Proteomics and Genomics of Biobutanol Production from *Clostridium beijerinckii*

by

Timothy Eric Cargal

Submitted in Partial Fulfillment of the Requirements

For the Degree of

Master of Science

in the

Biological Sciences

Program

Youngstown State University

August, 2015

Proteomics and Genomics of Biobutanol Production from Clostridium beijerinckii

Cargal, Timothy Eric

Signature:

_____

Timothy Eric Cargal, Student Date

Approvals:

_____

Dr. Gary Walker, Thesis Advisor Date

_____

Dr. David Asch, Committee Member Date

_____

Dr. Jonathan Caguiat, Committee Member Date

_____

Dr. Sal Sanders, Dean of School of Graduate Studies & Research Date

Abstract

The world consumption of energy is increasing at an exponential rate. Furthermore, the availability of our fuel of choice, fossil fuels, is declining, with estimates that reserves will be gone within fifty years. As such, the search for alternative fuel sources has begun. One promising source of energy is biomass, solar energy stored in the form of sugars by plants. Currently, the focus is on ethanol as a biomass energy source, but it is not without its problems. An emerging alternative to ethanol is butanol, which is chemically more similar to gasoline. Exploiting the Acetone-Butanol-Ethanol Fermentation pathway found in certain species of bacteria can produce butanol. However, this has its own problems in that the butanol is itself toxic to the bacteria. We have identified four potential targets for gene manipulation in order to improve the bacterial resistance to butanol in order to facilitate the production of more butanol. These four genes were G3P, CWBR, PF, and *cspA*. These targets have been analyzed at both the protein expression level and gene expression level. Of these four genes we tested, G3P showed expression levels as expected, PF did not, and the other two were inconclusive due to statistical outliers.

Acknowledgements

I would like to thank my family and girlfriend for their invaluable support and encouragement. I would also like to thank my advisor Dr. Walker and committee members Dr. Asch and Dr. Caguiat for their dedicated support and guidance. Finally, I would like to thank the rest of the Masters students at YSU for their camaraderie.

Table of Contents

Abstract iii

Acknowledgements iv

Table of Contents v

List of Figures vii

List of Tables viii

Introduction 1

- Energy Consumption and Traditional Sources 2
  - Fossil Fuels 2
  - Renewable Energy 5
- Biomass as Energy 6
  - Ethanol 7
  - Butanol 8
- Acetone-Butanol-Ethanol Fermentation (ABE Fermentation) 10
  - Biomass Substrate 11
  - *Clostridium acetylbutylicum* and *C. beijerinckii* 12
  - Biobutanol Yield 13
  - Reaction Environment Effects on Clostridia Growth 14
  - Understanding Yield Through Proteomics and Genomics 16

Specific Aims 19

Materials and Methods 20

- Chemicals 20
- Organism 20
- Media 20
- Cell Culturing and Collecting 21
- Protein Analysis 22
- Nucleic Acid Analysis 23

Results 25

- Protein Analysis 25

Nucleic Acid Analysis 26
Discussion 30
Appendix A – Energy Conversion Factors 33
Appendix B – Protein Gels 33
Appendix C – Detailed qPCR Methodology 36
Appendix D – G3P Gene Sequence 36
Appendix E – PF Gene Sequence 37
Appendix F – CWBR Gene Sequence 38
Appendix G – *cspA* Gene Sequence 39
Appendix H – qPCR Curves 40
Bibliography 47

List of Figures

Total Final Consumption 3

Chemical Properties of Butanol 8

Biochemical Pathway of Biobutanol Production in *C. beijerinckii* 15

qPCR Primer Check 27

qPCR Average Results 29

One Dimensional Gel Electrophoresis Days 1-5 35

One Dimensional Gel Electrophoresis Days 6-10 36

Two Dimensional Gel Electrophoresis Day 3 36

Two Dimensional Gel Electrophoresis Day 5 37

Two Dimensional Gel Electrophoresis Day 7 37

All qPCR Graphs 40

All G3P qPCR Graphs 41

All CWBR qPCR Graphs 41

All PF qPCR Graphs 42

All cspA qPCR Graphs 42

G3P Day 3 qPCR Graphs 43

G3P Day 7 qPCR Graphs 43

CWBR Day 3 qPCR Graphs 44

CWBR Day 7 qPCR Graphs 44

PF Day 3 qPCR Graphs 45

PF Day 7 qPCR Graphs 45

cspA Day 3 qPCR Graphs 46

cspA Day 7 qPCR Graphs 46

List of Tables

qPCR Primers 26
qPCR Data 29

1. INTRODUCTION

Around the world, energy consumption is rising. In the year 2011, the United States alone consumed over 97 quadrillion British Thermal Units (BTUs) of energy. Around the world, nearly 520 quadrillion BTUs of energy were expended that same year. Of that amount, a significant proportion was derived from fossil fuels in some form: nearly 80 quadrillion BTU (82% of total) in the United States, and roughly 425 quadrillion BTU (81.6% of total) in the world. Renewable energy sources, including biofuels, make up approximately 10% of the energy consumed (5, 12). If energy consumption continues at this rate and proportion, it will not be long before we have insufficient quantities of fossil fuels available for energy use, and the quantities of non-fossil fuel derived energy will be unable to support energy demand.

One of the more promising alternative sources of energy is the chemical butanol, C_4H_9OH. Butanol is a colorless organic solvent with a distinct smell that has similar potential as a fuel compared to gasoline, more so than the current biofuel of choice, ethanol. In fact, butanol can be used in gasoline engines without modification, whereas ethanol additives to gasoline have been shown to cause damage to engines. Compared to ethanol, butanol has a greater energy content, but is less volatile, less hydroscopic, and less corrosive (15, 20).

Up until the first half of the twentieth century, fermentative butanol production was one of the largest biotechnological processes, second only to the production of ethanol from *Saccharomyces cerevisiae*. Molasses was used as a substrate for butanol production by utilizing the Acetone-Butanol-Ethanol

fermentation pathway (ABE fermentation) found in some bacterial species such as *Clostridium beijerinckii*. Shortly after the Second World War, unfavorable economic conditions and the growing petrochemical industry led to the phasing out of biological production in favor of large scale petrochemical production (6, 11, 20).

As a part of the environmental friendly movement in industry, many groups are returning to the green production of butanol from bacteria and the ABE fermentation pathway. Clostridia are capable of producing butanol from a variety of substrates, including food stocks such as corn and sugar cane and non-food stocks such as switch grass and woodchip extract (4, 14). However, the same problems in biotechnological butanol production that allowed the petrochemical production to be more efficient in the mid-twentieth century still persist, primarily various forms of product inhibition. This has led to many research groups attempting new methods to optimize the biological production through a wide variety of methods targeting various steps of the production pathway (10, 29).

1.1. Energy Consumption and Traditional Sources:

1.1.1. Fossil Fuels:

The primary source of energy around the world is fossil fuel, which comes in three forms: coal, petroleum (or oil), and natural gas. Together, they constitute the primary source of nearly two thirds of all energy consumed in the world, and three fourths or more just a few decades ago (see Figure 1). All three of these originate

from organic matter that has been buried and compressed under high pressure for million of years. Although they come from natural sources, they are considered non-renewable because the amount of time to produce them is not practical on a human time scale, taking millions of years. Furthermore, because they come from natural sources, they contain some significant amount of sulfur, nitrogen, and phosphorus, and traces of many other elements, in addition to any possible contaminants from the surrounding rock. As a result, when fossil fuels are burned, they will release significant amounts of pollutants (23).

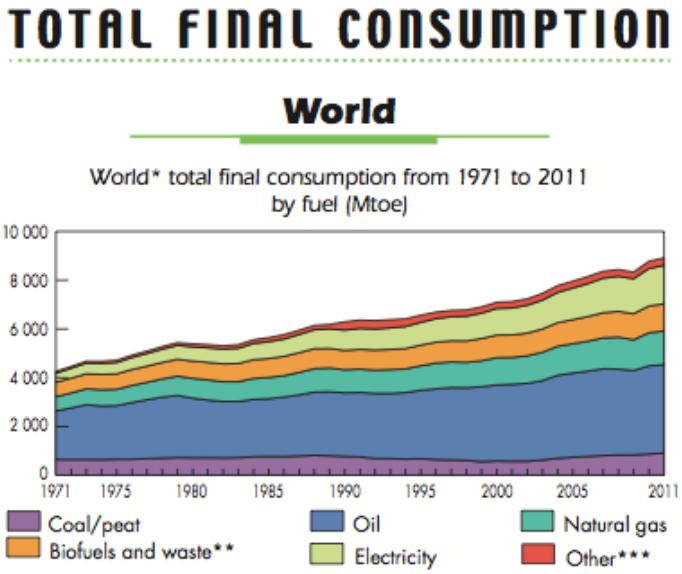


Figure 1 *World includes international aviation and marine bunkers; ** Biofuel data before 1994 is estimated; ***Other includes geothermal, solar, and wind energy (12)

A further complication with fossil fuels is that not all sources are readily available where they can be retrieved; some deposits are buried in locations where the cost of extraction is prohibitive. Although the advancement of technology has previously facilitated fuel extraction, allowing previously unreachable deposits to become attainable while still keeping costs to a minimum, this is no longer the case. World energy consumption is now increasing at a rate faster than the advancements in mining technology, causing the cost to rise. In addition, total reserves of fossil fuels are drying up. While estimates on how much remains vary, most independent

studies indicate that all fossil fuels will have been expended within a few decades (17).

Of the three types of fossil fuels, coal is the most harmful to the environment. Coal deposits are typically tainted with various fluorides and heavy metals, including radioactive isotopes such as uranium. When any fossil fuel is burned for energy, the reaction of importance is the combustion of hydrocarbons into carbon dioxide and water, with trace amounts of carbon monoxide as a result of incomplete combustion. Unfortunately, the other contaminants will also react with oxygen when burned, resulting in the production of toxic gases such as sulfur dioxide, nitrogen oxides, and fluorides, in addition to releasing the heavy metals and any radioactive particles. In the United States and other well-developed countries, there are laws that regulate how much pollution may be released from coal burning factories, and equipment such as scrubbers, catalytic converters, and electrostatic precipitators are installed in the smoke stacks for this purpose. Unfortunately, as coal also tends to be the most easily accessible and readily available of the three fossil fuels, coal is used to a large extent in developing countries, whose laws governing pollution may be more lax or less strictly enforced, leading to large scale pollution. A prominent example of this is China, whose recent technological boom combined with overpopulation has resulted in one of the highest pollution levels in the world (3, 26, 27).

The most commonly used of the three fossil fuels (referring back to Figure 1) is petroleum, typically in the refined form known as gasoline or petrol. Gasoline is primarily composed of hydrocarbons ranging in size from five to twelve carbons,

may or may not be saturated, and may be linear or cyclical in structure. The proportion of these constituents is monitored to some degree, as they will affect the vapor pressure of the mixture, which will in turn affect the efficacy of the engine. If the vapor pressure is too low in a cold environment, then the gasoline will have difficulty combusting and starting the engine. On the other hand, a high vapor pressure in a warm environment will cause the release of significantly more pollutants (8, 21).

1.1.2. Renewable Energy:

Renewable energy can be grouped into three general categories: tidal power, geothermal energy, and solar energy. All three of these involve tapping the energy released by inorganic sources that are beyond human control and will be renewed for as long as the earth exists. Tidal power is a result of the moon's revolution around the earth and the associated gravitational effect on water levels, and can be harnessed by turbines that are turned as large amounts of water move past them. While it is possibly the most reliable of the renewable energy sources, its use is limited by the necessity of a larger differential between high and low tides than is found in most parts of the world. Geothermal energy comes from the internal heat energy of the earth produced in part by the decay of radioactive particles, and is typically harnessed by using the heat to boil water, generating steam that turns a turbine. While also a reliable source of energy, it does require the presence of a geothermal reservoir, which does not always leave evidence on the surface so that

we can find and tap them. They are also typically found around tectonic plate boundaries, which are known for being areas of geographic instability (18).

Solar energy is a result of the sun's radiation heating portions of the earth's surface and comes in many forms, including wind power (being the result of hot air rising and cold air flowing in to take it's place) and biomass (being the result of photosynthetic organisms using light to convert carbon dioxide and oxygen into organic materials). Despite being the most easily tapped and most widely distributed, solar power is the least reliable of renewable energy sources as weather conditions and celestial events such as solar eclipses can prevent solar radiation from reaching the earth's surface where it can be utilized. Solar energy also needs large areas in order to be harvested efficiently; a single wind turbine is not going to generate significant amounts of energy, but many wind turbines will collect useable amounts of energy. Even solar plants that collect the energy directly require large amounts of space and must be able to efficiently collect sunlight from many directions. These limits to availability and productivity are why fossil fuels are more widely used than renewable energy sources, even though renewable energy sources tend to produce negligible amounts of pollution (18).

1.2. Biomass as Energy:

Biomass is a collective term for solar energy that has been stored in carbohydrates by plants and other photosynthetic organisms. The most common and oldest biomass fuel source is wood, but large quantities are typically reserved

for other applications. Most biomass fuels today are refined or extracted from other sources.

1.2.1. Ethanol:

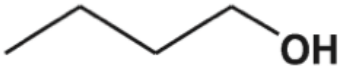
Ethanol is a colorless two-carbon alcohol (C_2H_5OH) that is currently frequently used as an additive to gasoline because of its ability to raise the octane rating (a measure of how much the fuel can be compressed before combustion). The biological production of ethanol from food sources by fermentation from *Saccharomyces cerevisiae* is the oldest known and largest biotechnological process. A wide variety of food stocks can be used to produce ethanol, including sugarcane, sugar beets, and wheat, but the most common in the United States is corn (25).

In addition to being used as an additive, ethanol can be used as a straight fuel source for automobiles and other devices that run on gasoline. Although significantly cleaner, ethanol is at a disadvantage as compared to gasoline as any engines that currently run on gasoline would have to be altered in order to accept ethanol as a fuel source without causing damage or corrosion (most engines can only handle ethanol concentrations up to around ten percent). As a further complication, the amount of corn required to produce enough ethanol to fill the gas tank of a sport utility vehicle (240 kilograms of corn for 100 liters of ethanol) can feed a person for a year. In fact, the development of the American biofuel production industry caused a significant spike in the price of corn (25).

A

Properties	Butanol
Melting point (°C)	−89.3
Boiling point (°C)	117.7
Ignition temperature (°C)	35
Flash point (°C)	365
Density at 20°C (g/mL)	0.8098
Critical pressure (hPa)	48.4
Critical temperature (°C)	287

Butanol structure



B

		Properties of fuels		
	Butanol	Gasoline	Ethanol	Methanol
Energy density (MJ/L)	29.2	32	19.6	16
Air-fuel ratio	11.2	14.6	9	6.5
Heat of vaporization (MJ/kg)	0.43	0.36	0.92	1.2
Research octane number	96	91–99	129	136
Motor octane number	78	81–89	102	104

Figure 2 Chemical Properties of Butanol (15) A) Butanol specifications; B) Comparison to other fuels

1.2.2. Butanol:

Butanol is a colorless four-carbon alcohol (C_4H_9OH) that is being investigated as a possible biofuel. Butanol has four isomeric forms: the straight chain form with the alcohol group on one of the terminal carbons, known as *n*-butanol or 1-butanol; the straight chain form with the alcohol group on one of the inner carbons, known as *sec*-butanol or 2-butanol; the branched form (where one carbon is connected to all three of the others) with the alcohol group on one of the terminal carbons, known as isobutanol or 2-methyl-1-propanol; and the branched form with the alcohol group attached to the central carbon, known as *tert*-butanol or 2-methyl-2-propanol (16).

When talking about butanol as a biofuel, this is generally referring to 1-butanol, which has some significant advantages over ethanol, some of which are shown in Figure 2 above. First, since it has more carbon-carbon bonds per unit mass than ethanol, butanol contains significantly more energy that can be released

through combustion reactions. Butanol also has a higher vapor pressure, is less hygroscopic, and has a lower heat of vaporization (MJ/kg) than ethanol. These factors together mean that butanol is less corrosive than ethanol and can thus be stored and transported in the same manner as gasoline, and can be used as a fuel source (100% replacement of gasoline with butanol) without making any modifications to existing machinery or infrastructure (6). As a final bonus, butanol fuel produces far less emissions than gasoline based fuel. In 2005, David Ramey and his team drove a 1992 Buick over 8,000 miles and underwent emission tests in nine states, with average hydrocarbon and carbon monoxide levels at five percent or less of the permissible limit (20).

In the chemical butanol industry, there are three primary reactions used. The first, oxo synthesis, utilizes propylene and reacts it with carbon monoxide and hydrogen gas in the presence of a metal catalyst (either cobalt, rhodium, or ruthenium). This produces the aldehydes butyraldehyde (also known as butanal) and isobutyraldehyde, which are then converted to 1-butanol and isobutanol by catalytic hydrogenation. The isomeric ratio of the two forms of butanol can be manipulated by controlling the pressure and temperature of the reaction, as well as through the choice of catalyst. The second method for chemical butanol synthesis, the Reppe method, also utilizes propylene and carbon monoxide with a metal catalyst, but replaces hydrogen gas with water to directly produce butanol, in both 1-butanol and isobutanol isoforms, and carbon dioxide at low temperature and pressure. Although effective, this method has had limited commercial success due to the expensive equipment necessary. The final chemical butanol reaction,

crotonaldehyde hydrogenation, begins with acetaldehyde undergoing aldol condensation to form 3-hydroxybutanal, which is then dehydrated to but-2-enal, more commonly known as crotonaldehyde, and then hydrogenated to 1-butanol (15). Both the oxosynthesis and Reppe method are dependent upon petrol and produce toxic unwanted byproducts. While greener (acetaldehyde can be dehydrogenated from ethanol), crotonaldehyde rehydration also produces toxic intermediates.

1.3. Acetone-Butanol-Ethanol Fermentation (ABE Fermentation):

ABE fermentation was the original primary source of butanol, was replaced by chemical production from petrol, and is now making a comeback as part of the biobutanol industry. While Louis Pasteur recorded ABE fermentation as early as 1862 from a mixed culture, ABE fermentation from pure cultures was not accomplished until later that century by independent scientists Albert Fitz and Matrinus Beijerinck. The first case of ABE fermentation on an industrial scale was 1915, towards the beginning of the First World War, when Chaim Weizmann used *Clostridium acetobutylicum* for the production of acetone, the bulk chemical needed for cordite, a component of smokeless gunpowder. At the time, the butanol was simply stored as waste. This changed during the American Prohibition; the banning of ethanol production also limited the amount of amyl alcohol, which was used to produce amyl acetate, a component of quick-drying lacquers for industrial purposes. Butanol was then used to produce butyl acetate, an adequate substitute (6).

1.3.1. Biomass Substrate:

The initial step with any fermentation pathway is finding a satisfactory substrate, one that is cheap and readily available, but can be used by the fermenting organism. In the simplest form, various forms of sugar would be used as feedstock for the bacteria to produce biofuel; however, as seen with the attempt to use corn as a feedstock for bioethanol production, this can have dramatic effects on the availability and cost of the sugar source. As a result, the search for a food stock for biobutanol has turned away from food stocks such as corn to other forms of biomass, including the non-edible portions of food crops such as corn husks (25).

Non-edible or lignocellulosic biomass is primarily composed of the polysaccharides cellulose, a linear chain of β(1→4) linked D-glucose, and hemicellulose, a branched chain of many different pentose and hexose sugars. In most plants, the predominating form of hemicellulose is xylan, long chains of the pentose sugar D-xylose. As cellulose and hemicellulose are both structural polysaccharides, they are tightly bundled (cellulose more so than hemicellulose) and tend to be difficult for bacteria or other organisms to break down, resulting in the need to use destructive industrial processes such as acid washing to breakdown the polysaccharides into shorter oligosaccharides that can be digested by microorganisms. Unfortunately, these industrial processes tend to produce many toxic byproducts, including furans such as furfural. A more ideal situation would be to use bacteria that could perform ABE fermentation, as well as break down

hemicellulose without industrial assistance beyond pulling out the cellulose for use in the making of paper goods (2,14).

1.3.2. *Clostridium acetylbutylicum* and *C. beijerinckii:*

Clostridia species are the most common bacteria used for butanol production, although other species have been employed such as *Butyribacterium methylotrophicum* and *Hyperthermus butylicus*. Clostridia are spore forming, rod-shaped Gram negative bacteria, and are strict anaerobes. Because of this, they have certain disadvantages as compared to bioethanol production by *Saccharomyces cerevisiae*; namely, they require a more stringently controlled environment, and production yields tend to be significantly lower (6). Furthermore, Clostridia tend to have slower reproduction rates, and can be easily out-competed by other obligate and facultative anaerobes should purity or sterility be compromised (1).

Although *C. acetylbutylicum* has been the standard for biobutanol fermentation, *C. beijerinckii* has been receiving increased interest. Originally styled as a strain of *C. acetylbutylicum*, *C. beijerinckii* has a genome that is roughly half again as large as that of *C. acetylbutylicum*. Within this enlarged genome is believed to be the gene for 1,4-beta-xylosidase, an enzyme that would allow for *C. beijerinckii* to break down hemicellulose, and more specifically xylan, to release the pentose sugars and produce butanol. If true, this would allow for the use of waste biomass as a substrate for *C. beijerinckii* fermentation without the need for industrial processing, thereby reducing costs and pollutants (6, 9, 14).

1.3.3. Biobutanol Yield

As with all industrial processes, one concern with biobutanol production is product yield. As a general rule, purely biological processes are less efficient than purely industrial. Part of the problem is that not all of the substrate will be used to make the desired product; unlike an industrial process or chemical reaction, some of the substrate will be used simply to maintain the reaction. In the case of Clostridia and ABE fermentation from hemicellulose, only some of the hemicellulose is going to be converted into butanol and the other fermentation products, where as some of the hemicellulose might be incorporated into the bacterial cell walls, while some sugars may be used for *de novo* amino and nucleic acid synthesis. This detracts from the final amount of butanol produced as compared to a pure industrial technique where all of the available hydrocarbons might be converted to butanol (29).

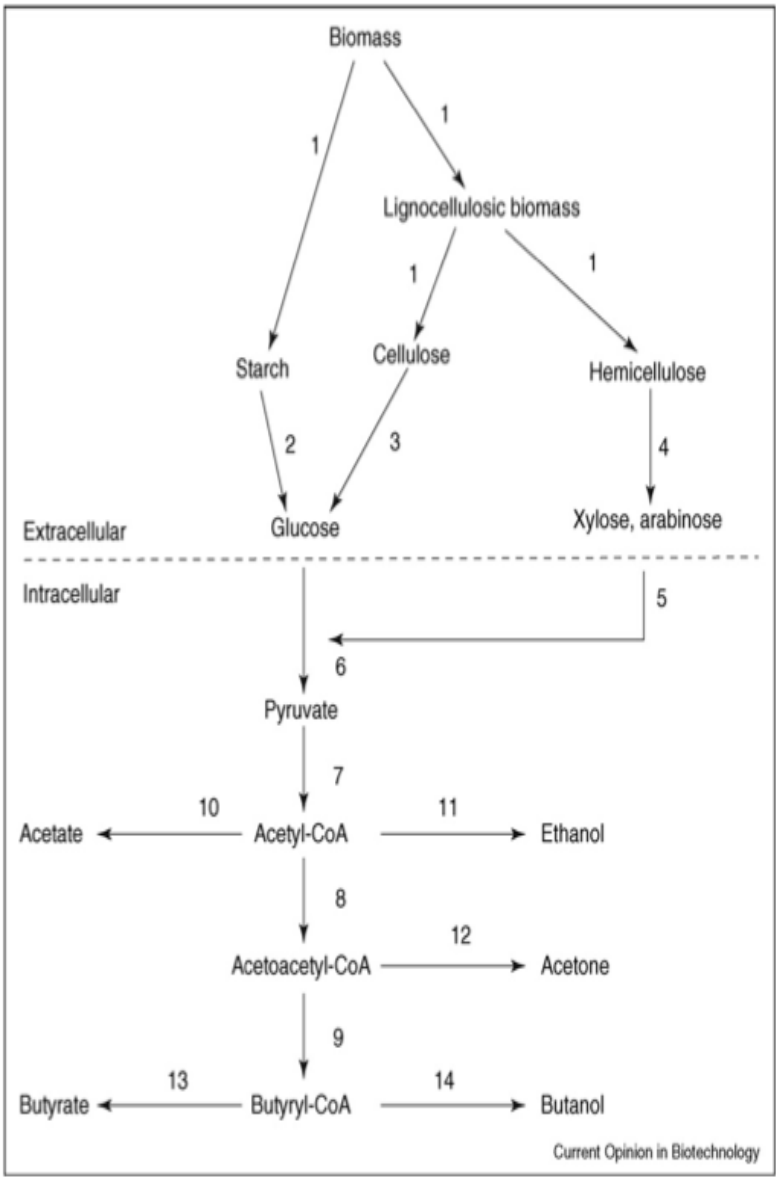
Another problem with the biological synthesis is end product toxicity; the butanol has a negative effect on the bacteria, and at some point the amount of butanol present exceeds the amount of butanol that the Clostridia can tolerate. The bacteria begin to die, and no more butanol will be produced. This problem can be diminished by using strains or species with higher tolerance for the products (such as *C. beijerinckii* in comparison to *C. acetylbutylicum*), but this is still just a temporary measure. For true industrial scale biological production to be effective, one would need to be able to extract the butanol (as well as the other waste

products), without disrupting the bacterial population. Currently, several methods for this are under investigation (29).

1.3.4. Reaction Environment Effects on Clostridia Growth:

End product toxicity is an important factor when dealing with ABE fermentation by Clostridia species. Full understanding of this requires understanding how ABE fermentation fits into the bacterial growth curve. While the bacterial growth curve has four phases (lag, exponential growth, stationary, and cell death), ABE fermentation is broken down into two phases: acidogenic and solventogenic. During the acidogenic phase, which takes place during the lag and exponential growth phases of bacterial growth, the bacteria break sugars down into acetic and butyric acids. These seem to go together as the production of the organic acids is accompanied by the production of ATP. Once the acid concentrations cross a certain threshold, the bacteria shift into solventogenic phase and stationary growth phase. This is when the Clostridia start converting the acetic and butyric acids into acetone, butanol, and ethanol, which are unfortunately toxic. This in turn leads to the death phase of the bacterial growth curve, which is accompanied in *C. beijerinckii* by sporulation. Under controlled pH conditions, it is possible to maintain a steady state culture in acidogenic or solventogenic phase without sporulation (13, 29).

Beyond just the end product toxicity, other chemicals present in the reaction chamber can cause inhibition of butanol production. For example, fiber rich



Simplified metabolism of biomass by solventogenic clostridia. 1, Pretreatment of corn and lignocellulose; 2, starch hydrolysis (α-amylase, β-amylase, pullulanase, glucoamylase, α-glucosidase); 3, cellulose hydrolysis (cellulases, β-glucosidase); 4, hemicellulose hydrolysis (see Figure 2); 5, xylose/arabinose uptake and subsequent breakdown via the transketolase-transaldolase sequence producing fructose 6-phosphate and glyceraldehydes 3-phosphate with subsequent metabolism by the Embden-Meyerhof-Parnas (EMP) pathway; 6, glucose uptake by the phosphotransferase system (PTS) and conversion to pyruvate by the EMP pathway; 7, pyruvate-ferrodoxin oxidoreductase; 8, thiolase; 9, 3-hydroxybutyryl-CoA dehydrogenase, crotonase and butyryl-CoA dehydrogenase; 10, phosphate acetyltransferase and acetate kinase; 11, acetaldehyde dehydrogenase and ethanol dehydrogenase; 12, acetoacetyl-CoA:acetate/butyrate:CoA transferase and acetoacetate decarboxylase; 13, phosphate butyltransferase and butyrate kinase; 14, butyraldehyde dehydrogenase and butanol dehydrogenase.

Figure 3 Biochemical Pathway of Biobutanol Production in *C. beijerinckii*(9)

agricultural residues tend to be hydrolyzed prior to usage as bacterial feedstock in order to break down tough cellulose and hemicellulose. However, this process also causes the production of furfural, organic acids, and phenolic compounds, all of which can inhibit bacterial growth. Further studies have shown that furfural and its modified form hydroxymethyl furfural, in low concentrations, can actually promote cell growth in *C. beijerinckii*, but the other chemicals are all potent inhibitors (11,19). HPLC and spectrometric studies have indicated that *Clostridium sp* are capable of converting the furfural, an aldehyde, into furfuryl alcohol. However, while cell growth is promoted by furfural in low concentrations, total ABE fermentation products may experience a decrease (28).

1.3.5. Understanding Yield Through Proteomics and Genomics:

Comparatively few studies have been done on the proteomics and genomics of *C. beijerinckii*, but there have been many studies to try and elucidate the molecular machinery of butanol production in other organisms, most notably related species *C. acetylbutylicum* but also strains of yeast and genetically engineered *Escherichia coli*. From these studies, we know that there are a variety of types of protein that are involved in solvent tolerance, including efflux pump response proteins, heat shock proteins, and ATPases (10, 24). A semi-comprehensive study was performed on *C. acetobutylicum*, looking at 6.5% of all proteins, or 21% of cytosolic proteins. Only intracellular soluble proteins with a pI

between 4 and 7 were examined, and proteins with low yields may have not been elucidated since the 2D gels were stained with Coomasie. However, several proteins with differential expression between a steady-state chemostat culture in the acidogenic phase and another in the solventogenic phase were identified, including a ATP phosphoribosyltransferase, an electron transfer flavoprotein, aldehyde-alcohol dehydrogenase, and an endoglucanase (13).

Previously in our own laboratory, a study was performed in which the proteome of the acidogenic phase and the solventogenic phase were compared by one-dimensional gel electrophoresis. This data showed a number of bands with significant deviation between the two time points. Some of these bands were cut out and sent to Ohio State University for identification. The four proteins received back were *cspA*, glyceraldehyde-3-phosphate dehydrogenase type I, a pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein, and a cell wall binding repeat-containing protein.

Beyond the differences between the acidogenic and solventogenic phases, it is important to understand the differences in the proteomics and genomics between different strains, particularly the hyper-butanol producing mutants. One microarray study compared the serotype strain *C. beijerinckii* NCIMB 8052 with the hyper-butanol producing mutant BA101 and found that BA101 had increased levels of expression for butyryl-CoA and butanol formation genes during solventogenesis. The same study also found reduced levels of induction for genes related to spore formation, corresponding with a morphological study showing that the hyper-butanol producing strain produced fewer spores (22). This could be indicative of

increased expression for unknown genes that improve the bacterial cell's resistance to butanol.

2. SPECIFIC AIMS

The aim of this project is to analyze the differences in the proteome and genome of *Clostridium beijerinckii* during biobutanol production from a standard growth media supplemented with xylose. Xylose has been chosen for use as the substrate in order to mimic the hemicellulose xylan without having to worry about the presence of possible inhibitory agents such as furfural and other furans. We will be observing the genetic expression of proteins that see significant shifts in expression, primarily *cspA*, glyceraldehyde-3-phosphate dehydrogenase type I, a pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein, and a cell wall binding repeat-containing protein, all of which were identified as proteins of interest in previous work performed in our laboratory. The end goal is to identify genes that may be targets for up-regulation in order to improve the production of biobutanol from *Clostridium beijerinckii*. If time and equipment permits, over the course the experiment, the levels of butanol and other products will be recorded by HPLC and correlated with changes in the proteome observed by 2DGE.

3. MATERIALS AND METHODS

3.1. Chemicals

D-(+)-Xylose was obtained through Sigma-Aldrich. Difco Luria-Bertani (LB) Broth, Miller was obtained through Becton, Dickinson and Company. All Polymerase Chain Reaction (PCR) reagents were obtained through QIAGEN. Agarose and TAE buffer were obtained through Sigma-Aldrich. EZVision Dye was obtained through BioRad. All quantitative PCR (qPCR) reagents were also obtained through BioRad. Water was filtered through a Millipore water filtration system.

3.2. Organism

The *Clostridium beijerinckii* used for this thesis was obtained from the American Type Culture Collection (ATCC) bioresource center and has been designated as *C. beijerinckii* ATCC 35702. The bacteria was received lyophilized in a sealed glass vial. They were seeded into a 50 mL Erlenmeyer flask containing 25 mL of xylose supplemented LB Broth.

3.3. Media

The media was originally prepared via the manufacturers directions; 25 grams of Difco LB Broth were mixed in 1 liter of purified water and autoclaved at

121˚ Celsius for 15 minutes. Once the broth was cooled to room temperature, 40 grams of Xylose were mixed in the broth, and the resulting solution was filtered through a Millipore Stericup Filter Unit. When not in use, the media was stored in a refrigerator at 4˚ Celsius.

3.4. Cell Culturing and Collecting

The *C. beijerinckii* bacteria from the ATCC were grown in a 1 liter screw-top Erlenmeyer flask with sidearm with 250 milliliters of xylose supplemented media. The environment was purged of oxygen by holding under a nitrogen vent for 15 seconds before screwing the cap closed and further sealing with Parafilm. Once the bacteria were in the logarithmic growth phase, as determined by turbidity measurements with a Klett Colorimeter from Scienceware, multiple 1 milliliter aliquots were collected by pipette, mixed with 50% glycerol, and stored under liquid nitrogen until further use.

The aliquots were used to begin experimental cultures. One aliquot would be added to 250 milliliters of xylose supplemented media in a 1 liter screw-top Erlenmeyer flask with sidearm. The environment was purged of oxygen by holding under a nitrogen vent for 15 seconds before screwing the cap closed and sealing with Parafilm. For ten days, the seal was broken every 24 hours in order to collect three 5 milliliter samples; one for protein analysis, one for DNA analysis, and one for HPLC analysis. These samples were collected by pipette, stored in screw-top

centrifuge tubes from Fisherbrand, and frozen at -80˚ Celsius until they were analyzed.

3.5. Protein Analysis

To prepare the samples for protein analysis, the cells were lysed open in an electrophoresis buffer using a Microson ultrasonic cell disruptor under the following procedure: 10 seconds on 4 watts, chill on ice, 10 seconds on 4 watts, chill, 20 seconds on 4 watts, chill, 30 seconds on 4 watts, chill, 10 seconds on 8 watts, chill. While sonicating, care was taken to avoid the production of bubbles, as proteins may be degraded at the liquid-air interface.

For one dimensional gel electrophoresis analysis, samples were run alongside different concentrations of bovine serum albumin (BSA) on Criterion TGX precast SDS-PAGE gels, 26 well, 4-20% polyacrylamide concentration gradient from BioRad. They were run at 40 amps using a PowerPac 3000 power box from BioRad for approximately one hour, until a bromophenol blue loading dye had reached the bottom of the gel. The gel was then stained with Coomasie Blue dye for one hour and destained with an acetone-methanol mixture until excess dye was removed. Afterwards, the gel was scanned with a PharosFX plus molecular imager from BioRad and analyzed with the software ImageJ made freely available by the NIH. Using the BSA concentration gradient, it was possible to determine the protein concentration for each of the daily protein samples.

Once the one dimensional gels were analyzed, select two dimensional gel electrophoresis samples were run. Three samples were run; Day 3 was selected to represent the acidogenic phase, Day 5 was selected to represent the transitional period, and Day 7 was selected to represent the solventogenic phase. Equal concentrations of protein were electrophoresed onto ReadyStrip™ ipg strips, 11 cm, pH 3-10 from BioRad using a Protean IEF cell from BioRad. Once the proteins were aligned, they were equilibrated for running on Criterion TGX precast SDS-PAGE gels, ipg+1 wells, 4-20% polyacrylamide concentration gradient from BioRad. Samples were run at 40 amps for approximately one hour, until a bromophenol blue loading dye had reached the bottom of the gel. The gel was then stained with Coomasie Blue dye for one hour and destained with an acetone-methanol mixture until excess dye was removed. Afterwards, the gel was scanned with a PharosFX plus molecular imager.

3.6. Nucleic Acid Analysis

Based on the two dimensional gel electrophoresis and work by a prior graduate student, four genes were chosen for study. These genes were: *cspA*; glyceraldehyde-3-phosphate dehydrogenase type I, designated G3P; a pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein, designated PF; and a cell wall binding repeat-containing protein, designated CWBR. The sequences for these genes were retrieved from the National Center of Biological Information (NCBI) and entered into the Integrated DNA Technologies (IDT) PrimerQuest

software to create DNA primers for amplification of sections of those genes. The primer sequences were then compared against the *C. beijerinckii* whole genome using NCBIs BLAST software to confirm that each primer bound to only the one location in the bacterial genome. The primers were then ordered from IDT.

Once the primers were received, they were first checked for efficacy with a PCR. DNA was purified using a kit from QIAGEN following the manufacturers directions, then concentration of DNA was determined through the use of a NanoDrop 2000c spectrophotometer from Thermo Scientific. One sample for each of the four genes from day 5 were prepared for PCR using a kit from QIAGEN according to the manufacturers directions. Once the samples were prepared, they were run on a iCycler thermocycler from BioRad according to a program dictated by the QIAGEN kit manufacturer's directions. Afterwards, 10 milliliters of each sample were run on 2% agarose gels with 2 milliliters of EZVision Dye at 100 watts for approximately one hour, until the first dye front from the EZVision reached the bottom of the gel. Samples were then visualized under UV light.

After confirming that the primers worked, qPCR was carried out with reagents from BioRad. Each reaction tube contained 0.25 microliters of reverse transcriptase mix, 2 microliters of a 10 micromolar concentration of each primer, 3.75 microliters of nuclease free water, 10 microliters of Sybr green, and either 2 microliters of RNA or an additional 2 microliters of nuclease free water. Samples were then run for 40 cycles in an iCycler thermocycler with iQ5 multicolor real time PCR deterction system from BioRad. Results were analyzed using iQ5 software from BioRad.

4. RESULTS

4.1. Protein Analysis

The first step in protein analysis was to run one dimensional gels in order to visualize the proteins present and determine the concentrations of protein within each days sample. To do this, two 26 well gels were run. In both gels, the end most lanes and one of the middle lanes were loaded with molecular weight standard in order to determine the relative sizes of protein bands. In the left half of both gels, BSA was loaded in a gradient across five lanes; the first lane held 200 micrograms of BSA, the second 400 micrograms, the third 600 micrograms, the fourth 800 micrograms, and the fifth 1,000 micrograms. The right half of the gels was loaded with five protein samples, days 1 through 5 for the first gel and days 6 through 10 for the second. These gels are shown in Figures 6 and 7 in Appendix B. The program ImageJ was then used to determine the protein concentrations for days 3, 5, and 7, which have been labeled in the figures.

Once the protein concentrations for each of the three samples was calculated, this was used to determine the amount of each sample that was needed to run a two-dimensional gel electrophoresis. Three IPG strips were used, one for each of the three samples that were selected. Once those had been run and equilibrated, three IPG+1 well gels were run, one for each sample. The IPG strips were placed in the long well with the acidic end towards the narrow well, which was loaded with molecular weight standards. These gels are shown in Figures 8 through 10 in

Appendix B. Each spot represents an individual protein or protein modification; a protein that can undergo multiple phosphorylations will appear as multiple spots separated by a small but predictable space. The horizontal streaks in Figure 9 indicate that the IPG strips could have been run for a little longer before equilibrating and placing in the polyacrylamide gel.

By overlaying the two dimensional protein gels over top one another, it would be possible to determine which spots, and therefore which proteins, underwent shifts in either expression or post-translational modification over time and environmental shift. However, this would have required gels with better resolution and was not the focus of our experiment as much as the nucleic acid analysis of the four genes identified in the previous studies.

4.2. Nucleic Acid Analysis

Using the three two-dimensional protein gels and the previous work of

Table 1 qPCR Primers

Gene	Forward Primer	Reverse Primer
G3P	TGCTAAGCCTGAAAACTTACCATGGGG	ACAACTTTCTTAGCACCAGCTGCGAT
CWBR	CCTAGCGGTGGAACAGTGAA	GCAATGCAACAGCAGTACTATC
PF	ACTGGATGGGTAAAGATAGACAATAG	ACCTGACGTACCATCCAGATA
cspA	CGGAGGAAAGAGCGATGATAATA	CCGTCTTGTAACCATCCAGTAG

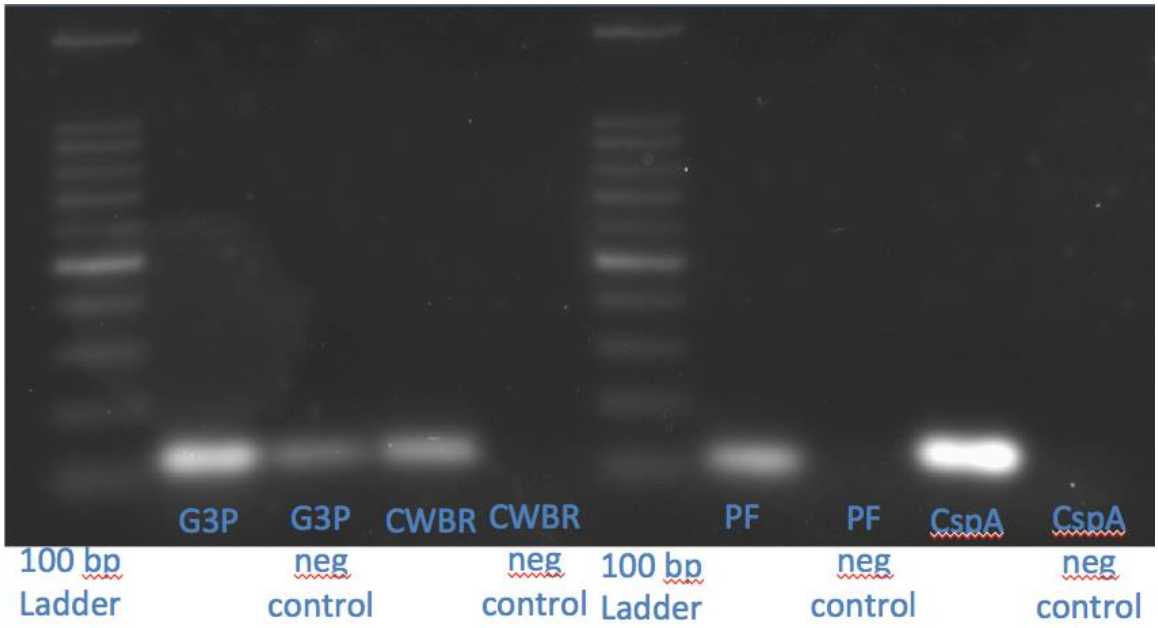


Figure 4 qPCR Primer check. This gel confirms that the primers selected bind to the genome and that the replicated bands are of the expected size given the bioinformatics data.

another student (work not shown) we selected four genes for study: G3P, CWBR, PF, and *cspA*. The sequences for these genes were pulled from the NCBI website (see Appendices D-G) and entered into the IDT PrimerQuest software to develop the primers shown in Table 1. Once the primers were received, they were checked for efficacy through PCR. The samples were incubated in the thermocycler with the primers and then run on an agarose gel, shown in figure 9. The strong, bright bands in the gene lanes indicates that the primers are working to duplicate a section of the genome; their placement between the 100 base pairs and 200 base pairs bands on the 100 base pair ladder indicates that they are of the expected size as was given by IDT. The absence of bands in the negative controls indicates that the primers do not bind with one another to form primer dimers. There is a band in the G3P negative control lane of the same size as the primer constructs, but this can probably be ignored as spill over from one of the other lanes.

Having confirmed that the primers work, the next step was to perform qPCR. Three trials were performed for each gene and day combination. The samples were prepared and run in the thermocycler, generating the graphs below. The important part of each line graph is where the sample lines cross the horizontal green line; this represents the threshold cycle where the concentration of DNA exceeds a certain amount and is indicative of the amount of that gene product was present in the sample before amplification. The initial graph is presented in Appendix G as figure 11. The individual line graphs can then be separated out and grouped together in charts as desired. Figures 12 through 15 are charts representing each of the four genes, while Figures 16 through 23 have been further subdivided by day. In addition to displaying the data as a graph, the threshold cycle is also presented by the thermocycler as a raw number. These numbers can then be transferred to a statistical analysis program in order to determine the average threshold cycle for each gene and day pairing, as well as a determine if there is a statistically significant difference between the two day data points for each gene. This data is presented in numerical form in Table 2 and graphically in Figure 5, where the error bars represent standard deviation. Table 2 also includes the statistical analysis of the data, a T-Test performed comparing the two days for each of the four genes.

Table 2 qPCR Data. T 1, 2, and 3 represent the three trials performed for each gene/day pairing. The numbers in those columns represent the threshold cycle at which the fluorescence, and therefore concentration of DNA, exceeded the threshold value. The Average values and the Standard Deviation for those values is then shown to the right. Finally, at the far right of the table, is the result of a Student's T-Test for significance. The value shown in this column is the percentage chance that there is no statistical difference between the two sets of numbers (gene day 3 vs gene day 7). Values of less than 0.1 are considered to be indicative of a statistically significant difference between the two sets of values.

	Day 3					Day 7					
	T 1	T 2	T 3	Avg	St Dev	T 1	T 2	T 3	Avg	St Dev	T-Test
G3P	20.2	20.8	22.0	21.0	0.9	21.9	23.5	22.8	22.7	0.8	0.064
CWBR	24.9	23.2	24.3	24.1	0.8	35.4	22.0	22.3	26.6	7.6	0.637
PF	24.2	23.7	23.9	23.9	0.3	21.9	21.7	21.8	21.8	0.1	0.002
cspA	17.8	18.0	19.1	18.3	0.7	21.2	21.0	28.5	23.6	4.3	0.164

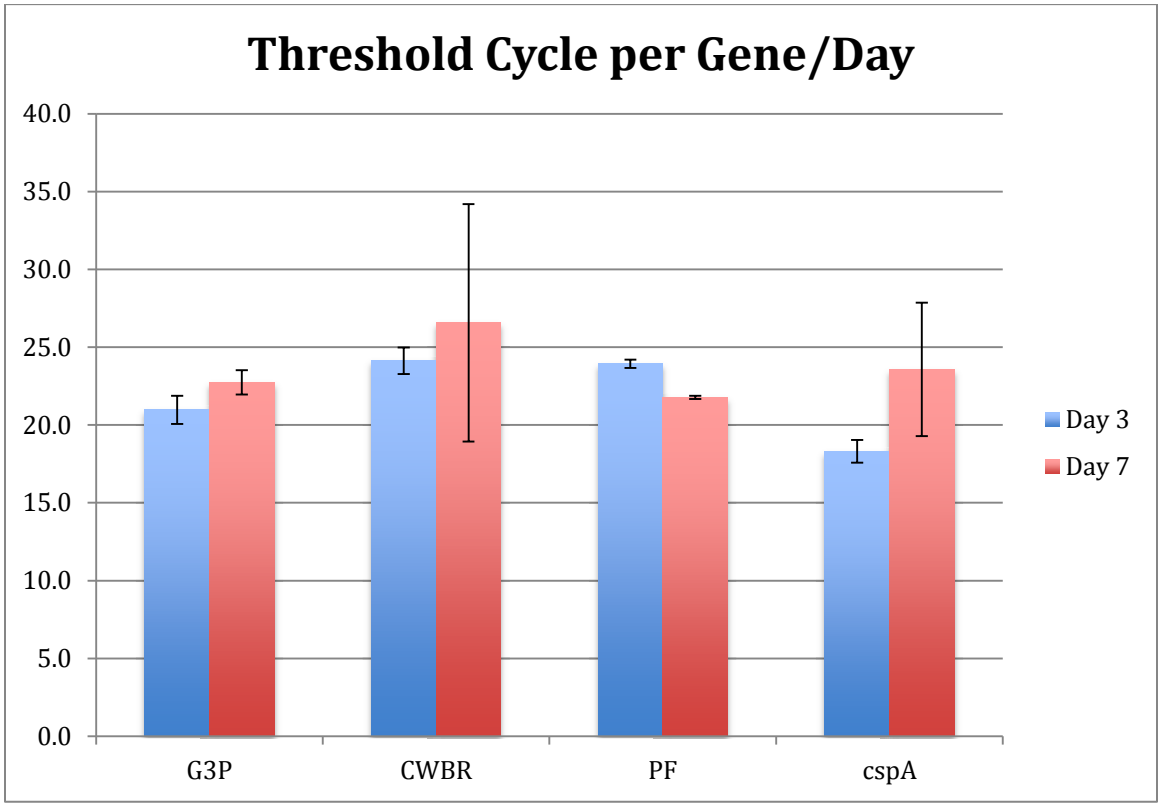


Figure 5 qPCR Average Results. The bar height represents the average threshold cycle that fluorescence, and therefore DNA concentration, exceeded the threshold value. Error bars represent the standard deviation.

5. DISCUSSION

According to the T-Test results, the differences in the expression levels of the genes at different time points was only significant for G3P (at an α level of 0.1) and PF (at an α level of 0.01); the differences between the time points were not significant for CWBR or *cspA*. However, each of these two genes had a significant outlier. If these outliers are removed, then the differences in time points also become significant at α levels of 0.05 for these two genes. Further trials should be performed in order to minimize the effects of these outliers. G3P agreed with the previous proteomic work from our lab, whereas the PF did not agree.

With regards to the two genes lacking outliers and for which the statistical analysis was significant, G3P behaved as expected, whereas PF did not. With regards to the previous proteomic work done in our lab, G3P agreed with the previous work, whereas the PF did not agree. Both of these genes code for proteins involved in metabolic reactions and were expected to express at higher levels on day 3, when sugar breakdown was still happening, and be expressed at lower levels on day 7, when the toxic solvents of acetone, butanol, and ethanol were being formed and sporulation was starting to occur. G3P had an earlier threshold cycle on day 3 than day 7, indicating that it was expressed at higher levels on day 3 than day 7, so it behaved as expected. However, PF had a later threshold cycle on day 3 than day 7, indicating that it was expressed at a higher level later in the time frame. However, it is important to keep in mind that these samples were standardized by total RNA rather than by bacterial concentration. It is possible that PF expression

levels could have dropped, but not by as much as the expression of other genes. This would cause there to be more PF gene product to be present relative to total gene product, and thus a shift towards an earlier threshold cycle. It is also important to note that gene expression does not always correlate with gene product concentration.

Another possible source of error is the frequent exposure of the culture to oxygen during the bacterial cell growth portion of the experiment. As *C. beijerinckii* is a strict anaerobe, it would become stressed and alter its gene and protein expression or die when exposed to normal atmosphere. While care was taken to minimize the exposure to oxygen during growth, the culture had to be exposed to oxygen when samples were collected. It is also possible that not all oxygen was purged from the system when resealing the reaction chamber, or that the chamber was not completely sealed. If possible, further experiments should be done in a bioreactor, which would allow for the collection of samples without exposing the culture environment to oxygen.

In addition to the future work to be carried out on the genes studied in this project, there are several more genes that remain. Ideally, each of the roughly 5,200 genes in *C. beijerinckii*, but the most important genes to study are those directly related to butanol production and sporulation, as well as any genes that can be linked to tolerance of increased butanol levels. Of particular interest would be butyraldehyde dehydrogenase and butanol dehydrogenase, the final steps in the production of butanol (see Figure 3 on page 15).

Beyond studying the genes for future manipulation of *C. beijerinckii* in order to increase butanol production, there are mechanical and physical approaches that are mainly focused on removing the butanol from the environment as it is produced, leaving the *C. beijerinckii* in its productive environment to continue producing butanol. While the equipment to do this is not available at Youngstown State University, it would be interesting to monitor the proteome and genome under these conditions to see if they remain stable or if there are still subtle shifts in gene expression.

In conclusion, there appears to be some correlation between gene expression levels and stage in butanol production for some genes. With additional experimentation, both on the genes analyzed in this study, particularly for CWBR and *cspA*, and on other genes of importance, such as butyraldehyde dehydrogenase and butanol dehydrogenase, further or improved correlation between gene expression and stage in butanol production. If possible, the bacterial growth portion of these experiments should be performed in a bioreactor for improved yield and reduced chance of error. It might also be useful to perform qPCR analysis on additional time points to obtain a more complete picture in the changes in gene expression over time.

Appendix A – Energy Conversion Factors

General conversion factors for energy

To:	TJ	Gcal	Mtoe	MBtu	GWh
From:	multiply by:				
TJ	1	238.8	2.388×10^{-5}	947.8	0.2778
Gcal	4.1868×10^{-3}	1	10^{-7}	3.968	1.163×10^{-3}
Mtoe	4.1868×10^{4}	10^{7}	1	3.968×10^{7}	11 630
MBtu	1.0551×10^{-3}	0.252	2.52×10^{-8}	1	2.931×10^{-4}
GWh	3.6	860	8.6×10^{-5}	3 412	1

Appendix B – Protein Gels

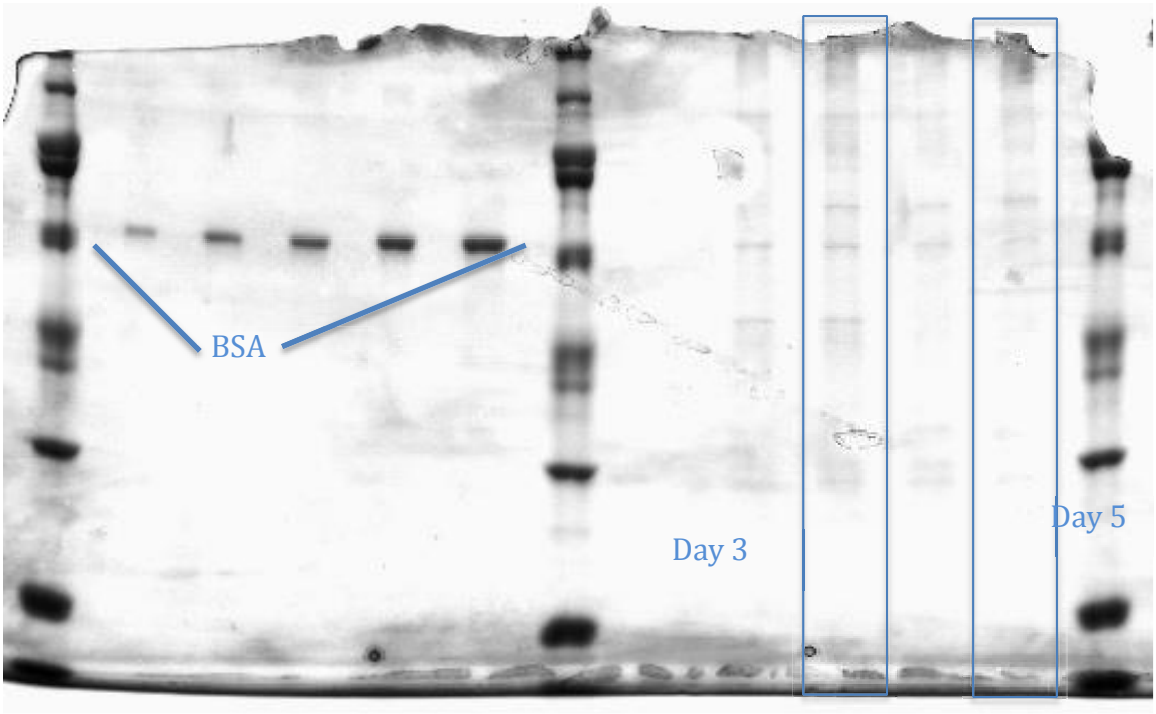


Figure 6 One Dimensional Gel Electrophoresis Days 1-5. BSA is in increasing increments of 200 micrograms and was used as a concentration standard. Molecular weight standards were loaded on each side and down the middle, between the BSA concentration standard and the daily samples. The days that were used in further experimentation have been boxed.

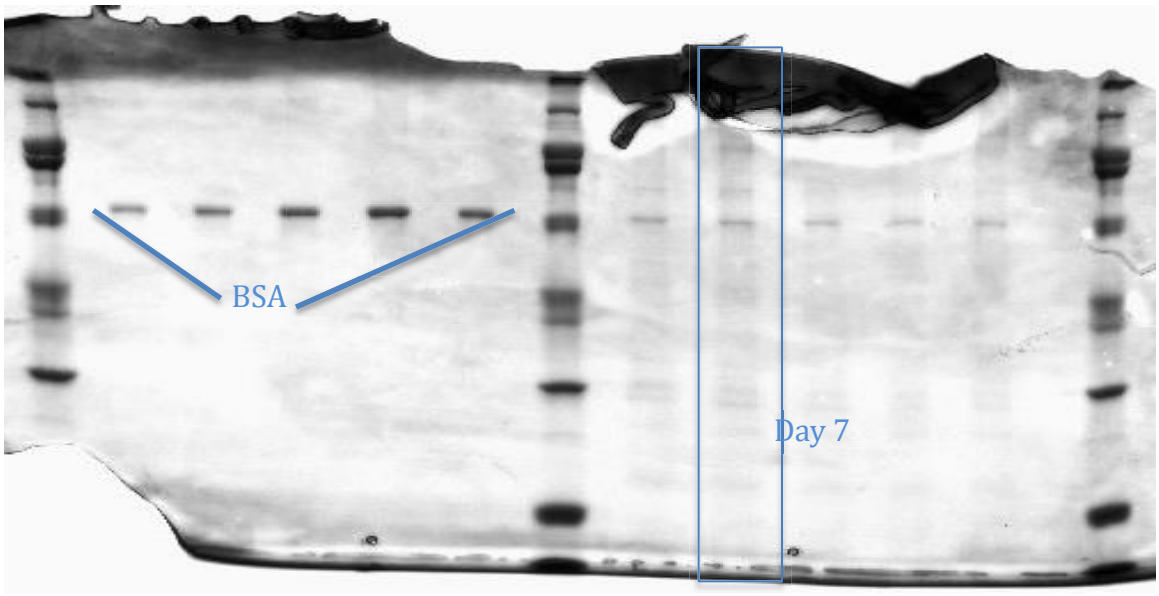


Figure 7 One Dimensional Gel Electrophoresis Days 6-10. BSA is in increasing increments of 200 micrograms and was used as a concentration standard. Molecular weight standards were loaded on each side and down the middle, between the BSA concentration standard and the daily samples. The days that were used in further experimentation have been boxed.

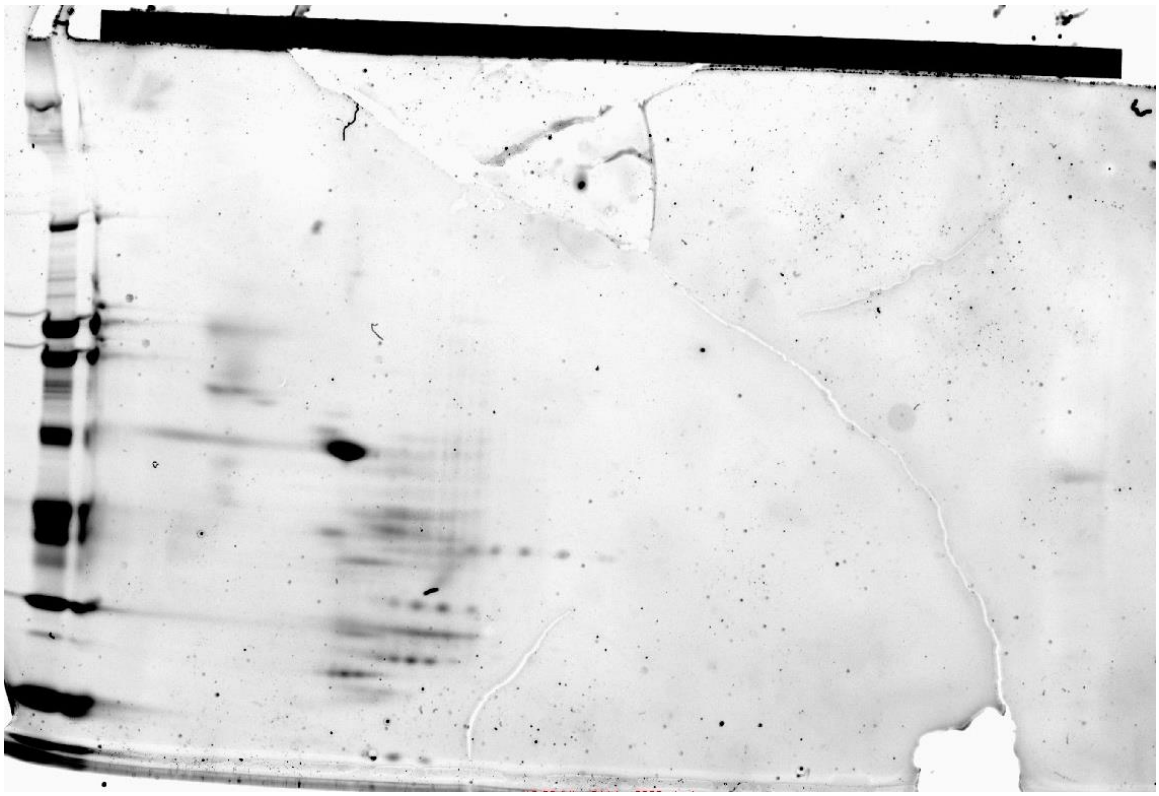


Figure 8 Two Dimensional Gel Electrophoresis Day 3. Down the left hand side is a molecular weight standard ladder. Across the rest of the gel, each dot represents a different protein or protein modification. Spots further to the left have a lower pI and are more acidic.

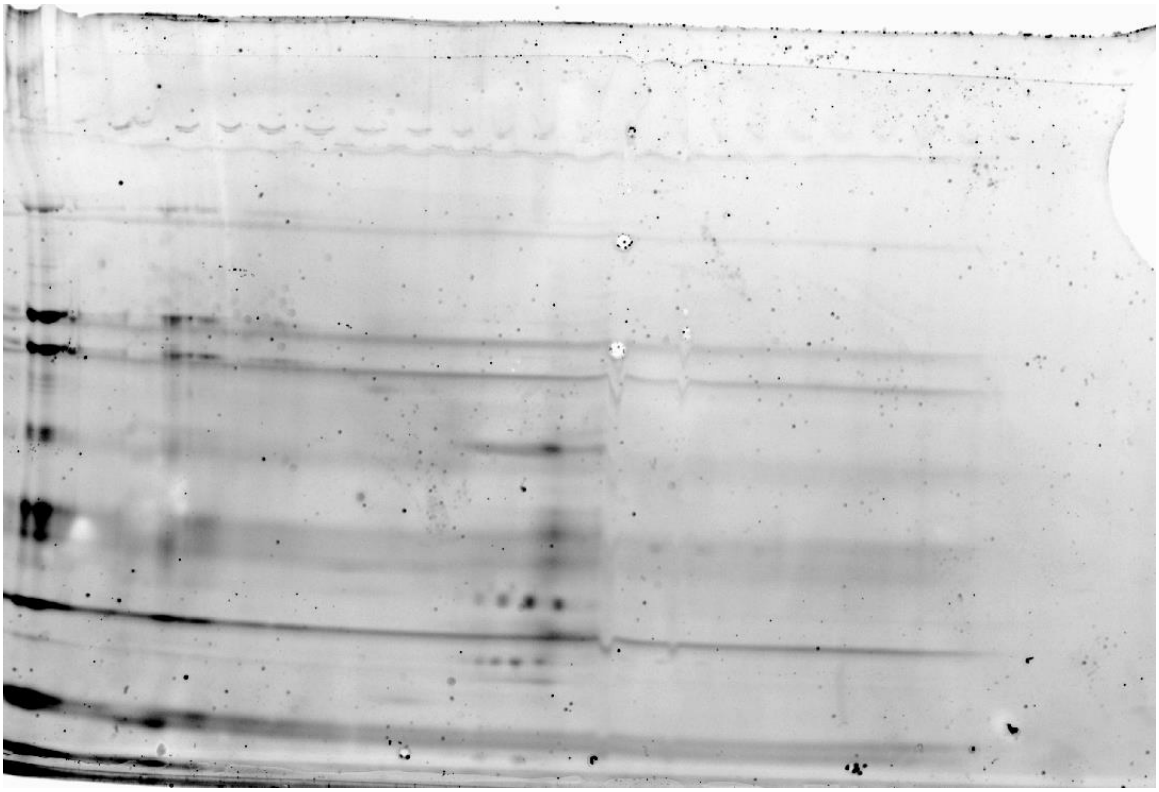


Figure 9 Two Dimensional Gel Electrophoresis Day 5. Down the left hand side is a molecular weight standard ladder. Across the rest of the gel, each dot represents a different protein or protein modification. Spots further to the left have a lower pI and are more acidic.

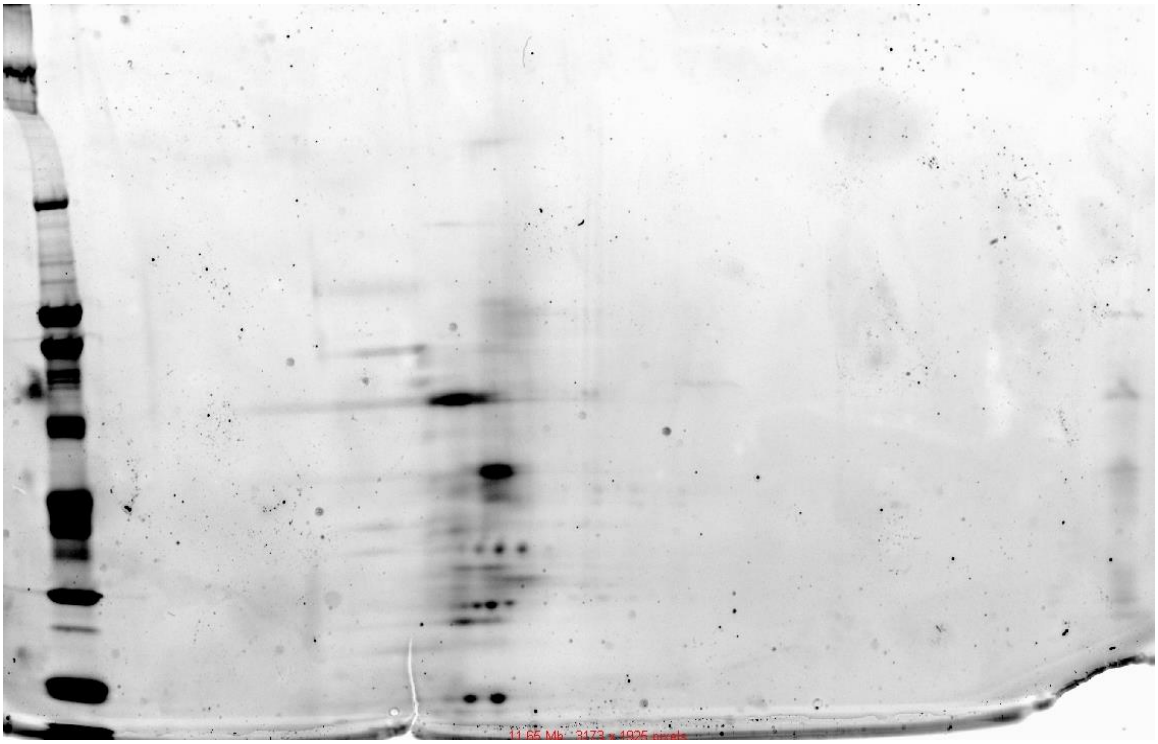


Figure 10 Two Dimensional Gel Electrophoresis Day 7. Down the left hand side is a molecular weight standard ladder. Across the rest of the gel, each dot represents a different protein or protein modification. Spots further to the left have a lower pI and are more acidic.

Appendix C – Detailed qPCR Methodology

Cycle 1:	50° C for 10 minutes	
Cycle 2:	95° C for 5 minutes	
Cycle 3:	Step 1:	95° C for 10 seconds
	Step 2:	60° C for 30 seconds
	Repeat 45 times	
Cycle 4:	95° C for 1 minute	
Cycle 5:	55° C for 1 minute	
Cycle 6:	55° C for 10 seconds	
	Repeat 81 times	

Appendix D – G3P Gene Sequence

ATGAATTTCAATAAGAAAACAATTGAAGATATTGAAGTGAGTGGAAAAAAAGTACTAGT
AAGATGTGACTTTAACGTACCACTTAAAGATGGAGTTATAACTGATGAAAATAGATTGG
TAGGAGCTCTTCCAACTATTAAATATTTAGTAGAAAAAGGTGCAAAGGTTATTCTTTGT
TCACATTTAGGAAAAGATGCTTCAAAATCATTAGCACCAGTAGCTACAAGATTAAGCGA
AATGCTAGGGAAAGAAGTTGTTTTTGCTAGGGATGAAGAAGTTGTTGGCGAAAACGCTA
AAAAAGCAGTTTCAGAAATGAAAGATGGAGACATCGTTTTATTAGAAAACACAAGATGC
AGAAAAGAAGAAACTAAGAACATTCCAGAATTCTCTAAAGAATTAGCTTCATTAGCTGA
TGTATTTGTTAATGACGCATTTGGAACAGCTCATAGAGCTCACTGTTCAACTGTTGGTGT
AACTGATTACTTAGATACAGCTGTATGTGGATACTTAATCCAAAAAGAATTAAAATTCT
TAGGAAATGCAGTTGAAAGTCCAGTAAGACCATTCGTTGCTATCTTAGGTGGAGCTAAA
GTTTCTGATAAAATCGCTGTTATAAACAATCTTTTAGATAAAGTTAACACAATTATTAT
CGGTGGAGGAATGGCTTATACATTCTTAAAAGCTCAAGGATATGAAATCGGAACTTCAT
TAGTTGAAGAAGATAGACTTGAATATGCTAAAGAAATGGTTGCTAAGGCTGCTGAAAAA

GGAGTTAAATTCTTATTACCAGTAGATCACAGAGTTGCTGCAGAATTTAAAGATGTAGA
AGCTACAATTACTGAAGATCAAAATATTCCAGTAGGAAACATGGGATTGGATATTGGAC
CAAAGACAGAAACTTTATATGCTGATGCAATTAAAGATGCTAAGACTGTAATTTGGAAT
GGACCAATGGGTGTATTTGAATTCGAAAACTATAATAAAGGAACAATTGCAGTAGCTAA
GGCTATGGCTGATGCAGATGCTACTACTATAATTGGTGGTGGAGATTCAGCTGCTGCTGT
TAATATTTTAGGATTTGGGGATAAGATGACTCACATTTCAACAGGTGGTGGAGCATCAT
TAGAATTCTTAGAAGGTAAAGTATTACCAGGAATCGCTGCATTAAACGACTAA

Appendix E – PF Gene Sequence

ATGAGAAAAATGAAAACTATGGACGGTAATACTGCAGCAGCTCATATATCTTATGCATT
TACAGAAGTTTCTGCAATATTCCCAATTACACCATCATCACCAATGGCAGAACATGTAGA
TGAATGGGTAGCACAAGGAAGAAAGAACATATTTGGACAACCTGTAAAAGTTATGGAAA
TGCAATCAGAAGCTGGAGCAGCTGGTGCAGTTCACGGATCTTTACAAGCTGGAGCTTTAA
CAACAACTTATACAGCTTCACAAGGTTTATTATTAATGATACCAAACATGTACAAAATT
TCTGGTGAAATGTTACCAGGAGTATTCCACGTATCAGCTAGAGCATTAGCTACTTCTGCT
TTAAACATATTTGGAGATCACCAAGATGTTATGGCAGCAAGACAAACTGGTTTTGCAAT
GCTTGCAGAAGGATCAGTTCAAGAAGTTATGGATTTAGCAGCAGTAGCACATTTAACTG
CTATTAAAACAAGAATTCCATTCTTAAACTTCTTTGATGGATTCAGAACTTCTCATGAAG
TACAAAAAATTGAAGTTTTAGAATATGATGAATTAGCAAGTTTACTTGATTGGGATTCA
GTTAAAGCTTTCAGAGAAAGAGCATTAAACCCAAATCATCCTGTAACTAGAGGAACTGC
TCAAAATGCAGATATCTATTTCCAAGAAAGAGAATCTGTTAATAAATTCTATAATGAAT
TACCAGAAACAGTTGAAAATTACATGGCTGAAATCACTAAGTTAACTGGTAGAGAATAT
CACTGTTTCGATTACTATGGAGCACCAGATGCTGATAGAGTAATTATAGCTATGGGTTC
TGCTACAGACGTTTGTGAAGAAACTATAGATTACTTAAATGCTAATGGACAAAAAGTTG
GTGTTATTAAAGTAAGATTATTCAGACCATTCTCAAATGAAAGATTATTAGCTGCTATT
CCAAAGACAGTTAAGAAAATTGCTGTTTTAGATAGAACTAAGGAACCAGGATCAACTGG
AGAACCATTATACTTAGATGTAAGAAATGCATTCTATGGACAAGCTAATGCACCACTTA
TTGTTGGTGGAAGATTTGGTTTAGGTTCAAAAGATCCAAATCCAGGACATATTGCTGCA
GTTTATGCTAACTTAGCACAAGATGCTCCTAAGAACGGATTCACAATCGGAATCGTTGAT
GACGTTACAAATACTTCATTAGAAGTAACTGAAGATATAGATGCTACTCCAGAGGGAAC
TACATCTTGTAAGTTCTGGGGATTAGGATCAGATGGTACTGTTGGAGCAAACAAGAGTG
CAATCAAGATCATTGGAGATAATACAGACATGTATGCTCAAGCGTACTTCTTCTATGAT
TCAAAGAAATCAGGTGGAATTACAGTATCTCACTTAAGATTTGGTAAGAAAGCAATTAA
GTCACCATACTTAATAAACAAAGCAGACTTTGTTTCATGTTCTAACCAATCATACATCCA
CAAATACAATGTACTTGAAGGTTTAAAACCAGGTTCTACTTTCTTATTAAATACTATCT
GGTCTCCAGAAGATTTAGAAGAAAAATTACCAGCTTCATATAAGAGATTCTTAGCAAAC
AACAACATTAAGTTCTACACTATCAATGCTGTAGGTATTGCTCAAGAAATCGGTCTTGG
TGGAAGAATCAACATGATAATGCAATCAGCTTTCTTCAAGTTAGCTAACATTATTCCAG
TTGAAGATGCTATTAAGCACTTAAAAGATTCAGTTGTAACTTCTTACGGTAAGAAGGGT
GAAAAAGTTGTTAACATGAACAACGCTGCTATCGATAAAGGTGTTGAATCAATTGTTGC

AATAGACATTCCAGAAGCTTGGAAGACAGTTGCAGATGAAGCTCCAGTAGAAATTAAAC
ATGCTACTAAGTTTGTTAAAGATATAGTTATTCCAATGAACAGACAAGAAGGAGATCAA
CTTCCAGTTTCAGCATTTGCAGGTATGGAAGATGGTACTTTTGAAAATGGTACTGCTGCT
TTCGAAAAGAGAGGAATTGCAGTAAATGTTCCTGAATGGGATAAAGATAAATGTATTCA
ATGTAACCAATGTTCAATGGTATGTCCACATGCTTCTATAAGACCATTCTTATTAACTGA
AGCAGAAAAGAATGCTGCGCCAAGTGCAAATAAGGCTGTAGCTGCTAAAGGATTAAAGA
CAGAAGAACCATTATTCTATACAATGGGTGTAACACCACTTGACTGTTCAGGTTGTGGA
AATTGTGCTCAAGTATGTCCAGCACCAGGAAAAGCATTAGTTATGAAACCACAAGAATC
TCAACATGATCAAATAGAAGCTTGGGATTACTTAACTCATGATATATCAGTTAAGAAGA
ACCCAATGAACAAGAAGACAGTTAAAGGTAGCCAATTTGAGCAACCATTGCTTGAGTTC
TCAGGAGCTTGTGCAGGTTGTGGAGAAACTCCATATGTTAAAGCTATAACTCAATTAGT
TGGTGATAGAATGATGGTTGCTAATGCAACTGGATGTACATCAATATGGGGAGGATCAG
CTCCTTCAACTCCATACACTAAGAATAAAAATGGACATGGTCCAGCTTGGGCTAACTCAT
TATTCGAAGATAATGCTGAATATGGATTAGGTATGTTCTTAGGAGTTAAGGCTATAAGA
GAAAGAATTGCAGAAAAAGCTGAAGCTGCTATAGCTGCAAATGATCCAGCAAAAGCTGA
ATTACAAGAATGGTTAGATAACGTAAATGAAGGTGAAGGAACTAGAGAAAGAGCTGATA
AATTAACAGCTGCTTTAGAAGCATCTAACACTGAATTAGCTAAAGAAATATTAGCTGAA
AAAGACTACTTCGTTAAGAGATCACAATGGATCTTCGGAGGAGACGGATGGGCTTACGA
TATCGGATACGGTGGAGTTGACCATGTATTAGCTTCAGGAGAAGATGTGAATATATTAG
TAGTAGATACAGAAGTTTACTCAAACACTGGTGGTCAATCATCTAAAGCTACTCCAACTG
CTGCAATCGCTAAGTTTGCTGCAAGTGGTAAGAAGACTAAGAAGAAAGATCTTGGAATG
ATGGCTATGAGTTATGGTTATGTATATGTAGCTCAAATTAACATGGGTGCTGATAAGAA
CCAAGTTATGAAAGCTATTGCAGAAGCTGAAGCTTATAAAGGACCATCATTAATAATAG
CTTATGCACCATGTATAAATCATGGATTAAGAATTGGTATGGGTAACAGCCAAGAAGAA
GCTAAGAGAGCTACTGCATGTGGTTACTGGCAAATGTACAGATTCAACCCAGAATTAAA
AGATGCTGGAAAGAATCCATTCTCATTAGATTCAAAAGAACCAACTGCAGACTTCAAGG
AATTCTTAATGGGAGAAGTTAGATACTCTTCATTAGCTAAGGCATTCCCAGAACAAGCT
GATGCTTTATTTGAAAAGACTAAGAAAGATGCTATGGAAAGATTAGAAGGATACAAGAA
ATTAGCTAATCAACAATAG

Appendix F – CWBR Gene Sequence

ATGATAAGAGGCATGGGAAAGGTAACATCACTATTAGTAGCAGCAGCAACAGTTGCTTC
ATTAGTACCTTTTAGTGGCGCTAATGCTGCTGAGGTGAAAAGAATTAGTTCTGATGATG
GAACTATCTATAATGCAATAGCATATAAAGACGGTAAAGCTTATATCGATGGAGAAATT
AATGATGACGAAGAAGCTTACTATTTATCCAACGGAAAATTCAATAAATTAGACGATGT
TGATTCAGGAGATAGCGCAGCTCTATTTGGAGAAAAATATTTAGATATATCAGATGGAG
ACTATACTGTTGACTTAGATAAAGGTAGCGTAACTGATGATGATATTAAAGGTGATACT
GAAGATGATGCCGCTGCTGCTTTAAGAAAAAAAAATCAAAGATGATACAGATGATAGGTA
TAATGAAACTGAAGCAAATACTATTAAAGATTCAAATCATGGTGACCTATTCGATTTAA
TTCCAGGAGCAAAGTATAATAAGGTTTGGTATTATACACAATATAAAGCCGCTCAAAAA
TCTATAGATAAAAATGTCAATGGATTAAATGGATTGGATTCAGCCCATCAAATATTTAA
TGTATTTACTGATGATAAAGGCAATTATATTGATGCGGACTATAACTTAGGTAAAGTTA
AGGTTACAACAACAGCTTCAAGTGCTTCTGGTACTACATTAACAAAGACAGATACAATA
GAGAACACAAACGACGCTTACGATGCAGCTGATGGAATTATCAATGGAACTAGCGTTAG

CGGCTCTGATAAATTAAGTGCAAGTGTTGTTCAAGATAGAGTATTAACCCAAGATAAAG
ATTATATATATAGACTTGCAACTGTAAAAATAACTATTACCACTGGTGCAGCTGCTACA
ATTAGTGAAATAAATGGTGTTAAAGTTGATCCAAATAATAGCAACGATATCTTCAAAGT
TGAAAATAACGGACAAGTTGTATCTTTCAAAGCTATTCAAAAAATATCAAAAACTCAAG
CTTCTGGTGATATTGATGACGCTAAATATGCTAAAACTGTAACTACTTATGCACTTTCTG
ATAAAGATGGTAAAAAGTTAGATGCTGAAGATTTATTTATAAATACTTCAGGCAACGTA
GTCACTACAACTAACTATACTGTAGGTTCTGGAAAGCTTATAGCTTACAATTCTGAAAT
TAACAATAACGACAAAGTTACTGTTAGAGCTTATACATTAAAATCAAGTAGTGGATTCT
ATTATGCTGATGAGGAAGATCAAAGCAAAGAAGATTGTGAAAACAGCAAAAACCAAGGT
GCTGCTGTTCAAACAGATGTTGACGGAAATCTATGGAGATTAGATGGTGGATATATCTA
CAAATTTGATAATACTGATGACTGGGACAAAGTATACAGAGTAGACGGATCATTTGATG
AATTCTCAGTTTATGATAAAGATAATATCGTTGCTTGGAGCGAAGATGATGATGTTTAT
TCATTAATAGGTGGAAAACAATCTAATACTGACCCTGATGATACTCCAGTAGTAAAAAC
AGGATGGGTACAAGCTTCAGATGGAACTTGGTCTTATATTAAAGCTGACGGAAATAAGG
CGACAGGTTGGGTTCAAGATGGATCAACTTGGTACTATTTTAAAGCTGATGGATCTATG
GCTACTGGTTGGGTTCAAGATGGATCAACTTGGTATTACTTCCAATCATGGGGTGGAAT
GCAAACTGGTTGGCTAAATAATAACGGTACTTGGTATTACTTACAATCATGGGGCGGAA
TGAAGACTGGTTGGCTAAATGATAACGGCAATTGGTACTACTTACAATCATGGGGCGGA
ATGCAAACTGGTTGGTTTAATGATAATGGTACTTGGTACTACCTATACTCAAATGGTGT
TATGGCCGCTAACACAGTAGTTAATGGATACAACTTATCTGCAAGCGGTGCTTGGGTAT
AA

Appendix G – *cspA* Gene Sequence

AATATTAACTAAAAAAATATATTTATATATTTTTTAAAAAAAGTACCCAAAGTGAATTT
TGACATTGTATTATTTAGTGATATTACAGTAAATACTAGGAAGAACAGTAAATTATGAA
AGTAATAGCATTTTATGTACAAGCATTATGTAATGTGGTATCATGTTATAGATTAATGT
AAATGTGTAATAAATTACATGTGATTTATTTGATTACTAAGAAATTTATACAAAAATAT
ACATTAGTAGTTATTATATAACTTTACATAGTTGTACAATAATGCTATAATTTATCTGA
ATTTTAAAATTTAGGAGGAAAAGAAATTATGTTTAAAAGAGCAAACAAAATTACATCTT
TATTAGTAGCTGCTGCTTCAGTTATGGCTTTAGTGCCTGCTTATGCTGCAGACGTAAAGA
AAGTTGATTCAGAAGATGGTACTGTATACAATGCAGTAGCATACAAAGATGGTAAATAC
TTTGTTGATGGAGAAATCAACGATGATGAAGAAGCTTACTATGTAGCTGACGGAAAATT
CAACAAATTAGAAGATGTTGATTCAGGAGATGAAGCAGTTCTATTTGGAGAAAAATACT
TAGATGTATCAGACGGAGACTACACTGTTGATTTAGATAAAGGTAGTGTAACTGATGAT
GACGTTAAAGGTGATACAGCAGACGATGCTGCAGCAGCTTTAAGAAAGAAAATTAAAGA
TGATACAGATGATAGATATTTAGAAAGTGAAGCAGAAGGCGTTAAAAATGAAGATGATC
TTGCTATAATTGGAGGAGCAAAGTATGACAAGCCTTGGTATGCTACAACATATACTGCT
TCTGCAAAAGCTATAGGTGATGTTAATGGATTAACAGCAACAAATAATAAGTTCAATGT
TTATACTGATACAAATGGTGCTTACATTGATGCAGATTATAACTTAGGAAAAGTAAAAG
TTACAACAACTGCTGACGGTGCTACAAAAGAAGTTACTGTAGAAAACACAAACGATACT
TATGATGCAGCTGGTTCTAGCGATTCAGGAAAAGATAAAGTAAGTGCAAGTGTTAAACA
AACTGCAGTTTTAACACAAGATAAGGATAATATCTACAGACTTGTAGAAGTAACTGTTA
AAACAGGTCATGGTGTTATAACAGAAATTAATGGCGTAAAAATAGCTGACATTGGTGCT
AATACAGTATTTGGTGGTACAAACACTGAAGTAACATTCCCAGCTATCCAAAAGATATC

AAAAGCACAAGCTTCTGATGATGTTGATGGAGCTAAATATGCTAAAACAGTAACTACTT
ATGCTTTATCTGATGATTCAGGAAATAAATTAGATGAACAAAGTTTATTTGTAAAAACT
GATGGTACTGCTGTTACTACAACTAAGTATACTGTAGTCAATGGAAAGCTTATAGCTTA
CAATACTGATATTAATGACAATAAAAAAGTAACTGTTGATGCATATACATTAAAGACAA
AAGGTGGATACTACTACGCTGATGAAGAAGATAAGAGCGAAGAAGATTGCGAAGTAAGT
GCTCAAGATAAAACAACAGCAGCTGTTCAAACAGACGTTGATGGAAATCTTTGGAGATT
AGATGGTGGATATATCTACAAATTCGATAACACTGATGATTGGGATAAAGTTTACAAAG
TAGACGGATCTTTCGATGAATTATCAGTTTACGATAAAGACAACATGGTTGCTTGGAGC
GAAGATGATGATGTATATTCATTAATCGGAGGAAAGAGCGATGATAATAAGGGTGATGA
CCAAGGAACAACTCCTGTAGTTAAAGCTGGTTGGGCTCAAACTTCAGCAGGATGGACTTA
TGTTAAAGCTGATGGAACTAAAGCTACTGGATGGTTACAAGACGGTGGTGCTTGGTACT
ACTTAAAAGCTGATGGTACAATGGCTACAGGTTGGATTCAAGATGGAGCAACTTGGTAC
TACTTAAACGGATCAGGTGCTATGCAAACTGGTTGGTTAAATGATAACGGAACTTTCTA
CTACTTAAATGGATCAGGTGCTATGTTATCTAACACAACAACTCCTGATGGATACTATGT
AGGAGCTAATGGAGCTTGGGTTAAATAGTTTTAATCTGAACTTATTAGATAAATAAAAA
AGGACAAGGCTTAGCCTTGTTCTTTTTTTATCAATAATATTATTATATAATGAATTTAC
ATTTGAAACTCTAATGGGCTAAGTGCTCAATTAAGATTAGAAAGTAGAAAATCAATTGT
TATATATTAGATAGGAAAAGATTTTATGTATATGAATGCTTTGAAGAATATAAGAAAAA
ATAGAGTTGATCCTCCTATGTAGAGGTTTGATTATTTAATTTACTTATATTTAGTATAT
AATAGTCTATATGAGTTTAAA

Appendix H – qPCR curves

For each of the following graphs, the Y-axis is a measurement of the fluorescence, while the X-axis shows the number of replication cycles performed by the thermocycler. The important data is where the curves cross the horizontal green line at 384.66 RFU.

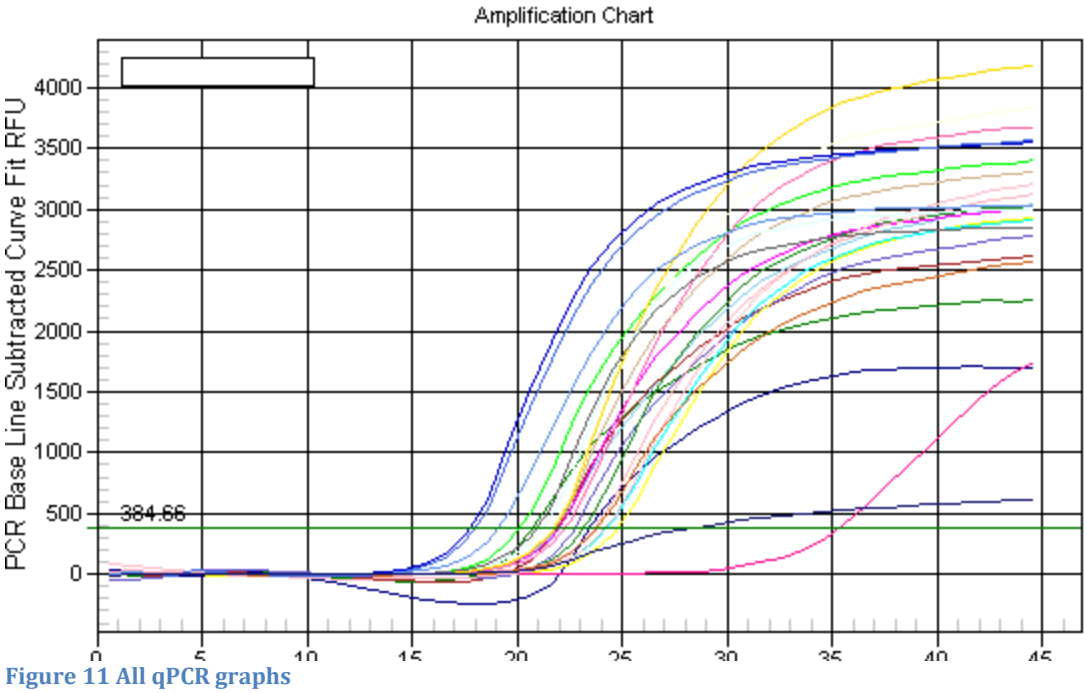


Figure 11 All qPCR graphs

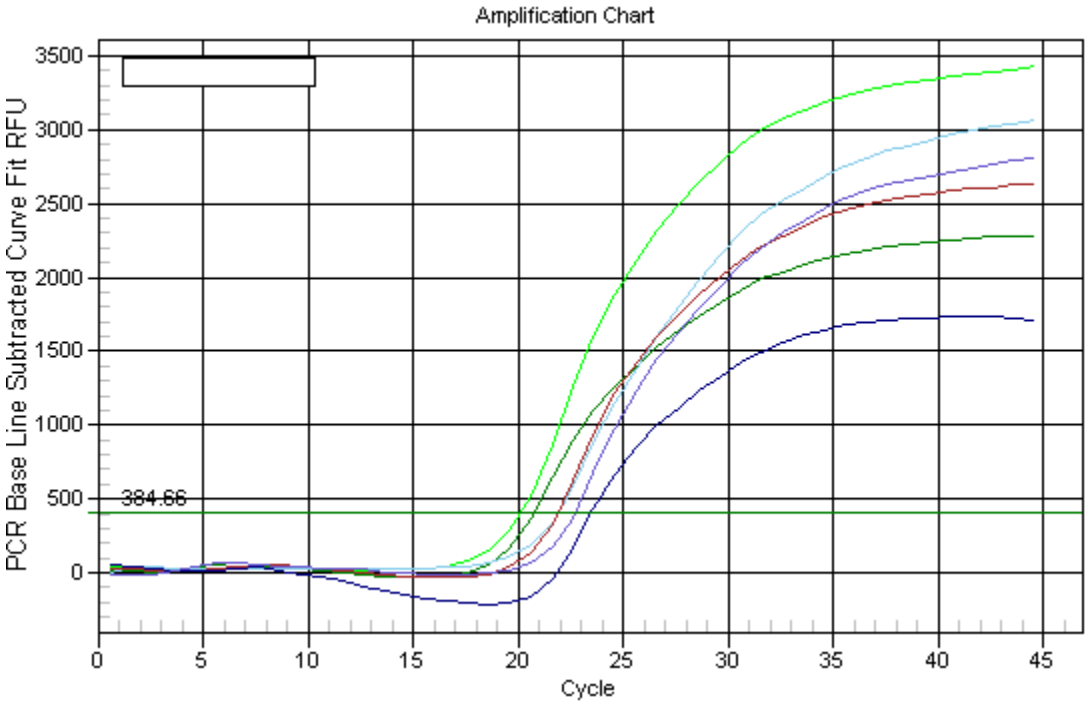


Figure 12 All G3P qPCR graphs

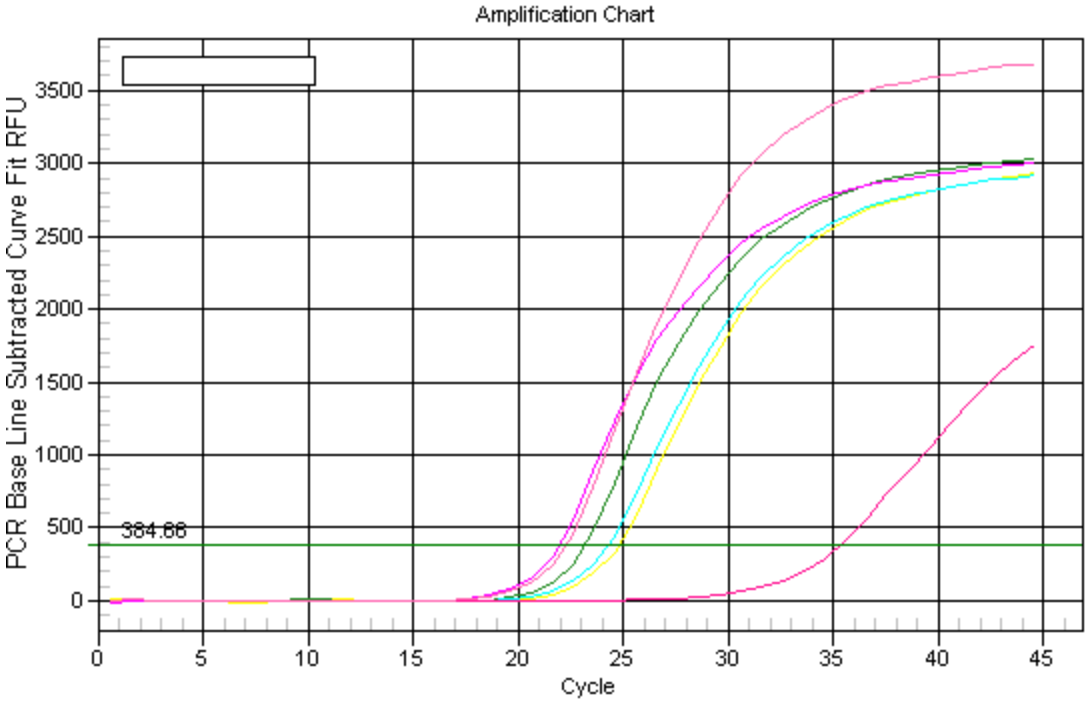


Figure 13 All CWBR qPCR graphs

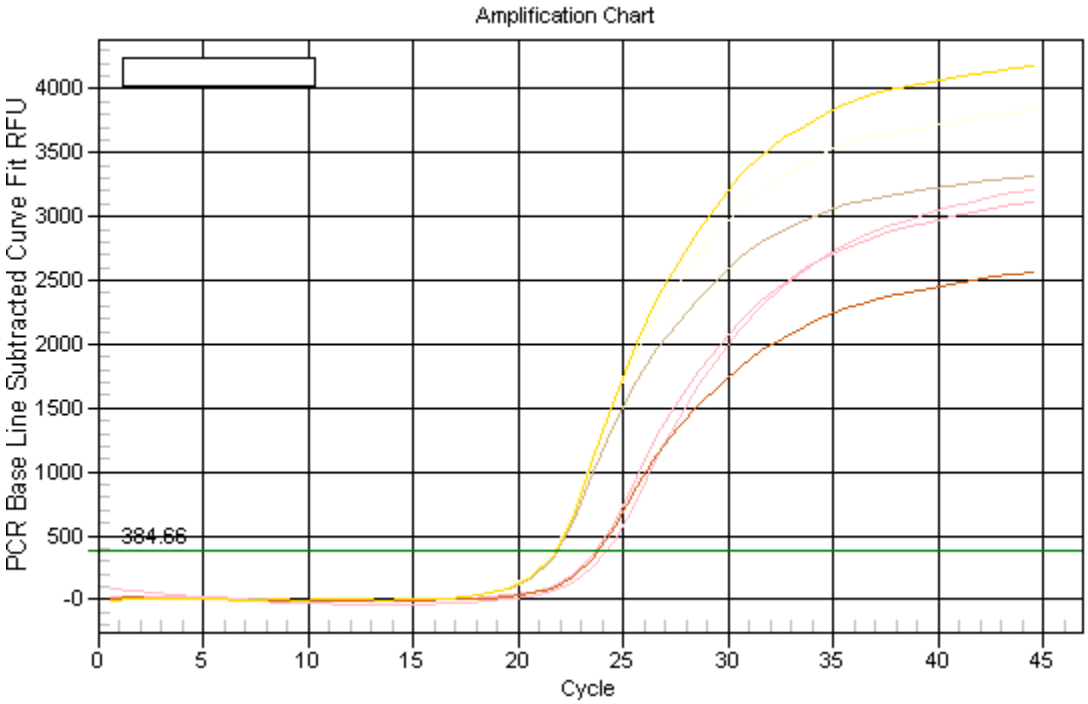


Figure 14 All PF qPCR graphs

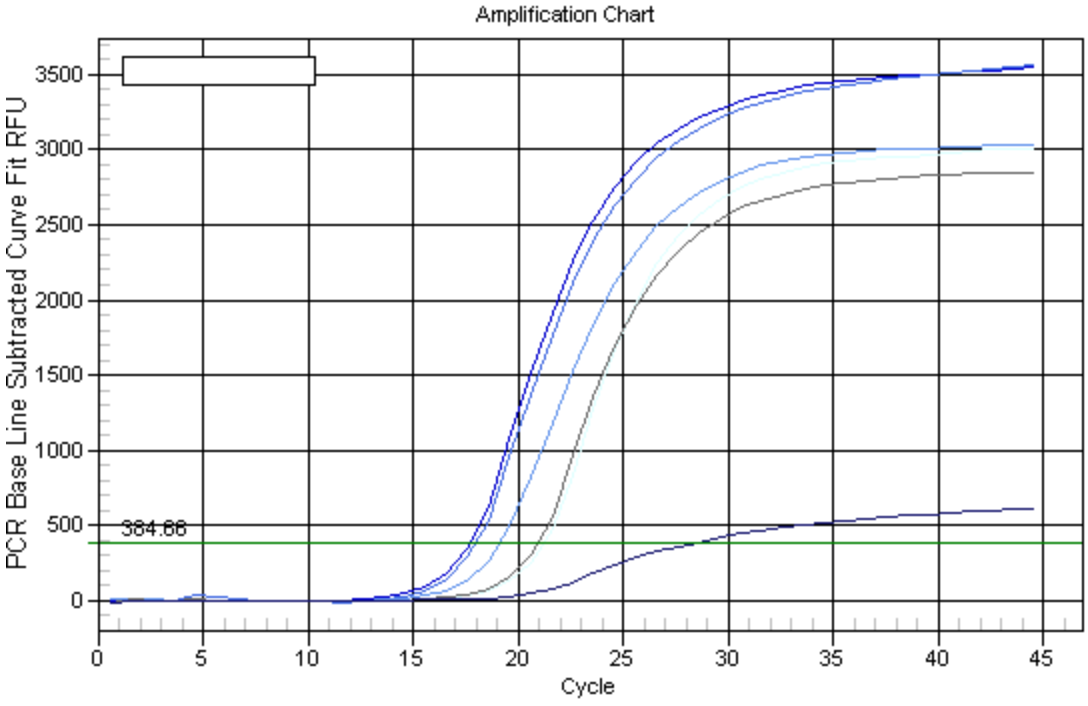


Figure 15 All cspA qPCR graphs

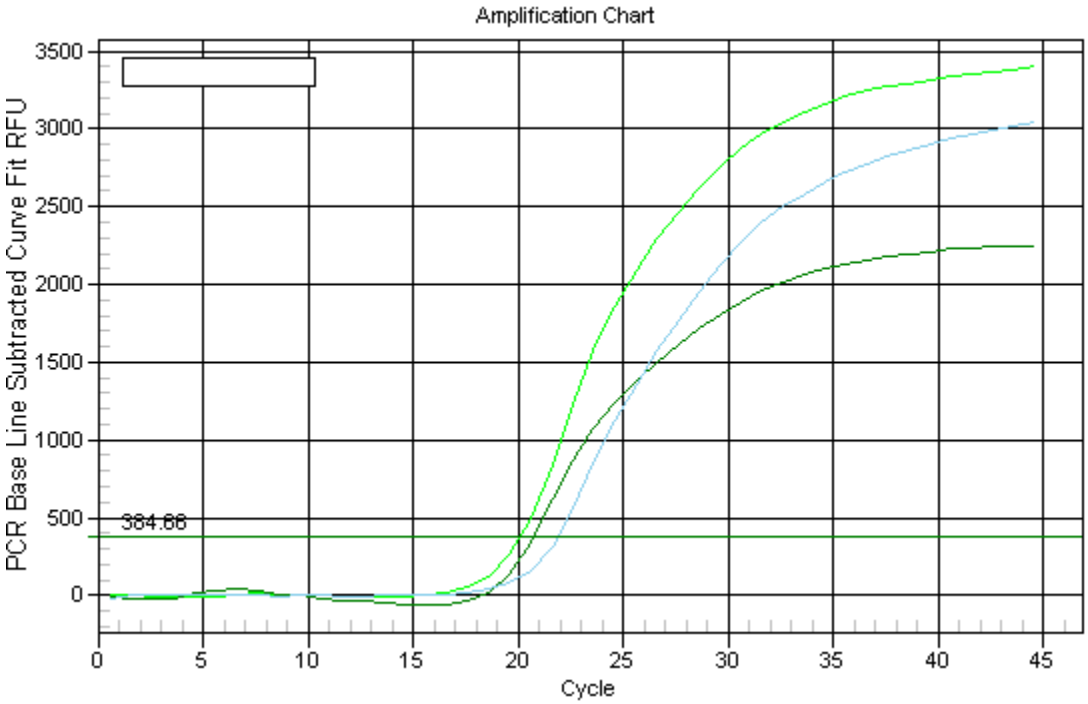


Figure 16 G3P Day 3 qPCR graphs

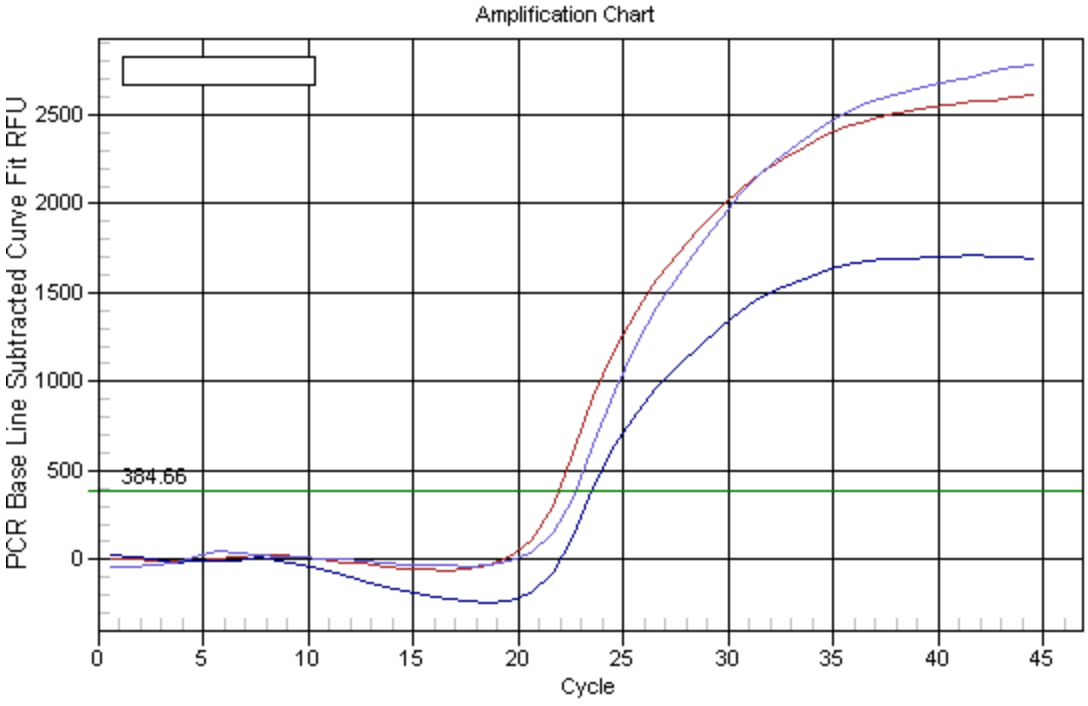


Figure 17 G3P Day 7 qPCR graphs

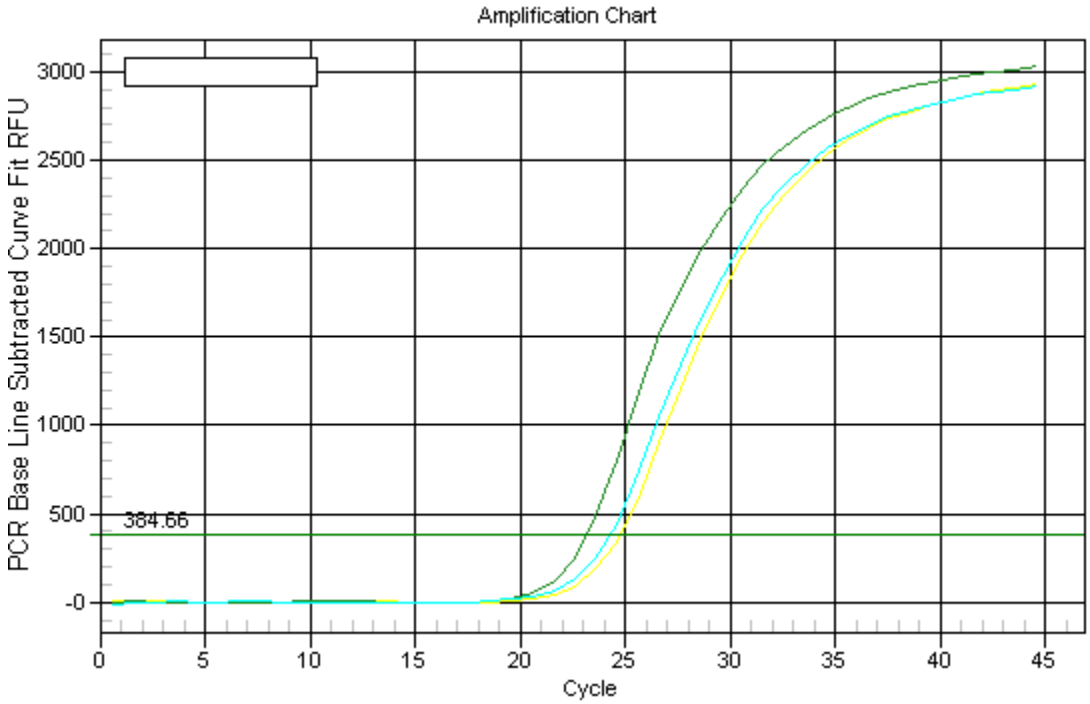


Figure 18 CWBR Day 3 qPCR graphs

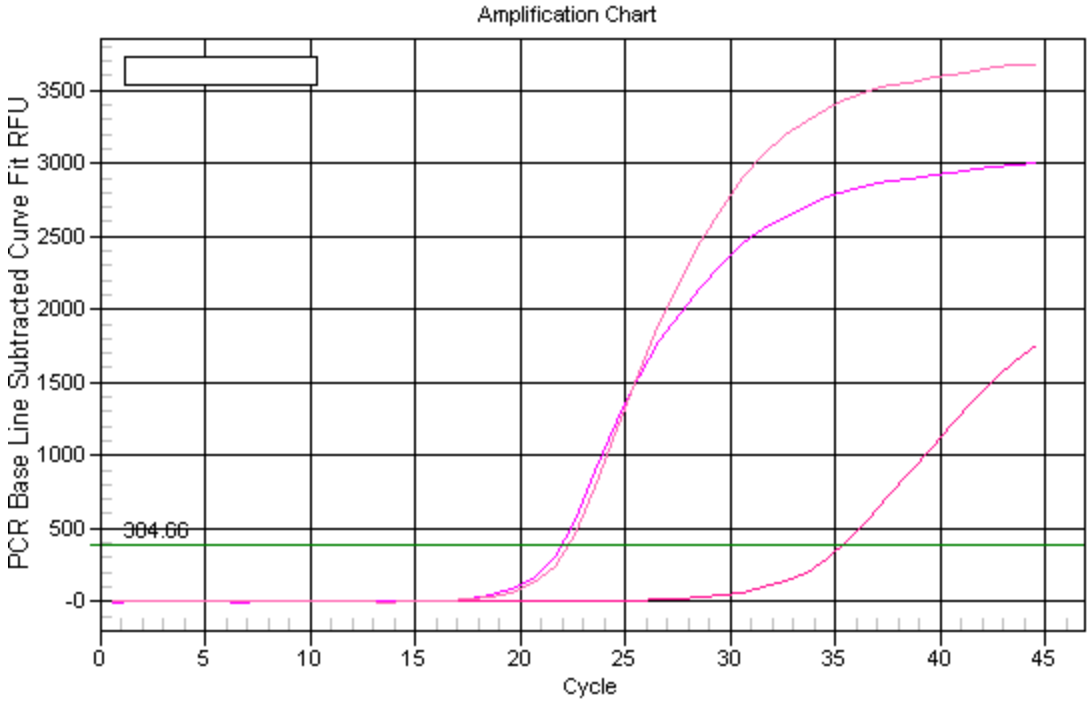


Figure 19 CWBR Day 7 qPCR graphs

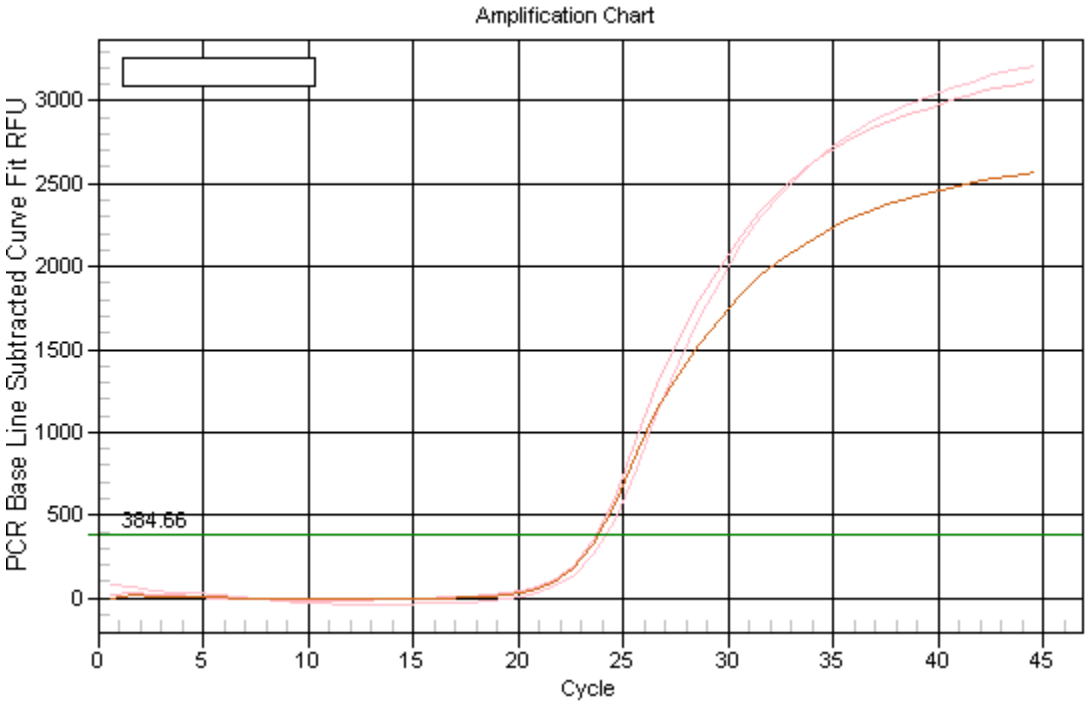


Figure 20 PF Day 3 qPCR graphs

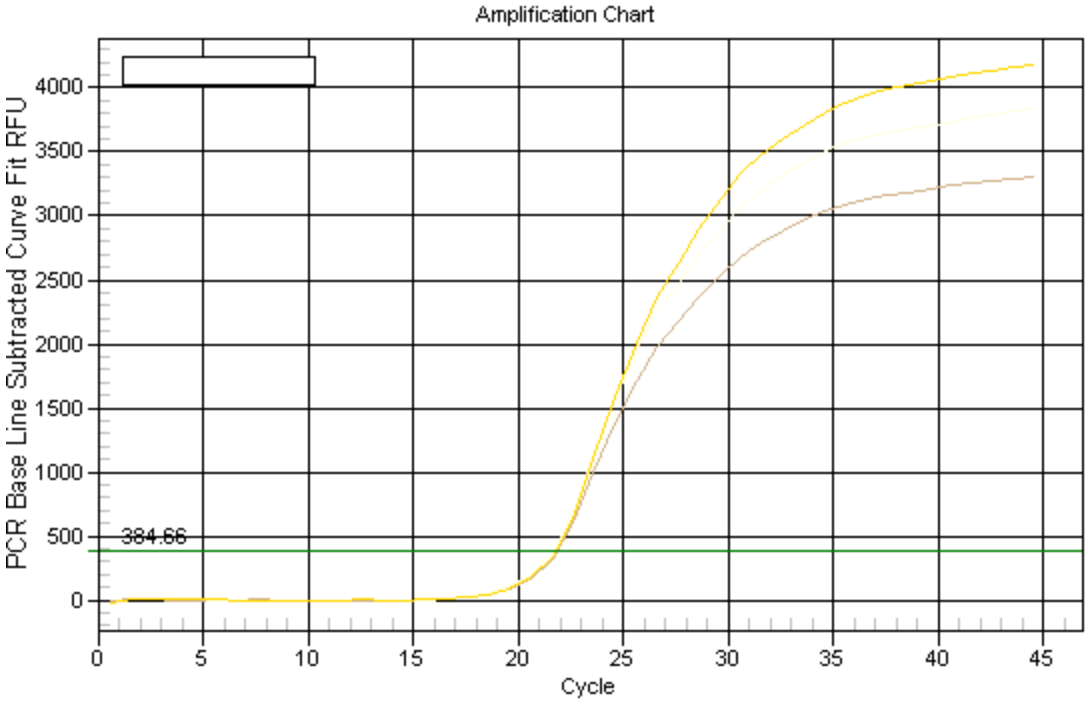


Figure 21 PF Day 7 qPCR graphs

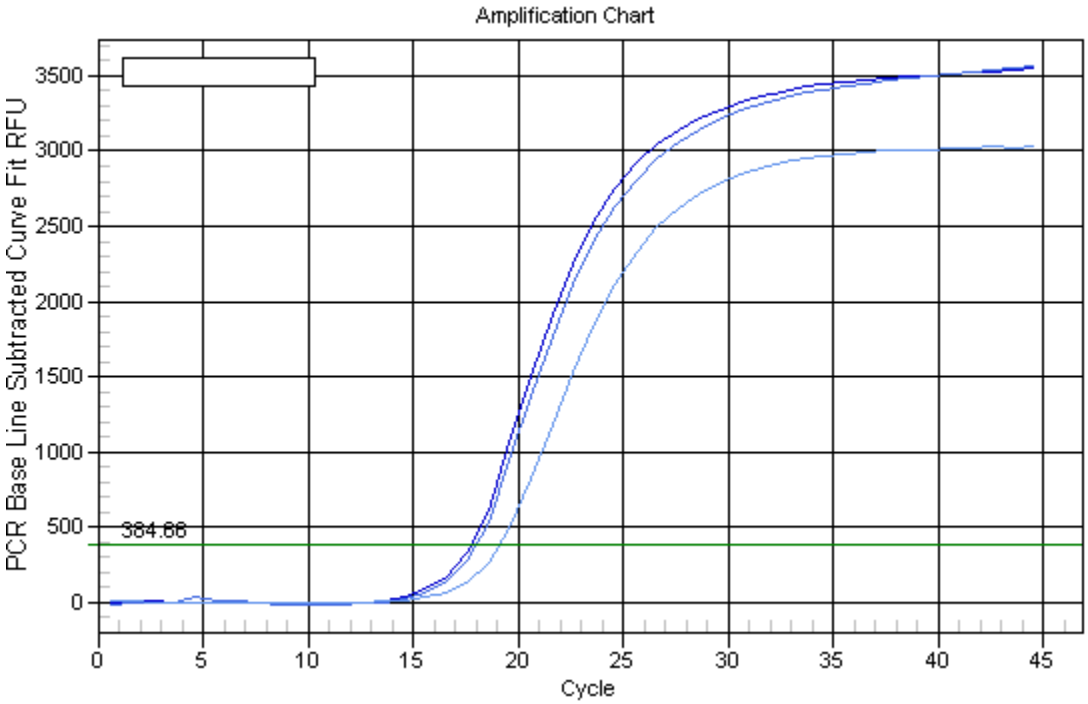


Figure 22 cspA Day 3 qPCR graphs

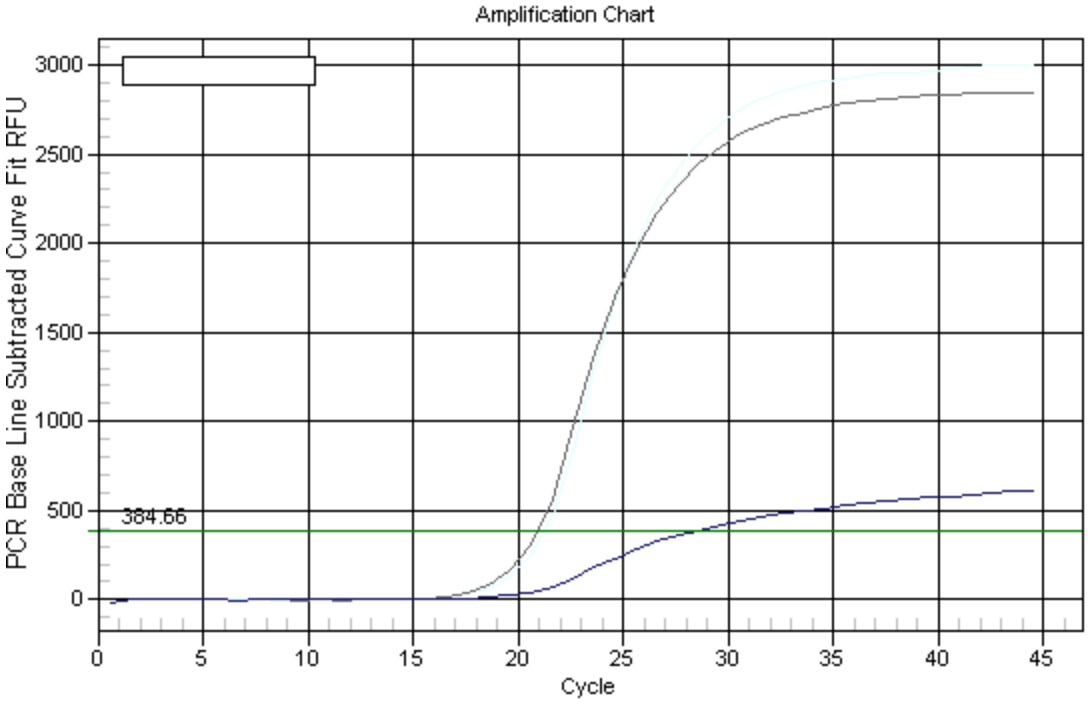


Figure 23 cspA Day 7 qPCR graphs

Bibliography

1. Ahmed I, Ross RA, Mathur VK, Chesbro WR. Growth rate dependence of solventogenesis and solvents produced byClostridium beijerinckii. Applied Microbiology and Biotechnology 1988;28(2):182-187.
2. Almeida JRM, Modig T, Petersson A, Hähn-Hägerdal B, Lidén G, Gorwa-Grauslund MF. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae. Journal of Chemical Technology & Biotechnology 2007;82(4):340-349.
3. Bauman A, Horvat D. The impact of natural radioactivity from a coal-fired power plant. Science of The Total Environment 1981;17(1):75-81.
4. Boddiger D. Boosting biofuel crops could threaten food security. The Lancet 2007;370(9591):923-924.
5. DOE. EIA - Total Energy - Energy Perspectives. In: Department of Energy, editor: Energy Information Administration; 2011.
6. Dürre P. Fermentative Butanol Production. Annals of the New York Academy of Sciences 2008;1125(1):353-362.
7. EIA. How Much Coal Is Left – Energy Explained, Your Guide to Understanding Energy In: Energy Information Administration editor2013.
8. EPA. Study of Boutique Fuels and Issues Relating to Transition from Winter to Summer Gasoline. In: Environmental Protection Agency, editor2001.

9. Ezeji TC, Qureshi N, Blaschek HP. Bioproduction of butanol from biomass: from genes to bioreactors. Current Opinion in Biotechnology 2007;18(3):220-227.

10. Ezeji T, Milne C, Price N, Blaschek H. Achievements and perspectives to overcome the poor solvent resistance in acetone and butanol-producing microorganisms. Applied Microbiology and Biotechnology 2010;85(6):1697-1712.

11. Ezeji T, Qureshi N, Blaschek HP. Butanol production from agricultural residues: Impact of degradation products on Clostridium beijerinckii growth and butanol fermentation. Biotechnology and Bioengineering 2007;97(6):1460-1469.

12. IEA. Key World Energy Statistics 2013. In: International Energy Association, editor2013.

13. Janssen H, Döring C, Ehrenreich A, Voigt B, Hecker M, Bahl H, Fischer R-J. A proteomic and transcriptional view of acidogenic and solventogenic steady-state cells of Clostridium acetobutylicum in a chemostat culture. Applied Microbiology and Biotechnology 2010;87(6):2209-2226.

14. Koukiekolo R, Cho H-Y, Kosugi A, Inui M, Yukawa H, Doi RH. Degradation of Corn Fiber by Clostridium cellulovorans Cellulases and Hemicellulases and Contribution of Scaffolding Protein CbpA. Applied and Environmental Microbiology 2005;71(7):3504-3511.

15. Lee SY, Park JH, Jang SH, Nielsen LK, Kim J, Jung KS. Fermentative butanol production by clostridia. Biotechnology and Bioengineering 2008;101(2):209-228.

16. Moss JT, Berkowitz AM, Oehlschlaeger MA, Biet J, Warth V, Glaude P-A, Battin-Leclerc F. An Experimental and Kinetic Modeling Study of the Oxidation of the Four Isomers of Butanol. The Journal of Physical Chemistry A 2008;112(43):10843-10855.

17. Owen NA, Inderwildi OR, King DA. The status of conventional world oil reserves—Hype or cause for concern? Energy Policy 2010;38(8):4743-4749.

18. Quaschning V. Understanding Renewable Energy Systems. Earthscan; 2005.

19. Qureshi N, Bowman MJ, Saha BC, Hector R, Berhow MA, Cotta MA. Effect of cellulosic sugar degradation products (furfural and hydroxymethyl furfural) on acetone–butanol–ethanol (ABE) fermentation using Clostridium beijerinckii P260. Food and Bioproducts Processing 2012;90(3):533-540.

20. Ramey D, Yang S-T. Production of Butyric Acid and Butanol from Biomass. US Department of Energy2004.

21. Sawyer RF. Trends in Auto Emissions and Gasoline Composition. Environmental Health Perspectives 1993;101:5-12.

22. Shi Z, Blaschek HP. Transcriptional Analysis of Clostridium beijerinckii NCIMB 8052 and the Hyper-Butanol-Producing Mutant BA101 during the

Shift from Acidogenesis to Solventogenesis. Applied and Environmental Microbiology 2008;74(24):7709-7714.

23. Tabak J. Geology and Chemistry of Oil. Science Online. Facts on File, Inc.2009.

24. Taylor M, Tuffin M, Burton S, Eley K, Cowan D. Microbial responses to solvent and alcohol stress. Biotechnology Journal 2008;3(11):1388-1397.

25. World Bank. World Development Report 2008 – Agriculture for Development. In: World Bank, editor2008.

26. Xu X, Chen C, Qi H, He R, You C, Xiang G. Development of coal combustion pollution control for SO2 and NOx in China. Fuel Processing Technology 2000;62(2–3):153-160.

27. Zhang J, Smith KR. Household Air Pollution from Coal and Biomass Fuels in China: Measurements, Health Impacts, and Interventions. Environmental Health Perspectives 2007;115(6):848-855.

28. Zhang Y, Han B, Ezeji TC. Biotransformation of furfural and 5-hydroxymethyl furfural (HMF) by Clostridium acetobutylicum ATCC 824 during butanol fermentation. New Biotechnology 2012;29(3):345-351.

29. Zheng Y-N, Li L-Z, Xian M, Ma Y-J, Yang J-M, Xu X, He D-Z. Problems with the microbial production of butanol. Journal of Industrial Microbiology & Biotechnology 2009;36(9):1127-1138.