.* 5, 769. doi:10.3389/fmicb.2014.00769.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8, 186–194.

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8, 175–185. doi:10.1101/gr.8.3.175.

Fargione, J., Hill, J., Tilman, D., Polasky, S., and Hawthorne, P. (2008). Land clearing and the biofuel carbon debt. *Science (80-.).* 319, 1235–1239.

Felsenstein, J. (1981). Evolutionary Trees from DNA Sequences : A Maximum Likelihood Approach. *J. Mol. Evol.* 17, 368–376.

Felsenstein, J. (1989). PHYLIP- Phylogeny Inference Package (Version 3.2). *Cladistics* 5, 164–166.

Fernández-Fueyo, E., Ruiz-Dueñas, F. J., Martínez, M. J., Romero, A., Hammel, K. E., Medrano, F. J., and Martínez, A. T. (2014). Ligninolytic peroxidase genes in the oyster mushroom genome: heterologous expression, molecular structure, catalytic and stability properties, and lignin-degrading ability. *Biotechnol. Biofuels* 7, 2.

Fortenbery, T. R., and Park, H. (2008). The Effect of Ethanol Production on the US Corn Price.

Freudenburg, K., and Neish, A. C. (1968). *Constitution and biosynthesis of lignin*.

Gibbons, J. G., Janson, E. M., Hittinger, C. T., Johnston, M., Abbot, P., and Rokas, A. (2009). Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. *Mol. Biol. Evol.* 26, 2731–2744. doi:10.1093/molbev/msp188.

Haggag, W. M., and Timmusk, S. (2008). Colonization of peanut roots by biofilm-forming *Paenibacillus polymyxa* initiates biocontrol against crown rot disease. *J. Appl. Microbiol.* 104, 961–969. doi:10.1111/j.1365-2672.2007.03611.x.

Hahn-Hägerdal, B., Karhumaa, K., Fonseca, C., Spencer-Martins, I., and Gorwa-Grauslund, M. F. (2007). Towards industrial pentose-fermenting yeast strains. *Appl. Microbiol. Biotechnol.* 74, 937–53. doi:10.1007/s00253-006-0827-2.

Han, M. V, and Zmasek, C. M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* 10, 356. doi:10.1186/1471-2105-10-356.

Hara, H., Masai, E., Katayama, Y., and Fukuda, M. (2000). The 4-Oxalomesaconate Hydratase Gene, Involved in the Protocatechuate 4, 5-Cleavage Pathway, Is Essential to Vanillate and Syringate Degradation in *Sphingomonas paucimobilis* SYK-6. *J. Bacteriol.* 182, 6950–6957.

Hara, H., Masai, E., Miyauchi, K., Katayama, Y., and Fukuda, M. (2003). Characterization of the 4-Carboxy-4-Hydroxy-2-Oxoadipate Aldolase Gene and Operon Structure of the Protocatechuate 4, 5-Cleavage Pathway Genes in *Sphingomonas paucimobilis* SYK-6. *J. Bacteriol.* 185, 41–50. doi:10.1128/JB.185.1.41.

Henrissat, B., and Daviest, G. (1997). Structural and sequenc-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644.

Hirsch, A. M. (2015). Developmental biology of legume nodulation. *New Phytol.* 122, 211–237.

Hofrichter, M. (2002). Review: lignin conversion by manganese peroxidase (MnP). *Enzyme Microb. Technol.* 30, 454–466.

Holl, F. B., and Chanway, C. P. (1992). Rhizosphere colonization and seedling growth promotion of lodgepole pine by *Bacillus polymyxa*. *Can. J. Microbiol.* 38, 303–308.

Hunt, D. E., Klepac-Ceraj, V., Acinas, S. G., Gautier, C., Bertilsson, S., and Polz, M. F. (2006). Evaluation of 23S rRNA PCR Primers for Use in Phylogenetic Studies of Bacterial Diversity. *Appl. Environ. Microbiol.* 72, 2221–2225. doi:10.1128/AEM.72.3.2221.

Hurt, R. A., Brown, S. D., Podar, M., Palumbo, A. V, and Elias, D. A. (2012). Sequencing Intractable DNA to Close Microbial Genomes. *PLoS One* 7, e41295. doi:10.1371/journal.pone.0041295.

Janusz, G., Kucharzyk, K. H., Pawlik, A., Staszczak, M., and Paszczynski, A. J. (2013). Fungal laccase, manganese peroxidase and lignin peroxidase: gene expression and regulation. *Enzyme Microb. Technol.* 52, 1–12. doi:10.1016/j.enzmictec.2012.10.003.

Kanehisa, M., and Goto, S. (2000). KEGG : Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids rResearch* 40, D109–114. doi:10.1093/nar/gkr988.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, 199–205. doi:10.1093/nar/gkt1076.

Khan, Z., Kim, S. G., Jeon, Y. H., Khan, H. U., Son, S. H., and Kim, Y. H. (2008). A plant growth promoting rhizobacterium, *Paenibacillus polymyxa* strain GBR-1, suppresses root-knot nematode. *Bioresour. Technol.* 99, 3016–3023. doi:10.1016/j.biortech.2007.06.031.

Kim, J. F., Jeong, H., Park, S.-Y., Kim, S.-B., Park, Y. K., Choi, S.-K., Ryu, C.-M., Hur, C.-G., Ghim, S.-Y., Oh, T. K., et al. (2010). Genome sequence of the polymyxin-producing plant-probiotic rhizobacterium *Paenibacillus polymyxa* E681. *J. Bacteriol.* 192, 6103–6104. doi:10.1128/JB.00983-10.

Kim, S., and Dale, B. E. (2005). Life cycle assessment of various cropping systems utilized for producing biofuels: Bioethanol and biodiesel. *Biomass and Bioenergy* 29, 426–439. doi:10.1016/j.biombioe.2005.06.004.

Kim, S. J. U. N., and Shoda, M. (1999). Purification and Characterization of a Novel Peroxidase from *Geotrichum candidum* Dec 1 Involved in Decolorization of Dyes. *Appl. Environ. Microbiol.* 65, 1029–1035.

Kim, S.-B., and Timmusk, S. (2013). A Simplified Method for Gene Knockout and Direct Screening of Recombinant Clones for Application in *Paenibacillus polymyxa*. *PLoS One* 8, e68092. doi:10.1371/journal.pone.0068092.

Klappenbach, J. A., Dunbar, J. M., and Schmidt, T. M. (2000). rRNA Operon Copy Number Reflects Ecological Strategies of Bacteria. *Appl. Environ. Microbiol.* 66, 1328–1333.

Klinke, H. B., Thomsen, A. B., and Ahring, B. K. (2004). Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Appl. Microbiol. Biotechnol.* 66, 10–26. doi:10.1007/s00253-004-1642-2.

Kricka, W., Fitzpatrick, J., and Bond, U. (2014). Metabolic engineering of yeasts by heterologous enzyme production for degradation of cellulose and hemicellulose from biomass: a perspective. *Front. Microbiol.* 5, 174. doi:10.3389/fmicb.2014.00174.

Labes, M., and Finan, T. M. (1993). Negative regulation of sigma 54-dependent dctA expression by the transcriptional activator DctD. *J. Bacteriol.* 175, 2674–2681.

Labes, M., Rastogi, V., Watson, R., and Finan, T. M. (1993). Symbiotic nitrogen fixation by a nifA deletion mutant of *Rhizobium meliloti*: the role of an unusual ntrC allele. *J. Bacteriol.* 175, 2662–2673.

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi:10.1093/nar/gkm160.

Lal, S., and Tabacchioni, S. (2009). Ecology and biotechnological potential of *Paenibacillus polymyxa*: a minireview. *Indian J. Microbiol.* 49, 2–10. doi:10.1007/s12088-009-0008-y.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Folke, C., Bruce, J. W., Coomes, O. T., Dirzo, R., George, P. S., et al. (2001). The causes of land-use and land-cover change : moving beyond the myths. *Glob. Environ. Chang.* 11, 261–269.

Langille, M. G. I., and Brinkman, F. S. L. (2009). IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* 25, 664–665. doi:10.1093/bioinformatics/btp030.

Li, C., Knierim, B., Manisseri, C., Arora, R., Scheller, H. V, Auer, M., Vogel, K. P., Simmons, B. A., and Singh, S. (2010). Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification. *Bioresour. Technol.* 101, 4900–4906. doi:10.1016/j.biortech.2009.10.066.

Li, J., Wang, W., Ma, Y., and Zeng, A. (2013). Medium optimization and proteome analysis of (R, R)-2, 3-butanediol production by *Paenibacillus polymyxa* ATCC 12321. *Appl. Microbiol. Biotechnol.* 97, 585–597. doi:10.1007/s00253-012-4331-6.

Lin, S.-H., and Liao, Y.-C. (2013). CISA: contig integrator for sequence assembly of bacterial genomes. *PLoS One* 8, e60843. doi:10.1371/journal.pone.0060843.

Littlewood, J., Guo, M., Boerjan, W., and Murphy, R. J. (2014). Bioethanol from poplar: a commercially viable alternative to fossil fuel in the European Union. *Biotechnol. Biofuels* 7, 113. doi:10.1186/1754-6834-7-113.

Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, 490–495. doi:10.1093/nar/gkt1178.

López-Guerrero, M. G., Ormeño-Orrillo, E., Rosenblueth, M., Martinez-Romero, J., and Martinez-Romero, E. (2013). Buffet hypothesis for microbial nutrition at the rhizosphere. *Front. Plant Sci.* 4, 188. doi:10.3389/fpls.2013.00188.

Lu, L., Zeng, G., Fan, C., Zhang, J., Chen, A., Chen, M., Jiang, M., Yuan, Y., Wu, H., Lai, M., et al. (2014). Diversity of two-domain laccase-like multicopper oxidase genes in *Streptomyces* spp.: identification of genes potentially involved in extracellular activities and lignocellulose degradation during composting of agricultural waste. *Appl. Environ. Microbiol.* 80, 3305–14. doi:10.1128/AEM.00223-14.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1, 18.

Ma, M., Wang, C., Ding, Y., Li, L., Shen, D., Jiang, X., Guan, D., Cao, F., Chen, H., Feng, R., et al. (2011). Complete genome sequence of *Paenibacillus polymyxa* SC2, a strain of plant growth-promoting Rhizobacterium with broad-spectrum antimicrobial activity. *J. Bacteriol.* 193, 311–312. doi:10.1128/JB.01234-10.

Machczynski, M. C., Vijgenboom, E., Samyn, B., and Canters, G. W. (2004). Characterization of SLAC : A small laccase from *Streptomyces coelicolor* with unprecedented activity. *Protein Sci.* 13, 2388–2397. doi:10.1110/ps.04759104.and.

MacLean, D., Jones, J. D. G., and Studholme, D. J. (2009). Application of "next-generation" sequencing technologies to microbial genetics. *Nat. Rev. Microbiol.* 7, 287–296. doi:10.1038/nrmicro2088.

Majumdar, S., Lukk, T., Solbiati, J. O., Bauer, S., Nair, S. K., Cronan, J. E., and Gerlt, J. A. (2014). Roles of Small Laccases from *Streptomyces* in Lignin Degradation. *Biochemistry* 53, 4047–4058.

Mansfield, S. D. (2009). Solutions for dissolution--engineering cell walls for deconstruction. *Curr. Opin. Biotechnol.* 20, 286–94. doi:10.1016/j.copbio.2009.05.001.

Mardis, E., McPherson, J., Martienssen, R., Wilson, R. K., and McCombie, W. R. (2002). What is finished, and why does it matter. *Genome Biol.* 12, 669–671. doi:10.1101/gr.032102.

Markowitz, V. M., Chen, I. A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Woyke, T., Huntemann, M., et al. (2014). IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* 42, 560–567. doi:10.1093/nar/gkt963.

Martinez, A. T. (2002). Molecular biology and structure-function of lignin-degrading heme peroxidases. *Enzyme Microb. Technol.* 30, 425–444.

Martínez, A. T., Ruiz-Dueñas, F. J., Martínez, M. J., Del Río, J. C., and Gutiérrez, A. (2009). Enzymatic delignification of plant cell wall: from nature to mill. *Curr. Opin. Biotechnol.* 20, 348–57. doi:10.1016/j.copbio.2009.05.002.

Masai, E., Momose, K., Hara, H., Nishikawa, S., Katayama, Y., and Fukuda, M. (2000). Genetic and Biochemical Characterization of 4-Carboxy-2- Hydroxymuconate-6-Semialdehyde Dehydrogenase and Its Role in the Protocatechuate 4 , 5-Cleavage Pathway in *Sphingomonas paucimobilis* SYK-6. *J. Bacteriol.* 182, 6651–6658.

Masai, E., Shinohara, S., Hara, H., Nishikawa, S., Katayama, Y., and Fukuda, M. (1999). Genetic and Biochemical Characterization of a 2-Pyrone-4, 6- Dicarboxylic Acid Hydrolase Involved in the Protocatechuate 4, 5-Cleavage Pathway of *Sphingomonas paucimobilis* SYK-6. *J. Bacteriol.* 181, 55–62.

Masai, E., Yamamoto, Y., Inoue, T., Takamura, K., Hara, H., Kasai, D., Katayama, Y., and Fukuda, M. (2007). Characterization of ligV Essential for Catabolism of Vanillin by *Sphingomonas paucimobilis* SYK-6. *Biosci. Biotechnol. Biochem.* 71, 2487–2492. doi:10.1271/bbb.70267.

Mazzoli, R., Lamberti, C., and Pessione, E. (2012). Engineering new metabolic capabilities in bacteria: lessons from recombinant cellulolytic strategies. *Trends Biotechnol.* 30, 111–9. doi:10.1016/j.tibtech.2011.08.003.

McSpadden Gardener, B. B. (2004). Ecology of *Bacillus* and *Paenibacillus* spp. in Agricultural Systems. *Phytopathology* 94, 1252–1258.

Medini, D., Serruto, D., Parkhill, J., Relman, D. A., Donati, C., Moxon, R., Falkow, S., and Rappuoli, R. (2008). Microbiology in the post-genomic era. *Nat. Rev. Microbiol.* 6, 419–430. doi:10.1038/nrmicro1901.

Msadek, T., Kunst, F., Klier, A., and Rapoport, G. (1991). DegS-DegU and ComP-ComA modulator-effector pairs control expression of the *Bacillus subtilis* pleiotropic regulatory gene degQ. *J. Bacteriol.* 173, 2366–2377.

Murray, E. J., Kiley, T. B., and Stanley-Wall, N. R. (2009). A pivotal role for the response regulator DegU in controlling multicellular behaviour. *Microbiology* 155, 1–8. doi:10.1099/mic.0.023903-0.

Naik, S. N., Goud, V. V, Rout, P. K., and Dalai, A. K. (2010). Production of first and second generation biofuels: A comprehensive review. *Renew. Sustain. Energy Rev.* 14, 578–597. doi:10.1016/j.rser.2009.10.003.

Nishida, H. (2012). Comparative Analyses of Base Compositions, DNA Sizes, and Dinucleotide Frequency Profiles in Archaeal and Bacterial Chromosomes and Plasmids. *Int. J. Evol. Biol.* 2012, 342482. doi:10.1155/2012/342482.

Niu, B., Rueckert, C., Blom, J., Wang, Q., and Borriss, R. (2011). The genome of the plant growth-promoting rhizobacterium *Paenibacillus polymyxa* M-1 contains nine sites dedicated to nonribosomal synthesis of lipopeptides and polyketides. *J. Bacteriol.* 193, 5862–5863. doi:10.1128/JB.05806-11.

Niu, B., Vater, J., Rueckert, C., Blom, J., Lehmann, M., Ru, J., Chen, X., Wang, Q., and Borriss, R. (2013). Polymyxin P is the active principle in suppressing phytopathogenic *Erwinia* spp. by the biocontrol rhizobacterium *Paenibacillus polymyxa* M-1. *BMC Microbiol.* 13, 137. doi:10.1186/1471-2180-13-137.

Ochman, H., Gerber, A. S., and Hartl, D. L. (1988). Genetic Applications of an Inverse Polymerase Chain Reaction. *Genetics* 120, 621–623.

Oldroyd, G. E. D., and Dixon, R. (2014). Biotechnological solutions to the nitrogen problem. *Curr. Opin. Biotechnol.* 26, 19–24. doi:10.1016/j.copbio.2013.08.006.

Paliwal, R., Rawat, a P., Rawat, M., and Rai, J. P. N. (2012). Bioligninolysis: recent updates for biotechnological solution. *Appl. Biochem. Biotechnol.* 167, 1865–89. doi:10.1007/s12010-012-9735-3.

Papoutsakis, E. T. (2008). Engineering solventogenic clostridia. *Curr. Opin. Biotechnol.* 19, 420–9. doi:10.1016/j.copbio.2008.08.003.

Peng, X., Masai, E., Kasai, D., Miyauchi, K., Katayama, Y., and Fukuda, M. (2005). A Second 5-Carboxyvanillate Decarboxylase Gene, ligW2, Is Important for Lignin-Related Biphenyl Catabolism in *Sphingomonas paucimobilis* SYK-6. *Appl. Environ. Microbiol.* 71, 5014–5021. doi:10.1128/AEM.71.9.5014.

Peng, X., Masai, E., Kitayama, H., Harada, K., Katayama, Y., and Fukuda, M. (2002). Characterization of the 5-Carboxyvanillate Decarboxylase Gene and Its Role in Lignin-Related Biphenyl Catabolism in *Sphingomonas paucimobilis* SYK-6. *Appl. Environ. Microbiol.* 68, 4407–4415. doi:10.1128/AEM.68.9.4407.

Peng, X. U. E., Egashira, T., Hanashiro, K., Masai, E., Nishikawa, S., Katayama, Y., Kimbara, K., and Fukuda, M. (1998). Cloning of a *Sphingomonas paucimobilis* SYK-6 Gene Encoding a Novel Oxygenase That Cleaves Lignin-Related Biphenyl and Characterization of the Enzyme. *Appl. Environ. Microbiol.* 64, 2520–2527.

Peng, X. U. E., Masai, E., and Katayama, Y. (1999). Characterization of the meta-Cleavage Compound Hydrolase Gene Involved in Degradation of the Lignin-Related Biphenyl Structure by *Sphingomonas paucimobilis* SYK-6. *Appl. Environ. Microbiol.* 65, 2789–2793.

Petrosillo, N., Giannella, M., Antonelli, M., Antonini, M., Barsic, B., Belancic, L., Inkaya, C., De Pascale, G., Grilli, E., Tumbarello, M., et al. (2014). Colistin-glycopeptide combination in critically ill patients with Gram negative infection: the clinical experience. *Antimicrob. Agents Chemother.* 58, 851–858. doi:10.1128/AAC.00871-13.

Pimentel, D., and Patzek, T. W. (2005). Ethanol Production Using Corn, Switchgrass, and Wood; Biodiesel Production Using Soybean and Sunflower. *Nat. Resour. Res.* 14, 65–76. doi:10.1007/s11053-005-4679-8.

Pollegioni, L., Tonin, F., and Rosini, E. (2015). Lignin-degrading enzymes: a review. *FEBS J.* 282, 1190–1213. doi:10.1111/febs.13224.

Rahmanpour, R., and Bugg, T. D. H. (2015). Characterisation of Dyp-type peroxidases from *Pseudomonas fluorescens* Pf-5: Oxidation of Mn(II) and polymeric lignin by Dyp1B. *Arch. Biochem. Biophys.* 574, 93–8. doi:10.1016/j.abb.2014.12.022.

Ricker, N., Qian, H., and Fulthorpe, R. R. (2012). The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. *Genomics* 100, 167–175. doi:10.1016/j.ygeno.2012.06.009.

Rissman, A. I., Mau, B., Biehl, B. S., Darling, A. E., Glasner, J. D., and Perna, N. T. (2009). Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics* 25, 2071–2073. doi:10.1093/bioinformatics/btp356.

Roberts, J. N., Singh, R., Crigg, J., Murphy, M., Bugg, T. D. H., and Eltis, L. D. (2011). Characterization of Dye-Decolorizing Peroxidases from *Rhodococcus jostii* RHA1. *Biochemistry* 50, 5096–107. doi:10.1021/bi101892z.

Ross, P., Mayer, R., and Benziman, M. (1991). Cellulose Biosynthesis and Function in Bacteria. *Microbiol. Rev.* 55, 35–58.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., and Barrel B (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945.

Ryu, C., Kim, J., Choi, O., Kim, S. H., and Park, C. S. (2006). Improvement of biological control capacity of *Paenibacillus polymyxa* E681 by seed pelleting on sesame. *Biol. Control* 39, 282–289. doi:10.1016/j.biocontrol.2006.04.014.

Ryu, S.-H., Cho, M.-K., Kim, M., Jung, S.-M., and Seo, J.-H. (2013). Enhanced lignin biodegradation by a laccase-overexpressed white-rot fungus *Polyporus brumalis* in the pretreatment of wood chips. *Appl. Biochem. Biotechnol.* 171, 1525–1534. doi:10.1007/s12010-013-0412-y.

Saier, M. H., Reddy, V. S., Tamang, D. G., and Vastermark, A. (2014). The transporter classification database. *Nucleic Acids Res.* 42, 251–258. doi:10.1093/nar/gkt1097.

Sanchez, O. J., and Cardona, C. A. (2008). Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresour. Technol.* 99, 5270–5295. doi:10.1016/j.biortech.2007.11.013.

Sannigrahi, P., Pu, Y., and Ragauskas, A. (2010). Cellulosic biorefineries - unleashing lignin opportunities. *Curr. Opin. Environ. Sustain.* 2, 383–393. doi:10.1016/j.cosust.2010.09.004.

Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, 686–689. doi:10.1093/nar/gki366.

Scheller, H. V., and Ulvskov, P. (2010). Hemicelluloses. *Annu. Rev. Plant Biol.* 61, 263–89. doi:10.1146/annurev-arplant-042809-112315.

Schwarz, W. H. (2001). The cellulosome and cellulose degradation by anaerobic bacteria. *Appl. Microbiol. Biotechnol.* 56, 634–649. doi:10.1007/s002530100710.

Searchinger, T., Heimlich, R., Houghton, R. A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D., and Yu, T. (2008). Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions from Land-Use Change. *Science (80-.).* 319, 1238–1241.

Shaheen, M., Li, J., Ross, A. C., Vederas, J. C., and Jensen, S. E. (2011). *Paenibacillus polymyxa* PKB1 Produces Variants of Polymyxin B-Type Antibiotics. *Chem. Biol.* 18, 1640–1648. doi:10.1016/j.chembiol.2011.09.017.

Shao, Y., He, X., Harrison, E. M., Tai, C., Ou, H.-Y., Rajakumar, K., and Deng, Z. (2010). mGenomeSubtractor: a web-based tool for parallel in silico subtractive hybridization analysis of multiple bacterial genomes. *Nucleic Acids Res.* 38, W194–200. doi:10.1093/nar/gkq326.

Sharma, P., Goel, R., and Capalash, N. (2006). Bacterial laccases. *World J. Microbiol. Biotechnol.* 23, 823–832. doi:10.1007/s11274-006-9305-3.

Shin, S. H., Kim, S., Kim, J. Y., Song, H., Cho, S., Kim, D. R., Lee, K. I., Lim, H. K., Park, N. J., Hwang, I. T., et al. (2012). Genome sequence of P*aenibacillus terrae* HPL-003, a xylanase-producing bacterium isolated from soil found in forest residue. *J. Bacteriol.* 194, 1266. doi:10.1128/JB.06668-11.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi:10.1038/msb.2011.75.

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–36. doi:10.1093/nar/gkj014.

Slade, R., Bauen, A., and Shah, N. (2009). The greenhouse gas emissions performance of cellulosic ethanol supply chains in Europe. *Biotechnol. Biofuels* 2, 15. doi:10.1186/1754-6834-2-15.

Song, H. Y., Lim, H. K., Kim, D. R., Lee, K. I., and Hwang, I. T. (2014). A new bi-modular endo-β-1,4-xylanase KRICT PX-3 from whole genome sequence of *Paenibacillus terrae* HPL-003. *Enzyme Microb. Technol.* 54, 1–7. doi:10.1016/j.enzmictec.2013.09.002.

Sonoki, T., Obi, T., Kubota, S., and Higashi, M. (2000). Coexistence of Two Different O Demethylation Systems in Lignin Metabolism by *Sphingomonas paucimobilis* SYK-6: Cloning and Sequencing of the Lignin Biphenyl-Specific O-Demethylase (LigX) Gene. *Appl. Environ. Microbiol.* 66, 2125–2132.

Sorda, G., Banse, M., and Kemfert, C. (2010). An overview of biofuel policies across the world. *Energy Policy* 38, 6977–6988. doi:10.1016/j.enpol.2010.06.066.

Sullivan, A. C. O. (1997). Cellulose: the structure slowly unravels. *Cellulose* 4, 173–207.

Sun, Y., and Cheng, J. (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review q. *Bioresour. Technol.* 83, 1–11.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi:10.1093/molbev/mst197.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V, Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., et al. (2003). The COG database : an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41.

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V, Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D., and Koonin, E. V (2001). The COG database : new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29, 22–28.

Thevenot, M., Dignac, M.-F., and Rumpel, C. (2010). Fate of lignins in soils: A review. *Soil Biol. Biochem.* 42, 1200–1211. doi:10.1016/j.soilbio.2010.03.017.

Timmusk, S., Grantcharova, N., Gerhart, E. H., and Wagner, E. G. (2005). *Paenibacillus polymyxa* Invades Plant Roots and Forms Biofilms. *Appl. Environ. Microbiol.* 71, 7292–7300. doi:10.1128/AEM.71.11.7292.

Timmusk, S., van West, P., Gow, N. A., and Huffstutler, R. P. (2009). *Paenibacillus polymyxa* antagonizes oomycete plant pathogens *Phytophthora palmivora* and *Pythium aphanidermatum*. *J. Appl. Microbiol.* 106, 1473–1481. doi:10.1111/j.1365-2672.2009.04123.x.

Tong, Y., Ji, X., Liu, L., Shen, M., and Huang, H. (2013). Genome sequence of *Paenibacillus polymyxa* ATCC 12321, a promising strain for optically active (R, R)-2, 3-butanediol production. *Genome Announc.* 1, e00572–13. doi:10.1128/genomeA.00572-13.Copyright.

Tsai, I. J., Otto, T. D., and Berriman, M. (2010). Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol.* 11, R41.

U.S. Department of Energy (2012). Biofuels Strategic Plan.

U.S. Enviornmental Protection Agency (2010). Renewable Fuel Standard Program (RFS2) Regulatory Impact Analysis.

Ude, S., Arnold, D. L., Moon, C. D., Timms-Wilson, T., and Spiers, A. J. (2006). Biofilm formation and cellulose expression among diverse environmental Pseudomonas isolates. *Environ. Microbiol.* 8, 1997–2011. doi:10.1111/j.1462-2920.2006.01080.x.

Udvardi, M., and Poole, P. S. (2013). Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* 64, 781–805. doi:10.1146/annurev-arplant-050312-120235.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 40, e115. doi:10.1093/nar/gks596.

Velkov, T., Thompson, P. E., Nation, R. L., and Li, J. (2010). Structure-activity relationships of polymyxin antibiotics. *J. Med. Chem.* 53, 1898–1916. doi:10.1021/jm900999h.

Wang, L., Zhang, L., Liu, Z., Zhao, D., Liu, X., Zhang, B., Xie, J., Hong, Y., Li, P., Chen, S., et al. (2013). A Minimal Nitrogen Fixation Gene Cluster from *Paenibacillus* sp . WLY78 Enables Expression of Active Nitrogenase in *Escherichia coli*. *PLoS Genet.* 9, e1003865. doi:10.1371/journal.pgen.1003865.

Von der Weid, I., Paiva, E., Nóbrega, A., van Elsas, J. D., and Seldin, L. (2000). Diversity of *Paenibacillus polymyxa* strains isolated from the rhizosphere of maize planted in Cerrado soil. *Res. Microbiol.* 151, 369–381.

Wetzel, J., Kingsford, C., and Pop, M. (2011). Assessing the benefits of using mate-pairs to resolve repeats in *de novo* short-read prokaryotic assemblies. *BMC Bioinformatics* 12, 95. doi:10.1186/1471-2105-12-95.

Yang, H., Yan, R., Chen, H., Lee, D. H., and Zheng, C. (2007). Characteristics of hemicellulose, cellulose and lignin pyrolysis. *Fuel* 86, 1781–1788. doi:10.1016/j.fuel.2006.12.013.

Yu, B., Sun, J., Bommareddy, R. R., Song, L., and Zeng, A. (2011). Novel (2R, 3R)-2, 3-Butanediol Dehydrogenase from Potential Industrial Strain *Paenibacillus polymyxa* ATCC 12321. *Appl. Environ. Microbiol.* 77, 4230–4233. doi:10.1128/AEM.02998-10.

Yurgel, S. N., and Kahn, M. L. (2004). Dicarboxylate transport by rhizobia. *FEMS Micrbiology Rev.* 28, 489–501. doi:10.1016/j.femsre.2004.04.002.

Zeng, Y., Zhao, S., Yang, S., and Ding, S. (2014). Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels. *Curr. Opin. Biotechnol.* 27, 38–45. doi:10.1016/j.copbio.2013.09.008.

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi:10.1101/gr.074492.107.

Zhang, Y.-H. P., and Lynd, L. R. (2004). Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnol. Bioeng.* 88, 797–824. doi:10.1002/bit.20282.

Zhao, X., Cheng, K., and Liu, D. (2009). Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis. *Appl. Microbiol. Biotechnol.* 82, 815–27. doi:10.1007/s00253-009-1883-1.

Zhou, A., and Thomson, E. (2015). The development of biofuels in Asia. *Appl. Energy* 86, S11–S20. doi:10.1016/j.apenergy.2009.04.028.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: A Fast Phage Search Tool. *Nucleic Acids Res.* 39, 347–352. doi:10.1093/nar/gkr485.

Appendices

Appendix 1. Locations of genomic islands and encoded genes

Start	End	Size	Locus ID	Product
733265	737356	4091	X809_03395	adenine glycosylase
			X809_03400	thermostable monoacylglycerol lipase
			X809_03405	Pseudogene
			X809_03410	hypothetical protein
			X809_03415	Pseudogene
970825	980949	10124	X809_04465	iron-enterobactin transporter ATP-binding protein
			X809_04470	hypothetical protein
			X809_04475	hypothetical protein
			X809_04480	hypothetical protein
			X809_04485	hypothetical protein
			X809_04490	achromobactin biosynthetic protein AcsC
			X809_04495	siderophore biosynthesis protein SbnG
			X809_04500	diaminopimelate decarboxylase
			X809_04505	IucA/IucC
1090008	1094654	4646	X809_05015	NifK
			X809_05020	NifE
			X809_05025	NifN
			X809_05030	NifX
1104459	1108844	4385	X809_05075	MFS transporter
			X809_05080	chemotaxis protein CheY
			X809_05085	membrane protein
			X809_05090	LacI family transcriptional regulator
1177296	1181298	4002	X809_05410	hypothetical protein
			X809_05415	hypothetical protein
			X809_05420	sugar-binding protein
1226634	1264518	37884	X809_05655	peptide synthetase
			X809_05660	peptide ABC transporter
			X809_05665	excinuclease ABC subunit A
			X809_05670	O-methyltransferase
			X809_05675	hypothetical protein
			X809_05680	hypothetical protein
			X809_05685	hypothetical protein
			X809_05690	polyketidesynthase
			X809_05695	aspartate aminotransferase

			X809_05700	peptide synthetase
			X809_05705	hypothetical protein
			X809_05710	nodU
			X809_05715	tyrocidine synthetase III
1377540	1391268	13728	X809_06155	DNA integrase
			X809_06160	repressor
			X809_06165	XRE family transcriptional regulator
			X809_06170	hypothetical protein
			X809_06175	antirepressor
			X809_06180	hypothetical protein
			X809_06185	hypothetical protein
			X809_06190	hypothetical protein
			X809_06195	hypothetical protein
			X809_06200	hypothetical protein
			X809_06205	hypothetical protein
			X809_06210	hypothetical protein
			X809_06215	hypothetical protein
			X809_06220	hypothetical protein
			X809_06225	hypothetical protein
			X809_06230	hypothetical protein
			X809_06235	hypothetical protein
			X809_06240	single-stranded DNA-binding protein
			X809_06245	hypothetical protein
			X809_06250	DNA damage-indicible protein DnaD
			X809_06255	hypothetical protein
			X809_06260	hypothetical protein
			X809_06265	hypothetical protein
			X809_06270	hypothetical protein
			X809_06275	hypothetical protein
			X809_06280	hypothetical protein
			X809_06285	hypothetical protein
			X809_06290	hypothetical protein
			X809_06295	hypothetical protein
			X809_06300	hypothetical protein
			X809_06305	hypothetical protein
			X809_06310	hypothetical protein
			X809_06315	DNA methyltransferase
1394287	1419648	25361	X809_06360	hypothetical protein
			X809_06365	hypothetical protein
			X809_06370	hypothetical protein
			X809_06375	hypothetical protein

			X809_06380	hypothetical protein
			X809_06385	hypothetical protein
			X809_06390	hypothetical protein
			X809_06395	hypothetical protein
			X809_06400	hypothetical protein
			X809_06405	hypothetical protein
			X809_06410	hypothetical protein
			X809_06415	hypothetical protein
			X809_06420	hypothetical protein
			X809_06425	terminase
			X809_06430	portal protein
			X809_06435	head protein
			X809_06440	hypothetical protein
			X809_06445	hypothetical protein
			X809_06450	scaffold protein
			X809_06455	hypothetical protein
			X809_06460	hypothetical protein
			X809_06465	hypothetical protein
			X809_06470	hypothetical protein
			X809_06475	hypothetical protein
			X809_06480	hypothetical protein
			X809_06485	hypothetical protein
			X809_06490	hypothetical protein
			X809_06495	hypothetical protein
			X809_06500	phage portal protein
			X809_06505	phage portal protein
			X809_06510	hypothetical protein
			X809_06515	hypothetical protein
			X809_06520	hypothetical protein
			X809_06525	peptidoglycan-binding protein LysM
			X809_06530	hypothetical protein
			X809_06535	hypothetical protein
			X809_06540	hypothetical protein
			X809_06545	baseplate J protein
			X809_06550	phage portal protein
			X809_06555	hypothetical protein
1780934	1792521	11587	X809_08315	hypothetical protein
			X809_08320	hypothetical protein
			X809_08325	phosphotransferase
			X809_08330	hypothetical protein
			X809_08335	hypothetical protein
			X809_08340	hypothetical protein

			X809_08345	hypothetical protein
			X809_08350	hypothetical protein
			X809_08355	endonuclease
			X809_08360	hypothetical protein
			X809_08365	hypothetical protein
			X809_08370	DNA-binding protein
			X809_08375	hypothetical protein
			X809_08380	hypothetical protein
			X809_08385	hypothetical protein
2633151	2640620	7469	X809_12115	hypothetical protein
			X809_12120	hydroxylase
			X809_12125	hypothetical protein
			X809_12130	transcriptional regulator
			X809_12135	esterase
			X809_12140	hypothetical protein
			X809_12145	type 12 methyltransferase
			X809_12150	alpha/beta hydrolase
2648552	2654352	5800	X809_12200	alpha/beta hydrolase
			X809_12205	hypothetical protein
			X809_12210	Pseudogene
			X809_12215	methionine aminopeptidase
			X809_12220	Pseudogene
			X809_12225	hypothetical protein
			X809_12230	esterase
2711714	2717405	5691	X809_12430	PTS sugar transporter
			X809_12435	PTS cellobiose transporter subunit IIB
			X809_12440	PTS cellobiose transporter subunit IIA
			X809_12445	PTS cellobiose transporter subunit IIC
			X809_12450	6-phospho-beta-glucosidase
2813715	2822924	9209	X809_12895	3-ketoacyl-ACP reductase
			X809_12900	hypothetical protein
			X809_12905	hypothetical protein
			X809_12910	nickel ABC transporter substrate-binding protein
			X809_12915	nickel ABC transporter permease
			X809_12920	nickel ABC transporter permease
			X809_12925	nickel ABC transporter ATP-binding protein
			X809_12930	nickel ABC transporter ATP-binding protein
			X809_12935	hypothetical protein

2933019	2950227	17208	X809_13400	hypothetical protein
			X809_13405	AraC family transcriptional regulator
			X809_13410	sugar ABC transporter permease
			X809_13415	sugar ABC transporter permease
			X809_13420	sugar ABC transporter substrate-binding protein
			X809_13425	glycosyl hydrolase
			X809_13430	glycoside hydrolase
			X809_13435	hypothetical protein
			X809_13440	beta-galactosidase
			X809_13445	hypothetical protein
2959513	3025698	66185	X809_13500	PTS cellbiose transporter subunit IIC
			X809_13505	hypothetical protein
			X809_13510	transcriptional regulator
			X809_13515	hypothetical protein
			X809_13520	transcriptional regulator
			X809_13525	hypothetical protein
			X809_13530	hypothetical protein
			X809_13535	hypothetical protein
			X809_13540	ATPase AAA
			X809_13545	hypothetical protein
			X809_13550	DNA integrase
			X809_13555	hypothetical protein
			X809_13560	IS4 - disrupted
			X809_13565	short-chain dehydrogenase
			X809_13570	gramicidin dehydrogenase
			X809_13575	tyrocidine synthetase III
			X809_13580	phenylalanine racemase
			X809_13585	hypothetical protein
			X809_13590	diadenosine tetraphosphatase
			X809_13595	lytic transglycosylase
			X809_13600	Pseudogene
			X809_13605	isochorismatase
			X809_13610	hypothetical protein
			X809_13615	hypothetical protein
			X809_13620	hypothetical protein
			X809_13625	hypothetical protein
			X809_13630	cation transporter
			X809_13635	hypothetical protein
			X809_13640	hypothetical protein
			X809_13645	hypothetical protein

			X809_13650	hypothetical protein
			X809_13655	PTS sugar transporter subunit IIA
			X809_13660	6-phospho 3-hexuloisomerase
			X809_13665	iditol 2-dehydrogenase
			X809_13670	ribulose-phosphate 3-epimerase
			X809_13675	PTS galactitol transporter subunit IIC
			X809_13680	PTS galactitol transporter subunit IIB
			X809_13685	oxidoreductase
			X809_13690	TetR family transcriptional regulator
			X809_13695	NAD(P)H nitroreductase
			X809_13700	TetR family transcriptional regulator
			X809_13705	DNA integrase
3110386	3121035	10649	X809_14070	DNA-binding protein
			X809_14075	transposase
			X809_14080	hypothetical protein
			X809_14085	hypothetical protein
			X809_14090	NAD-dependent dehydratase
			X809_14095	oxidoreductase
			X809_14100	NADPH dehydrogenase
			X809_14105	MarR family transcriptional regulator
			X809_14110	resolvase
			X809_14115	hypothetical protein
3151095	3159256	8161	X809_14255	hypothetical protein
			X809_14260	PTS sugar transporter
			X809_14265	CoA transferase
			X809_14270	MFS transporter
			X809_14275	2-nitropropane dioxygenase
			X809_14280	diguanylate cyclase
			X809_14285	hypothetical protein
3285590	3291474	5884	X809_14885	membrane protein
			X809_14890	hypothetical protein
			X809_14895	hypothetical protein
			X809_14900	hypothetical protein
			X809_14905	Pseudogene
			X809_14910	hypothetical protein
			X809_14915	galactose oxidase
			X809_14920	hypothetical protein
			X809_14925	hypothetical protein
3379397	3384415	5018	X809_15340	aspartate oxidase
			X809_15345	hypothetical protein

			X809_15350	galactose oxidase
			X809_15355	hypothetical protein
			X809_15360	hypothetical protein
4096559	4124749	28190	X809_18460	methyltransferase type 11
			X809_18465	hypothetical protein
			X809_18470	histone acetyltransferase
			X809_18475	hypothetical protein
			X809_18480	hypothetical protein
			X809_18485	oxidoreductase
			X809_18490	hypothetical protein
			X809_18495	hypothetical protein
			X809_18500	hypothetical protein
			X809_18505	hypothetical protein
			X809_18510	GNAT family acetyltransferase
			X809_18515	hypothetical protein
			X809_18520	transcriptional regulator
			X809_18525	hypothetical protein
			X809_18530	hypothetical protein
			X809_18535	hypothetical protein
			X809_18540	hypothetical protein
			X809_18545	Zn-finger containing protein
			X809_18550	hypothetical protein
			X809_18555	hypothetical protein
			X809_18560	hypothetical protein
			X809_18565	XRE family transcriptional regulator
			X809_18570	hypothetical protein
			X809_18575	hypothetical protein
			X809_18580	hypothetical protein
			X809_18585	hypothetical protein
			X809_18590	hypothetical protein
			X809_18595	hypothetical protein
			X809_18600	XRE family transcriptional regulator
			X809_18605	hypothetical protein
			X809_18610	hypothetical protein
			X809_18615	hypothetical protein
			X809_18620	hypothetical protein
			X809_18625	hypothetical protein
			X809_18630	hypothetical protein
			X809_18635	hypothetical protein
			X809_18640	hypothetical protein
			X809_18645	hypothetical protein
			X809_18650	hypothetical protein
			X809_18655	hypothetical protein

			X809_18660	hypothetical protein
			X809_18665	hypothetical protein
4136370	4151975	15605	X809_18715	N-acetyltransferase GCN5
			X809_18720	urea carboxylase
			X809_18725	urea carboxylase
			X809_18730	allophanate hydrolase
			X809_18735	hypothetical protein
			X809_18740	ABC transporter permease
			X809_18745	macrolide ABC transporter ATP-binding protein
			X809_18750	hypothetical protein
			X809_18755	hypothetical protein
			X809_18760	cysteine hydrolase
			X809_18765	cysteine hydrolase
			X809_18770	peptidase M20
			X809_18775	hypothetical protein
			X809_18780	GCN5 family acetyltransferase
			X809_18785	hypothetical protein
			X809_18790	hypothetical protein
4257273	4262931	5658	X809_19305	hypothetical protein
			X809_19310	hypothetical protein
			X809_19315	hypothetical protein
			X809_19320	hypothetical protein
			X809_19325	hypothetical protein
			X809_19330	hypothetical protein
			X809_19335	hypothetical protein
4266813	4271353	4540	X809_19345	hypothetical protein
			X809_19350	hypothetical protein
			X809_19355	hypothetical protein
			X809_19360	hypothetical protein
			X809_19365	hypothetical protein
			X809_19370	zinc-binding protein
4355246	4361538	6292	X809_19740	chemotaxis protein CheY
			X809_19745	Pseudogene
			X809_19750	hypothetical protein
			X809_19755	hypothetical protein
			X809_19760	hypothetical protein
			X809_19765	hypothetical protein
			X809_19770	hypothetical protein
5129029	5175885	46856	X809_23445	hypothetical protein
			X809_23450	hypothetical protein
			X809_23455	hypothetical protein
			X809_23460	tyrocidine synthetase III - disrupted

			X809_23465	bacitracin synthase - disrupted
			X809_23470	multidrug ABC transporter permease
			X809_23475	multidrug ABC transporter permease
			X809_23480	peptide synthetase
			X809_23485	peptide synthetase
			X809_23490	gramicidin synthetase
			X809_23495	hypothetical protein
			X809_23500	hypothetical protein
			X809_23505	hypothetical protein
			X809_23510	maturase
			X809_23520	hypothetical protein
			X809_23525	hypothetical protein
			X809_23530	hypothetical protein
			X809_23535	hypothetical protein
5206724	5211731	5007	X809_23720	hypothetical protein
			X809_23725	hypothetical protein
			X809_23730	hypothetical protein
			X809_23735	hypothetical protein
			X809_23740	hypothetical protein
			X809_23745	hypothetical protein
			X809_23750	hypothetical protein
			X809_23755	hypothetical protein
5214587	5236057	21470	X809_23795	tail protein
			X809_23800	baseplate J protein
			X809_23805	lysozyme
			X809_23810	hypothetical protein
			X809_23815	hypothetical protein
			X809_23820	phage late control protein
			X809_23825	phage tail protein
			X809_23830	hypothetical protein
			X809_23835	hypothetical protein
			X809_23840	hypothetical protein
			X809_23845	tail protein
			X809_23850	phage tail protein
			X809_23855	hypothetical protein
			X809_23860	hypothetical protein
			X809_23865	hypothetical protein
			X809_23870	hypothetical protein
			X809_23875	hypothetical protein
			X809_23880	hypothetical protein
			X809_23885	phage portal protein

			X809_23890	peptidyl-prolyl cis-trans isomerase
			X809_23895	terminase
			X809_23900	hypothetical protein
			X809_23905	hypothetical protein
			X809_23910	hypothetical protein
			X809_23915	transcriptional regulator
			X809_23920	hypothetical protein
			X809_23925	hypothetical protein
			X809_23930	hypothetical protein
5239500	5248697	9197	X809_23970	hypothetical protein
			X809_23975	hypothetical protein
			X809_23980	hypothetical protein
			X809_23985	hypothetical protein
			X809_23990	hypothetical protein
			X809_23995	hypothetical protein
			X809_24000	hypothetical protein
			X809_24005	hypothetical protein
			X809_24010	hypothetical protein
			X809_24015	hypothetical protein
			X809_24020	hypothetical protein
			X809_24025	hypothetical protein
			X809_24030	Rha family transcriptional regulator
			X809_24035	hypothetical protein
			X809_24040	hypothetical protein
			X809_24045	XRE family transcriptional regulator
			X809_24050	hypothetical protein
			X809_24055	hypothetical protein
			X809_24060	integrase
5320533	5328150	7617	X809_24460	spore coat protein
			X809_24465	TDP-4-oxo-6-deoxy-D-glucose aminotransferase
			X809_24470	membrane protein
			X809_24475	hypothetical protein
			X809_24480	family 2 glycosyl transferase
			X809_24485	hypothetical protein
			X809_24490	hypothetical protein
5943786	5967303	23517	X809_27260	methyltransferase type 12
			X809_27265	hypothetical protein
			X809_27270	hypothetical protein
			X809_27275	Pseudogene
			X809_27280	elongation factor G
			X809_27285	hypothetical protein
			X809_27290	hypothetical protein

X809_27295	hypothetical protein
X809_27300	hypothetical protein
X809_27305	hypothetical protein
X809_27310	transcriptional regulator
X809_27315	hypothetical protein
X809_27320	hypothetical protein
X809_27325	copper amine oxidase
X809_27330	hydrolase
X809_27335	hypothetical protein
X809_27340	DNA repair protein RadC
X809_27345	hypothetical protein
X809_27350	resolvase
X809_27355	resolvase

Appendix 2 Insertion sequences identified in *P. polymyxa* CR1

	IS Family	Group	Origin	Score (bits)	E-value
ISBcy1	IS1182		Bacillus cytotoxicus	105	6.00E-19
ISHaha5	IS110		Halobacillus halophilus	66	5.00E-07
ISPaen2	IS5	IS5	Paenibacillus sp.	60	3.00E-05
ISArsp6	Tn3		Arthrobacter sp.	54	0.002

Appendix 3. Identified antimicrobial compounds encoded by *P. polymyxa* CR1

Gene cluster type	Gene cluster genes	Gene cluster gene accessions

Bacteriocin	X809_03455;X809_03460;
	X809_03465;X809_03470;
	X809_03475;X809_03480;
	X809_03485;X809_03490;
	X809_03495;X809_03500;
	X809_03505;X809_03510;
	X809_03515
		YP_008910200.1;YP_008910201.1;
		YP_008910202.1;YP_008910203.1;
		YP_008910204.1;YP_008910205.1;
		YP_008910206.1;YP_008910207.1;
		YP_008910208.1;YP_008910209.1;
		YP_008910210.1;YP_008910211.1;
		YP_008910212.1

Siderophore	X809_04455;X809_04460;
	X809_04465;X809_04470;
	X809_04475;X809_04480;
	X809_04485;X809_04490;
	X809_04495;X809_04500;
	X809_04505;X809_04510;
	X809_04515;X809_04520;
		YP_008910397.1;YP_008910398.1;
		YP_008910399.1;YP_008910400.1;
		YP_008910401.1;YP_008910402.1;
		YP_008910403.1;YP_008910404.1;
		YP_008910405.1;YP_008910406.1;
		YP_008910407.1;YP_008910408.1;
		YP_008910409.1;YP_008910410.1;

	X809_04525;X809_04530	YP_008910411.1;YP_008910412.1
Bacteriocin	X809_05280;X809_05285;X809_05290;X809_05295;X809_05300;X809_05305;X809_05310;X809_05315;X809_05320;X809_05325;X809_05330;X809_05335;X809_05340	YP_008910550.1;YP_008910551.1;YP_008910552.1;YP_008910553.1;YP_008910554.1;YP_008910555.1;YP_008910556.1;YP_008910557.1;YP_008910558.1;YP_008910559.1;YP_008910560.1;YP_008910561.1;YP_008910562.1
NRP-PK hybrid	X809_05535;X809_05540;X809_05545;X809_05550;X809_05555;X809_05560;X809_05565;X809_05570;X809_05575;X809_05580;X809_05585;X809_05590;X809_05595;X809_05600;X809_05605;X809_05610;X809_05615;X809_05620;X809_05625;X809_05630;X809_05635;X809_05640;X809_05645;X809_05650;X809_05655;X809_05660;X809_05665;X809_05670;X809_05675;X809_05680;X809_05685;X809_05690;X809_05695;X809_05700;X809_05705;X809_05710;X809_05715;X809_05720;X809_05725;X809_05730;X809_05735;X809_05740;X809_05745;X809_05750;X809_05755;X809_05760;X809_05765;X809_05770;X809_05775;X809_05780;X809_05795;X809_05800	YP_008910601.1;YP_008910602.1;YP_008910603.1;YP_008910604.1;YP_008910605.1;YP_008910606.1;YP_008910607.1;YP_008910608.1;YP_008910609.1;YP_008910610.1;YP_008910611.1;YP_008910612.1;YP_008910613.1;YP_008910614.1;YP_008910615.1;YP_008910616.1;YP_008910617.1;YP_008910618.1;YP_008910619.1;YP_008910620.1;YP_008910621.1;YP_008910622.1;YP_008910623.1;YP_008910624.1;YP_008910625.1;YP_008910626.1;YP_008910627.1;YP_008910628.1;YP_008910629.1;YP_008910630.1;YP_008910631.1;YP_008910632.1;YP_008910633.1;YP_008910634.1;YP_008910635.1;YP_008910636.1;YP_008910637.1;YP_008910638.1;YP_008910639.1;YP_008910640.1;YP_008910641.1;YP_008910642.1;YP_008910643.1;YP_008910644.1;YP_008910645.1;YP_008910646.1;YP_008910647.1;YP_008910648.1;YP_008910649.1;YP_008910650.1;YP_008910651.1;YP_008910652.1

Lantipeptide	X809_08095;X809_08100;	YP_008911100.1;YP_008911101.1;
	X809_08105;X809_08110;	YP_008911102.1;YP_008911103.1;
	X809_08115;X809_08120;	YP_008911104.1;YP_008911105.1;
	X809_08125;X809_08130;	YP_008911106.1;YP_008911107.1;
	X809_08135;X809_08140;	YP_008911108.1;YP_008911109.1;
	X809_08145;X809_08150;	YP_008911110.1;YP_008911111.1;
	X809_08155;X809_08160;	YP_008911112.1;YP_008911113.1;
	X809_08165;X809_08170;	YP_008911114.1;YP_008911115.1;
	X809_08175;X809_08180;	YP_008911116.1;YP_008911117.1;
	X809_08185;X809_08190;	YP_008911118.1;YP_008911119.1;
	X809_08195;X809_08200;	YP_008911120.1;YP_008911121.1;
	X809_08205;X809_08210;	YP_008911122.1;YP_008911123.1;
	X809_08215	YP_008911124.1
Other	X809_09710;X809_09715;	YP_008911416.1;YP_008911417.1;
	X809_09720;X809_09725;	YP_008911418.1;YP_008911419.1;
	X809_09730;X809_09735;	YP_008911420.1;YP_008911421.1;
	X809_09740;X809_09745;	YP_008911422.1;YP_008911423.1;
	X809_09750;X809_09755;	YP_008911424.1;YP_008911425.1;
	X809_09760;X809_09765;	YP_008911426.1;YP_008911427.1;
	X809_09770;X809_09775;	YP_008911428.1;YP_008911429.1;
	X809_09780;X809_09785;	YP_008911430.1;YP_008911431.1;
	X809_09790;X809_09795;	YP_008911432.1;YP_008911433.1;
	X809_09800;X809_09805;	YP_008911434.1;YP_008911435.1;
	X809_09810;X809_09815;	YP_008911436.1;YP_008911437.1;
	X809_09820;X809_09825;	YP_008911438.1;YP_008911439.1;
	X809_09830;X809_09835;	YP_008911440.1;YP_008911441.1;
	X809_09840;X809_09845;	YP_008911442.1;YP_008911443.1;
	X809_09850;X809_09855;	YP_008911444.1;YP_008911445.1;
	X809_09860;X809_09865;	YP_008911446.1;YP_008911447.1;
	X809_09870;X809_09875;	YP_008911448.1;YP_008911449.1;
	X809_09880	YP_008911450.1
NRPS	X809_11790;X809_11795;	YP_008911822.1;YP_008911823.1;
	X809_11800;X809_11805;	YP_008911824.1;YP_008911825.1;
	X809_11810;X809_11815;	YP_008911826.1;YP_008911827.1;
	X809_11820;X809_11825;	YP_008911828.1;YP_008911829.1;
	X809_11830;X809_11835;	YP_008911830.1;YP_008911831.1;
	X809_11840;X809_11845;	YP_008911832.1;YP_008911833.1;
	X809_11850;X809_11855;	YP_008911834.1;YP_008911835.1;
	X809_11860;X809_11865;	YP_008911836.1;YP_008911837.1;
	X809_11870;X809_11875;	YP_008911838.1;YP_008911839.1;

	X809_11880;X809_11885; X809_11890	YP_008911840.1;YP_008911841.1; YP_008911842.1
NRPS-Type1PK hybrid	X809_13470;X809_13475; X809_13480;X809_13485; X809_13490;X809_13495; X809_13500;X809_13505; X809_13510;X809_13515; X809_13520;X809_13525; X809_13530;X809_13535; X809_13540;X809_13545; X809_13550;X809_13555; X809_13565;X809_13570; X809_13575;X809_13580; X809_13585;X809_13590; X809_13595;X809_13605; X809_13610;X809_13615; X809_13620;X809_13625; X809_13630;X809_13635; X809_13640;X809_13645; X809_13650;X809_13655; X809_13660;X809_13665; X809_13670;X809_13675; X809_13680;X809_13685; X809_13690;X809_13695	YP_008912136.1;YP_008912137.1; YP_008912138.1;YP_008912139.1; YP_008912140.1;YP_008912141.1; YP_008912142.1;YP_008912143.1; YP_008912144.1;YP_008912145.1; YP_008912146.1;YP_008912147.1; YP_008912148.1;YP_008912149.1; YP_008912150.1;YP_008912151.1; YP_008912152.1;YP_008912153.1; YP_008912154.1;YP_008912155.1; YP_008912156.1;YP_008912157.1; YP_008912158.1;YP_008912159.1; YP_008912160.1;YP_008912161.1; YP_008912162.1;YP_008912163.1; YP_008912164.1;YP_008912165.1; YP_008912166.1;YP_008912167.1; YP_008912168.1;YP_008912169.1; YP_008912170.1;YP_008912171.1; YP_008912172.1;YP_008912173.1; YP_008912174.1;YP_008912175.1; YP_008912176.1;YP_008912177.1; YP_008912178.1;YP_008912179.1
PK	X809_17110;X809_17115; X809_17120;X809_17125; X809_17130;X809_17135; X809_17145;X809_17150; X809_17155;X809_17160; X809_17165;X809_17170; X809_17175;X809_17180; X809_17185;X809_17190	YP_008912834.1;YP_008912835.1; YP_008912836.1;YP_008912837.1; YP_008912838.1;YP_008912839.1; YP_008912840.1;YP_008912841.1; YP_008912842.1;YP_008912843.1; YP_008912844.1;YP_008912845.1; YP_008912846.1;YP_008912847.1; YP_008912848.1;YP_008912849.1

Type 2 PK	X809_17205;X809_17210;	YP_008912850.1;YP_008912851.1;
	X809_17215;X809_17220;	YP_008912852.1;YP_008912853.1;
	X809_17225;X809_17230;	YP_008912854.1;YP_008912855.1;
	X809_17235;X809_17240;	YP_008912856.1;YP_008912857.1;
	X809_17250;X809_17255;	YP_008912858.1;YP_008912859.1;
	X809_17260;X809_17265;	YP_008912860.1;YP_008912861.1;
	X809_17270;X809_17275;	YP_008912862.1;YP_008912863.1;
	X809_17280;X809_17285;	YP_008912864.1;YP_008912865.1;
	X809_17290;X809_17295;	YP_008912866.1;YP_008912867.1;
	X809_17300;X809_17305;	YP_008912868.1;YP_008912869.1;
	X809_17310;X809_17315;	YP_008912870.1;YP_008912871.1;
	X809_17320;X809_17325;	YP_008912872.1;YP_008912873.1;
	X809_17330;X809_17335;	YP_008912874.1;YP_008912875.1;
	X809_17340;X809_17345;	YP_008912876.1;YP_008912877.1;
	X809_17350;X809_17355	YP_008912878.1;YP_008912879.1
NRPS	X809_23470;X809_23475;	YP_008914041.1;YP_008914042.1;
	X809_23480;X809_23485;	YP_008914043.1;YP_008914044.1;
	X809_23490;X809_23495;	YP_008914045.1;YP_008914046.1;
	X809_23500;X809_23505;	YP_008914047.1;YP_008914048.1;
	X809_23510;X809_23520;	YP_008914049.1;YP_008914050.1;
	X809_23525;X809_23530;	YP_008914051.1;YP_008914052.1;
	X809_23535;X809_23540;	YP_008914053.1;YP_008914054.1;
	X809_23545;X809_23550;	YP_008914055.1;YP_008914056.1;
	X809_23555;X809_23560;	YP_008914057.1;YP_008914058.1;
	X809_23565;X809_23570;	YP_008914059.1;YP_008914060.1;
	X809_23575;X809_23580;	YP_008914061.1;YP_008914062.1;
	X809_23585;X809_23590	YP_008914063.1;YP_008914064.1
Phosphonate	X809_26780;X809_26785;	YP_008914670.1;YP_008914671.1;
	X809_26790;X809_26795;	YP_008914672.1;YP_008914673.1;
	X809_26800;X809_26805;	YP_008914674.1;YP_008914675.1;
	X809_26810;X809_26815;	YP_008914676.1;YP_008914677.1;
	X809_26820;X809_26825;	YP_008914678.1;YP_008914679.1;
	X809_26830;X809_26835;	YP_008914680.1;YP_008914681.1;
	X809_26840;X809_26845;	YP_008914682.1;YP_008914683.1;
	X809_26850;X809_26855;	YP_008914684.1;YP_008914685.1;
	X809_26860;X809_26865;	YP_008914686.1;YP_008914687.1;
	X809_26870;X809_26875;	YP_008914688.1;YP_008914689.1;
	X809_26880;X809_26885;	YP_008914690.1;YP_008914691.1;
	X809_26890;X809_26895;	YP_008914692.1;YP_008914693.1;
	X809_26900;X809_26905;	YP_008914694.1;YP_008914695.1;

	X809_26910;X809_26915;	YP_008914696.1;YP_008914697.1;
	X809_26920;X809_26925;	YP_008914698.1;YP_008914699.1;
	X809_26930;X809_26935;	YP_008914700.1;YP_008914701.1;
	X809_26940;X809_26945;	YP_008914702.1;YP_008914703.1;
	X809_26950;X809_26955;	YP_008914704.1;YP_008914705.1;
	X809_26960;X809_26965;	YP_008914706.1;YP_008914707.1;
	X809_26970;X809_26975;	YP_008914708.1;YP_008914709.1;
	X809_26980;X809_26985;	YP_008914710.1;YP_008914711.1;
	X809_26990;X809_26995;	YP_008914712.1;YP_008914713.1;
	X809_27000;X809_27005	YP_008914714.1;YP_008914715.1

Appendix 4. Transporter classification for sequenced *P. polymyxa* strains.

SUPER FAMILIES	CR1	E681	M1	SC2
Major Facilitator Superfamily (MFS)	67	60	71	70
ATP-binding Cassette (ABC) Superfamily	409	351	367	361
Drug/Metabolite Transporter (DMT) Superfamily	21	15	22	20
Resistance-Nodulation-Cell Division (RND) Superfamily	6	4	5	5
Multidrug/Oligosaccharidyl-lipid/Polysaccharide (MOP) Flippase Superfamily	13	12	11	11
Bacterial Flagellar Motor/ExbBD Outer Membrane Transport Energizer (Mot/Exb) Superfamily	2	4	4	4
P-type ATPase (P-ATPase) Superfamily	7	8	7	7
H+- or Na+-translocating F-type, V-type and A-type ATPase (F-ATPase) Superfamily	9	10	10	10
Iron/Lead Transporter (ILT) Superfamily	2	2	2	2
Family	**CR1**	**E681**	**M1**	**SC2**
10 TMS Putative Sulfate Exporter (PSE) Family	1	1	1	1
2-Hydroxycarboxylate Transporter (2-HCT) Family	1	2	1	1
2-Keto-3-Deoxygluconate Transporter (KDGT) Family	1	1	1	1
Alanine or Glycine:Cation Symporter (AGCS) Family	2	2	2	2
Amino Acid-Polyamine-Organocation (APC) Family	3	4	4	4
Ammonia Transporter Channel (Amt) Family	1	1	1	1
Aromatic Acid Exporter (ArAE) Family	2	2	2	2
Arsenite-Antimonite (ArsAB) Efflux Family	1	1	2	1
Autoinducer-2 Exporter (AI-2E) Family (Formerly PerM Family, TC #9.B.22)	6	6	6	6
Auxin Efflux Carrier (AEC) Family	4	3	4	4
Bacterial Competence-related DNA Transformation Transporter (DNA-T) Family	1	1	1	1

Branched Chain Amino Acid Exporter (LIV-E) Family	4	2	2	2
Branched Chain Amino Acid:Cation Symporter (LIVCS) Family	1	1	1	1
C4-Dicarboxylate Uptake (Dcu) Family	0	0	1	1
Ca2+:Cation Antiporter (CaCA) Family	1	1	1	1
Camphor Resistance (CrcB) Family	2	2	1	2
Cation Channel-forming Heat Shock Protein-70 (Hsp70) Family	1	1	1	1
Cation Diffusion Facilitator (CDF) Family	5	4	4	4
Chloroplast Envelope Protein Translocase (CEPT or Tic-Toc) Family	2	2	2	1
Concentrative Nucleoside Transporter (CNT) Family	1	1	1	1
Copper Resistance (CopD) Family	1	1	1	1
CorA Metal Ion Transporter (MIT) Family	3	3	3	3
Cph1 Holin (Cph1 Holin) Family	0	0	0	1
MPA1-C Family	2	2	2	2
DedA or YdjX-Z (DedA) Family	7	8	5	5
Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS) Family	2	2	2	2
Dipicolinic Acid Transporter (DPA-T) Family	2	2	2	2
Disulfide Bond Oxidoreductase B (DsbB) Family	1	1	1	1
Disulfide Bond Oxidoreductase D (DsbD) Family	1	1	1	1
Ethanolamine Facilitator (EAF) Family	0	1	0	0
Ferrous Iron Uptake (FeoB) Family	1	1	1	1
Formate-Nitrite Transporter (FNT) Family	2	2	2	2
General Secretory Pathway (Sec) Family	7	7	8	8
Gluconate:H+ Symporter (GntP) Family	4	3	2	2
Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter Family	5	4	5	5
Hly III (Hly III) Family	1	1	1	1
HlyC/CorC (HCC) Family	4	4	4	4
Hydroxy/Aromatic Amino Acid Permease (HAAAP) Family	1	0	2	1
Inorganic Phosphate Transporter (PiT) Family	1	1	1	1
K+ Transporter (Trk) Family	2	2	2	2
Large Conductance Mechanosensitive Ion Channel (MscL) Family	1	1	1	1
L-Lysine Exporter (LysE) Family	1	2	1	2
LrgA Holin (LrgA Holin) Family	1	1	1	1
Major Intrinsic Protein (MIP) Family	1	1	1	1
Metal Ion (Mn2+-iron) Transporter (Nramp) Family	1	1	1	1
Minor Capsid Protein, gp7 of Baccilus subtilis Phage SPP1 (gp7) Family	1	0	2	1
Monovalent Cation:Proton Antiporter-2 (CPA2) Family	2	2	2	2
Na+-transporting Carboxylic Acid Decarboxylase (NaT-DC) Family	2	1	0	0
Neurotransmitter:Sodium Symporter (NSS) Family	1	1	1	1

Nisin (Nisin) Family	0	1	0	0
Nucleobase:Cation Symporter-2 (NCS2) Family	2	2	2	2
Outer Membrane Factor (OMF) Family	4	4	3	3
Outer Membrane Protein Secreting Main Terminal Branch (MTB) Family	0	0	1	0
Phosphate:Na+ Symporter (PNaS) Family	1	1	1	1
Phosphotransferase System Enzyme I (EI) Family	2	2	2	2
Pore-forming ESAT-6 Protein (ESAT-6) Family	7	9	8	9
Prokaryotic Molybdopterin-containing Oxidoreductase (PMO) Family	5	5	5	5
Prokaryotic Succinate Dehydrogenase (SDH) Family	3	3	3	3
Proposed Fatty Acid Transporter (FAT) Family	0	1	1	2
PTS Fructose-Mannitol (Fru) Family	1	3	3	3
PTS Galactitol (Gat) Family	2	1	1	1
PTS Glucose-Glucoside (Glc) Family	21	13	16	16
PTS Lactose-N,N'-Diacetylchitobiose-beta-glucoside (Lac) Family	4	2	3	3
PTS Mannose-Fructose-Sorbose (Man) Family	5	5	5	5
Putative Arginine Transporter (ArgW) Family	2	2	2	2
Putative Bacterial Murein Precursor Exporter (MPE) Family	2	3	3	3
Putative Heme Handling Protein (HHP) Family	1	1	1	1
Putative Inorganic Carbon (HCO3-) Transporter/O-antigen Polymerase (ICT/OAP) Family	1	0	0	0
Putative Mg2+ Transporter-C (MgtC) Family	2	2	2	2
Putative Permease Duf318 (Duf318) Family	1	1	1	1
Resistance to Homoserine/Threonine (RhtB) Family	4	2	4	4
SdpC (Peptide-Antibiotic Killer Factor) Immunity Protein, SdpI (SdpI) Family	1	1	1	1
SecDF-associated Single Transmembrane Protein, YajC (YajC) Family	1	1	1	1
Sensor Histidine Kinase (SHK) Family	5	5	5	5
Septal DNA Translocator (S-DNA-T) Family	2	2	3	2
Small Conductance Mechanosensitive Ion Channel (MscS) Family	1	1	1	1
Staphylococcus aureus Putative Quorum Sensing Peptide Exporter, AgrB (AgrB) Family	1	1	1	1
Stomatin/Podocin/Band 7/Nephrosis.2/SPFH (Stomatin) Family	3	3	3	4
Sulfate Permease (SulP) Family	2	2	2	2
Tellurium Ion Resistance (TerC) Family	4	4	4	5
Testis-Enhanced Gene Transfer (TEGT) Family	0	0	0	1
Threonine/Serine Exporter (ThrE) Family	1	1	1	1
Tricarboxylate Transporter (TTT) Family	1	1	1	1
Twin Arginine Targeting (Tat) Family	2	2	2	2
Type III (Virulence-related) Secretory Pathway (IIISP) Family	11	11	11	11
Type IV (Conjugal DNA-Protein Transfer or VirB) Secretory Pathway (IVSP) Family	0	0	2	1

YaaH (YaaH) Family	1	1	1	1
YebN (YebN) Family	2	2	2	2
YggT or Fanciful K+ Uptake-B (FkuB; YggT) Family	1	1	1	1
Zinc (Zn2+)-Iron (Fe2+) Permease (ZIP) Family	1	1	1	1
RD1 or ESX-1/Snm Protein Secretion System (RD1) Family	7	9	8	9
	CR1	**E681**	**M1**	**SC2**
TOTAL TRANSPORTERS	751	672	711	704

All transporter classifications were obtained from the Transporter Classification Database. Categorization of putative transporter genes was performed using available tools on the JGI IMG database

Appendix 5. CaZY profile of sequenced *P. polymyxa* strains.

GH Family	Number of CDS				PL Family	Number of CDS			
	E681	M1	SC2	CR1		E681	M1	SC2	CR1
1	8	10	8	17	1	4	3	3	5
2	4	4	4	7	3	1	1	1	1
3	4	4	4	7	9	2	2	2	1
4	3	3	3	3	10	1	1	1	1
5	7	7	7	5	11	11	2	2	2
6	1	1	1	1	NC	0	1	1	0
10	1	3	1	2	CE Family	Number of CDS			
11	1	2	1	2		E681	M1	SC2	CR1
13	9	9	9	10	1	2	3	2	3
14	1	1	1	1	2	1	1	1	1
16	2	3	2	1	4	15	13	12	13
18	3	3	3	3	7	2	2	2	2
23	2	3	2	3	8	2	2	2	2
24	1	1	1	0	9	1	1	1	1

25	2	3	2	3
26	5	5	5	4
27	1	1	1	1
28	1	1	1	1
30	2	2	2	2
31	1	1	1	1
32	10	10	10	9
35	1	1	1	1
36	4	4	4	3
38	0	0	0	1
42	3	3	3	5
43	10	10	10	9
44	1	1	1	1
46	1	1	1	1
48	1	1	1	0
51	3	3	3	3
52	1	1	1	2
53	1	1	1	2
65	1	1	1	1
67	1	1	1	1
68	1	1	1	1
74	1	1	1	1
78	1	1	1	3
81	1	1	1	0
84	1	1	1	1

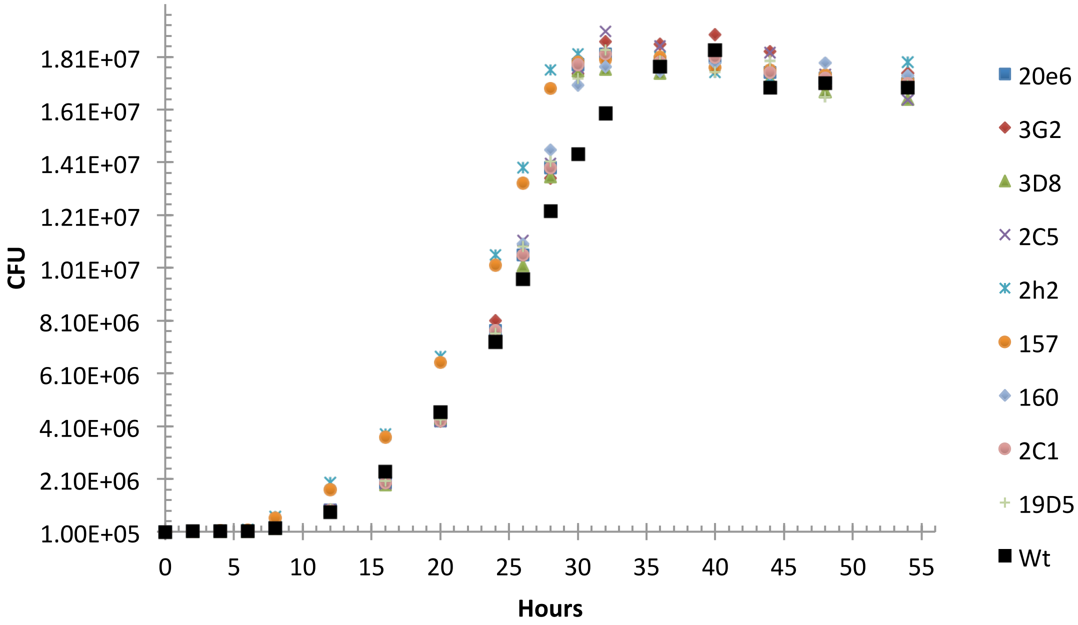
12	4	4	4	4
14	1	1	1	1

CBM Family	Number of CDS			
	E681	M1	SC2	CR1
3	5	5	5	4
6	2	2	2	2
12	0	0	0	1
13	1	3	3	2
16	0	1	1	0
22	1	2	2	2
25	2	2	1	2
26	0	1	0	1
32	2	3	3	2
34	1	1	1	1
35	4	3	3	1
36	2	3	2	3
38	0	3	3	1
41	1	1	1	1
46	1	1	1	1
48	2	2	2	2
50	3	6	4	7
56	1	1	0	0
59	0	1	1	0
61	0	0	0	1
63	0	1	1	1

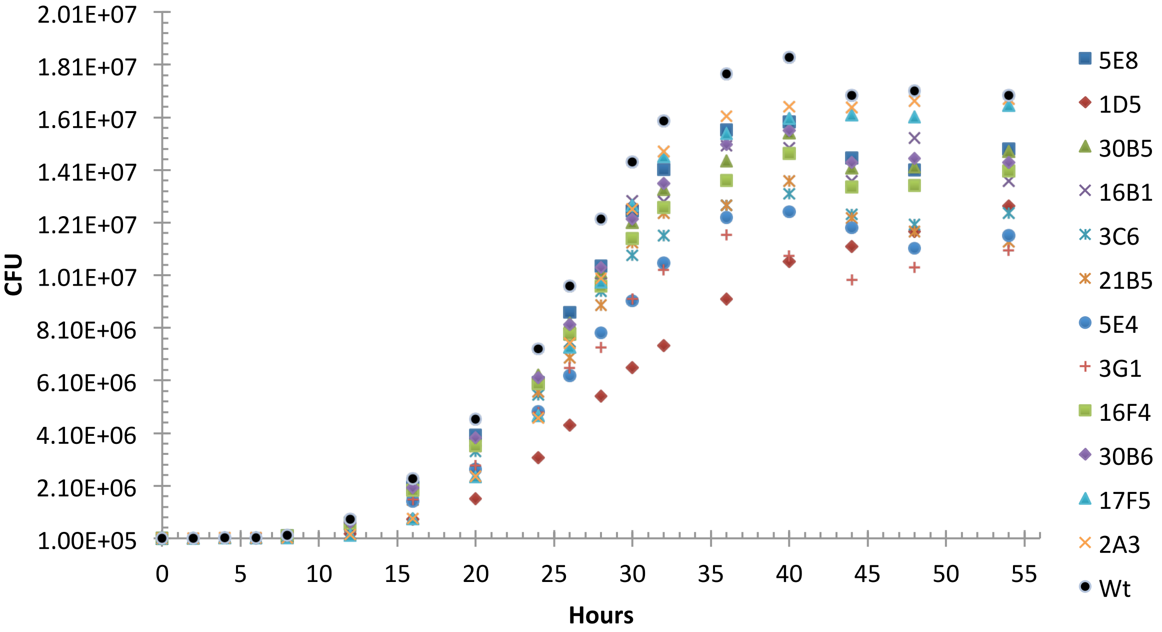
88	1	1	1	1	66	3	3	3	7
94	2	2	2	3					
95	1	1	1	1					
105	3	3	3	3					
112	1	1	1	1					
115	1	1	1	1					
120	0	0	0	1					
127	1	1	1	1					
130	2	2	2	2					
NC	1	1	1	0					

Data was obtained from the CAzY Database. GH – Glycoside Hydrolase, PL – Pectin Lyase, CE- Carbohydrate Esterase, CBM – Carbohydrate Binding Motif

Appendix 6. Growth curves of Tn5 isolates displaying increased growth on lignin.



Appendix 7. Growth curves of Tn5 isolates displaying decreased growth.



Curriculum Vitae

Name:	Alexander William Eastman
Post-secondary Education and Degrees:	University of Western Ontario London, Ontario, Canada 2009-2013 BMSc – Honours Biochemistry *with distinction*
	The University of Western Ontario London, Ontario, Canada 2013-2015 M.Sc - Microbiology
Honours and Awards:	Dean's List 2009-2013
	Western Graduate Research Scholarship 2013-2015
Related Work Experience :	Research Assistant Agriculture and Agri-Food Canada 2013-2015

Publications:

Eastman, A.W., Nathoo, N., Weselowski, B., and Yuan, Z-C. (2014) Complete Genome Sequence of *Paenibacillus polymyxa* CR1, a Plant Growth-Promoting Bacterium Isolated from the Corn Rhizosphere Exhibiting Potential for Biocontrol, Biomass Degradation, and Biofuel Production. *Genome Announc.* **2**(1)**:**e01218-13

Eastman, A.W., Heinrichs, D.E., and Yuan, Z-C. (2014) Comparative and genetic analysis of the four sequenced *Paenibacillus polymyxa* genomes reveals a diverse metabolism and conservation of genes relevant to plant-growth promotion and competitiveness. *BMC Genomics*. **15:**851

Eastman, A.W., and Yuan, Z-C. (2015) Development and validation of an rDNA operon based primer walking strategy applicable to *de novo* bacterial genome finishing. *Front. Microbiol.* **5:**769

Eastman, A.W., and Yuan, Z-C. (2015) Cellulosic biofuels – challenges and opportunities (submitted)

Hassan, I., Eastman, A.W., Weselowski, B., and Yuan, Z-C. (2015) Genome sequencing and biomass metabolism by *Arthrobacter arilaitensis* FG1, isolated from degrading plant residues. (in preparation)

Weselowski, B., Chou, N, Nathoo, N., Eastman, A.W., and Yuan, Z-C. (2015) Isolation,

identification and characterization of *Paenibacillus polymyxa* CR1 with potential for bio-fertilization, biomass degradation and fuel production. (in review)