ABSTRACT

MARKS, TIMOTHY JAMES. Development of a Genetic Toolbox for *Geobacillus kaustophilus* using Novel Bacteriophages GBK1 and GBK2. (Under the direction of Dr. Paul Hamilton).

Thermophiles are microorganisms that are capable of growth at temperatures above 45 ˚C. Enzymes from thermophiles are often functional at these higher temperatures, which is of interest from an industrial standpoint because many processes require steps at elevated temperatures. Thermophilic organisms can be used as whole-cell biocatalysts in processes such as bioremediation and bioconversion. Some thermophilic organisms, like *Clostridium thermocellum*, are well-studied and have a variety of tools available for genetic manipulation. Other thermophilic organisms, like *Geobacillus* spp., have been characterized more recently and there is a lack of genetic manipulation tools for these organisms that needs to be addressed in order to be viable hosts for industrial processes.

Geobacillus spp. are thermophilic (optimal growth between 37 ˚C and 80 ˚C), Gram-positive, facultatively anaerobic bacteria that have industrial potential as whole-cell biocatalysts as well as producers of thermophilic enzymes of interest, including DNA polymerases, lipases, and amylases. *Geobacillus* spp. are also capable of growth on a variety of C5 and C6 sugars and they can survive unfavorable conditions due to their ability to form endospores. There are challenges with working with *Geobacillus* spp. as a host, including difficulties in DNA uptake due to the thick peptidoglycan layer and restriction-modification systems present in several host strains that degrade exogenous DNA. Additionally, most antibiotics are not thermostable at ideal growth temperatures for geobacilli and most antibiotic resistance markers lack thermostability.

In this research, two novel bacteriophages (named GBK1 and GBK2) were isolated, sequenced, annotated, and characterized. Potential tools from these bacteriophages were identified and developed for use in *Geobacillus* spp. GBK2 is a lytic bacteriophage isolated from a backyard compost pile in Cary, NC that infects *Geobacillus kaustophilus* ATCC 8005. It has a circularly permuted genome with 62 open reading frames (ORFs) revealed upon annotation. GBK2 shares close homology (12 ORFs) with *Bacillus subtilis* phage SPP1, revealing evolutionary similarities with this mesophilic phage. ORFs of interest from GBK2 include ORF 51 (homology to a thymidylate synthase) and ORF 39 (homology to *recT*). The thymidylate synthase could be used to create a marker-free selection system (see details for GBK1) of a *thyA* mutant *Geobacillus* strain and *recT* could be used to create a recombineering system in geobacilli similar to the RecET system in *E. coli*.

GBK1 is a linear lysogenic bacteriophage isolated from a backyard compost pile in Cary, NC that infects *G. kaustophilus* ATCC 8005. Analysis of the genome sequence of GBK1 identified 56 ORFs, including a homolog to a tyrosine integrase and a 22 bp *attP* site that has a matching *attB* site in several *Geobacillus* strains. This *attP/*integrase was used to create a shuttle plasmid capable of integration into *G. thermoglucosidasius* as well as *G thermoleovorans*. Another ORF of interest in GBK1 is ORF 37, which has homology to a thymidylate synthase. The thymidylate synthase was used to create a shuttle vector capable of auxotrophic complementation of a *thyA*- mutant of *G. thermoleovorans.*

The lack of tools for genetic manipulation of *Geobacillus* spp. is a deterrent to using these potentially useful organisms for industrial processes. We describe a shuttle vector capable of replication in *E. coli* that stably integrates into *G. thermoglucosidasius* (as well as other strains of *Geobacillus*) and a marker-free antibiotic selection system based on

complementation of a deficient thymidylate synthase gene in a mutant host strain of *G. thermoleovorans.* The results generated by these studies provide new tools to expand the genetic manipulation resources of *Geobacillus* spp. for future use in industrial processes such as bioremediation and biofuel production.

Development of a Genetic Toolbox for *Geobacillus kaustophilus* using Novel Bacteriophages GBK1 and GBK2

by
Timothy James Marks

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Microbiology

Raleigh, North Carolina

2020

APPROVED BY:

_____	_____
Dr. Paul Hamilton Committee Chair	Dr. Amy Grunden
_____	_____
Dr. Eric Miller	Dr. Gary Gilleskie

DEDICATION

To my wife, Rebecca, for her unwavering patience, understanding, love and support. Those that have have ever lived with a scientist, particularly one attempting to finish a thesis, know it is not for the faint of heart. The emotional roller coaster of engaging in research is very real and is not easy to leave in the lab. Thank you for lending an ear for a project that you know nothing about. Thank you for the encouragement when things fell apart time after time. Thank you for the tough love when needed. And above all, thank you for being the driving force to keep me looking at the big picture. I could not have done it without you.

To my children, Camden, Kara, and Tyler, for their unconditional love. It's not easy for young kids to understand the sacrifice necessary for an undertaking like this, but it makes it a lot easier when you have kids like I do. Thanks for always being willing to come with me to Campbell on the weekends to set up experiments or check results and making the road trips fun.

To Dr. Paul Hamilton for his expert guidance during this unconventional undertaking. It wasn't an easy situation for either of us, but I couldn't think of anyone better suited to be an adviser in this situation. Thank you for always responding to my emails and phone calls and for your guidance in getting me to think outside the box and expand my horizons as a scientist.

To my parents, Steve and Barb Marks, for their encouragement and willingness to listen, especially when I was having doubts about being able to see this project through to completion.

To my grandfather, Dr. Dale Butz, the inspiration for this Ph.D. His doctorate in Agricultural Economics from the University of Minnesota inspired me to pursue the highest degree in my own area of interest. I can think of no one better to emulate.

BIOGRAPHY

Timothy James Marks was born in 1981 in Kankakee, Illinois. He spent most of his childhood in Bloomington-Normal, Illinois. Tim received his bachelor's degree from Butler University in 2003. In addition to meeting his wife there, Tim also began working in the lab of Dr. James Shellhaas in 2002. It was here that he discovered his love for microbiology. After five years of playing professional baseball and one year of teaching high school earth science, Tim decided to pursue a Master's of Microbial Biotechnology (MMB) degree from North Carolina State University in Raleigh. Between his first and second years in the MMB program, he did an internship with Novozymes in Franklinton, NC under the direction of Dr. Nathan Kreel. Tim's work in the Research and Development department concerning the production of second generation biofuels furthered his interest in industrial microbiology. Upon graduation from the MMB program in 2010, Tim gained employment at Campbell University as a Program Administrator/Instructor in Pharmaceutical Sciences. In an effort to move into a tenure-track position within the Campbell Pharmaceutical Sciences position, Tim applied for admission to the doctorate program in Plant and Microbial Biology at North Carolina State University under the direction of Dr. Paul Hamilton, who was also his advisor in the MMB program at NCSU. In this program he was given the opportunity to work on a phage characterization project involving developing genetic manipulation tools for the thermophilic microorganism *Geobacillus kaustophilus*. Tim has pursued his doctorate since 2012 while working full-time as an Assistant Research Professor at Campbell. During this time, Tim has taught courses in molecular biology, upstream/downstream bioprocessing, scientific and technical writing, interpersonal skills, and a seminar in scientific literature. He has also served as a faculty adviser, mentor, and principle investigator to over thirty students

in the Campbell Pharmaceutical Sciences program in addition to serving as the chair of the Curriculum Committee. He has been an active participant in the redesign of the curriculum for the Master of Science in Pharmaceutical Sciences (MSPS) program and looks forward to expanding the impact of biotechnology on the MSPS students. Tim continues to be fascinated with all of the things microorganisms can do and plans to continue research and teaching.

ACKNOWLEDGMENTS

I would like to thank Dr. Paul Hamilton for his mentorship over the last 9 years. I will forever appreciate Dr. Hamilton's willingness to answer my questions and provide guidance in both research and teaching. I would also like to thank my former mentor at Campbell, Dr. Timothy Bloom, for his mentorship, especially in my role as an instructor and assistant professor. I would be remiss if I didn't thank the unending support from my department at Campbell. The Pharmaceutical Sciences department (specifically Dr. Diliberto and Dr. Eagerton) were very supportive of this endeavor, both in time and monetary resources. I would also like to thank my committee members Dr. Amy Grunden, Dr. Eric Miller, Dr. Nathaniel Hentz, and Dr. Gary Gilleskie for their thoughtful questions and experimental suggestions. Thank you to everyone who provided feedback on methods and use of equipment and supplies: Dr. Stephanie Mathews, Dr. Caroline Smith-Moore, Dr. Rebecca Kitchener, Jennifer Greenstein, Dr. Stephen Sharkady, Dr. Stephen Holly, Dr. Brianne Raccor, and Dr. Dorothea Thompson. Thank you to NCSU staff who provided support including Cindy Whitehead, Dwayne Barnes, Robert Davis, and Vicki Lemaster. Lastly, thank you Rebecca, Mom, Dad, and Katie for your encouragement and patience.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1. Literature Review: Biotechnological Potential of the Genus *Geobacillus*

CHAPTER 2. Characterization of a thermophilic bacteriophage of *Geobacillus kaustophilus*

CHAPTER 3. The characterization of a novel lysogenic bacteriophage of the thermophile *Geobacillus kaustophilus*: source of genetic tools and a glimpse at phage evolution

CHAPTER 4. Development of an integrating suicide vector and counterselection system for *Geobacillus* spp.

LIST OF FIGURES

CHAPTER 1. Literature Review: Biotechnological Potential of the Genus *Geobacillus*

CHAPTER 2. Characterization of a thermophilic bacteriophage of *Geobacillus kaustophilus*

CHAPTER 3. The characterization of a novel lysogenic bacteriophage of the thermophile *Geobacillus kaustophilus*: source of genetic tools and a glimpse at phage evolution

CHAPTER 4. Development of an integrating suicide vector and counterselection system for *Geobacillus* spp.

CHAPTER 1

Literature Review

Biotechnological Potential of the Genus *Geobacillus*

Marks, Timothy J.[AB] and Hamilton, Paul T.[A]

[A]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695

[B]Department of Pharmaceutical Sciences, Campbell University, Buies Creek, NC 27506

Abstract

Geobacillus spp. are thermophilic, Gram-positive bacteria that are of biotechnological potential because of their catabolic versatility and ability to express recombinant proteins. Furthermore, these facultative anaerobes are being studied for their potential to produce thermostable enzymes as well as to serve as biocatalysts for the conversion of biomass into biofuels. Thermophiles provide some key advantages over mesophilic counterparts during fermentation, including lower energy expenditures, a lower risk of contamination by common mesophilic microorganisms, and easier separation of volatile products like ethanol. However, there are a lack of tools for genetic manipulation of *Geobacillus* spp., including a lack of reliable DNA uptake protocols and difficulties in proliferation of exogenous DNA in geobacilli. This review outlines the current tools available for *Geobacillus* spp., including methods for DNA transfer, positive selection markers that are thermostable, commonly used strains of *Geobacillus* spp., vectors, antibiotic-free selection systems, reporter genes, recombinant gene expression, protein secretion systems, potential bacteriophage tools, and CRISPR.

1.1 *Geobacillus* spp.: A Background

Thermophiles are heat-loving organisms that grow and thrive at temperatures above 45 ˚C. Thermophilic hosts can allow for microbial bioprocesses to be carried out at high temperatures (high temperature microbial processes [HTMPs]), which can offer advantages over more traditional mesophilic bioprocesses that require moderate temperatures, 20-45 ˚C (Suzuki et al., 2013). For example, HTMPs allow for low energy expenditure for agitation and cooling, a reduced risk of contamination and thus reduced need for antibiotics, and easy removal of volatile products, such as ethanol (Weigel and Ljungdahl, 1986).

Members of the genus *Geobacillus* are aerobic and facultatively anaerobic, Gram-positive, thermophilic bacilli that were reclassified in 2001 from the genus *Bacillus* (Nazina et al., 2001) and grow optimally between 50 °C and 60 °C, with some strains able to grow between 37 °C and 80 °C. They also form endospores, which can be advantageous because they can survive unfavorable conditions (Kananavičiūtė and Čitavičius, 2015; Blanchard et al., 2014). Spores from *Geobacillus* spp. are extremely resistant to drying out, damage from ultraviolet (UV) light, and inactivation by heat, much like their close relatives in the *Bacillus* genus (Setlow, 2006). Sporulation is commonly observed in these organisms with 67 of the 75 core sporulation genes from *Bacillus* spp. found in *Geobacillus* spp. (Galperin et al., 2012; Zeigler, 2014).

Geobacillus spp. have been isolated from a wide range of higher-temperature and temperate environments, including hot springs, deep geothermal vents, compost heaps, cool soils, and low-temperature deep sea water environments (Hussein et al., 2015; Zeigler, 2014). The worldwide distribution of *Geobacillus* spp. is most likely due to their spores, which are transient and can travel large distances in the atmosphere (Zeigler, 2014). Since 2006, several

genome sequences of *Geobacillus* spp. have been characterized, helping with the comparisons of genome dynamics as well as deepening the understanding of metabolic capabilities within the genus (Studholme, 2014). G+C content varies widely between species (a range of 43.9% in the facultative anaerobes to 51.9% to 55% in *G. thermoleovorans, G. vulcani, G. lituanicus, G. kaustophilus,* and *G. thermocatenulatus)* (Hussein et al., 2015). Aerobic growth of *Geobacillus* spp. is very similar to *E. coli,* leading to a high growth rate and, consequently, high cell densities (Suzuki et al., 2013). Growth can occur on a mineral salts medium as well (Hussein et al., 2015).

1.1.1 Thermostable Enzyme Production using *Geobacillus* spp. Recently, members of *Geobacillus* spp. have been studied for their advantages over other commonly used organisms and their biotechnological potential due to their ability to break down a variety of carbon sources, their ability to survive extreme environmental conditions by way of endospore formation, and their thermophilic nature (Coorevits et al., 2012; Nazina et al., 2001; Omokoko et al., 2008; Zeigler, 2014; Blanchard et al., 2014). The production of thermostable enzymes by *Geobacillus* spp., including proteases, lipases (McMullan et al., 2004), amylases, DNA polymerase (Mead et al., 1991) and reverse transcriptase (Vellore et al., 2004) is of industrial interest. Proteases are popular, industrially useful enzymes used in detergents, applications in the food industry, leather processing, and peptide synthesis, amongst other biotechnological applications (Haki and Rakshit, 2003). Advantages for thermostable proteases include high catalytic efficiencies, compatibility with processes functional at higher temperatures due to reduced viscosity, and a reduction in mesophilic microbial contamination (Hussein et al., 2015). A number of proteases from *Geobacillus* species have been isolated and characterized as having widespread putative proteolytic

capabilities (Chen et al., 2004; Hawumba et al., 2002; Iqbal et al., 2015; Zhu et al., 2007). Lipases (EC 3.1.1.3) have a preference for substrates that are insoluble in water and are instrumental in the hydrolysis of carboxyl esters while hydrolyzing long-chain triglycerides (Bornscheuer et al., 2002). They generally interact more readily with hydrophobic substrates because they have a hydrophobic domain that covers the active site; furthermore, they are more resistant to organic solvents than carboxylesterases (Patel, 2006). Lipases are routinely used in the cosmetics industry to improve the appearance and glide of personal care products, for the synthesis of sugar esters (Khanniri et al., 2015), for cheese making, for the synthesis of emollient esters in the cosmetics industry (Ansorge-Schumacher and Thum, 2013), and for applications in synthetic chemistry (Pandey et al., 1999). Lipases produced by thermophiles have the same advantages as thermostable proteases when compared to their mesophilic counterparts. Several *Geobacillus* spp., including *G. thermodenitrificans* IBRL-nra (Balan et al., 2012; Balan et al., 2010), *G. thermodenitrificans* AZ1 (Abdel-Fattah et al., 2012), *G. stearothermophilus* JC (Jiang et al., 2010), *G. stearothermophilus* L1 (Kim et al., 1998), *G. stearothermophilus* strain 5 and 6 (Berekaa et al., 2009; Sifour et al., 2010), *G. thermoleovorans* CCR11 (Quintana-Castro et al., 2009), *G. thermoleovarans* Toshki (Abdel-Fattah et al., 2008), *Geobacillus zalihae* (Nurbaya et al., 2014) as well as many other *Geobacillus* isolates (Hamid et al., 2003; Leow et al., 2007; Li and Zhang, 2005; Mahadevan and Neelagund, 2014; Wang et al., 1995; Zhu et al., 2014) produce thermostable lipases that have been isolated. Carboxylesterases catalyze the cleavage of carboxylic esters that are water-soluble, contrary to lipases (Hussein et al., 2015). Carboxylesterases have been characterized from *G. stearothermophilus* ATCC12980 and ATCC 7954 (Ewis et al., 2004), *G. thermoleovorans* YN (Soliman et al., 2014), *G. thermodenitrificans* T2 (Yang et al.,

2013), *G. thermodenitrificans* CMB-A2 (Charbonneau et al., 2010), *G. kaustophilus* HTA426 (Montoro-Garcia et al., 2009), *G. caldoxylosilyticus* TK4 (Yildirim et al., 2009), *G. thermoglucosidasius* EAEC and many other *Geobacillus* isolates (Ayna, 2013; Özbek et al., 2014; Tekedar and Sanu-Mohamed, 2011; Zhu et al., 2012). Acetylxylan esterases (EC 3.1.1.72), a subfamily of carboxylesterase of interest in second-generation biofuels, deacetylate xylans and xylo-oligosaccharides (Hussein et al., 2015) and enhance the accessibility of xylans in hemicellulose mixtures (Zhang et al., 2011).

The current standard for production of these enzymes uses mesophilic hosts with well-established expression systems; however, expression of heterologous thermophilic proteins in mesophilic hosts can be problematic due to differences in codon usage and incorrect folding of thermophilic proteins at lower temperatures (Suzuki et al., 2013; Turner et al., 2007). Some genes require sequence modification to match the codon frequencies of the host strain, either with the expression of thermophilic proteins in mesophilic bacteria or with the expression of mesophilic proteins in thermophilic bacteria (Blanchard et al., 2014; van Zyl et al., 2014). This demonstrates the need for an established host protein expression system that is functional and efficient in the native *Geobacillus* spp.

1.1.2 Biofuel Production using *Geobacillus* spp. First-generation bioethanol (C_2H_5OH) is a well-established biofuel currently produced using corn/wheat-derived glucose or cane sucrose (Hussein et al., 2015). Second-generation processes will focus on the production of bioethanol from lignocellulosic feedstocks, and *Geobacillus* spp. have great potential as an industrial producer because of their aforementioned ability to use many different carbohydrate sources that can be used as feedstock, including pentoses, hexoses, and di- and oligo-saccharides found in lignocellulosic biomass (Turner et al., 2007). Most

Geobacillus spp. have acquired a genomic island of genes for hemicellulose utilization from an organism with a lower G+C content (De Maayer et al., 2014). *Geobacillus* spp. are capable of being utilized as whole-cell biocatalysts because they can produce a large variety of enzymes needed for the production of industrially valuable products (Xiao et al., 2012). Some *Geobacillus* spp. secrete a minimal set of enzymes, including some glycoside hydrolases (Shulami et al., 2014) that hydrolyze noncrystalline polymeric substrates (Hussein et al., 2015). The addition of genes encoding for these glycoside hydrolase enzymes could increase the range of poly-saccharide degradation in other *Geobacillus* strains. While there is no direct evidence of the ability of these organisms to degrade crystalline cellulose, there is evidence of extracellular endoglucanase activity that has been detected (Aspeborg et al., 2012). Bartosiak-Jentys et al. (2013) created their own cassette for expression of heterologous endoglucanase in *G. thermoglucosidasius* for this purpose. Many *Geobacillus* spp. are also able to produce thermostable α-amylases and 1-6 specific pullanases for starch breakdown (Offen et al., 2015; Kuriki et al., 1988). Engineering *Geobacillus* spp. for increased breakdown of various oligo- and poly-saccharides found in feedstock biomass could greatly improve fermentation processes (Turner et al., 2007).

Most biotechnological applications depend on fermentation processes to produce second-generation biofuels. In addition to the enzyme production listed previously, members of this genus have been used to produce acetate, formate, lactate, ethanol, and succinate (Fig 1-1) as a result of mixed acid fermentation resulting from anaerobic metabolism (Hussein et al., 2015). Additionally, some strains are also able to produce R, R 2, 3-butanediol (Xiao et al., 2012).

To date, *G. thermoglucosidasius* has been studied for second-generation biofuels production by such European companies such as TMO Renewables and Argrol Ltd. There are several advantages of bioprocesses that function at temperatures greater than 50 °C, including the reduction of energy expenditures for cooling, the minimization of the threat for microbial contamination, and the ease of downstream recovery for volatile products (like ethanol) (Kananavičiūtė and Čitavičius, 2015). Since ethanol's boiling point is 78 °C, continuous removal is possible in these processes that are run at higher growth temperatures while also reducing the concern for contamination from mesophilic sources (Cripps et al., 2009). Ethanol is not the main fermentation product of the mixed-acid producing *Geobacillus* spp., however. Further genetic manipulation in *G. thermoglucosidasius* was required to increase ethanol yield by knocking out some the metabolic pathway responsible for lactate production. Lactate dehydrogenase is found in all known *Geobacillus* spp. (Hussein et al., 2015). Both FAD and NAD-linked enzymes are present, with NAD-linked lactate dehydrogenase being responsible for lactate production under oxygen-limited conditions (Hussein et al., 2015). Knocking out the L-lactate dehydrogenase pathway would presumably cause fermentation products to be funneled through the pyruvate formate lyase (PFL) pathway (Cripps et al., 2009). Because pyruvate dehydrogenase (Pdh) is still active in the presence of oxygen, the PFL pathway led to higher yields of ethanol than expected (Cripps et al., 2009). A homoethanolic pathway was developed by making deletions in the PFL pathway and increasing expression of Pdh (Cripps et al., 2009). *G. thermoglucosidasius* TM242, a triple mutant (Δ*ldh*, Δ*pflB*, pdh_{up}) strain, produced ethanol yields of up to 90% of the theoretical value from glucose, reaching yields of up to 3.2 g/L h on cellobiose and 2.85 g/L

h on glucose (Cripps et al., 2009). A gas stripping model with continuous fermentation predicts *G. thermoglucosidasius* TM242 could lead to 5.2 g/L h of ethanol (Niu et al., 2005).

Efforts have also been made to express mesophilic enzymes responsible for ethanol production in a *G. thermoglucosidasius* host, but yields were not as expected (Taylor et al., 2008; Thompson et al., 2008), presumably because of incompatibility of codon usage between the source of the pyruvate decarboxylase (*pdc*) gene and the host (Van Zyl et al., 2014). As a response to this, Van Zyl et al. (2014) attempted to reconcile codon usage for *pdc* from *Gluconobacter oxydans* into *G. thermoglucosidasius* (*ldh*$^{-}$ mutant) and achieved yields of 0.35 g ethanol/g glucose, albeit there was a limit on temperatures only up to 48 ˚C. Thus, further research is needed for optimization of ethanol production utilizing the PDC-dependent pathway in *Geobacillus* spp.

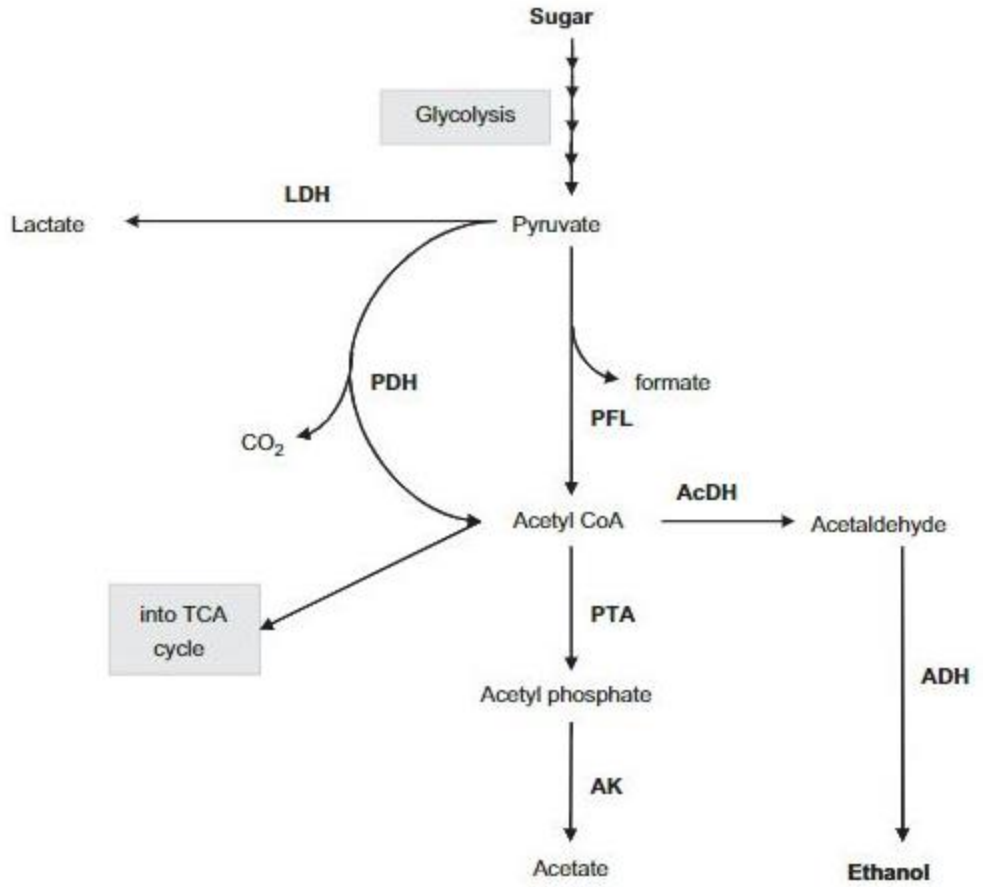

Figure 1-1 Major metabolic pathways involved in ethanol production by fermentation in *Geobacillus* spp. LDH, lactate dehydrogenase; PFL, pyruvate formate lyase; PDH, pyruvate dehydrogenase; AcDH, acetaldehyde dehydrogenase; ADH, alcohol dehydrogenase; PTA, phosphotransace-tylase; AK, acetate kinase; TCA cycle, tricarboxylic acid cycle (Adapted from Cripps et al., 2009).

1.2 Genetic Manipulation Challenges in Geobacilli

There are three tenets to a reliable genetic modification system: a host amenable to DNA uptake, a methodology for uptake of the exogenous DNA into this host, and a vector for recombinant gene transfer and expression (Inoue and Sako, 2013). Several genetic modification systems for mesophilic non-model microorganisms have been developed recently using synthetic biology approaches (Schiel-Bengelsdorf and Dürre, 2012; Nikel et al., 2014; Papon and Clastre, 2014; Markley et al., 2015; Mimee et al., 2015), but synthetic biology is lagging behind when it comes to thermophilic organisms because the establishment of methodologies for transformation of exogenous DNA into new species is time consuming and challenging. Geobacilli are phylogenetically similar to well-studied members of the genus *Bacillus,* which could be useful in the development of a DNA transfer system development based on *Bacillus* methods (Kananavičiūtė and Čitavičius, 2015). Further sequencing and characterization of genomes can aid in developing host-vector systems needed for genetic manipulation (Studholme, 2015).

In addition to the lack of reliable vectors in *Geobacillus* spp., the Gram-positive nature of these organisms paired with restriction-modification systems pose problems when it comes to DNA uptake. Gram-positive bacteria are typically resistant to exogenous DNA transformation due to the lack of permeability of the plasma membrane and the thick peptidoglycan layer (Lin and Xu, 2013). Additionally, a host restriction-modification (R-M) system that cleaves specific double-stranded DNA sites exists in most *Geobacillus* spp. that complicates the transformation of foreign DNA (Hussein et al., 2015). These R-M systems probably evolved as a natural defense to bacteriophage infection in these organisms. Unmethylated DNA is targeted by restriction enzymes in the host cell while host DNA is

protected by methylation of specific sites by any of the four types of restriction modification systems that have been found.

1.3 Current Genetic Tools for *Geobacillus* spp.

Genetic manipulation tools in *Geobacillus* spp. genomes are largely based on the myriad of tools developed in the well-studied, close relative *B. subtilis.* However, *B. subtilis* tools (and other mesophilic tools) are of limited utility for thermophiles due to the low mesophilic protein stability and the breakdown of commonly used antibiotics at higher temperatures (Hussein et al., 2015). Therefore, the development of new thermostable tools in these organisms requires use of genetic machinery from the thermophilic organisms themselves (Hussein et al., 2015).

In order to genetically manipulate geobacilli, it was necessary to come up with efficient protocols for DNA uptake to overcome some of the challenges listed previously. Protoplast transformation (Imanaka et al., 1982; Liao et al., 1986; Wu and Welker, 1989), electroporation (Cripps et al., 2009; Zeigler, 2001), and conjugation (Suzuki et al., 2013) are the three main methods that have been developed for the introduction of exogenous DNA into various *Geobacillus* spp. Table 1-1 shows the transformation efficiencies of each of these methods in their respective host strains. There are other methods of DNA transfer available for different bacteria that could be applied for *Geobacillus* spp. (Kananavičiūtė and Čitavičius, 2015). Several *Bacillus* spp. use natural competence for DNA uptake, and it is possible that this method could be used for *Geobacillus* spp. as natural competence genes are found in the genomes of *Geobacillus* spp. that have been sequenced (Kovács et al., 2009; Spizizen, 1958; Wiegand et al., 2013). To date, however, conditions for DNA uptake by

natural competence have not been described for geobacilli. Various vectors and selection methods have been utilized following the characterization of these DNA transfer protocols.

Table 1-1

Transformed *Geobacillus* strains, DNA transfer methods, and DNA transfer efficiencies (Adapted from Kananavičiūtė et al. 2014).

Strain	**Method**	**Plasmids**	**Transformation Efficiency; CFU of transformants/ µg of plasmid DNA, unless otherwise stated**	**Reference**
G. stearothermophilus NUB36 and its derivatives (NUB3621 and NUB3621R)	Protoplast transformation	pTHT15 [a] pLW05 [b] pRP9 pSTE12	4 x 10^8 2 x 10^7 4-6 x 10^5 -	Wu and Welker (1989) Wu and Welker (1989) De Rossi et al. (1994) Couñago and Shamoo (2005)
	Electroporation	pNW33N pUCG18	- 1.4 x 10^2	Blanchard et al. (2014) Kananavičiūtė et al. (2014)
G. thermoglucosidasius NCIMB 11955 (aka DSM 2542^T) and its derivatives	Electroporation	pTMO19 pUCG3.8 pNW33N pUCG18 pG1K pG1AK	- 2.8 x 10^5 - 4.9 x 10^3 5.3 x 10^4 5.8 x 10^3	Cripps et al. (2009) Bartosiak-Jentys et al. (2013) Lin et al. (2014) Reeve et al. (2016) Reeve et al. (2016)
G. kaustophilus HTA426 and its derivatives	Conjugative Transfer	pUCG18T pSTE33T pGAM46	10^{-5}-10^{-3} recipient^{-1} 10^{-7}-10^{-6} recipient^{-1} -	Suzuki and Yoshida (2012) Suzuki and Yoshida (2012) Suzuki et al. (2012)

[a] Plasmid source is same strain.
[b] This plasmid is derivative of mesophilic plasmid pPL401 (3 kb, contains chloramphenicol resistance marker).

1.3.1 DNA Transfer. The first protocol developed for DNA transfer into geobacilli was by way of protoplasts into *G. stearothermophilus*. Protoplasts involve the removal of the peptidoglycan layer by treatment in an osmotically buffered media with lysozyme (Imanaka et al., 1982). Polyethylene glycol enables transfer of the plasmid into the protoplast, and transformants are grown on selective media (Imanaka et al., 1982; Liao et al., 1986; Wu and Welker, 1989). This protocol was originally developed for *B. subtilis* and later adapted to *G. stearothermophilus*. This procedure was used consistently for over 10 years because high

transformation frequencies can be attained (Table 1-1) (Imanaka et al., 1982). This method is inconvenient, however, because cells must be treated for each new round of transformation, and protoplasts require a 24-48 hr regeneration step. Additionally, the method is unreliable because the protoplasts are fragile due to the removal of the peptidoglycan layer. This method has only been successful in *G. stearothermophilus* to date.

Electroporation is a much easier and time-saving method when compared with protoplast preparation for transformation, but electroporation in Gram-positive organisms is much more complicated than in their Gram-negative counterparts due to the nature of the thick peptidoglycan cell wall in Gram-positive bacteria. Electroporation is the action or process of introducing DNA or chromosomes into bacteria or other cells using a pulse of electricity to briefly open the pores in the cell membranes. Many of the electroporation protocols developed for *Geobacillus* spp. come from protocols developed for *Bacillus* spp. in the 1990s. The drawback to the electroporation protocols that exist for *Geobacillus* spp. is that transformation efficiencies tend to be lower and are greatly reduced with increasing plasmid sizes. For example, the transformation frequency of a commonly used shuttle plasmid, pUCG18 (6331 bp), into *G. thermoglucosidaisius* DL44 was 9.8×10^3 CFU of transformants/µg of plasmid DNA (Taylor et al., 2008). Electroporation has replaced protoplast preparation as a preferred method of transformation because the ease of use of electroporation is much simpler than the methods involved with protoplast formation, but only in cases where transformation efficiency is not crucial (Hussein et al., 2015). The ease of long-term storage of electrocompetent cell preparations is beneficial as well. Transformation by electroporation involves three steps: preparation of electrocompetent cells, electroporation of exogenous DNA into those recipient cells, and regeneration with

selection of recipient cells. The preparation stage is typically accomplished by allowing cells to grow to the appropriate stage in log growth, centrifuging the cells to remove growth medium, and then washing the cells in electrotransformation medium (ETM) at least a couple times. Studies in *B. subtilis* have shown that the appropriate stage in log growth (varies between early to late stage) for electrocompetence is strain-dependent (Lu et al., 2012).

Variabilities in electroporation conditions and electrocompetent cell preparation techniques demonstrate the need for further exploration into optimized conditions for transformations into *Geobacillus* spp. Once optimized for a species, electroporation is highly reproducible and can attain high efficiencies, reaching up to 10^9 CFU per µg DNA in *E. coli* and 10^6 CFU per µg DNA in *B. subtilis* (Aune and Aachmann, 2010; Lu et al., 2012; Xue et al., 1999). Only one example of optimization for a *Geobacillus* spp. exists- Narumi et al. (1992) reported optimal conditions for electroporation into *Geobacillus thermodenitrificans* K1041, previously known as *Bacillus stearothermophilus.* In this case, the highest electroporation efficiency (10^5 CFU of transformants/µg DNA) was attained when cells were grown to an OD600 of 0.95, a 10% glycerol solution was used as ETM, and 10 kV/cm was used (Narumi et al., 1992). However, this transformation method did not work in other *Geobacillus* strains. As with other Gram-positive bacteria, electroporation could be improved by increasing the field strength (between 4 and 12.5 kV/cm) of the applied electrical current, which in turn can decrease cell viability (Bellevue and Trevors, 1989; Trevors et al., 1992; Xue et al., 1999). One way to increase the cells' tolerance for a higher electrical field strength (up to 25 kV/cm) is to apply a high osmolarity method in which high concentrations of mannose or sucrose are added to the media used for washing the cells, thus increasing the survival rate of electroshocked cells as well as enhanced electroporation efficiencies in

Bacillus spp. (Lu et al., 2012; Xue et al., 1999). This method of supplementing ETM with compatible solutes has been successful in several *Geobacillus* strains, including *G. thermoglucosidasius* DL33, NCIMB 11955 (Cripps et al., 2009) and *G. stearothermophilus* NUB3621R (Kananavičiūtė et al., 2014). An additional consideration when electroporation is used is the size of the plasmid to be transferred. The best transformation efficiency using electroporation was reached using *G. thermoglucosidasius* strain NCIMB 11955, ETM supplemented with 0.5M sorbitol and 0.5M mannitol, a field strength of 25 kV/cm and plasmid pUCG3.8 (with a size of 3.8 kb), which reached 2.8 x 10^6 CFU of transformants/µg of plasmid DNA (Bartosiak-Jenys et al., 2013). As previously noted, there is only one study with one strain of *Geobacillus* that looks at optimization of electroporation; thus, more research is needed on a variety of *Geobacillus* strains as well as media, ETM composition, and electrical field strength for electroporation into a wider range of strains.

A third protocol for DNA uptake in geobacilli is conjugative transfer, which has the advantage that it maintains efficient DNA transfer with increasing plasmid size. Conjugative transfer requires a specific conjugative donor of exogenous DNA for introduction into a recipient (Kananavičiūtė and Čitavičius, 2015). Recently, conjugation using donor *E. coli* cells harboring transfer (*tra*) genes found on helper plasmids (pRK2013 and pUB307) or directly in the chromosome (*E. coli* S-17-1) and recipient *Geobacillus* spp. cells have been performed (Suzuki and Yoshida, 2012; Suzuki et al., 2013). Additionally, the plasmid being transferred must also have an inserted *oriT/mob* genes in order to be conjugated successfully. Donor and recipient cells are mixed and co-incubated at 37 ˚C overnight after they are separately grown to mid-log (OD of 0.5). One challenge with conjugation is distinguishing between donor and recipient cells. This is easily resolved in the last step of the conjugation

from *E. coli* to *Geobacillus* spp. because all cells are plated on selective medium and grown at a temperature that only allows for the growth of the *Geobacillus* cells. Thus, the recipient cells are easily distinguished from the *E. coli* donor cells due to the temperature requirements for each (Suzuki and Yoshida, 2012; Suzuki et al., 2012). Transfer efficiencies of 1.2×10^{-3} exconjugants per recipient *G. kaustophilus* HTA426 have been reported (Suzuki and Yoshida, 2012). In this case, Suzuki and Yoshida (2012) used a host-mimicking system to overcome restriction modification issues in *G. kaustophilus* HTA426 by first propagating plasmid DNA in an *E. coli* host expressing the DNA methyl-transferase characteristics of *G. kaustophilus* HTA426 (Suzuki and Yoshida, 2012). Additionally, Suzuki et al. (2013) created *G. kaustophilus* strain MK72 (a derivative of *G. kaustophilus* HTA426) by deleting the type I and type IV R-M system gene clusters resulting in strain MK244, (Suzuki et al., 2013) which was able to take up the pUCG18T DNA conjugated from *E. coli* DH5α with an efficiency of $1.3 \text{ X } 10^{-5}$ recipient $^{-1}$ in contrast to the wild-type HTA426 that was unable to be transformed (Suzuki et al., 2013). These protocols were breakthroughs because *G. kaustophilus* had previously been shown to be resistant to electroporation (Suzuki et al., 2012; Suzuki et al., 2013).

1.3.2 Positive Selection Markers. There are challenges with using established selection markers and antibiotics when using thermophilic organisms due to the effects of higher temperature on antibiotics. Most commonly used selection markers and antibiotics are thermolabile at temperatures above 60 ˚C (Table 1-2). For example, some of the more commonly used plasmids in *Geobacillus* spp. are pNW33N, pBST22, and pRP9, and all of these use the chloramphenicol acetyltransferase gene (Cam^R) derived from the pC194 plasmid (*Staphylococcus aureus*) (Zeigler, 2001; Liao and Kanikula, 1990; De Rossi et al.,

1991). Chloramphenicol is a bacteriostatic antibiotic, and the Cam^R marker is only functional in *Geobacillus* spp. at temperatures up to 60 °C (Blanchard et al., 2014); however, above 60 °C, both the antibiotic and the Cam^R marker are subject to degradation (Hussein et al., 2015). In addition, transformation is not as efficient at 48 ˚C with chloramphenicol selection when compared to tetracycline, decreasing from 1.4 x 10^5/µg DNA using tetracycline to 7.2 x 10^4/µg DNA using chloramphenicol when plasmid pIH41 is introduced into *G. thermodenitrificans* K1041 by electroporation (Liao and Kanikula, 1990; Narumi et al., 1992). Kanamycin is one of the more stable antibiotics at higher temperatures, which makes it one of the more popular antibiotics used when selecting for *Geobacillus* spp.

Table 1-2

Degradation constants (Avg. *k* value (SD)[a]) of commonly-used antibiotics in *Geobacillus* spp. incubated in anaerobic medium at pH 7.3. The degradation constant (*k*) = 1/half-life. Half-lives were determined in a modified MIC test with *Escherichia coli*, *Staphylococcus aureus*, and *Bacillus megaterium* as indicator strains. (Adapted from Peteranderl et al., 1990)

Antibiotic	72˚C	50˚C
Kanamycin	0 (0)	ND
Chloramphenicol	-0.59 (0.06)	+0.22 (0.04)
Streptomycin	-1.34 (0.06)	-0.45 (0.10)
Ampicillin	-7.26 (0.81)	-1.56 (0.17)

[a] The change of activity over time was determined separately for all possible time intervals in each experiment. Negative values indicate a loss of potency and positive values indicate an increase of potency of the antibiotic during incubation; ND indicates that the compound did not show observable decay at 72°C and was therefore not tested at the same pH at 50°C.

1.3.3 Strains of *Geobacillus* Most Commonly Used for Transformation. Three strains of *Geobacillus* spp. utilizing three separate DNA uptake methods (see 1.3.1) are commonly employed: *Geobacillus stearothermophilus* NUB36 (protoplast), *Geobacillus thermoglucosidasius* NCIMB 11955 (electroporation), and *Geobacillus kaustophilus* HTA426 (conjugation). Table 1-1 outlines all three of these strains as well as the method of

transformation, plasmids used, and transformation efficiency. *G. stearothermophilus* NUB36 strains grow best at 65 °C, with a range of growth temperatures between 40 °C to 75 °C (Chen et al., 1986; Wu and Welker, 1991). A number of mutant strains of *G. stearothermophilus* have been developed, including NUB3621, which is rifampin resistant and lacks a restriction modification system to increase transformation efficiencies (Wu and Welker, 1989; Chen et al., 1986). *Geobacillus thermoglucosidasius* NCIMB 11955 (also referred to as DSM 2542^{T}) is the type strain and has been extensively studied because of its potential with second-generation biofuels production (Cripps et al., 2009). An electroporation protocol has been developed (Taylor et al., 2008) for this strain, which is lacking with other strains of interest in this genus. There has also been the development of an integrative vector system (Cripps et al., 2009) and an expression system in *G. thermoglucosidasius* NCIMB 11955 (Bartosiak-Jentys et al., 2013; Lin et al., 2014). *Geobacillus kaustophilus* HTA426 is a promising host for genetic manipulation because it can grow at temperatures between 42 and 74 °C and in NaCl concentrations up to 3%, it grows as fast as *E. coli*, and it reaches cell densities on par with *E. coli* (Suzuki and Yoshida, 2012). A conjugation protocol was developed by Suzuki et al. (2012, 2013, 2014) using conjugative transfer from an *E. coli* to a *G. kaustophilus* HTA426 recipient. Suzuki et al. (2013) have developed other tools for *G. kaustophilus* HTA426, including the development of an auxotrophic strain (see 1.3.5), the creation of gene libraries, and the classification of thermophilic enzymes by *in vivo* functional screening as well.

1.3.4 Plasmid Vectors. Genetic manipulation efforts in *Geobacillus* spp. have been dependent on the classification and development of plasmids that (1) have a compatible origin of replication for growth in *Geobacillus* spp. and (2) have selection markers for

plasmid maintenance across multiple generations (Hussein et al., 2015). Most plasmids have been constructed for propagation and maintenance in both *E. coli* and *Geobacillus* spp. as well as a thermostable selection marker for reliable selection at higher temperatures (Kananavičiūtė et al., 2014). This combination of features functional in *E. coli* as well as *Geobacillus* spp. typically results in larger plasmids, which can be a challenge if electroporation is used for DNA uptake. Plasmids commonly used in *Geobacillus* spp. are listed in Table 1-3. The segregational and structural stability, copy number, and host range of the vector are largely dependent on which type of mechanism of replication is present (Kananavičiūtė et al., 2014). Two mechanisms of plasmid replication are found in vectors: theta-type and rolling circle (RC). Vectors with theta-replication use similar mechanisms as bacterial chromosomes (Kananavičiūtė et al., 2014). This mechanism gives them more structural stability and allows for insertions of larger DNA inserts (Kiewiet et al., 1993).

Table 1-3

Vectors used for the transformation of *Geobacillus* spp. (Adapted from Kananavičiūtė et al. 2014 and Reeve et al., 2016).

Name of Vector	**Components of Vector**		**Size, kb**	**Remarks**	**References**
	***E. coli* vector (or its elements)**	**Thermostable Marker and *Geobacillus* origin**			
pBST22	pUC19	Kan^R (*TK101*), Cam^R (from pC194), pBST1[a] origin	7.6	Stable up to 68 ˚C with antibiotic selection. No functional blue/white selection system	Liao and Kanikula (1990)
pUCG18	pUC19	Kan^R (*TK101*), pBST1[a] origin	6.3	Stable up to 68 ˚C with antibiotic selection.	Taylor et al. (2008)
pUCG3.8	pUC18 (without Amp^R)	Kan^R (*TK101*), pBST1[a] origin	3.8	Compared to pUCG18, enhanced transformation efficiency	Bartosiak-Jentys et al. (2013)
pG1K	ColE1 origin	Kan^R (*TK101*), pBST1[a] origin	3.7	Compared to pUCG3.8, enhanced transformation efficiency	Reeve et al. (2016)
pG1AK	ColE1 origin, *bla* (Amp^R)	Kan^R (*TK101*), pBST1[a] origin, *sfGFP* under constitutive promoter	4.7	pG1K with *sfGFP* reporter gene (expressed in both *E. coli* and *Geobacillus* spp. Amp^R marker for higher transformation efficiency in *E. coli*. Larger size leads to reduced transformation efficiency in *Geobacillus* spp.	Reeve et al. (2016)
pSTE33	pUC19	Kan^R (*TK101*), pSTK1[b] origin	5.7	Stable at 67 ˚C even without selection	Narumi et al. (1993)
pRP9	pUC19 (only MCS)	Cam^R (from pC194), pBC1[c] origin	2.9	Low transformation efficiencies and/or stability in *E. coli*	De Rossi et al. (1994), Taylor et al. (2008)
pNW33N	pUC19	Cam^R (from pC194), pBC1[c] origin	3.9	Stable at 60 ˚C	Zeigler (2001)
pTMO31	pUC19, *bla* (Amp^R)	Kan^R (*TK101*), pUB110[d] origin	5.1	Temperature sensitive suicide vector	Cripps et al. (2009)

[a] pBST1 origin from *G. stearothermophilus* NRTL 1102; θ-type
[b] pSTK1 origin from *G. stearothermophilus* TK015; RC-type
[c] pBC1 origin from *B. coagulans*; RC-type; broad host range
[d] pUB110 origin from *Staphylococcus aureus*; RC-type; Kan^R

A primary source of plasmids with stability at the higher temperatures required for growth in *Geobacillus* spp. come from geobacilli themselves. However, only a small number of plasmids have been isolated, sequenced, and characterized from *Geobacillus* spp. to date (Kananavičiūtė and Čitavičius, 2015). Plasmids pSTK1, pGS18, pGTG5, and pGTD7 have all been isolated, sequenced, and characterized (Kananavičiūtė et al., 2014; Nakayama et al., 1993; Stuknyte et al., 2008). Plasmids pBST1 (Taylor et al., 2008) and pTB19 (Oskam et al., 1991; Van der Lelie et al., 1989) have been partially sequenced. Tools have been further developed that incorporate origins of replication suitable for propagation in *Geobacillus* spp. from some of these naturally occurring native plasmids. pGS18 (Stuknyte et al., 2008) and pBST1 (Taylor et al., 2008) both use theta-based replication mechanisms. pBST1 is a very large plasmid (over 80 kb) isolated from *G. stearothermophilus* NRRL 1102 and can be found in cells growing at temperatures up to 70 ˚C (Liao et al., 1986). The replicon from pBST1 is commonly found in many constructs for *Geobacillus* spp. (Table1-3). For instance, the pBST1 replicon is found in the construction of pUCG18, a popular choice for genetic manipulation in geobacilli (Taylor et al., 2008). Likewise, pBST22 was developed (Liao et al., 1986) from pBST1 and has been used for transformation of *G. stearothermophilus* protoplasts (Liao and Kanikula, 1990). As previously mentioned, kanamycin has the highest thermostability of the commonly used antibiotics in *Geobacillus* spp. (Peteranderl et al., 1990). Liao et al. (1986) selected a variant of the kanamycin nucleotidyltransferase gene (KNT-ase), granting resistance to kanamycin (which is bactericidal) in temperatures of up to 70 ˚C, making kanamycin an even more attractive option for selection at temperatures of 60 ˚C and above. This group used a mutator strain of *E. coli* (mutD5) to create point mutations into the mesophilic KNT-ase resistance marker from pUB110 (Liao et al., 1986). This

thermostable marker (known as TK101) has been shown to function in at 37 ˚C (in *E. coli*) and up to 70 ˚C (in *Geobacillus* spp.) (Bartosiak-Jentys et al., 2013; Taylor et al., 2008). The transformation efficiencies using the TK101 selection system is much lower in *E. coli*, however (Narumi et al., 1993; Kananavičiūtė et al., 2014; Reeve et al., 2016). The pBST22 plasmid was the first to utilize a thermostable kanamycin marker (TK101) (Liao and Kanikula, 1990), but pBST22 lacks a suitable multiple cloning site (MCS) and blue-white screening capabilities in *E. coli*. pUCG18 was developed to address the issues of having an accessible MCS while still retaining the TK101 thermostable resistance marker (Taylor et al., 2008). Furthermore, Bartosiak-Jentys et al. (2013) used pUCG18 as a backbone and developed pUCG3.8 to decrease the size of pUCG18 from 6 kb to 3.8 kb, effectively increasing the transformation efficiency of the plasmid into *G. thermoglucosidasius* NCIMB 11955. One problem with pUCG18, pBST22, and pUCG3.8 is that they are not maintained by the host cells under nonselective conditions (Liao and Kanikula, 1990; Suzuki and Yoshida, 2012).

Reeve et al. (2016) created a toolkit for genetically engineering *G. thermoglucosidasius* that consists of modular components that can be swapped in and out to establish systematized plasmids with greater flexibility and transformation efficiencies. Two plasmids of differing sizes have been developed for use in various *Geobacillus* spp.: pG1AK and pG1K (Reeve et al., 2016). Each of these plasmids are shuttle vectors with the ColE1 origin of replication and *bla* gene (amp^R) for propagation in *E. coli* because it results in higher transformation efficiencies (Reeve et al., 2016). For propagation in *G. thermoglucosidasius* pG1K and pG1AK contain the repBST1 origin from pBST1 (Liao and Kanikula, 1990) and the TK101 thermostable kanamycin resistance gene. Efficiencies of

transformation by electroporation into *G. thermoglucosidasius* are shown in Table 1-4 (adapted from Reeve et al., 2016). Plasmid pG1K was also shown to transform other *Geobacillus* species (*G. thermoleovorans, G. kaustophilus,* and *G. thermodenitrificans*) with efficiencies between 10^2 and 10^3 CFU/µg DNA (Reeve et al., 2016). A transcriptional terminator after the kanR gene was added in pG1AK and pG1K, which was absent in other vectors that have the TK101 marker (Reeve et al., 2016). Transformation efficiencies using the TK101 marker in *Geobacillus* spp. were significantly higher than the chloramphenicol marker used in other vectors, likely because of the higher thermostability of the TK101 marker (Reeve et al., 2016). Copy numbers (estimated by qPCR) in *G. thermoglucosidasius* were approximately 160 for repBST1 at 55 ˚C (Reeve et al., 2016). Plasmids containing the pBST1 replicon were stably maintained for 5 days growth under antibiotic selection (Reeve et al., 2016).

Table 1-4

Efficiency of transformation by electroporation of *Geobacillus* plasmids compared to the most efficient existing vectors, pUCG18 and pUCG3.8. Percent error shown is the standard deviation of three biological replicates (adapted from Reeve et al., 2016).

Plasmid	**Size/kpb**	**Antibiotic Selection**	**CFU/µg DNA in *G.thermoglucosidasius***	**CFU/µg DNA in *E. coli***
pUCG18	6.3	Kan Amp	$4.9x10^3$ +/-22% using Kan	$1.6x10^6$ +/-7% using Kan $4.4 x 10^6$ +/- 19% using Amp
pUCG3.8	3.8	Kan	$5.2x10^3$ +/-27% using Kan	$1.9x10^6$ +/-6% using Kan
pG1K	3.7	Kan	$5.3x10^4$ +/-23% using Kan	$3.4x10^6$ +/-7% using Kan
pG1AK	4.7	Kan Amp	$5.8x10^3$ +/-49% using Kan	$3.0x10^6$ +/-9% using Kan $7.4x10^6$ +/-9% using Amp

Vectors that have rolling circle origins of replication typically have a more extensive host range and higher copy number when compared to vectors with theta-replicating mechanisms, leading to continuing attempts to isolate and characterize novel plasmids from *Geobacillus* spp. that utilize RC replication (Kananavičiūtė et al., 2014). Higher copy

numbers typically means a higher level of production of recombinant proteins (Suzuki et al., 2013); however, RC replicons are less stable than theta-based plasmids and are limited in the size of insert that they can carry (Kiewiet et al., 1993). *Geobacillus* sp. plasmids pSTK1 (Nakayama et al., 1993), pGTG5, and pGTD7 (Kananavičiūtė et al., 2014) utilize the RC-type replicon. The replicon from pSK1 has been commonly used for construction of shuttle vectors with RC-type mechanisms in *Geobacillus* spp. (Table 1-3). This is the replicon that was used for the construction of pSTE33. This shuttle vector has shown high structural stability in *G. thermodenitrificans* K1041 (Narumi et al., 1993) and *G. kaustophilus* HTA426 (Suzuki and Yoshida, 2012). Another commonly used vector in *Geobacillus* spp., pNW33N, is derived from pBC1 which was isolated from a thermophilic variant of *B. coagulans* and is a small RC-plasmid (Zeigler, 2001). A drawback to pNW33N, however, is that it uses chloramphenicol as a selection marker, which is only thermostable up to 60 ˚C as previously mentioned (Taylor et al., 2008).

The stable integration of genes into the bacterial chromosome can sometimes be more desirable than maintaining an extrachromosomal plasmid. Methods for integrating genetic components involve utilizing transitory DNA uptake into the genome of the host in a site-specific manner via homologous recombination (Hussein et al., 2015). Integrative vectors have been constructed for use in *Geobacillus* spp. that either have no replicative stability (pGAM plasmids) or are only maintained under certain temperature (pTM031) or selective pressure (pSTE12) (Hussein et al., 2015). The vectors with no replicative stability are known as “suicide vectors.” Integrative vectors lacking a replicon that is functional in *Geobacillus* spp. (pGAM46 and pGAM47) have recently been introduced to *Geobacillus kaustophilus*

HTA426 using the increased transformation efficiencies afforded by conjugative transfer (Suzuki et al., 2012).

Plasmids with replicons that are only maintained (pSTE12 and pTM031) in *Geobacillus* spp. under certain conditions have also been developed for integrative genetic manipulation. The pSTE12 suicide vector replaced a gene via homologous recombination in *G. stearothermophilus* NUB3621-R at a specific locus (Couñago and Shamoo, 2005). Selective pressure must be applied for this plasmid to be maintained, which allows for the loss of the non-integrated plasmid and enables growth of only those transformants that have been incorporated into the host genome (Couñago and Shamoo, 2005). Vector pTM031 utilized the mesophilic replicon and kanamycin resistance marker from pUB110, a plasmid capable of replication in a variety of Gram-positive bacteria, including several different *Geobacillus* spp. (Cripps et al., 2009; Gryczan et al., 1978). However, this plasmid must be used at temperatures lower than 55 ˚C because the kanamycin marker is ineffective at higher temperatures (Imanaka et al., 1982; Matsumura et al., 1984). Thus, pTM031 can be eliminated in *Geobacillus* spp. at temperatures above 55 ˚C (Cripps et al., 2009). Marker-free gene deletion/insertion strategies have been used to knock out the *ldh* gene in *G. thermoglucosidasius* NCIMB 11955, which is significant because the same vector with the same marker can be used for further genetic manipulation after successful targeted homologous recombination has been utilized (Cripps et al., 2009).

1.3.5 Antibiotic-free Selection Systems. Selection systems that function without the use of antibiotics are often more suitable for industrial uses due to high scale-up costs. Auxotrophic selectable markers encode for essential genes in a metabolic pathway that are lacking in a mutant strain (Kananavičiūtė et al., 2015). It is possible to complement these

deficient genes to create a selection system based on auxotrophy. Markers for the synthesis of certain amino acids and nucleotides have been described, including genes responsible for thymidine, tryptophan, and uracil synthesis in a variety of microorganisms (Inoue and Sako, 2013; Taylor et al., 2011). Thermophilic marker-free systems are lacking (especially for *Geobacillus* spp.), but there is one key example that has been developed in these organisms: a prokaryotic homolog of the eukaryotic ura3-FOA (5-fluoorootic acid) that was developed in *Geobacillus kaustophilus* HTA426 (Suzuki et al., 2012). The bacterial equivalent of ura3-FOA is *pyrF*, responsible for orotidine 5'-phosphate decarboxylase. This enzyme handles synthesis and metabolism of pyrimidines (UMP, UDP, UTP) as well as converting 5-FOA into toxic metabolites (Hussein et al., 2015). The conversion of 5-FOA into metabolites that are toxic to the host cell was useful for the development of a counterselection system allowing for selection of a *pyrF* mutant strain of *G. kaustophilus* HTA426 (Hirokazu, 2012; Suzuki et al., 2012; Suzuki et al., 2013; Suzuki et al., 2013). The pGAM vectors use the *pyrF* counterselection system and are conjugated into the mutant *G. kaustophilus* strain MK72 (*ΔpyrF ΔpyrR*) that is uracil-auxotrophic and also resistant to 5-FOA (Suzuki et al., 2012). In this system, a plasmid-borne copy of *pyrF* complements the *pyrF* knockout in MK72, resulting in mutants that are prototrophic for uracil. A second crossover event occurs after multiple passages that leads to the loss of the supplemented *pyrF* gene (Suzuki et al., 2012). This results in mutants that are uracil-auxotrophic and 5-FOA resistant when uracil is present while being 5-FOA sensitive when uracil is absent (Boeke et al., 1984; Suzuki et al., 2012). This process has been used for chromosomal integration and heterologous expression in *G. kaustophilus* of *bgaB* (encoding β-galactosidase) in *G. stearothermophilus* IAM11011 and *amyE* (encoding α-amylase) in *G. stearothermophilus* CU21 (Suzuki et al., 2012). However,

half of the resulting uracil-prototrophic colonies were false positives (5-FOA sensitive when uracil was not present) (Suzuki et al., 2012). The *pyrF* system is the only auxotrophic complementation system that has been published for *Geobacillus* spp. Recently, another auxotrophic complementation system has been developed in a mutant strain of *Geobacillus thermoleovorans* DSM 5366 based on complementation of a deficient *thyA* (thymidylate synthase for the synthesis of thymine) gene (Marks and Hamilton, unpublished results). In this case, mutant strains were unable to grow in minimal media without added thymine while transconjugants were able to grow with a complement *thyA* gene from GBK1, a bacteriophage that infects *G. kaustophilus* (Marks and Hamilton, unpublished results). These systems are promising as many mutant auxotrophic strains of *G. stearothermophilus* NUB36 have been characterized, so there is potential for the development of other antibiotic-free selection systems in other *Geobacillus* spp. (Vallier and Welker, 1990).

1.3.6 Reporter Genes. Noncoding elements like promoters and ribosome-binding sites play a large role in the expression of genes. The efficient expression of genes *in vivo* has led to a rapid expansion and characterization of promoters, particularly inducible promoters for conditional expression. Reporter genes that are readily assayed are also very useful when looking at promoter strength. Recently, a thermostable variant of green fluorescent protein (GFP) designated superfolder GFP (sfGFP) (Pédelacq et al., 2006) has been expressed in thermophilic organisms, including some *Geobacillus* spp. and *Thermus* spp. (Blanchard et al., 2014). sfGFP shows strong intensity and thermostability, maintaining strong fluorescence at temperatures up to 70 ˚C *in vivo* (Lesnik et al., 2001). Additionally, sfGFP is functionally expressed by *E. coli* at 37 ˚C which adds to its utility as a reporter gene in shuttle plasmids (Reeve et al., 2016). Unfortunately, sfGFP cannot be used in environments in which oxygen

is absent (i.e. fermentative conditions) because the expression of the fluorescent chromophore is reliant upon molecular oxygen (Hussein et al., 2015). Alternatively, *G. stearothermophilus* DSM 6285 *pheB* was developed as a reporter gene in geobacilli undergoing fermentation (Bartosiak-Jentys et al., 2012). The gene codes for a thermostable catechol 2,3-dioxygenase enzyme that forms a yellow-colored 2-hydroxymuconic semialdehyde (detectable at an absorbance of 375 nm) when catechol was added (Bartosiak-Jentys et al., 2012). Other carbohydrate metabolism enzymes (α-amylase, β-galactosidase, and α-galactosidase) have also been used as reporter genes in *Geobacillus* spp. (Blanchard et al., 2014; Lin et al., 2014; Suzuki et al., 2012).

1.3.7 Recombinant Gene Expression. Homologous gene expression is dependent on certain noncoding elements, including promoters that drive expression in the host, a suitable ribosome binding site (RBS) upstream of the start codon, and a downstream transcriptional terminator (Kananavičiūtė et al., 2015). *Geobacillus* spp. promoters are under development but are still lacking somewhat. A strong constitutive promoter in *Geobacillus kaustophilus* HTA426, P_{sigA}, was discovered upstream of two constitutively-expressed housekeeping genes (Suzuki et al., 2012) and has been utilized in β-galactosidase expression assays. Another strong constitutive promoter in *G. stearothermophilus* NUB3621 for ribonuclease H III, P_{RHIII}, has been tested for expression using the *sfGFP* reporter gene (Blanchard et al., 2014). P_{ldh}, a promoter found in *G. stearothermophilus* NCA1503 and *G. thermodenitrificans* DSM465^T for lactate dehydrogenase, has been used to drive ethanol and isobutanol production in *G. thermoglucosidasius* (Cripps et al., 2009; Lin et al., 2014). It was later discovered that the P_{ldh} promoter is potentially induced in the transition between aerobic and anaerobic growth and drives greater gene expression when oxygen is limited (Bartosiak-

Jentys et al., 2013). The P_{ldh} promoter does not lead to expression in *E. coli* (Reeve et al., 2016). Reeve et al. (2016) include the *G. thermoglucosidasius* P_{Rpls} constitutive promoter in their modular plasmid set for *Geobacillus* spp. P_{Rpls} promotes transcription of the ribosomal protein RplS, observed in other bacilli to be strongly expressed in all growth conditions. Reeve et al. (2016) showed that P_{Rpls} exhibited strong expression of sfGFP in both *G. thermoglucosidasius* and *E. coli*. Expression levels in *G. thermoglucosidasius* were found to be higher than the commonly used P_{ldh} promoter (Reeve et al., 2016).

There has been a recent emphasis on inducible promoters that can be utilized under certain conditions, such as anaerobic growth during fermentation processes. Promoters inducible under controlled conditions are required for expression of heterologous proteins that can be toxic to the host cell (Hussein et al., 2015). Several *G. kaustophilus* HTA426 inducible promoters have recently been tested, including P_{gk704} (maltose induction), P_{gk1859} (lactose induction), P_{gk1894} (myoinositol induction), and P_{gk2150} (D-galactose induction) (Suzuki et al., 2013). P_{gk704} increased protein expression by over 4-fold with the addition of soluble starch and 12-fold with the addition of maltose, with target protein yields of up to 59 mg per liter being achieved (Suzuki et al., 2013). Some proteins that are insoluble in *E. coli* could be successfully produced in *Geobacillus* spp. using this approach (Suzuki et al., 2013). *pheB* expression was enhanced in the presence of cellobiose by $P_{\beta glu}$, which comes from the *G. thermoglucosidasius* NCIMB 11955 phosphotransferase system operon (cellobiose-specific); however, glucose and xylose also activated this inducible promoter as well (Bartosiak-Jentys et al., 2013). P_{surP}, from the sucrose-utilization operon in *G. stearothermophilus* NUB3621, increased α-galactosidase expression 5-fold when sucrose was added, although other sugars were not tested for induction (Blanchard et al., 2014).

1.3.8 Protein Secretion. The secretion of proteins from the cell can be desirable due to the ease of downstream bioprocessing of secreted proteins. Signal peptides that are recognized by the host are required for recombinant protein secretion from the cell. A thorough understanding of the extracellular secretion mechanisms of hydrolytic enzymes is required to increase the production capability of *Geobacillus* spp. for commercially useful proteins like lipases and amylases. However, there have only been a few studies to date that have looked at secretory expression of heterologous proteins (Bartosiak-Jentys et al., 2013; Suzuki et al., 2013). N-terminal signal peptides of variable length (as well as amino acid sequences) are what signal a cell to secrete a specific protein (Hussein et al., 2015). Suzuki et al. (2013) used native signal sequences in *G. kaustophilus* HTA426 to express and secrete a *G. stearothermophilus* α-amylase as well as a truncated-cellulase from *Pyrococcus horikoshii.* In another successful example of protein secretion in *Geobacillus* spp., Bartosiak-Jentys et al. (2013) were able to use a signal peptide from an *G. thermoglucosidasius* NCIMB 11955 endo-xylanase gene to secrete a *T. maritima* glycosyl hydrolase (Bartosiak-Jentys et al., 2013). However, the ideal sequence peptide for one protein can be ineffective for the secretion of another; thus, screening of multiple signal peptides in *Geobacillus* spp. is required for the proficient secretion of diverse proteins (Brockmeier et al., 2006).

1.3.9 Mining Thermophilic Bacteriophage for Tools. Historically, *E. coli* phage such as λ and T7 have been sources of many useful molecular biology tools, most notably with the development of expression systems using λ promoters P_L and P_R (Elvin et al., 1990) and the T7 promoter (Studier et al., 1986; Date et al., 1990). These promoters were developed to promote high-level overproduction of gene products in *E. coli* (Elvin et al., 1990). Integration tools have also been developed around the λ attB attachment site in *E. coli*

(Diederich et al., 1992). A recently discovered lysogenic phage that infects *G. kaustophilus* ATCC 8005 has been utilized in much the same way. A suicide vector, pTM2, has been integrated into *G. thermoglucosidasius* NCIMB 11955 (Marks and Hamilton, unpublished results). Plasmid pTM6, a suicide vector that expresses a phage-derived thymidylate synthase and can be integrated into a *thyA*- mutant strain of *G. thermoleovorans*, has also been developed, thus avoiding the need for antibiotic selection (e.g. kanamycin; Marks and Hamilton, unpublished results).

Recombineering is another useful tool for genetic manipulation that was first explored in *E. coli.* The traditional strategy for making recombinant DNA relies upon using restriction endonucleases that recognize and cleave at specific sites. DNA ligase then facilitates the joining of DNA strands to make recombinant DNA. However, the repeated occurrence of restriction sites in larger pieces of DNA limits the size of engineered DNA fragments to less than 20 kb (Zhang et al., 1998). With large DNA molecules, even restriction endonucleases that have rare cleavage sites have many occurrences that make cloning complicated. New systems of DNA engineering have been developed to circumvent the issues associated with using restriction endonucleases. One of these methods is recombineering (recombination-mediated genetic manipulation), which is the precise exchange of genetic information between two homologous DNA molecules through areas of identical sequence (Court et al., 2002). Synthesis of single- and double-stranded DNA oligonucleotides that provide homology to the host is required *in vitro* for targeting recombineering events (Court et al., 2002). Typically, single-stranded oligonucleotides or double-stranded PCR products generated with ~50-bp ends of homology flank the region that is to be inserted into the host. Recombineering is built around the premise of using

bacteriophage proteins that mediate recombination specifically around regions of homology. The bacteriophage λ Red system and the Rac prophage RecET system have been exploited in *E. coli* for homologous recombineering (Zhang et al., 1998; Lee et al., 2001; Murphy et al., 1998; Muyrers et al., 1999; Yu et al., 2000). Systems have been developed in additional bacterial hosts as well, including *M. tuberculosis* (van Kessel and Hatfull, 2007) and *B. subtilis* (Sun et al., 2015). Recently sequenced phage GBK2 (Marks and Hamilton, 2014), which infects *G. kaustophilus* HTA426, contains a homolog to Bacillus phage SPP1 protein GP35, a RecT-type recombinase, which is the protein that Sun et al. (2015) used to develop their recombineering system in *B. subtilis*.

Phage tools that have been developed in other model organisms could provide genetic manipulation mechanisms in *Geobacillus* spp. as well. There have been a few prophage sequences that have been identified in *G. kaustophilus*, *G. thermoleovorans*, and *G. thermodenitrificans* (Hussein et al., 2015); however, more work is needed to identify novel phage that infect *Geobacillus* spp. in an effort to mine potential tools for genetic manipulation.

1.3.10 Clustered Regularly InterSpaced Repeats (CRISPR). CRISPR and the CRISPR-associated (Cas) proteins provide bacteria and other prokaryotes with a kind of adaptive immunity against invading bacteriophages (Brouns et al., 2008; Barrangou et al., 2007; Wright et al., 2016; Mohanraju et al., 2016). CRISPR systems initiate targeted double-stranded breaks in DNA based on ribonucleoprotein complexes (RNP) formed by a guide RNA and a single Cas endonuclease (Mougiakos et al., 2017). CRISPR systems have recently been developed as genetic manipulation tools in several bacteria, (Mougiakos et al., 2016; Yan et al., 2017; Jiang et al., 2017) mammalian (Komor et al., 2017; Zetsche et al.,

2016), plant (Puchta, 2017; Tang et al., 2017), and fruit fly (Xu et al., 2015) models. Cas9 from *Streptococcus pyogenes* (known as SpCas9) is the most widely used Cas9 for genome editing (Mougiakos et al., 2017). Success of the SpCas9 system has driven development of CRISPR/Cas systems in other organisms, including *Geobacillus* spp.

While several species of *Geobacillus* encode for a CRISPR-associated helicase (*G. kaustophilus* HTA426, *G. thermoleovorans* CCB_US.3_UF5, *G. thermodenitrificans* NG80-2), the highest number of CRISPR motifs is found in the *G. thermoglucosidasius* C56-YS93 genome, with over 100 CRISPR-associated proteins characterized (Hussein et al., 2015). *G. thermoglucosidasius* has no known prophage sequences, presumably because of the many CRISPR-associated sequences and Cas genes found in the genome (Hussein et al. 2015). Recently, a Cas9 protein (ThermoCas9) from *G. thermodenitrificans* T12 was used for the deletion of a specific gene and silencing of transcription in *B. smithii* at 55 ˚C, which is a temperature at which SpCas9 cannot function (Mougiakos et al., 2017). ThermoCas9 has been found to be active between 20 and 70 ˚C, a much wider range than SpCas9 (Mougiakos et al., 2017). ThermoCas9 also shows more fidelity as it tolerates fewer spacer-protospacer mismatches than its SpCas9 counterpart (Mougiakos et al., 2017). Another thermostable Cas9 (GeoCas9) protein from *G. stearothermophilus* was developed as an effective tool for genome editing in both mesophilic and mammalian genomes (when delivered as a ribonucleoprotein complex) at temperatures up to 70 ˚C (Harrington et al., 2017). The GeoCas9 protein catalyzes RNA-guided DNA cleavage and was found to have increased lifetime in human plasma when compared with SpCas9 (Harrington et al., 2017).

1.4 Conclusion

Thermophilic bacteria are microorganisms capable of growth at a wide range of elevated temperatures. Growth at elevated temperatures offers advantages for certain industrial processes, including a reduction in issues with mesophilic contamination, the potential for a continuous process whereby volatile products can be separated from the cells, and a reduction in cooling costs. Additionally, thermophiles are a great source of thermostable enzymes, including DNA polymerases, lipases, amylases, and reverse transcriptases. *Geobacillus* spp. are Gram-positive, facultative anaerobic thermophiles that are capable of growth on several different C-5 and C-6 sugars, including cellulosic precursors. They also form spores, which can be beneficial for long-term storage of these organisms because it allows for survival under harsh conditions. However, there are a lack of genetic manipulation tools developed for *Geobacillus* spp. due in part to challenges with DNA transfer and the lack of reliable thermostable antibiotics/antibiotic resistance markers at the higher temperatures required for growth of these organisms. Most work in *Geobacillus* spp. has centered on *G. stearothermophilus, G. thermoglucosidasius,* and *G. kaustophilus* species. Current tools have focused on generating shuttle vectors that can work in *E. coli* and these *Geobacillus* spp., including some suicide vectors that integrate into *Geobacillus* spp. hosts. Recombinant protein expression using promoters (both constitutive and inducible), ribosome binding site characterization, and protein secretion systems are also being explored but are lacking. Tools from alternative sources (like bacteriophages and CRISPR) are just starting to be investigated for use in *Geobacillus* spp.

This work describes the annotation and characterization of two novel bacteriophage (named GBK1 and GBK2) that infect *Geobacillus kaustophilus* ATCC 8005. Sequencing and

subsequent annotation of GBK1 and GBK2 ORFs revealed genes that could potentially fill a need regarding expanding the available tools for genetic manipulation of *Geobacillus* spp. Chapter 2 describes GBK2 as a lytic, circularly permuted bacteriophage that has two ORFs of interest. ORF 39 has homology to *recT* and could be used to develop a recombineering system in *Geobacillus* spp. ORF 51 has homology to a thymidylate synthase and could be applicable as an antibiotic-free auxotrophic selection marker. Chapter 3 describes GBK1 as a lysogenic, linear bacteriophage that has two ORFs of interest. Chapter 3 also describes the integration and sequencing of GBK1 into the *G. kaustophilus* type strain (ATCC 8005), which had not been sequenced previously. ORF 37 from GBK1 has homology to a thymidylate synthase which we used to create a functional antibiotic-free auxotrophic selection system in a *thyA*$^{-}$ mutant of *G. thermoleovorans*. ORF 43 has homology to a tyrosine recombinase which we used (along with the *attP* sequence) to construct an integrating suicide vector in *G. thermoglucosidasius* and *G. thermoleovorans*. Both of these tools are explained in Chapter 4. In addition, we look at phage evolutionary relationships between GBK1/GBK2 and closely related mesophilic phages that infect *Bacillus* spp. All bacteriophages and strains used in this study are described in Table 1-5.

Table 1-5

Bacteriophages and *Geobacillus* strains used in this study.

Full Description	Comments
Bacteriophage GBK1	Linear, lysogenic phage, source of *attP/int* and *thyA*
Bacteriophage GBK2	Circularly permuted, lytic phage
Geobacillus kaustophilus ATCC 8005	Type strain, host for GBK1 and GBK2
Geobacillus thermoglucosidasius NCIMB 11955	Type strain, starch hydrolyzing glucosidase activity. Engineered for ethanol production. Can grow on D-xylitol, cellobiose, inositol
Geobacillus thermoleovorans DSM 5366	Type strain, Capable of utilizing hydrocarbons (C13-C20 n-alkanes), acetate, butyrate, pyruvate, cellobiose, glucose, mannose, casein, ribose. Created a *thyA*- mutant strain for marker-free selection

References

Abdel-Fattah, Y., & Gaballa, A. (2008). Identification and over-expression of a thermostable lipase from *Geobacillus thermoleovorans* Toshki in *Escherichia coli*. Microbiol Res, 163(1), 13-20. doi:10.1016/j.micres.2006.02.004

Abdel-Fattah, Y., Soliman, N., Yousef, S., & El-Helow, E. (2012). Application of experimental designs to optimize medium composition for production of thermostable lipase/esterase by *Geobacillus thermodenitrificans* AZ1. Genetic Engineering and Biotechnology, 10, 193-200.

Ansorge-Schumacher, M., & Thum, O. (2013). Immobilised lipases in the cosmetics industry. Chem Soc Rev, 42(15), 6475-6490. doi:10.1039/c3cs35484a

Anya, C. (2013). Purification and characterization of a pH and heat stable esterase from *Geobacillus* sp. TF17. Turkish J of Biochem, 38, 329-336.

Aspeborg, H., Coutinho, P., Wang, Y., Brumer, H., & Henrissat, B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol Biol, 12, 186. doi:10.1186/1471-2148-12-186

Balan, A., Ibrahim, D., Abdul Rahim, R., & Ahmad Rashid, F. (2012). Purification and characterization of a thermostable lipase from *Geobacillus thermodenitrificans* IBRL-nra. Enzyme Res, 2012, 987523. doi:10.1155/2012/987523

Balan, A., Magalinggam, N., Ibrahim, D., & Rahim, R. (2010). Thermostable lipase production by *Geobacillus thermodenitrificans* in a 5-L stirred-tank bioreactor. Internet J of Microbiol, 8, 9.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science, 315(5819), 1709-1712. doi:10.1126/science.1138140

Bartosiak-Jentys, J., Eley, K., & Leak, D. (2012). Application of pheB as a reporter gene for *Geobacillus* spp., enabling qualitative colony screening and quantitative analysis of promoter strength. Appl Environ Microbiol, 78(16), 5945-5947. doi:10.1128/AEM.07944-11

Bartosiak-Jentys, J., Hussein, A., Lewis, C., & Leak, D. (2013). Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. Microbiology, 159(Pt 7), 1267-1275. doi:10.1099/mic.0.066332-0

Berekaa, M., Zaghloul, T., Abdel-Fattah, Y., Saeed, H., & Sifour, M. (2009). Production of a novel glycerol-inducible lipase from thermophilic *Geobacillus stearothermophilus* strain-5. World J of Microbiol and Biotechnol, 25, 287-294.

Blanchard, K., Robic, S., & Matsumura, I. (2014). Transformable facultative thermophile *Geobacillus stearothermophilus* NUB3621 as a host strain for metabolic engineering. Appl Microbiol Biotechnol, 98(15), 6715-6723. doi:10.1007/s00253-014-5746-z

Boeke, J., LaCroute, F., & Fink, G. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Molecular and General Genetics, 197, 345-346.

Bornscheuer, U., Bessler, C., Srinivas, R., & Krishna, S. (2002). Optimizing lipases and related enzymes for efficient application. Trends Biotechnol, 20(10), 433-437.

Brockmeier, U., Caspers, M., Freudl, R., Jockwer, A., Noll, T., & Eggert, T. (2006). Systematic screening of all signal peptides from *Bacillus subtilis*: a powerful strategy

in optimizing heterologous protein secretion in Gram-positive bacteria. J Mol Biol, 362(3), 393-402. doi:10.1016/j.jmb.2006.07.034

Brouns, S., Jore, M., Lundgren, M., Westra, E., Slijkhuis, R., Snijders, A., van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science, 321(5891), 960-964. doi:10.1126/science.1159689

Charbonneau, D., Meddeb-Mouelhi, F., & Beauregard, M. (2010). A novel thermostable carboxylesterase from *Geobacillus thermodenitrificans*: evidence for a new carboxylesterase family. J Biochem, 148(3), 299-308. doi:10.1093/jb/mvq064

Chen, X., Stabnikova, O., Tay, J., Wang, J., & Tay, S. (2004). Thermoactive extracellular proteases of *Geobacillus caldoproteolyticus*, sp. nov., from sewage sludge. Extremophiles, 8(6), 489-498. doi:10.1007/s00792-004-0412-5

Coorevits, A., Dinsdale, A., Halket, G., Lebbe, L., De Vos, P., Van Landschoot, A., & Logan, N. (2012). Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus, G. stearothermophilus, G. jurassicus, G. toebii, G. thermodenitrificans* and *G. thermoglucosidans* (nom. corrig., formerly 'thermoglucosidasius'); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. Int J Syst Evol Microbiol, 62(Pt 7), 1470-1485. doi:10.1099/ijs.0.030346-0

Court, D., Sawitzke, J., & Thomason, L. (2002). Genetic engineering using homologous recombination. Annu Rev Genet, 36, 361-388. doi:10.1146/annurev.genet.36.061102.093104

Couñago, R., Chen, S., & Shamoo, Y. (2006). In vivo molecular evolution reveals biophysical origins of organismal fitness. Mol Cell, 22(4), 441-449. doi:10.1016/j.molcel.2006.04.012

Couñago, R., & Shamoo, Y. (2005). Gene replacement of adenylate kinase in the gram-positive thermophile *Geobacillus stearothermophilus* disrupts adenine nucleotide homeostasis and reduces cell viability. Extremophiles, 9(2), 135-144. doi:10.1007/s00792-004-0428-x

Cripps, R., Eley, K., Leak, D., Rudd, B., Taylor, M., Todd, M., Atkinson, T. (2009). Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. Metab Eng, 11(6), 398-408. doi:10.1016/j.ymben.2009.08.005

Date, T., Tanihara, K., & Numura, N. (1990). Construction of *Escherichia coli* vectors for expression and mutagenesis: synthesis of human c-Myc protein that is initiated at a non-AUG codon in exon 1. Gene, 90(1), 141-144. doi:10.1016/0378-1119(90)90450-6

De Maayer, P., Brumm, P., Mead, D., & Cowan, D. (2014). Comparative analysis of the *Geobacillus hemicellulose* utilization locus reveals a highly variable target for improved hemicellulolysis. BMC Genomics, 15, 836. doi:10.1186/1471-2164-15-836

De Rossi, E., Brigidi, P., Rossi, M., Matteuzzi, D., & Riccardi, G. (1991). Characterization of gram-positive broad host-range plasmids carrying a thermophilic replicon. Res Microbiol, 142(4), 389-396. doi:10.1016/0923-2508(91)90108-m

De Rossi, E., Brigidi, P., Welker, N., Riccardi, G., & Matteuzzi, D. (1994). New shuttle vector for cloning in *Bacillus stearothermophilus*. Res Microbiol, 145(8), 579-583. doi:10.1016/0923-2508(94)90074-4

Diederich, L., Rasmussen, L., & Messer, W. (1992). New cloning vectors for integration in the lambda attachment site attB of the *Escherichia coli* chromosome. Plasmid, 28(1), 14-24. doi:10.1016/0147-619x(92)90032-6

Elvin, C., Thompson, P., Argall, M., Hendry, P., Stamford, N., Lilley, P., & Dixon, N. (1990). Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. Gene, 87(1), 123-126. doi:10.1016/0378-1119(90)90503-j

Ewis, H., Abdelal, A., & Lu, C. (2004). Molecular cloning and characterization of two thermostable carboxyl esterases from *Geobacillus stearothermophilus*. Gene, 329, 187-195. doi:10.1016/j.gene.2003.12.029

Galperin, M., Mekhedov, S., Puigbo, P., Smirnov, S., Wolf, Y., & Rigden, D. (2012). Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. Environ Microbiol, 14(11), 2870-2890. doi:10.1111/j.1462-2920.2012.02841.x

Gryczan, T. J., Contente, S., & Dubnau, D. (1978). Characterization of *Staphylococcus aureus* plasmids introduced by transformation into *Bacillus subtilis*. J Bacteriol, 134(1), 318-329.

Haki, G., & Rakshit, S. (2003). Developments in industrially important thermostable enzymes: a review. Bioresour Technol, 89(1), 17-34.

Hamid, N., Zen, G., Tein, O., Halifah, Y., Saari, N., & Bakar, F. (2003). Screening and identification of extracellular lipase-producing thermophilic bacteria from a Malaysian hot spring. World J of Micriobiol and Biotechnol, 19, 961-968.

Harrington, L., Paez-Espino, D., Staahl, B., Chen, J., Ma, E., Kyrpides, N., & Doudna, J. (2017). A thermostable Cas9 with increased lifetime in human plasma. Nat Commun, 8(1), 1424. doi:10.1038/s41467-017-01408-4

Hawumba, J., Theron, J., & Brözel, V. (2002). Thermophilic protease-producing *Geobacillus* from Buranga hot springs in Western Uganda. Curr Microbiol, 45(2), 144-150. doi:10.1007/s00284-001-0116-3

Hussein, A., Lisowska, B., & Leak, D. (2015). The genus *Geobacillus* and their biotechnological potential. Adv Appl Microbiol, 92, 1-48. doi:10.1016/bs.aambs.2015.03.001

Imanaka, T., Fujii, M., Aramori, I., & Aiba, S. (1982). Transformation of *Bacillus stearothermophilus* with plasmid DNA and characterization of shuttle vector plasmids between *Bacillus stearothermophilus* and *Bacillus subtilis*. J Bacteriol, 149(3), 824-830.

Inoue, T., & Sako, Y. (2013). Host-Vector Systems in Thermophiles. In Thermophilic Microbes in Environmental and Industrial Biotechnology: Springer, Dordrecht.

Iqbal, I., Aftab, M., Afzal, M., Ur-Rehman, A., Aftab, S., Zafar, A., Ul-Haq, I. (2015). Purification and characterization of cloned alkaline protease gene of *Geobacillus stearothermophilus*. J Basic Microbiol, 55(2), 160-171. doi:10.1002/jobm.201400190

Jiang, Y., Qian, F., Yang, J., Liu, Y., Dong, F., Xu, C., Yang, S. (2017). CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*. Nat Commun, 8, 15179. doi:10.1038/ncomms15179

Jiang, Y., Zhou, X., & Chen, Z. (2010). Cloning, expression, and biochemical characterization of a thermostable lipase from *Geobacillus stearothermophilus* JC. World J of Microbiol and Biotechnol, 26, 747-751.

Kananavičiūtė, R., Butaitė, E., & Citavičius, D. (2014). Characterization of two novel plasmids from *Geobacillus* sp. 610 and 1121 strains. Plasmid, 71, 23-31. doi:10.1016/j.plasmid.2013.10.002

Kananavičiūtė, R., & Čitavičius, D. (2015). Genetic engineering of *Geobacillus* spp. J Microbiol Methods, 111, 31-39. doi:10.1016/j.mimet.2015.02.002

Khanniri, E., Bagheripoor-Fallah, N., Sohrabvandi, S., Mortazavian, A., Khosravi-Darani, K., & Mohammadi, R. (2015). Application of liposomes in some dairy products. Critical Reviews in Food Science and Nutrition. doi:http://dx.doi.org/10.1080/10408398.2013.779571

Kiewiet, R., Kok, J., Seegers, J., Venema, G., & Bron, S. (1993). The mode of replication is a major factor in segregational plasmid instability in *Lactococcus lactis*. Appl Environ Microbiol, 59(2), 358-364.

Kim, H., Park, S., Lee, J., & Oh, T. (1998). Gene cloning and characterization of thermostable lipase from *Bacillus stearothermophilus* L1. Biosci Biotechnol Biochem, 62(1), 66-71. doi:10.1271/bbb.62.66

Komor, A. C., Badran, A. H., & Liu, D. R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell, 169(3), 559. doi:10.1016/j.cell.2017.04.005

Kovács, A., Smits, W., Mirończuk, A., & Kuipers, O. (2009). Ubiquitous late competence genes in *Bacillus* species indicate the presence of functional DNA uptake

machineries. Environ Microbiol, 11(8), 1911-1922. doi:10.1111/j.1462-2920.2009.01937.x

Kuriki, T., Okada, S., & Imanaka, T. (1998). New type of pullulanase from *Bacillus stearothermophilus* and molecular cloning and expression of the gene in *Bacillus subtilis*. J of Bacteriology, 170, 1554-1559.

Lee, E., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D., Court, D., Copeland, N. (2001). A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics, 73(1), 56-65. doi:10.1006/geno.2000.6451

Leow, T., Rahman, R., Basri, M., & Salleh, A. (2007). A thermoalkaliphilic lipase of *Geobacillus* sp. T1. Extremophiles, 11(3), 527-535. doi:10.1007/s00792-007-0069-y

Lesnik, E., Sampath, R., Levene, H., Henderson, T., McNeil, J., & Ecker, D. (2001). Prediction of rho-independent transcriptional terminators in *Escherichia coli*. Nucleic Acids Res, 29(17), 3583-3594. doi:10.1093/nar/29.17.3583

Li, H., & Zhang, X. (2005). Characterization of thermostable lipase from thermophilic *Geobacillus* sp. TW1. Protein Expr Purif, 42(1), 153-159. doi:10.1016/j.pep.2005.03.011

Liao, H., & Kanikula, A. (1990). Increased efficiency of transformation of *Bacillus stearothermophilus* by a plasmid carrying a thermostable kanamycin resistance marker. Current Microbiol, 21, 301-306.

Liao, H., McKenzie, T., & Hageman, R. (1986). Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. Proc Natl Acad Sci U S A, 83(3), 576-580. doi:10.1073/pnas.83.3.576

Lin, L., & Xu, J. (2013). Dissecting and engineering metabolic and regulatory networks of thermophilic bacteria for biofuel production. Biotechnol Adv, 31(6), 827-837. doi:10.1016/j.biotechadv.2013.03.003

Lin, P., Rabe, K., Takasumi, J., Kadisch, M., Arnold, F., & Liao, J. (2014). Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*. Metab Eng, 24, 1-8. doi:10.1016/j.ymben.2014.03.006

Mahadevan, G., & Neelagund, S. (2014). Thermostable lipase from *Geobacillus* sp. Iso5: bioseparation, characterization and native structural studies. J Basic Microbiol, 54(5), 386-396. doi:10.1002/jobm.201200656

Markley, A., Begemann, M., Clarke, R., Gordon, G., & Pfleger, B. (2015). Synthetic biology toolbox for controlling gene expression in the cyanobacterium *Synechococcus* sp. strain PCC 7002. ACS Synth Biol, 4(5), 595-603. doi:10.1021/sb500260k

Marks, T., & Hamilton, P. (2014). Characterization of a thermophilic bacteriophage of *Geobacillus kaustophilus*. Arch Virol, 159(10), 2771-2775. doi:10.1007/s00705-014-2101-8

Matsumura, M., Katakura, Y., Imanaka, T., & Aiba, S. (1984). Enzymatic and nucleotide sequence studies of a kanamycin-inactivating enzyme encoded by a plasmid from thermophilic bacilli in comparison with that encoded by plasmid pUB110. J Bacteriol, 160(1), 413-420.

McMullan, G., Christie, J., Rahman, T., Banat, I., Ternan, N., & Marchant, R. (2004). Habitat, applications and genomics of the aerobic, thermophilic genus *Geobacillus*. Biochem Soc Trans, 32(Pt 2), 214-217. doi:10.1042/bst0320214

Mead, D., McClary, J., Luckey, J., Kostichka, A., Witney, F., & Smith, L. (1991). Bst DNA polymerase permits rapid sequence analysis from nanogram amounts of template. Biotechniques, 11(1), 76-78, 80, 82-77.

Mohanraju, P., Makarova, K., Zetsche, B., Zhang, F., Koonin, E., & van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science, 353(6299), aad5147. doi:10.1126/science.aad5147

Montoro-García, S., Martínez-Martínez, I., Navarro-Fernández, J., Takami, H., García-Carmona, F., & Sánchez-Ferrer, A. (2009). Characterization of a novel thermostable carboxylesterase from *Geobacillus kaustophilus* HTA426 shows the existence of a new carboxylesterase family. J Bacteriol, 191(9), 3076-3085. doi:10.1128/JB.01060-08

Mougiakos, I., Bosma, E., de Vos, W., van Kranenburg, R., & van der Oost, J. (2016). Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. Trends Biotechnol, 34(7), 575-587. doi:10.1016/j.tibtech.2016.02.004

Mougiakos, I., Mohanraju, P., Bosma, E., Vrouwe, V., Finger Bou, M., Naduthodi, M., van der Oost, J. (2017). Characterizing a thermostable Cas9 for bacterial genome editing and silencing. Nat Commun, 8(1), 1647. doi:10.1038/s41467-017-01591-4

Murphy, K. (1998). Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. J Bacteriol, 180(8), 2063-2071.

Muyrers, J., Zhang, Y., Testa, G., & Stewart, A. (1999). Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res, 27(6), 1555-1557. doi:10.1093/nar/27.6.1555

Nakade, S., Yamamoto, T., & Sakuma, T. (2017). Cas9, Cpf1 and C2c1/2/3-What's next? Bioengineered, 8(3), 265-273. doi:10.1080/21655979.2017.1282018

Nakayama, N., Narumi, I., Nakamoto, S., & Kihara, H. (1993). Complete nucleotide sequence of pSTK1, a cryptic plasmid from *Bacillus stearothermophilus* TK015. Biotechnol Lett, 15, 1013-1016.

Narumi, I., Nakayama, N., Nakamoto, S., Kimura, T., Yanagkisawa, T., & Kihara, H. (1993). Construction of a new shuttle vector pSTE33 and its stabilities in *Bacillus stearothermophilus*, *Bacillus subtilis*, and *Escherichia coli*. Biotechnol. Lett, 15, 815-820.

Narumi, I., Sawakami, K., Nakamoto, S., Nakayama, N., Yanagisawa, T., Takahashi, N., & Kihara, H. (1992). A newly isolated *Bacillus stearothermophilus* K1041 and its transformation by electroporation. Biotech Techniques, 6, 83-86.

Nazina, T., Tourova, T., Poltaraus, A., Novikova, E., Grigoryan, A., Ivanova, A., Ivanov, M. (2001). Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. Int J Syst Evol Microbiol, 51(Pt 2), 433-446. doi:10.1099/00207713-51-2-433

Nikel, P., Martínez-García, E., & de Lorenzo, V. (2014). Biotechnological domestication of pseudomonads using synthetic biology. Nat Rev Microbiol, 12(5), 368-379. doi:10.1038/nrmicro3253

Niu, H., Leak, D., Shah, N., & Kontoravdi, C. (2015). Metabolic characterization and modeling of fermentation process of an engineered *Geobacillus thermoglucosidasius* strain for bioethanol production with gas stripping. Chem. Eng. Sci., 122, 138-149.

Nurbaya, O., Bakar, S., Raja, R., Chor, L., & Basri, M. (2014). *Pichia pastoris* as a host to overexpress the thermostable T1 lipase from Geobacillus zalihae. GSTF Journal of BioSciences (JBio), 3(1), 7-17.

Offen, W., Viksoe-Nielsen, A., Borchert, T., Wilson, K., & Davies, G. (2015). Three-dimensional structure of a variant Termamyl-like *Geobacillus stearothermophilus* α-amylase at 1.9 Å resolution. Acta Crystallographica Section F: Structural Biology Communications, 71(1), 66-70.

Omokoko, B., Jäntges, U., Zimmermann, M., Reiss, M., & Hartmeier, W. (2008). Isolation of the phe-operon from *G. stearothermophilus* comprising the phenol degradative meta-pathway genes and a novel transcriptional regulator. BMC Microbiol, 8, 197. doi:10.1186/1471-2180-8-197

Oskam, L., Hillenga, D., Venema, G., & Bron, S. (1991). The large *Bacillus* plasmid pTB19 contains two integrated rolling-circle plasmids carrying mobilization functions. Plasmid, 26(1), 30-39. doi:10.1016/0147-619x(91)90034-t

Ozbek, E., Kolcuoğlu, Y., Konak, L., Colak, A., & Oz, F. (2014). Partial purification and biochemical characterization of an extremely thermo- and pH-stable esterase with great substrate affinity. Turkish J of Chemistry, 38, 538-546.

Pandey, A., Benjamin, S., Soccol, C., Nigam, P., Krieger, N., & Soccol, V. (1999). The realm of microbial lipases in biotechnology. Biotechnol Appl Biochem, 29 (Pt 2), 119-131.

Papon, N., Courdavault, V., & Clastre, M. (2014). Biotechnological potential of the fungal CTG clade species in the synthetic biology era. Trends Biotechnol, 32(4), 167-168. doi:10.1016/j.tibtech.2013.10.006

Patel, R. (2006). Biocatalysis in the pharmaceutical and biotechnological industries: CRC Press.

Peteranderl, R., Shotts, E. B., & Wiegel, J. (1990). Stability of antibiotics under growth conditions for thermophilic anaerobes. Appl Environ Microbiol, 56(6), 1981-1983.

Puchta, H. (2017). Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. Curr Opin Plant Biol, 36, 1-8. doi:10.1016/j.pbi.2016.11.011

Pédelacq, J., Cabantous, S., Tran, T., Terwilliger, T., & Waldo, G. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol, 24(1), 79-88. doi:10.1038/nbt1172

Quintana-Castro, R., Díaz, P., Valerio-Alfaro, G., García, H., & Oliart-Ros, R. (2009). Gene cloning, expression, and characterization of the *Geobacillus thermoleovorans* CCR11 thermoalkaliphilic lipase. Mol Biotechnol, 42(1), 75-83. doi:10.1007/s12033-008-9136-6

Reeve, B., Martinez-Klimova, E., de Jonghe, J., Leak, D., & Ellis, T. (2016). The *Geobacillus* Plasmid Set: A Modular Toolkit for Thermophile Engineering. ACS Synth Biol, 5(12), 1342-1347. doi:10.1021/acssynbio.5b00298

Schiel-Bengelsdorf, B., & Dürre, P. (2012). Pathway engineering and synthetic biology using acetogens. FEBS Lett, 586(15), 2191-2198. doi:10.1016/j.febslet.2012.04.043

Setlow, P. (2006). Spores of *Bacillus subtilis*: their resistance to and killing by radiation, heat and chemicals. J Appl Microbiol, 101(3), 514-525. doi:10.1111/j.1365-2672.2005.02736.x

Shulami, S., Shenker, O., Langut, Y., Lavid, N., Gat, O., Zaide, G., Shoham, Y. (2014). Multiple regulatory mechanisms control the expression of the *Geobacillus stearothermophilus* gene for extracellular xylanase. J Biol Chem, 289(37), 25957-25975. doi:10.1074/jbc.M114.592873

Sifour, M., Zaghloul, T., Saeed, H., Berekaa, M., & Abdel-Fattah, Y. (2010). Enhanced production of lipase by the thermophilic *Geobacillus stearothermophilus* strain-5 using statistical experimental designs. N Biotechnol, 27(4), 330-336. doi:10.1016/j.nbt.2010.04.004

Soliman, N., Abdel-Fattah, Y., Mostafa, H., & Gaballa, A. Heterologous expression of thermostable esterase gene from *Geobacillus thermoleovorans* YN under different expression promoters. International J of Env Science and Technology, 11, 119-126.

Spizizen, J. (1958). Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. Proc Natl Acad Sci U S A, 44(10), 1072-1078. doi:10.1073/pnas.44.10.1072

Studholme, D. (2015). Some (bacilli) like it hot: genomics of *Geobacillus* species. Microb Biotechnol, 8(1), 40-48. doi:10.1111/1751-7915.12161

Studier, F., & Moffatt, B. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol, 189(1), 113-130. doi:10.1016/0022-2836(86)90385-2

Stuknyte, M., Guglielmetti, S., Mora, D., Kuisiene, N., Parini, C., & Citavicius, D. (2008). Complete nucleotide sequence of pGS18, a 62.8-kb plasmid from *Geobacillus stearothermophilus* strain 18. Extremophiles, 12(3), 415-429. doi:10.1007/s00792-008-0145-y

Sun, Z., Deng, A., Hu, T., Wu, J., Sun, Q., Bai, H., Wen, T. (2015). A high-efficiency recombineering system with PCR-based ssDNA in *Bacillus subtilis* mediated by the native phage recombinase GP35. Appl Microbiol Biotechnol, 99(12), 5151-5162. doi:10.1007/s00253-015-6485-5

Suzuki, H., Murakami, A., & Yoshida, K. (2012). Counterselection system for *Geobacillus kaustophilus* HTA426 through disruption of pyrF and pyrR. Appl Environ Microbiol, 78(20), 7376-7383. doi:10.1128/AEM.01669-12

Suzuki, H., Wada, K., Furukawa, M., Doi, K., & Ohshima, T. (2013). A ternary conjugation system for the construction of DNA libraries for *Geobacillus kaustophilus* HTA426. Biosci Biotechnol Biochem, 77(11), 2316-2318. doi:10.1271/bbb.130492

Suzuki, H., & Yoshida, K. (2012). Genetic transformation of *Geobacillus kaustophilus* HTA426 by conjugative transfer of host-mimicking plasmids. J Microbiol Biotechnol, 22(9), 1279-1287. doi:10.4014/jmb.1203.03023

Suzuki, H., Yoshida, K., & Ohshima, T. (2013). Polysaccharide-degrading thermophiles generated by heterologous gene expression in *Geobacillus kaustophilus* HTA426. Appl Environ Microbiol, 79(17), 5151-5158. doi:10.1128/AEM.01506-13

Tang, X., Lowder, L., Zhang, T., Malzahn, A., Zheng, X., Voytas, D., Qi, Y. (2017). A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. Nat Plants, 3, 17103. doi:10.1038/nplants.2017.103

Taylor, M., Esteban, C., & Leak, D. (2008). Development of a versatile shuttle vector for gene expression in *Geobacillus* spp. Plasmid, 60(1), 45-52. doi:10.1016/j.plasmid.2008.04.001

Taylor, M. P., van Zyl, L., Tuffin, I. M., Leak, D. J., & Cowan, D. A. (2011). Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts. Microb Biotechnol, 4(4), 438-448. doi:10.1111/j.1751-7915.2010.00246.x

Tekedar, H., & Sanlı-Mohamed, G. (2011). Molecular cloning, over expression and characterization of thermoalkalophilic esterases isolated from *Geobacillus* sp. Extremophiles, 15(2), 203-211. doi:10.1007/s00792-010-0344-1

Thompson, A., Studholme, D., Green, E., & Leak, D. (2008). Heterologous expression of pyruvate decarboxylase in *Geobacillus thermoglucosidasius*. Biotechnol Lett, 30(8), 1359-1365. doi:10.1007/s10529-008-9698-1

Turner, P., Mamo, G., & Karlsson, E. (2007). Potential and utilization of thermophiles and thermostable enzymes in biorefining. Microb Cell Fact, 6, 9. doi:10.1186/1475-2859-6-9

Vallier, H., & Welker, N. E. (1990). Genetic map of the *Bacillus stearothermophilus* NUB36 chromosome. J Bacteriol, 172(2), 793-801. doi:10.1128/jb.172.2.793-801.1990

van der Lelie, D., Bron, S., Venema, G., & Oskam, L. (1989). Similarity of minus origins of replication and flanking open reading frames of plasmids pUB110, pTB913 and pMV158. Nucleic Acids Res, 17(18), 7283-7294. doi:10.1093/nar/17.18.7283

van Kessel, J., & Hatfull, G. (2007). Recombineering in *Mycobacterium tuberculosis*. Nat Methods, 4(2), 147-152. doi:10.1038/nmeth996

Van Zyl, L. J., Taylor, M. P., Eley, K., Tuffin, M., & Cowan, D. A. (2014). Engineering pyruvate decarboxylase-mediated ethanol production in the thermophilic host *Geobacillus thermoglucosidasius*. Appl Microbiol Biotechnol, 98(3), 1247-1259. doi:10.1007/s00253-013-5380-1

Vellore, J., Moretz, S., & Lampson, B. (2004). A group II intron-type open reading frame from the thermophile *Bacillus (Geobacillus) stearothermophilus* encodes a heat-stable reverse transcriptase. Appl Environ Microbiol, 70(12), 7140-7147. doi:10.1128/AEM.70.12.7140-7147.2004

Wang, Y., Srivastava, K., Shen, G., & Wang, H. (1995). Thermostable alkaline lipase from a newly isolated thermophilic *Bacillus* strain A30-1 (ATCC 53841). J Fermentation and Bioengineering, 79, 433-438.

Weigel, J., & Ljungdahl, L. (1986). The importance of thermophilic bacteria in biotechnology. Crit. Rev. Biotechnol., 3, 39-108.

Wiegand, S., Rabausch, U., Chow, J., Daniel, R., Streit, W., & Liesegang, H. (2013). Complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase-secreting bacterium. Genome Announc, 1(2), e0009213. doi:10.1128/genomeA.00092-13

Wright, A., Nuñez, J., & Doudna, J. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell, 164(1-2), 29-44. doi:10.1016/j.cell.2015.12.035

Wu, L., & Welker, N. (1989). Protoplast transformation of *Bacillus stearothermophilus* NUB36 by plasmid DNA. J Gen Microbiol, 135(5), 1315-1324. doi:10.1099/00221287-135-5-1315

Xiao, Z., Wang, X., Huang, Y., Huo, F., Zhu, X., Xi, L., & Lu, J. (2012). Thermophilic fermentation of acetoin and 2,3-butanediol by a novel *Geobacillus* strain. Biotechnol Biofuels, 5(1), 88. doi:10.1186/1754-6834-5-88

Xu, J., Ren, X., Sun, J., Wang, X., Qiao, H., Xu, B., Ni, J. (2015). A Toolkit of CRISPR-Based Genome Editing Systems in *Drosophila*. J Genet Genomics, 42(4), 141-149. doi:10.1016/j.jgg.2015.02.007

Yan, M., Yan, H., Ren, G., Zhao, J., Guo, X., & Sun, Y. (2017). CRISPR-Cas12a-assisted recombineering in bacteria. Appl Environ Microbiol, 83(17). doi:10.1128/AEM.00947-17

Yang, Z., Zhang, Y., Shen, T., Xie, Y., Mao, Y., & Ji, C. (2013). Cloning, expression and biochemical characterization of a novel, moderately thermostable GDSL family esterase from *Geobacillus thermodenitrificans* T2. J Biosci Bioeng, 115(2), 133-137. doi:10.1016/j.jbiosc.2012.08.016

Yildirim, M., Colak, A., Col, M., & Canakci, S. (2009). A new recombinant phosphotriesterase homology protein from *Geobacillus caldoxylosilyticus* TK4: an extremely thermo- and pH-stable esterase. Process Biochemistry, 44, 1366-1373.

Yu, D., Ellis, H., Lee, E., Jenkins, N., Copeland, N., & Court, D. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc Natl Acad Sci U S A, 97(11), 5978-5983. doi:10.1073/pnas.100127597

Zeigler, D. (2001). The genus Geobacillus (7th ed.). In *Bacillus* Genetic Stock Center Catalog of Strains.

Zeigler, D. (2014). The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? Microbiology, 160(Pt 1), 1-11. doi:10.1099/mic.0.071696-0

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E., Zhang, F. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat Biotechnol, 35(1), 31-34. doi:10.1038/nbt.3737

Zhang, J., Siika-Aho, M., Tenkanen, M., & Viikari, L. (2011). The role of acetyl xylan esterase in the solubilization of xylan and enzymatic hydrolysis of wheat straw and giant reed. Biotechnol Biofuels, 4(1), 60. doi:10.1186/1754-6834-4-60

Zhang, Y., Buchholz, F., Muyrers, J., & Stewart, A. (1998). A new logic for DNA engineering using recombination in *Escherichia coli*. Nat Genet, 20(2), 123-128. doi:10.1038/2417

Zhu, W., Cha, D., Cheng, G., Peng, Q., & Shen, P. (2007). Purification and characterization of a thermostable protease from a newly isolated *Geobacillus* sp. YMTC 1049. Enzyme and Microbial Technology, 40, 1592-1597.

Zhu, Y., Li, H., Ni, H., Xiao, A., Li, L., & Cai, H. (2015). Molecular cloning and characterization of a thermostable lipase from deep-sea thermophile *Geobacillus* sp. EPT9. World J Microbiol Biotechnol, 31(2), 295-306. doi:10.1007/s11274-014-1775-0

Zhu, Y., Liu, G., Li, H., Liu, J., Bai, X., Guan, R., & Cai, H. (2012). Cloning and characterization of a thermostable carboxylesterase from inshore hot spring thermophile *Geobacillus* sp. ZH1. Acta Oceanologica Sinica, 31(6), 117-126.

CHAPTER 2

Characterization of a thermophilic bacteriophage of *Geobacillus kaustophilus*

Marks, Timothy J.[AB] and Hamilton, Paul T.[B]

[A]Department of Pharmaceutical Sciences, Campbell University, Buies Creek, NC 27506

[B] Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695

Citation: Marks, T., & Hamilton, P. (2014). Characterization of a thermophilic bacteriophage of Geobacillus kaustophilus. Arch Virol, 159(10), 2771-2775.

Abstract

GBK2 is a bacteriophage isolated from a backyard compost pile that infects the thermophile *Geobacillus kaustophilus*. GBK2 has a circularly permuted genome of 39,078 bp with a G+C content of 43%. Annotation of the genome reveals 62 putative open reading frames (ORFs), 25 of which (40.3%) show homology to known proteins and 37 of which (59.7%) are proteins with unknown functions. Twelve of the identified ORFs had the greatest homology to genes from the phage SPP1, a phage that infects the mesophile *Bacillus subtilis*. The overall genomic arrangement of GBK2 is similar to SPP1, with the majority of GBK2 SPP1-like genes coding for proteins involved in DNA replication and metabolism.

2.1 Introduction

Geobacillus species are gram-positive thermophilic bacteria that can ferment C-5 and C-6 sugars to mixed acids and ethanol and have potential for biofuel production [1, 2]. Some species are also able to degrade hydrocarbons, cellulose, and produce enzymes of industrial interest, including amylases, proteases, and lipases [3]. Additionally, these bacteria have drawn attention due to their thermostable enzymes that could be used for biotechnology applications [4, 5]. Unfortunately, there is a lack of gene transfer and protein expression vectors that are useful in *Geobacillus*. Therefore, there is a pressing need for additional biotechnological tools for these organisms [6]. The genetic diversity among phages represents a vast set of potential tools that can be developed to exchange, alter and express genes and gene products from an organism of interest. We have started to characterize bacteriophages that infect *Geobacillus* species in an effort to enhance the tools available to study and modify *Geobacillus* species for various industrial applications.

2.2 Methods

The *Geobacillus kaustophilus*-infecting bacteriophage GBK2 was isolated from a backyard compost pile in Cary, NC. The phage was enriched out of the compost in liquid culture with *G. kaustophilus* (ATCC 8005) at 55°C. GBK2 was purified and quantitated by the double-layer agar technique [7] and phage particles were concentrated by polyethylene glycol (PEG 8000) precipitation. Phage particles were suspended in phage buffer and lysed at 65°C with SDS (0.1% w/v) and EDTA (5mM). The DNA was purified using phenol/chloroform extraction and ethanol precipitation [8]. The DNA sequence for GBK2 was obtained using Illumina GAIIx (Illumina, San Diego, CA, USA) and assembled using CLC Bio software (CLC Bio, Cambridge, MA, USA). Open reading frames (ORFs) in the

GBK2 genome were predicted using VectorNTI (Invitrogen, Carlsbad, CA, USA) and HMM protocols with GeneMark [9] and Glimmer [10]. Annotation and analysis of predicted protein functions were performed using BLASTX [11].

2.3 Results and Discussion

The nucleotide sequence of GBK2 is 39,078 bp with a G+C content of 43%. The genomic sequence of the host, *G. kaustophilus* (ATCC 8005), has not been determined but its G+C content has been estimated to be 52-58% [12], and the sequences of the genomes of two other strains of *G. kaustophilus* (HTA426 & GBlys) have been shown to have a G+C content of 52% [13]. Restriction analysis of the GBK2 phage DNA with a variety of restriction endonucleases predicted to cut the GBK2 genome showed that the genome contains few if any modified nucleotides and is circularly permuted (data not shown). Analysis of the genome sequence of GBK2 identified 62 ORFs (Fig 2-1 and Table 2-1). The longest predicted gene is 2,570 nucleotides (unknown function; ORF 32), whereas the shortest gene included is 134 nucleotides (unknown function; ORF 27). Twenty-five ORFs (40.3%) encode predicted proteins of recognized function, while 37 (59.7%) were of unknown function. GBK2 ORFs that have homology to proteins of known function fall into four groups: phage packaging, phage structural proteins, DNA metabolism/replication and lysis. There were no ORFs with homology to integrase genes which is consistent with a lytic phage and the clear plaque morphology of GBK2. All of the predicted ORFs except ORF 24 (unknown function) are transcribed in the same direction along the genome. ORF 51 of GBK2 has homology to thymidylate synthases. We have cloned ORF 51 and shown it can complement growth of an *E. coli thyA* deletion mutant (Marks, Griffin and Hamilton, unpublished results). Isolation of

thyA mutants of *G. kaustophilus* and the development of GBK2 ORF 51 as a selectable marker in *G. kaustophilus* are in progress.

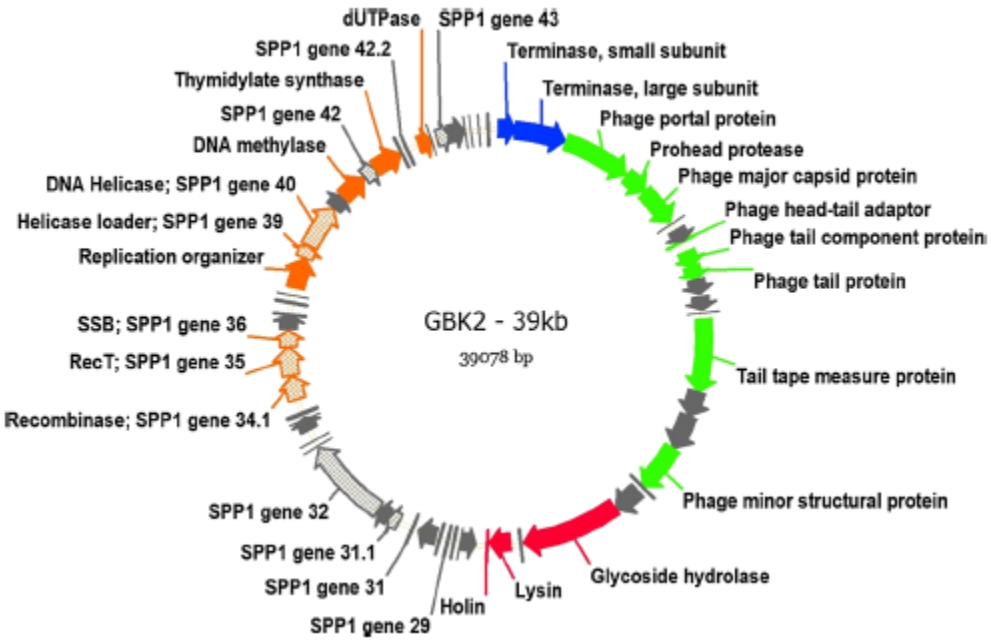

Figure 2-1 Genome arrangement of GBK2. The categories of functional ORFs are indicated by the following colors: blue, DNA packaging; green, structural; red, lysis; gray, unknown; orange, DNA replication/metabolism. SPP1 homologs are designated with a hatched pattern (color figure online). The GenBank file (accession no. KJ159566) starts with base 1 at the beginning of the terminase small subunit – ORF 1.

Table 2-1

The predicted ORFs of GBK2 and their presumed functions.

ORF[1]	Coordinates	Predicted Function[2]	Significant Match [Organism] (Protein Sequence ID)	E value[3]
1	1-507	Terminase- small subunit	Transposase [*Paenibacillus* sp. OSY-SE] (WP_019419039.1)	4E-34
2	500-2041	Terminase- large subunit	Terminase [*Orenia marismortui*] (WP_018250335.1)	0.0
3	2075-4132	Phage portal protein	HK97 family phage portal protein [*Desmospora* sp. 8437] (WP_009709190.1)	6E-138
4	4169-4870	Prohead protease	Peptidase U35 [*Clostridium bolteae*] (WP_002578668.1)	7E-44
5	4890-6065	Major capsid protein	Phage major capsid protein [*Desmospora* sp. 8437] (WP_009709188.1)	3E-111
8	6746-7066	Head-tail adaptor protein	Phage head-tail adaptor protein [*Actinobacillus succinogenes* 130Z] (YP_001344537.1)	1E-18
9	7071-7550	Tail component protein	Phage tail component protein [*Desmospora* sp. 8437] (WP_009709182.1)	8E-18
10	7547-7954	Tail protein	Phage tail protein p028 [*Bacillus* sp. 916] (WP_007408588.1)	3E-32
11	7973-8383	Unknown	Antigen A [*Bacillus* sp. 1NLA3E] (YP_007908532.1)	3E-40
12	8489-8875	Unknown	Hypothetical protein B1NLA3E_01250 [*Bacillus* sp. 1NLA3E] (YP_007908533.1)	3E-48
13	8968-9168	Unknown	Hypothetical protein B1NLA3EDRAFT_3741 [*Bacillus* sp. 1NLA3E] (YP_007908534.1)	2E-98
14	9168-11351	Tail tape measure protein	Tail tape measure protein [*Bacillus oceanisediminis*] (WP_019381550.1)	3E-136
15	11351-12097	Tail component protein	Hypothetical protein BH3517 [*Bacillus halodurans* C-125] (NP_244384.1)	2E-84
16	12113-13171	Unknown	Hypothetical protein GC56T3_0281 [*Geobacillus* sp. C56-T3] (YP_003669925.1)	3E-60
17	13184-14653	Phage minor structural protein	Hypothetical protein BH3515 [*Bacillus halodurans* C-125] (NP_244382.1)	0.0
18	14666-14860	Unknown	Hypothetical protein [*Laceyella sacchari*] (WP_022737485.1)	8E-15
19	14850-15704	Unknown	Hypothetical protein BH0016 [*Bacillus halodurans* C-125] (NP_240882.1)	6E-67
20	15701-18667	Glycoside hydrolase	Glycoside hydrolase family protein [*Geobacillus* sp. C56-T3] (YP_003669929.1)	0.0
21	18768-19058	Unknown	Hemolysin XhlA [*Brevibacillus brevis*] (WP_016741605.1)	2E-06
22	19073-19720	Lysin	Hypothetical protein- Peptidase [*Anoxybacillus* sp. DT3-1] (WP_009362064.1)	1E-61
23	19788-20027	Holin	Holin [*Bacillus licheniformis*] (WP_003185319.1)	3E-23
24	20542-20123 (Minus)	Unknown	Hypothetical protein [*Bacillus atrophaeus* 1942] (YP_003975304.1)	1E-07

Table 2-1 [continued]				
26	20993-21241	Unknown	Hypothetical protein; SPP1 gene 29 [*Bacillus* phage SPP1] (NP_690707.1)	6E-06
28	21355-21921	Unknown	Hypothetical protein [*Bacillus* sp. 123MFChir2] (WP_020061772.1)	1E-07
29	22109-22414	Unknown	Hypothetical protein; SPP1 gene 31 [*Bacillus* phage SPP1] (NP_690711.1)	1E-19
30	22550-22891	Unknown	Hypothetical protein; SPP1 gene 31.1 [*Bacillus* phage SPP1] (NP_690712.1)	1E-41
31	22896-23258	Unknown	Hypothetical protein [*Bacillus thuringiensis* serovar chinensis CT-43] (YP_005569893.1)	9E-19
32	23270-25840	Unknown	Hypothetical protein; SPP1 gene 32 [*Bacillus* phage SPP1] (NP_690714.1)	0.0
34	26120-26374	Unknown	Hypothetical protein [*Paenibacillus alvei*] (WP_005545873.1)	1E-07
35	26471-26860	Unknown	Hypothetical protein [*Bacillus* phage vB_BanS-Tsamsa] (AGI11844.1)	3E-62
38	27391-28182	Recombinase	Recombinase; SPP1 gene 34.1 [*Bacillus* phage SPP1] (NP_690722.1)	3E-99
39	28179-29042	recT	recT; SPP1 gene 35 [*Bacillus* phage SPP1] (NP_690727.1)	3E-113
40	29035-29505	SSB	SSB; SPP1 gene 36 [*Bacillus* phage SPP1] (NP_690727.1)	8E-56
42	29993-30208	Unknown	Hypothetical protein [*Sporosarcina newyorkensis*] (WP_009498071.1)	3E-06
43	30350-30586	Unknown	Helix-turn-helix domain-containing protein [*Bacillus cellulosilyticus* DSM 2522] (YP_004095737.1)	1E-07
45	30825-31754	Replication Organizer	Putative prophage replication protein O [*Bacillus* sp. 1NLA3E] (YP_007908506.1)	8E-86
46	31754-32134	Helicase loader	Helicase loader; SPP1 gene 39 [*Bacillus* phage SPP1] (NP_690732.1)	2E-17
47	32134-33468	Helicase	Helicase; SPP1 gene 40 [*Bacillus* phage SPP1] (NP_690733.1)	1E-163
49	33942-34751	DNA methylase	DNA modification methylase [*Clostridium* sp. BNL1100] (YP_005147202.1)	1E-100
50	34792-35214	Unknown	Hypothetical protein; SPP1 gene 42 [*Bacillus* phage SPP1] (NP_690738.1)	5E-17
51	35214-36083	Thymidylate synthase	Thymidylate synthase [*Bacillus* sp. L1(2012)] (WP_017728547.1)	1E-120
52	36127-36324	Unknown	Hypothetical protein [Deep-sea thermophilic phage D6E] (YP_007010960.1)	2E-16
53	36318-36614	Unknown	Hypothetical protein; SPP1 gene 42.2 [*Bacillus* phage SPP1] (NP_690740.1)	2E-24

Table 2-1 [continued]				
54	36692-37078	dUTPase	dUTPase [*Geobacillus thermoglucosidasius* C56-YS93] (YP_004587001.1)	5E-24
56	37277-37633	Unknown	Hypothetical protein; SPP1 gene 43 [*Bacillus* phage SPP1] (NP_690741.1)	8E-04
57	37360-38121	Unknown	Phage-like protein [*Bacillus licheniformis* DSM 13 = ATCC 14580] (YP_078639.1)	2E-22

[1] ORFs numbered consecutively. ORFs with no matches were not included in the table but all predicted ORFs can be found in GenBank under accession number KJ159566.
[2] Protein function based on amino acid sequence identity in BLAST analysis uning the NCBI prokaryotic genome annotation pipeline.
[3] The probability of obtaining a match by chance as determined by BLAST analysis. Values less than 10E-5 were considered significant.

Overall, GBK2 is more closely related to bacteriophage that infect mesophilic hosts than it is to other phage that infect thermophilic geobacilli or other thermophiles (results not shown). The genomic arrangement of GBK2 is similar to SPP1, a phage that infects the mesophile *Bacillus subtilis* [14], but there are only 14 ORFs that encode proteins that show homology. Twelve ORFs (shown as hatched arrows in Fig 2-1) have the greatest homology to proteins encoded by SPP1. Most of the 14 genes encoding known proteins with homology between SPP1 and GBK2 function in DNA replication; however, two ORF's predicted to encode structural proteins (ORF 8 and ORF 9) have similarities to SPP1 proteins, SPP1p027 (e value of 7E-12) and SPP1p028 (e value of 2E-31) respectively, that fall within our threshold. Among the DNA replication/metabolism genes, GBK2 encodes an exonuclease/recombinase (ORF 38), RecT-type recombinase (ORF 39), SSB protein (ORF 40), helicase loader (ORF 46), and helicase (ORF 47). However, ORF 45 encodes a replication organizer with homology to the phage replication protein O family and does not have homology to the SPP1 replication organizer gene 38. ORF 45 of GBK2 contains seven copies of a 17 bp inverted DNA repeat which is a potential origin of replication for GBK2. In addition, GBK2 has several ORFs encoding DNA metabolism enzymes (ORF 49, 51, & 54) inserted between the SPP1 homologs. It would appear that since the divergence of the

SPP1/GBK2 replication module it has lost/gained several functions. In SPP1, genes 42, 42.2 and 43 have unknown function and are non-essential for growth [14], but homologs of these genes are found in GBK2 which would indicate that while they may be non-essential they may play a significant role in the phage life cycles under certain conditions.

The genome sequence of GBK2, a unique bacteriophage that infects the thermophile *Geobacillus kaustophilus* provides useful building blocks for the development of tools to characterize and manipulate *Geobacillus* species, in addition to increasing our understanding of thermophiles. The complete genome sequence of GBK2 is available in the GenBank database under accession number KJ159566.

Acknowledgements

The authors would like to thank the Genomic Sciences Laboratory (GSL) at NCSU for assistance in sequence determination and assembly.

References

1. Taylor MP, Eley KL, Martin S, Tuffin MI, Burton SG, Cowan DA (2009). Thermophilic ethanologenesis: Future prospects for second-generation bioethanol production. Trends in Biotechnology. 27(7):398-405

2. Taylor MP, van Zyl L, Tuffin IM, Leak DJ, Cowan DA (2010). Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts. Microb Biotechnology. 4(4):438-448

3. Cuebasg M, Sannino D, Bini B (2011). Isolation and characterization of an arsenic resistant *Geobacillus kaustophilus* strain from geothermal soils. J of Basic Microbiology. 51:364 – 371

4. Feng L, Wang W, Cheng J, Ren Y, Zhao G, Gao C, Tang Y, Lui X, Han W, Peng X, Liu R, Wang L (2007) Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. Proc Natl Acad Sci. U.S.A. 104:5602-5607

5. McMullen G, Christie JM, Rahman TJ, Banat IM, Ternan NG, Marchant R (2004) Habitat, applications, and genomics of the aerobic, thermophilic genus *Geobacillus*. Biochem Soc Trans 32:214-217

6. Taylor MP, Esteban CD, Leak DJ (2008). Development of a versatile shuttle vector for gene expression in *Geobacillus spp*. Plasmid. 60:45-52.

7. Kropinski A, Mazzocco A, Waddell T, Lingohr E, Johnson R (2009) Enumeration of bacteriophages by double agar overlay plaque assay. In: Clokie M, Kropinski A (eds) Bacteriophages: methods and protocols. Humana Press, New York, pp 69-76

8. Sambrook JF and Russell DW (2001) Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press

9. Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29: 2607-2618

10. Delcher AL, Bratke KA, Powers EC, Salzber SL (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinforma 23: 673-679

11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. 215:403-410.

12. Nazina TN et al. (2001) Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus, G. thermoleovorans*, *G. kaustophilus, G. thermoglucosidasius* and *G. thermodenitrificans*. Int J Syst Evol Microbiol. Mar;51(Pt 2):433-46. PubMed PMID: 11321089.

13. Doi K, Mori K, Martono H, Nagayoshi Y, Fujino Y, Tashiro K, Kuhara S, Ohshima T (2013) Draft genome sequence of *Geobacillus kaustophilus* GBlys, a lysogenic strain with bacteriophage ϕOH2. Genome Announc. 1(4):e00634-13. doi:10.1128/genomeA.00634-13.

14. Alonso JC, Luder G, Stiege AC, Chai S, Weise F, Trautner TA (1997)The complete nucleotide sequence and functional organization of *Bacillus subtilus* bacteriophage SPP1. Gene 204:201-212

CHAPTER 3

The characterization of a novel lysogenic bacteriophage of the thermophile *Geobacillus kaustophilus*: source of genetic tools and a glimpse at phage evolution

Marks, Timothy J.[AB] and Hamilton, Paul T.[B]

[A]Department of Pharmaceutical Sciences, Campbell University, Buies Creek, NC 27506

[B] Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695

Abstract

Members of the genus *Geobacillus* are thermophilic organisms of great biotechnological potential because of their metabolic diversity and their production of industrially relevant thermostable enzymes and products. However, a lack of genetic manipulation tools has hindered efforts to use these organisms to their full potential. Here we present potential new tools developed from a lysogenic bacteriophage that infects *Geobacillus kaustophilus*. Additionally, we examine the evolutionary relationship between this thermophilic phage (named GBK1) and a mesophilic phage that infects *Bacillus* spp. GBK1 is a bacteriophage isolated from a backyard compost pile with 56 identified ORFs. ORF 44, which has high homology to several phage tyrosine integrase/recombinase genes, was used to develop an integrating suicide vector. A 22 base pair site upstream of ORF 44 has been identified as the *attP* site, which has homology with potential *attB* sites in several *Geobacillus* species. ORF 37, which has high homology to thymidylate synthase, complements *thyA*- mutants of *E. coli*. We isolated and determined the DNA sequence of a GBK1 lysogenic strain of *Geobacillus kaustophilus* ATCC 8005/NBRC 102445 using NextGen PacBio sequencing. This is the first time the entire genomic sequence of the type strain of *Geobacillus kaustophilus* has been completely determined, and analysis confirmed the presence of the GBK1 prophage at the *attP*/*attB* integration site.

3.1 Introduction

Geobacillus species are gram-positive thermophilic microorganisms that grow in a temperature range of 37-80˚C. They are of industrial interest due to their ability to grow on C5 as well as C6 sugars (Coorevits et al., 2012; Nazina et al., 2001; Omokoko et al., 2008; Zeigler, 2014) and their use in production of industrially relevant, thermostable enzymes, such as amylases, lipases (McMullan et al., 2004), DNA polymerase (Mead et al., 1991) and proteases. There has also been research published on the production of ethanol (Cripps et al., 2009) and isobutanol (Lin et al., 2014), as *Geobacillus* species are mixed acid fermenters. Unfortunately, there is a lack of tools for genetic manipulation purposes in these organisms due to difficulties with DNA transformation, the presence of host restriction modification systems, and a lack of plasmids (Kananavičiūtė and Čitavičius, 2015). In addition, many standard antibiotics are unreliable at the higher temperatures required for growth of these organisms, which makes antibiotic selection difficult (Kananavičiūtė and Čitavičius, 2015). Phage represent a potential resource for the development of tools for *Geobacillus* species in the form of potential promoters, DNA modifying enzymes, auxotrophic complementation systems, and the development of plasmids capable of integration into a specific host site using a phage integrase and homologous *attP*/*attB* sites. GBK1 is a novel lysogenic bacteriophage isolated from a backyard compost pile in Cary, NC. We have successfully isolated a strain of *G. kaustophilus* ATCC 8005/NBRC 102445 that contains the GBK1 prophage, as evidenced by NextGen PacBio sequencing and *de novo* assembly of the lysogen genome.

Geobacillus species were once classified in the *Bacillus* genus but were reclassified into their own genus in 2001 (Nazina et al., 2001). Several strains have been sequenced and

show an average genome size of 3.5-3.9 Mbp and a G + C content ranging from 45 to 55% (Hussein et al., 2015). Bezuidt et al. (2016) found that the core genome, genes conserved in all 29 *Geobacillus* genomes analyzed, were similar to that of *Bacillus* species, indicating that *Geobacillus* species evolved from *Bacillus* species (Bezuidt et al., 2016). This raises a question about the evolution of bacteriophage that infect *Geobacillus* spp. Did they evolve from phage that infect *Bacillus* species or from phage that infect other bacterial species?

This study describes the isolation, characterization, and complete genome analysis of GBK1 as well as a comparison of its *attP* site to potential attB sites in several *Geobacillus* species. We also discuss the GBK1 *thyA* gene and its potential as a marker-free selection system in *Geobacillus* spp. Additionally, we describe the evolutionary relationship between the thermophilic GBK1 and other bacteriophage.

3.2 Methods

3.2.1 Phage isolation and propagation. Bacteriophage GBK1 was enriched out of the compost in liquid culture with *G. kaustophilus* ATCC 8005/NBRC 102445 at 55 °C (Marks and Hamilton, 2014). After infection of the host cells, GBK1 was purified and quantitated by the double-layer agar technique (Kropinski et al., 2009). Phage particles were concentrated by polyethylene glycol (PEG 8000) precipitation and suspended in phage buffer (10 mM Tris pH7.5, 100 mM NaCl, 10 mM $MgSO_4$, 2 mM $CaCl_2$).

3.2.2 Phage DNA extraction and genome sequencing. Phage particles were lysed at 65 °C with SDS (0.1% w/v) and EDTA (5mM) and the phage DNA was purified using phenol/chloroform extraction and ethanol precipitation (Sambrook and Russell, 2001). The DNA sequence for GBK1 was obtained using Illumina GAIIx (Illumina, San Diego, CA,

USA) and assembled using CLC Bio software (CLC Bio, Cambridge, MA, USA) with an average coverage of 27,970.07.

3.2.3 Genome analysis. Open reading frames (ORFs) in the GBK1 genome were predicted using VectorNTI (Invitrogen, Carlsbad, CA, USA) and HMM protocols with GeneMark (Besemer et al., 2001) and Glimmer (Delcher et al., 2007). Annotation and analysis of predicted protein functions were performed using BLASTX (Altschul et al., 1990). A minimum cutoff of 100 bp was used for putative ORFs. The complete genome of GBK1 was deposited in the GenBank database with the accession number MN727882.

3.2.4 *G. kaustophilus* lysogen genome sequencing. *G. kaustophilus* ATCC 8005 was infected with GBK1 and produced turbid plaques. Cells from a turbid plaque were isolated. Bacterial chromosomal DNA was extracted using the GenElute Bacterial DNA Kit (Sigma-Aldrich, St. Louis, MO, USA). Samples were tested for the presence of GBK1 DNA using PCR and primers specific for GBK1. A 15-20 kb PacBio library prep of a GBK1 lysogen of *G. kaustophilus* was performed by the North Carolina State University Genomic Sciences Laboratory. Nucleotide sequencing was carried out using a PacBio Sequel SMRT cell (NCSU Genomic Sciences Laboratory, Raleigh, NC, USA). The genome was assembled *de novo* with CLC Genomics Workbench software v.6.0.1, with an average coverage of 150.52x. The *Geobacillus kaustophilus* NBRC 102445 /ATCC 8005 lysogen sequence was deposited in the GenBank database using their annotation pipeline with the accession number CP038860.

3.2.5 Phylogenetic analysis. The predicted amino acid sequences for phage terminase large subunit and integrase were used to conduct a phylogenetic analysis of the GBK1 bacteriophage. Predicted amino acid sequences of each protein were aligned with a

collection of homologous sequences using the program MegaX (Kumar et al., 2018). Those multiple alignments along with a maximum parsimony analysis were used to construct a phylogenetic tree, with bootstrap support (n = 1000 replicates).

3.2.6 Prophage search of the *G. kaustophilus* NBRC 102445 genome. PHASTER (Arndt et al., 2016) was used to predict likely prophage sequences in the *G. kaustophilus* NBRC 102445 genome. A minimum score of 70 was the set threshold for possible prophage sequences.

3.2.7 Genome Comparison of GBK1 and *Bacillus* Phage vB_BpsS-36. The progressiveMauve (Darling et al., 2004) alignment tool was used to compare the genomes of GBK1 and mesophilic *Bacillus* phage vB_BpsS-36 (Genbank accession MH884513, author Akhwale JK). Default seed weights and HOXD (default) scoring were used to compare the two genomes.

3.3 Results

3.3.1 Bacteriophage GBK1 characteristics. The nucleotide sequence of GBK1 is 45,439 bp with a G+C content of 43.6%. The genomic sequence of the host, *G. kaustophilus* (NBRC 102445/ATCC 8005), was also determined and its G+C content is 52%. Additionally, sequences of the genomes of four other strains of *G. kaustophilus* (HTA426, GBlys, Et2/3, and Et7/4) have been shown to have a G+C content of 52% (Doi et al., 2013). Restriction analysis of the GBK1 phage DNA with a variety of restriction endonucleases predicted to cut the GBK1 genome showed that the genome contains few if any modified nucleotides and is linear (data not shown). Analysis of the genome sequence of GBK1 identified 56 ORFs (Fig 3-1 and Table 3-1). The longest predicted gene is 3,148 nucleotides (homology to DNA polymerase III subunit alpha; ORF 29), whereas the shortest gene

identified is 101 nucleotides (unknown function; ORF 34). A limit of 100 bp was used for the analysis of ORFs. Twenty-six ORFs (46.4%) encode predicted proteins of recognized function, while 30 (53.6%) were of unknown function. GBK1 ORFs that have homology to proteins of known function fall into five groups: phage packaging, phage structural proteins, DNA metabolism/replication, integration and lysis. All of the predicted ORFs are transcribed in the same direction along the genome. ORF 37 of GBK1 has homology to thymidylate synthases. We have cloned ORF 37 and shown it can complement growth of an *E. coli thyA* deletion mutant (Marks, Griffin and Hamilton, unpublished results).

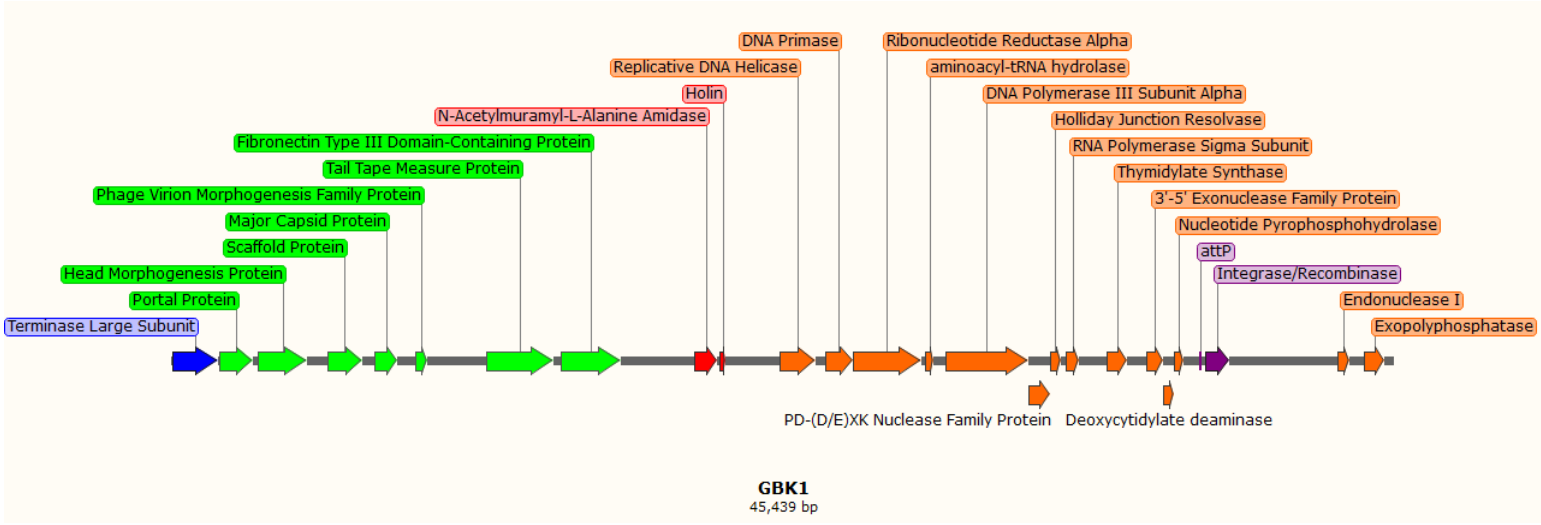

Figure 3-1 Genome arrangement of GBK1. The categories of functional ORFs are indicated by the following colors: blue, DNA packaging; green, structural; red, lysis; orange, DNA replication/metabolism; purple, integration. Only proteins with homology to proteins with known functions are shown. The GenBank file (accession no. MN727882) starts with base 1 at the beginning of the terminase large subunit – ORF 1. Map was generated using SnapGene from the annotated genome file.

Table 3-1

The predicted ORFs of GBK1 and their presumed functions.

ORF[1]	Coordinates	Predicted Function[2]	Significant Match [Organism] (Protein Sequence ID)	E value[3]
1	57-1698	Terminase- large subunit	Terminase [*Bacillus* phage vB_BpsS-36] (AYP68709.1)	0.0
2	1787-3055	Portal protein	Phage portal protein [*Bacillus* virus Riggi] (YP_008770568.1)	0.0
3	3211-5073	Head morphogenesis protein	Phage head morphogenesis protein [*Bacillus* phage vB_BpsS-36] (AYP68711.1)	0.0
4	5076-5372	Unknown	Hypothetical LG52_25 [*Geobacillus kaustophilus*] (KJE28512.1)	1E-61
5	5800-7110	Scaffold protein	Scaffold protein [*Bacillus* phage vB_BpsS-36] (AYP68714.1)	4E-121
6	7138-7551	Unknown	Hypothetical protein [*Enterococcus faecalis*] (WP_128712880.1)	2E-34
7	7578-8423	Major capsid protein	Hypothetical protein [*Bacillus* phage vB_BpsS-36] (AYP68716.1)	1E-145
8	8705-9100	Unknown	Hypothetical protein LG52_21 [*Geobacillus kaustophilus*] (KJE28465.1)	3E-87
9	9109-9543	Head morphogenesis protein	Phage head morphogenesis protein [*Geobacillus kaustophilus*] (KJE28708.1)	6E-84
10	9546-9977	Unknown	Hypothetical protein BpsS36_00015 [*Bacillus* phage vB_BpsS-36] (AYP68721.1)	3E-70
11	9994-10170	Unknown	Hypothetical protein LG52_18 [*Geobacillus kaustophilus*] (KJE26325.1)	8E-27
12	10170-11096	Unknown	Hypothetical protein LG52_17 [*Geobacillus kaustophilus*] (KJE28249.1)	0.0
13	11151-11498	Unknown	Hypothetical protein LG52_16 [*Geobacillus kaustophilus*] (KJE25723.1)	1E-76
14	11519-11740	Unknown	Hypothetical protein [*Geobacillus kaustophilus*] (WP_044730442.1)	1E-44
15	11749-14220	Tail tape measure protein	Hypothetical protein [*Bacillus megaterium*] (WP_098325516.1)	0.0
16	14502-16745	Unknown	Hypothetical protein [*Bacillus megaterium*] (WP_098325515.1)	0.0
17	16748-18847	Unknown	Hypothetical protein LG52_13 [*Geobacillus kaustophilus*] (KJE26789.1)	0.0
18	18847-19095	Unknown	Hypothetical protein LG52_12 [*Geobacillus kaustophilus*] (KJE26979.1)	4E-51
19	19111-19488	Unknown	Hypothetical protein LG52_11 [*Geobacillus kaustophilus*] (KJE28059.1)	6E-61

Table 3-1 [continued]				
20	19484-20329	N-acetylmuramyl-L-alanine amidase	N-acetylmuramyl-L-alanine amidase [*Halobacillus hunanensis*] (WP_079529067.1)	4E-88
21	20434-20646	Holin (Hypothesized)	Hypothetical protein LG52_9 [*Geobacillus kaustophilus*] (KJE25763.1)	1E-39
22	20879-21196	Unknown	Hypothetical protein LG52_8 [*Geobacillus kaustophilus*] (KJE26024.1)	9E-52
23	21198-21485	Unknown	Hypothetical protein LG52_7 [*Geobacillus kaustophilus*] (KJE27921.1)	8E-36
24	21678-22511	Phosphoesterase	Phosphoesterase [*Bacillus wiedmannii*] (WP_098115317.1)	5E-97
25	22644-23990	DNA Helicase	Replicative DNA helicase [*Bacillus* phage vB_BpsS-36] (AYP68738.1)	0.0
26	24340-25362	DNA Primase	DNA primase [*Bacillus* phage vB_BpsS-36] (AYP68739.1)	1E-147
27	25380-27938	Ribonucleoside-diphosphate reductase	Ribonucleoside-diphosphate reductase [*Geobacillus kaustophilus*] (KJE27215.1)	0.0
28	28063-28401	Aminoacyl-tRNA hydrolase	Hypothetical protein LG52_2969 [*Geobacillus kaustophilus*] (KJE26812.1)	3E-58
29	28817-31900	DNA polymerase III subunit alpha	DNA polymerase III subunit alpha [*Bacillus* phage vB_BpsS-36] (AYP68745.1)	0.0
30	31900-32730	PD-(D/E)XK nuclease	PD-(D/E)XK nuclease family protein [*Geobacillus kaustophilus*] (WP_044732478.1)	0.0
31	32726-33115	Resolvase	Holliday junction resolvase [*Geobacillus kaustophilus*] (KJE27036.1)	4E-77
32	32325-33789	RNA polymerase sigma factor	RNA polymerase sigma factor, sigma-70 family protein [*Geobacillus kaustophilus*] (KJE27454.1)	1E-102
33	33795-34100	Unknown	Hypothetical protein LG52_2975 [*Geobacillus kaustophilus*] (KJE27253.1)	1E-61
34	34097-34198	Unknown	Hypothetical protein LG52_2976 [*Geobacillus kaustophilus*] (KJE27477.1)	8E-13
35	34198-34389	Membrane protein	Putative Membrane Protein [*Geobacillus kaustophilus*] (KJE28344.1)	3E-14
36	34508-34819	Unknown	Hypothetical protein LG52_2978 [*Geobacillus kaustophilus*] (KJE27821.1)	3E-47
37	34835-35602	Thymidylate synthase	Thymidylate synthase [*Bacillus* phage vB_BpsS-36] (AYP68753.1)	3E-130
38	35602-35727	Unknown	Hypothetical protein LG52_2982 [*Geobacillus kaustophilus*] (KJE29169.1)	4E-19
39	35762-36304	Unknown	Hypothetical protein LG52_2983 [*Geobacillus kaustophilus*] (KJE27631.1)	2E-133
40	36304-36927	3'-5' exonuclease	3'-5' exonuclease family protein [*Geobacillus kaustophilus*] (KJE28783.1)	4E-102
41	36927-37346	Deoxycytidylate deaminase	Deoxycytidylate deaminase [*Geobacillus kaustophilus*] (WP_044732485.1)	1E-97

Table 3-1 [continued]				
42	37349-37672	mazG nucleotide pyrophosphohydrolase	Nucleotide pyrophosphohydrolase domain protein [*Geobacillus kaustophilus*] (KJE26390.1)	4E-58
43	37699-38055	Unknown	Bacterial PH domain protein [*Geobacillus kaustophilus*] (KJE28260.1)	7E-79
44	38485-39390	Integrase	Site-specific recombinase XerD [*Sporolactobacillus nakayamae*] (SFG68922.1)	4E-137
45	39527-39727	Unknown	Hypothetical protein LG52_41 [*Geobacillus kaustophilus*] (KJE28339.1)	9E-40
46	39847-40500	Unknown	Hypothetical protein LG52_40 [*Geobacillus kaustophilus*] (KJE27939.1)	2E-135
47	40533-40922	Unknown	Hypothetical protein LG52_39 [*Geobacillus kaustophilus*] (KJE28048.1)	1E-73
48	40915-41103	Unknown	Hypothetical protein LG52_38 [*Geobacillus kaustophilus*] (KJE27270.1)	2E-34
49	41099-41329	Unknown	Hypothetical protein LG52_37 [*Geobacillus kaustophilus*] (KJE25841.1)	1E-44
50	41606-41974	Unknown	Hypothetical protein LG52_36 [*Geobacillus kaustophilus*] (KJE27041.1)	4E-77
51	42448-42708	Unknown	Hypothetical protein LG52_35 [*Geobacillus kaustophilus*] (KJE26686.1)	5E-40
52	42698-42901	Unknown	Hypothetical protein LG52_34 [*Geobacillus kaustophilus*] (KJE25918.1)	2E-34
53	43406-43888	Endonuclease	Hypothetical protein LG52_33 [*Geobacillus kaustophilus*] (KJE27433.1)	1E-89
54	43888-44082	Unknown	Hypothetical protein LG52_32 [*Geobacillus kaustophilus*] (KJE27967.1)	1E-37
55	44039-44221	Unknown	Hypothetical protein LG52_31 [*Geobacillus kaustophilus*] (KJE26613.1)	2E-33
56	44430-45170	Exopolyphosphatase	Exopolyphosphatase domain protein [*Geobacillus kaustophilus*] (KJE25635.1)	2E-172

[1] ORFs numbered consecutively.
[2] Protein function based on amino acid sequence identity in BLAST analysis using the NCBI prokaryotic genome annotation pipeline.
[3] The probability of obtaining a match by chance as determined by BLAST analysis. Only values less than 10E-5 were considered significant.

3.3.2 *attP*/*attB* site in *G. kaustophilus* NBRC 102445. A 40 base pair site upstream of ORF 44 (integrase) with homology to a corresponding 40 base pair site in *Geobacillus kaustophilus* NBRC 102445 was also identified using PHASTER (Fig 3-2). This 40 base pair sequence in GBK1 is TTGGGGCATTAGCTCAGTTGGGAGAGCGTCGTGCTGGCAG. The same 40 base pair site is present in several other *Geobacillus* species (Table 3-2), including *G. thermodinitricans* T12, *G. thermoleovorans* SGAir0734, *G. stearothermophilus* 10, *G. kaustophilus* HTA426, and *Parageobacillus thermoglucosidasius* NCIMB 11955. In addition, other thermophilic non-*Geobacillus* species share the same homologous site as well, including *Anoxybacillus amylolyticus* DSM 15939 and *Bacillus caldolyticus* NEB414 (Table 3-2). The homologous site falls within a conserved tRNA-Ala gene in all cases and is flanked by a DnaD domain protein as well as a MgtC/SapB transporter protein in most cases (Table 3-2). Analysis of the *G. kaustophilus* NBRC 102445 site of GBK1 prophage integration revealed an *attP*/*attB* region of 22 base pairs, designated in black capital letters in Fig 3-2 as the overlap region.

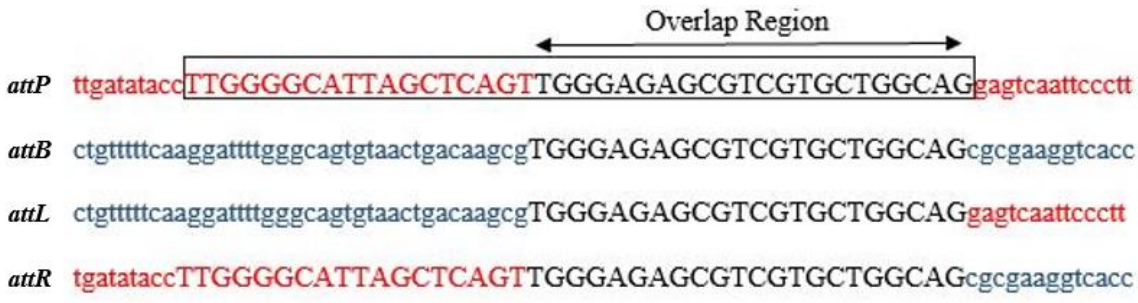

Figure 3-2 GBK1/Gbk NBRC 102445 DNA regions with *attP*, *attB*, *attL*, and *attR* sites as determined by sequencing. Homologous sequences where crossover occurs between phage and host are indicated by the overlap region, in upper case black letters. DNA sequences of phage origin and DNA sequences of bacterial origin are indicated by red and blue letters, respectively. The region designated by the black box in *attP* represents the 40 base pair sequence with homology to 21 other thermophilic organisms.

Table 3-2

Base pair homology between GBK1 and various thermophilic organisms as determined by a BLASTn search. All complements are designated with an asterisk.

Exact matches- 40/40 bp match (Accession number)	Homology Location	Upstream CDS to Homology	Downstream CDS to Homology
Geobacillus kaustophilus NBRC 102445 (CP038860.1)	tRNA-Ala	Hypothetical	DnaD Domain
**Geobacillus thermoleovorans* strain SGAir0734 (CP027303.2)	tRNA-Ala	DnaD Domain	MgtC/SapB Transporter
**Bacillus caldolyticus* strain NEB414 (CP025074.1)	tRNA-Ala	DnaD Domain	MgtC/SapB Transporter
**Geobacillus thermodenitrificans* strain ID-1 (CP017690.1)	tRNA-Ala	DnaD Domain	MgtC/SapB Transporter
Geobacillus lituanicus strain N-3 (CP017692.1)	tRNA-Ala	MgtC/SapB Transporter	Nuclease
Geobacillus thermodenitrificans strain T12 (CP020030.1)	tRNA-Ala	Hypothetical	Hypothetical
**Geobacillus thermodenitrificans* strain KCTC3902 (CP017694.1)	tRNA-Ala	DnaD Domain	MgtC/SapB Transporter
Geobacillus thermoleovorans strain FJAT-2391 (CP017071.1)	tRNA-Ala	MgtC/SapB Transporter	DnaD Domain
**Anoxybacillus amylolyticus* strain DSM 15939 (CP015438.1)	tRNA-Ala	Acetyltrans-ferase Domain	MgtC/SapB Transporter
**Geobacillus subterraneus* strain KCTC 3922 (CP014342.1)	tRNA-Ala	DnaD Domain	MgtC/SapB Transporter
Geobacillus thermoleovorans strain KCTC 3570 (CP014335.1)	tRNA-Ala	MgtC/SapB Transporter	DnaD Domain
Geobacillus sp. JS12, complete genome (CP014749.1)	tRNA-Ala	MgtC/SapB Transporter	Hypothetical
Geobacillus stearothermophilus 10 (CP008934.1)	tRNA-Ala	MgtC/SapB Transporter	DnaD Domain
Geobacillus sp. LC300 (CP008903.1)	tRNA-Ala	MgtC/SapB Transporter	Nuclease
Geobacillus sp. 12AMOR1 (CP011832.1)	tRNA-Ala	MgtC/SapB Transporter	SPBc2 prophage endonuclease YokF
Geobacillus sp. GHH01 (CP004008.1)	tRNA-Ala	Hypothetical	Hypothetical
Geobacillus thermoleovorans CCB_US3_UF5 (CP003125.1)	tRNA-Ala	MgtC/SapB Transporter	DnaD Domain
Geobacillus sp. Y412MC52 (CP002442.1)	tRNA-Ala	MgtC/SapB Transporter	Nuclease
**Geobacillus* sp. C56-T3 (CP002050.1)	tRNA-Ala	DnaD Domain	MgtC/SapB Transporter
Geobacillus sp. Y412MC61 (CP001794.1)	tRNA-Ala	MgtC/SapB Transporter	

Table 3-2 [continued]			
Geobacillus thermodenitrificans NG80-2 (CP000557.1)	tRNA-Ala	SapB	Hypothetical
Geobacillus kaustophilus HTA426 DNA (BA000043.1)	tRNA-Ala	Hypothetical	Hypothetical
Close matches- 39/40 bp match (Accession number)	Homology Location	Upstream CDS to Homology	Downstream CDS to Homology
Parageobacillus thermoglucosidasius strain TM242 (CP016916.1)	tRNA-Ala	MgtC/SapB Transporter	NUDIX hydrolase
Parageobacillus thermoglucosidasius strain NCIMB 11955 (CP016622.1)	tRNA-Ala	MgtC/SapB Transporter	NUDIX hydrolase
Parageobacillus thermoglucosidasius strain DSM 2542 (CP012712.1)	tRNA-Ala	MgtC/SapB Transporter	NUDIX hydrolase
**Parageobacillus thermoglucosidasius* C56-YS93 (CP002835.1)	tRNA-Ala	NUDIX hydrolase	MgtC/SapB Transporter
**Geobacillus* sp. Y4.1MC1 (CP002293.1)	tRNA-Ala	Hypothetical	MgtC/SapB Transporter
Geobacillus sp. WCH70 (CP001638.1)	tRNA-Ala	MgtC/SapB Transporter	Nuclease
Close matches- 38/40 bp match (Accession number)	Homology Location	Upstream CDS to Homology	Downstream CDS to Homology
Aeribacillus pallidus strain KCTC3564 (CP017703.1)	tRNA-Ala	Hypothetical	Hypothetical
Geobacillus genomosp. 3 strain JF8 (CP006254.2)	tRNA-Ala	Methyltransferase	Hypothetical
Close match- 38/41 bp match (Accession number)	Homology Location	Upstream CDS to Homology	Downstream CDS to Homology
**Anaerococcus prevotii* DSM 20548 (CP001708.1)	tRNA-Ala	Hypothetical	MATE efflux family

E values of all 40/40 exact matches are 6e-11. E values of all 39/40 close matches are 2e-10. E values of all 38/40 close matches are 7e-10. E values of 38/41 close match is 6e-06.

3.3.2 Phylogeny based on multiple genetic loci. The amino acid sequence of the large terminase subunit is commonly used to compare and classify phage and is considered one of the most universally conserved gene sequences in phage (Casjens, 2003); it was chosen for comparative analysis (Fig. 3-3). GBK1 shares closest homology to *Bacillus* phage vB BpsS-36 (Akhwale et al., 2019; accession MH884513.1) based on terminase large subunit similarities (Fig. 3-3). We also included phylogenetic analysis based on the tyrosine integrase found in GBK1 (Fig. 3-4). Similarities between tyrosine integrase genes indicate a close phylogenetic relationship between GBK1 and a prophage sequence in *Sporolactobacillus nakayamae* (accession WP_093673392.1).

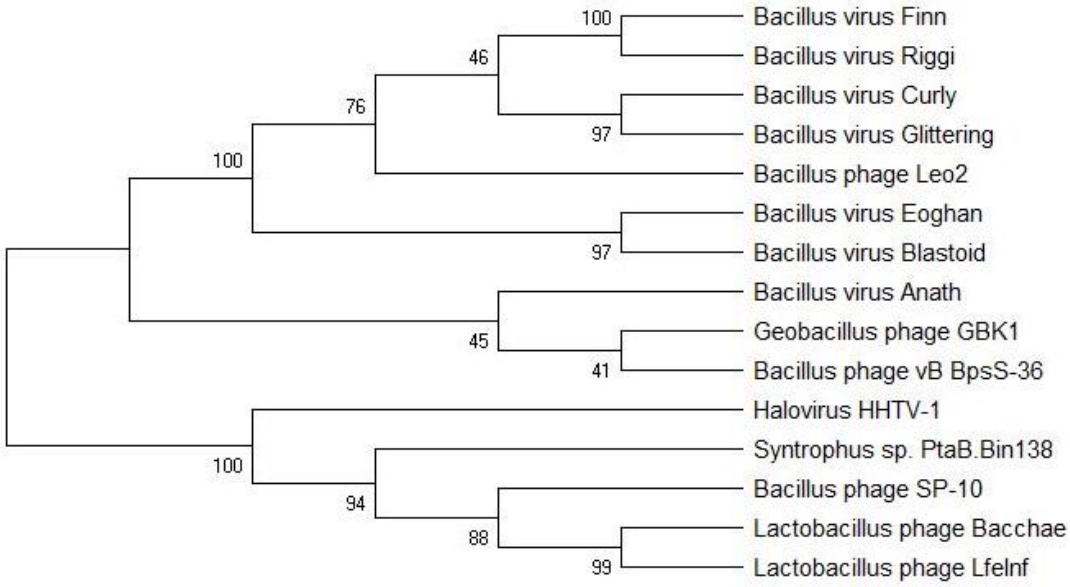

Figure 3-3 Rooted maximum parsimony tree based on the aligned amino acid sequences of the large terminase subunit gene of phage GBK1 and 14 other large terminase genes from diverse phage genomes. The numbers at the nodes represent bootstrap values based on 1,000 resamplings. Analysis was performed using the program MegaX (Kumar et al., 2018).

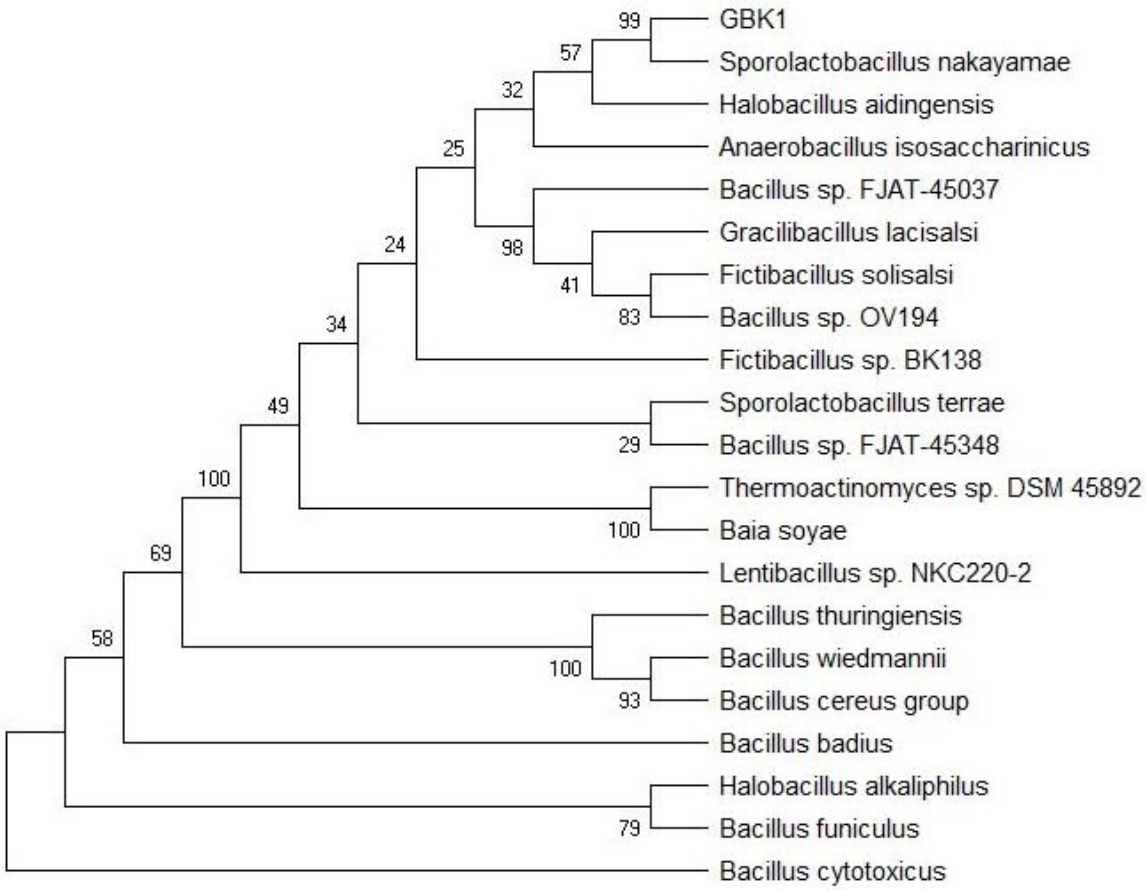

Figure 3-4 Rooted maximum parsimony tree based on the aligned amino acid sequences of the integrase gene of phage GBK1 and 20 other integrase genes from diverse prophage integrated in host bacterial genomes. The numbers at the nodes represent bootstrap values based on 1,000 resamplings. Analysis was performed using the program MegaX (Kumar et al., 2018).

3.3.3 Prophage present in *G. kaustophilus* NBRC 102445 genome. PHASTER identified two regions in the *G. kaustophilus* ATCC 8005/NBRC 102445 genome that correspond to prophage sequences, one of which is the GBK1 sequence with the anticipated *attL* and *attR* sequences shown in Fig. 3-2. Region 1 was considered by PHASTER to be a complete prophage (score of 150), with sequence homologs to phage recombinase, integrase, terminase, portal, head, capsid, tail, lysin, and transposase proteins. The *attL*/*attR* sequences identified by PHASTER were CCGCTTTTTATG with a homolog to a serine recombinase downstream of the *attL* sequence. An integrase with homology to other tyrosine integrases was located following the DNA replication/metabolism genes. The DNA packaging genes, phage structural genes, and lysis genes followed the replication/metabolism genes. Sixty total proteins were identified, with 13 of those having close homology to *Thermus* phage OH2 (Doi et al., 2013); accession AB823818). One CDS of note from this prophage had homology to the *Thermus* phage OH2 prophage antirepressor, which was not present in the GBK1 identified ORFs. Also of note is that the third highest match for this prophage (6 CDS) was a *Bacillus* phage, BtCS33 (accession NC_018085). This match included the small and large terminase subunits.

Regions 2 and 3 were prophages with scores under 70 and were disregarded. Region 4 represented the GBK1 sequence and was considered by PHASTER to be questionable (score of 70), with sequence homologs to phage integrase, terminase, and capsid. If the total number of CDS of the prophage region does not correlate with 100% to the phages found in the NCBI Virus database, which contains 9,222 complete viral genomes to date, PHASTER assigns an incomplete score. The score of 70 for GBK1 is based on 10 points each for having homology to a phage integrase, a phage terminase, and a phage capsid. Additionally, 10

points for each criteria were awarded because the prophage region is greater than 30 kb, there are at least 40 proteins in the region, and all of the phage-related proteins and hypothetical proteins constitute more than 70% of the total number of proteins in the region. Fifty-five total proteins were identified, with 23 of those having close homology to the following *Bacillus* phage: Andromeda (Lorenz et al., 2013; accession NC_020478), Eoghan (Lorenz et al., 2013; accession NC_020477), and Curly (Lorenz et al., 2013; accession NC_020479). *Bacillus* phage Finn (Lorenz et al., 2013; accession NC_020480), Riggi (Still et al., 2013; accession NC_022765), and Glittering (Matthew et al., 2013; accession NC_022766) also had homology of 22, 22, and 21 proteins respectively.

Another *G. kaustophilus* whole genome shotgun sequence in the Genbank database, Et7/4 (accession JYBP01000003.1), contains a prophage with 94% similarities to GBK1. PHASTER identified a prophage region from 748-35852 with close homology to the *Bacillus* phage Finn (Lorenz et al., 2013; accession NC_020480), similar to our results for GBK1 in *G. kaustophilus* ATCC 8005/NBRC 102445. Additionally, the 22 base pair overlap region found in GBK1 is present in the Et7/4 sequence flanking the integrase, which is further evidence that the Et7/4 strain contains GBK1 as a prophage.

3.3.4 Genome Comparison of GBK1 and *Bacillus* Phage vB_BpsS-36. Figure 3-5 shows the similarities between the thermophilic GBK1 and mesophilic *Bacillus* phage vB_BpsS-36. There was one Locally Collinear Block (LCB) identified with a weight of 47034. This LCB, which is shown in red, occurs between several genes, including at least parts of the terminase large subunit, portal protein, head morphogenesis protein, scaffold protein, major capsid protein, phage virion morphogenesis family protein, tail tape measure protein, N-acetylmuramyl-LAlanine amidase, holin (hypothesized), phosphoesterase,

replicative DNA helicase, DNA primase, ribonucleotide reductase alpha, DNA polymerase III subunit alpha, PD-(D/E)XK nuclease family protein, RNA polymerase, thymidylate synthase, 3'-5' exonuclease family protein, deoxycytidylate deaminase, nucleotide pyrophosphohydrolase, and exopolyphosphatase. In addition, the G+C content of GBK1 (43.6%) is similar to the G+C content of *Bacillus* phage vB_BpsS-36 (41.1%). Notably absent in the LCB is the GBK1 *attP*/integrase region, which does not have homology to any *Bacillus* phage vB_BpsS-36 genes. The GBK1 *attP*/integrase shows greatest homology to the site-specific recombinase XerD in *Sporolactobacillus nakayamae* (accession SFG68922.1).

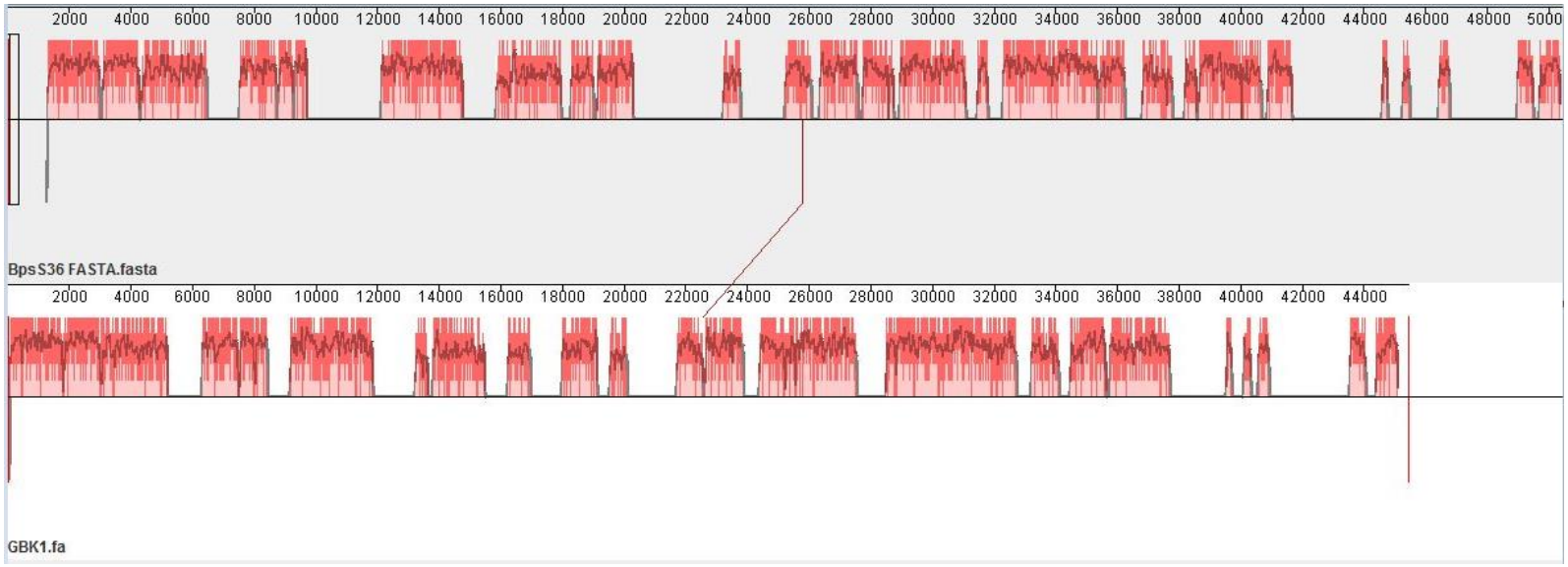

Figure 3-5 ProgressiveMauve comparison of *Bacillus* phage vB_BpsS-36 genome with GBK1 genome. Similarities between the two genomes are shown in red. Default seed weights and HOXD (default) scoring were used to compare the two genomes.

3.4 Discussion

This study described a novel *Geobacillus* bacteriophage GBK1 isolated from a compost pile in Cary, North Carolina, USA. This phage is lysogenic and infects *G. kaustophilus* ATCC 8005/NBRC 102445. Here we have described features of this lysogenic phage as well as the attachment site in the phage/host. Some ORFs of note are ORF 29 (DNA polymerase III), ORF 37 (Thymidylate synthase), and ORF 44 (Integrase) (Table 3-1). Additionally, we show that GBK1 is closely related to a mesophilic *Bacillus* phage, vB_BpsS-36 (Fig 3-5).

Geobacillus species have great potential as a source of thermostable enzymes as well as for production of industrially important products, including ethanol (Cripps et al., 2009). Some species of *Geobacillus* are also able to metabolize C5 sugars, including xylose (Cordova and Antoniewicz 2016; Cripps et al., 2009), cellobiose (Bartosiak-Jentys et al., 2013; Cripps et al., 2009; Lin et al., 2014; Assareh et al., 2012), and lignocellulose (Bhalla et al., 2015; Cripps et al., 2009; Potprommanee et al., 2017). The main drawback to using *Geobacillus* spp. for these processes is that there is a lack of tools for genetic manipulation (Kananavičiūtė and Čitavičius, 2015). There are numerous examples of integrases being used to develop suicide vectors (McShan et al., 1998; Du et al., 2015; Pargellis et al., 1988). This could be useful in *Geobacillus* spp. because there are very few vectors available with few origins of replication (Kananavičiūtė and Čitavičius, 2015). Development of a suicide vector would avoid use of an origin of replication in *Geobacillus* spp., allowing for the integration of a gene directly into the host genome. This would also allow for multiple genetic manipulations as a vector with one of the two commonly-used *Geobacillus* origin of replications could still be utilized in those cells that harbor the integrated suicide vector.

Future work will involve developing a vector capable of integrating into different *Geobacillus* host species based on the tyrosine integrase/recombinase and the *attP*/*attB* homologous sites. In addition, isolation of *thyA* mutants of *G. thermoleovorans*, *G. kaustophilus*, and *G. thermoglucosidasius* and the development of GBK1 ORF 37 as a selectable marker in these *Geobacillus* spp. are in progress (Marks and Hamilton, in preparation).

Several ORFs identified in GBK1 had closest homology to a mesophilic phage, *Bacillus* phage vB_BpsS-36 (Table 3-1: ORFs 1, 3, 5, 7, 10, 25, 26, 29, and 37). The order and arrangement of the genes in each phage is very similar, particularly in the DNA replication/modification region. When skipping the hypothetical ORFs, the central region in both phage is helicase-primase-ribonucleotide reductase-DNA polymerase III alpha subunit-PD-(D/E) XK nuclease-RNA polymerase sigma factor-thymidylate synthase. The order is not 100% identical, however. For example, GBK1 has the Holliday junction resolvase but vB_BpsS-36 does not. Most of the nonhomologous regions between phage vB_BpsS-36 are intragenic, which is not surprising because the noncoding regions do not have the same selective pressure as the coding regions. There are also several nonhomologous proteins, most of which are of unknown function, that GBK1 has most likely picked up from other sources. Additionally, most of the coding regions that are nonhomologous between vB_BpsS-36 and GBK1 are related to GBK1's ability to infect *Geobacillus kaustophilus*, including the *attP* and integrase regions (integration into the *G. kaustophilus* host), cell wall-binding domain of lysin (ability to bind to the *G. kaustophilus* cell wall), and changes in portions of two tail proteins: a portion of the tape measure protein (ORF 15) and a portion of

the tail fiber protein (ORF 16). These changes may represent a change in the phage tail proteins to allow it to bind to a receptor on the surface of *G. kaustophilus*.

Because genomes can undergo many changes over the course of evolution, it can be useful to look at local changes like nucleotide substitutions and indels as well as large-scale changes such as gene loss, duplication, rearrangement, and horizontal transfer. Mauve was designed to look at genome alignments in order to identify evolutionary changes by aligning homologous regions of sequence (Darling et al., 2004). Analysis using progressiveMauve (Darling et al., 2004) identified one LCB spanning several ORFs between the two phages (Fig 3-5).

Bezuidt et al. (2016) found that the core cluster of the *Geobacillus* genome, genes conserved in all the *Geobacillus* genomes currently available, displayed strong homology to the *Bacillus* core genome, indicating an evolutionary relationship between *Bacillus* species and *Geobacillus* species. Ecologic diversity in bacteria is typically linked with microevolutionary events such as horizontal gene transfer (Boto 2010). For *Geobacillus* species, ecological niche adaptation, including thermophilicity, appears to have involved HGT from a wide variety of bacterial and archael genes (Bezuidt et al., 2016). Key factors to the evolution of *Geobacillus* from *Bacillus* are thought to be due to environmental pressures and adaptation to niche environments (Alalouf et al., 2011). In analyzing the genome of GBK1, we were struck by the similarities with the genome of the *Bacillus* phage vB_BpsS-36. It would appear that the evolution of GBK1 mirrors the evolution of its host. As *Geobacillus kaustophilus* evolved from its *Bacillus* ancestor, GBK1 evolved in parallel from its ancestral *Bacillus* phage. Other *Geobacillus* and *Bacillus* phage show similar relationships, including GBK2 with *Bacillus* phage SPP1 (Marks and Hamilton, 2014) and

Geobacillus phage GVE3 with *Bacillus* anthracis phage vB_BanS_Tsamsa (van Zyl et al., 2015). However, this is not the only path for *Geobacillus* phage evolution. *Geobacillus* phages GVE2 and D6E share significant homology between their genomes (Wang and Zhang, 2010).

Two other ORFs of interest in GBK1 are the predicted lysin (ORF 20) and holin (ORF 21). ORF 20 in GBK1 has an acetylmuramyl-L-alanine amidase catalytic domain (Fig 3-6: 1-188 in green), a short linker and then a peptidoglycan-binding domain (SPOR) comprised of two tandem ~35 residue repeats (Fig 3-6: repeats are underlined; exact matches between repeats are highlighted in yellow and conserved amino acids are highlighted in grey). There are five types of functional lysin catalytic domains: Endo-β-N-acetylglucosaminidase, N-acetylmuramidase (lysozyme-like), endopeptidase, N-acetylmuramoyl-L-alanine amidase (T7-like), and γ-D-glutaminyl-L-lysine endopeptidase. ORF 20 falls under the N-actetylmuramoyl-L-alanine amidase classification, which hydrolyzes the amide bond connecting the glycan strand and peptide moieties (Loeffler et al., 2001). The amidase is responsible for degrading the peptidoglycan layer of the host. The two domains of the GBK1 lysin have different top hits for homology when run in BLAST, which is consistent with the theory of phage lysin evolution that describes modular swapping of lytic domains and cell wall-binding domains (Fischetti, 2008). Lysins have been developed as tools for lysing Gram-positive organisms (Fischetti, 2008), and while *Geobacillus* are not pathogenic for humans, there have been issues with Geobacilli in food storage and spoiling (Rigaux et al., 2013). Additionally, the peptidoglycan-binding domain may be useful in surface display systems.

VTKYIALDDGHGMETAGKRTPYIPSLGRQIKENEFNRAVVKFLDQELKRCGFRTL
LVAPTDVDTPLKERTDKANKAGVDAYISIHYNAFDGTFAGKNPEGFQAHVYLGHS
NKKAGKLAKCILKHLAGGTKQVNRGLHESNFHVLRETHMPAVLLELGFMDNEREA
LLMINTDFQKECAREIAQGICEYFGVKYVPEKKANPPKTTTPTKTFYRVVTGSFK
EKANAERRVAELKRKGFDSFIDYKNGWYRVITGSFQDRGNAKKRMAELKRAGFDS
FLESHKA

Figure 3-6 Amino acid sequence of the acetylmuramyl-L-alanine amidase in GBK1. The catalytic domain is shown in green, tandem repeats are underlined, exact matches between repeats are in yellow, and conserved amino acids are in grey.

The gene following the lysin is likely a holin (ORF 21). Holins do not typically share sequence similarity, although there are common characteristics between these genes amongst phage (Shi et al., 2012). Shi et al. (2012) identify three criteria for holins: most holin genes are adjacent to the endolysin gene, at least one hydrophobic transmembrane domain (TMD) occurs in all holins, and holins have highly charged, hydrophilic, C-terminal domains. For GBK1, ORF 21 immediately follows the lysin. In addition, ORF 21 has two predicted transmembrane domains according to TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) and PredictProtein (Yachdav et al., 2014). Finally, the last eight amino acids of ORF 21 in GBK1 are charged (K, K, H, E, K, E, Q, K). The function of holin is to insert in the membrane and form pores to allow the lysin to have access to the peptidoglycan.

The genome sequence of GBK1, a unique bacteriophage that infects the thermophile *Geobacillus kaustophilus,* provides useful building blocks for the development of genetic tools to characterize and manipulate *Geobacillus* species, in addition to increasing our understanding of thermophiles. The complete genome sequence of GBK1 is available in the GenBank database under accession number MN727882.

Acknowledgements

The authors would like to thank the Genomic Sciences Laboratory (GSL) at NCSU for assistance in sequence determination and assembly.

References

Akhwale JK, Rohde M, Rohde C, Bunk B, Spröer C, Klenk HP, Boga HI, Wittmann J (2019) Comparative genomic analysis of eight novel haloalkaliphilic bacteriophages from Lake Elmenteita, Kenya. PloS one.14;14(2):e0212102

Alalouf O, Balazs Y, Volkinshtein M, Grimpel Y, Shoham G, Shoham Y (2011) A new family of carbohydrate esterases is represented by a GDSL hydrolase/acetylxylan esterase from *Geobacillus stearothermophilus*. J Biol Chem. 286: 41993-42001. Doi:10.1074/jbc.M111.301051

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. 215:403-410

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic acids research, 44(W1), W16-W21

Assareh R, Zahiri HS, Noghabi KA, Aminzadeh S (2012) Characterization of the newly isolated *Geobacillus* sp. T1, the efficient cellulase-producer on untreated barley and wheat straws. Bioresource technology. 120:99-105

Bartosiak-Jentys J, Hussein AH, Lewis CJ, Leak DJ (2013) Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. Microbiology. 159(7):1267-75

Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29: 2607-2618

Bezuidt OK, Pierneef R, Gomri AM, Adesioye F, Makhalanyane TP, Kharroub K, Cowan DA (2016) The *Geobacillus* pan-genome: implications for the evolution of the genus. Front Microbiol. 7:723. Doi: 10.3389/fmicb.2016.00723

Bhalla A, Bischoff KM, Sani RK (2015) Highly thermostable xylanase production from a thermophilic *Geobacillus* sp. strain WSUCF1 utilizing lignocellulosic biomass. Frontiers in bioengineering and biotechnology. 3:84

Boto L (2010) Horizontal gene transfer in evolution: facts and challenges. Proc Biol Sci 277, 819-827. Doi: 10.1098/rspb.2009.1679

Casjens S (2013) Prophages and bacterial genomics: what have we learned so far? Molecular microbiology. 49(2):277-300

Coorevits A, Dinsdale AE, Halket G, Lebbe L, De Vos P, Van Landschoot A, Logan NA (2012) Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus, G. stearothermophilus, G. jurassicus, G. toebii, G. thermodenitrificans* and *G. thermoglucosidans* (nom. corrig., formerly 'thermoglucosidasius'); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. Int. J. Syst. Evol. Microbiol. 62, 1470–1485

Cordova LT, Antoniewicz MR (2016) 13C metabolic flux analysis of the extremely thermophilic, fast growing, xylose-utilizing *Geobacillus* strain LC300. Metabolic engineering.33:148-57

Cripps RE, Eley K, Leak DJ, Rudd B, Taylor M, Todd M, Boakes S, Martin S, Atkinson T (2009). Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. Metab. Eng. 11, 398–408

Cuebasg M, Sannino D, Bini B (2011). Isolation and characterization of an arsenic resistant *Geobacillus kaustophilus* strain from geothermal soils. J of Basic Microbiology. 51:364 – 371

Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple sequence alignment of conserved genomic sequences with rearrangements. Genome research 14(7), 1394-1403

Delcher AL, Bratke KA, Powers EC, Salzber SL (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinforma 23: 673-679

Doi K, Mori K, Martono H, Nagayoshi Y, Fujino Y, Tashiro K, Kuhara S, Ohshima T (2013) Draft genome sequence of *Geobacillus kaustophilus* GBlys, a lysogenic strain with bacteriophage ϕOH2. Genome Announc. 1(4):e00634-13. doi:10.1128/genomeA.00634-13

Du D, Wang L, Tian Y, Liu H, Tan H, Niu G (2015) Genome engineering and direct cloning of antibiotic gene clusters via phage ϕBT1 integrase-mediated site-specific recombination in *Streptomyces*. Scientific reports. 5:8740

Feng L, Wang W, Cheng J, Ren Y, Zhao G, Gao C, Tang Y, Lui X, Han W, Peng X, Liu R, Wang L (2007) Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. Proc Natl Acad Sci. U.S.A. 104:5602-5607

Fischetti V (2008) Bacteriophage lysins as effective antibacterials. Current Opinions in Micro. 11(5): 393-400

Hussein AH, Lisowska BK, Leak DJ (2015) The Genus *Geobacillus* and Their Biotechnological Potential. Adv App Microbiol 92: 1-48.

Kananavičiūtė R and Čitavičius D (2015). Genetic engineering of *Geobacillus* spp. Journal of microbiological methods, 111, 31-39

Kropinski A, Mazzocco A, Waddell T, Lingohr E, Johnson R (2009) Enumeration of bacteriophages by double agar overlay plaque assay. In: Clokie M, Kropinski A (eds) Bacteriophages: methods and protocols. Humana Press, New York, pp 69-76

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution, 35(6), 1547-1549

Lin PP, Rabe KS, Takasumi JL, Kadisch M, Arnold FH, Liao JC (2014). Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*. Metabolic engineering. 24:1-8

Loeffler JM, Nelson D, Fischetti VA (2001) Rapid killing of *Streptococcus pneumonia* with a bacteriophage cell wall hydrolase. Science. 294 (5549): 2170-2172. Doi:10.1126/science.1066869

Lorenz L, Lins B, Barrett J, Montgomery A, Trapani S, Schindler A, Christie GE, Cresawn SG, Temple L (2013) Genomic characterization of six novel *Bacillus pumilus* bacteriophages. Virology. 444(1-2):374-83

Marks TJ, Hamilton PT (2014) Characterization of a thermophilic bacteriophage of *Geobacillus kaustophilus*. Arch of virology. 159(10):2771-5

Matthew SP, Decker SL, Chamakura KR, Everett GF (2013) Complete genome of *Bacillus pumilus* siphophage Glittering. Genome Announce. 1(6):e00856-13

McMullen G, Christie JM, Rahman TJ, Banat IM, Ternan NG, Marchant R (2004) Habitat, applications, and genomics of the aerobic, thermophilic genus *Geobacillus*. Biochem Soc Trans 32:214-217

McShan WM, McLaughlin RE, Nordstrand A, Ferretti JJ (1998) Vectors containing streptococcal bacteriophage integrases for site-specific gene insertion. Methods for studying the genetics, molecular biology, physiology, and pathogenesis of the streptococci. pp. 51-57. Springer, Dordrecht

Mead DA, McClary JA, Luckey JA, Kostichka AJ, Witney FR, Smith LM (1991) Bst DNA polymerase permits rapid sequence analysis from nanogram amounts of template. Biotechniques 11, 76–78

Nazina TN, Tourova TP, Poltaraus AB, Novikova EV, Grigoryan AA, Ivanova AE, Lysenko AM, Petrunyaka VV, Osipov GA, Belyaev SS, Ivanov MV (2001) Taxonomic study

of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus, Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G. kaustophilus, G. thermoglucosidasius* and *G. thermodenitrificans*. Int. J. Syst. Evol. Microbiol. 51, 433–446

Omokoko B, Jäntges UK, Zimmermann M, Reiss M, Hartmeier W (2008) Isolation of the phe-operon from *G. stearothermophilus* comprising the phenol degradative meta-pathway genes and a novel transcriptional regulator. BMC Microbiol. 8, 197. http://dx.doi.org/10.1186/1471-2180-8-197

Pargellis CA, Nunes-Düby SE, De Vargas LM, Landy A (1988) Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. Journal of Biological Chemistry. 263(16):7678-85

Potprommanee L, Wang XQ, Han YJ, Nyobe D, Peng YP, Huang Q, Liu JY, Liao YL, Chang KL (2017) Characterization of a thermophilic cellulase from *Geobacillus* sp. HTA426, an efficient cellulase-producer on alkali pretreated of lignocellulosic biomass. PLoS one. 12(4):e0175004

Rigaux C, André S, Albert I, Carlin F (2014) Quantitative assessment of the risk of microbial spoilage in foods. Prediction of non-stability at 55 ˚C caused by *Geobacillus stearothermophilus* in canned green beans. Int J of Food Micro. 171:119-128

Sambrook JF and Russell DW (2001) Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press

Shi Y, Yan Y, Ji W, Du B, Meng X, Wang H, Sun J (2012) Characterization and determination of holin protein of *Streptococcus suis* bacteriophage SMP in heterologous host. Virology Journal. 9(1):70

Still EL, Riggi CF, Chamakura KR, Everett GF (2013) Complete genome of *Bacillus pumilus* siphophage Riggi. Genome Announce. 1(6):e00861-13

Taylor MP, Esteban CD, Leak DJ (2008). Development of a versatile shuttle vector for gene expression in *Geobacillus* spp. Plasmid. 60:45-52

Taylor MP, Eley KL, Martin S, Tuffin MI, Burton SG, Cowan DA (2009). Thermophilic ethanologenesis: Future prospects for second-generation bioethanol production. Trends in Biotechnology. 27(7):398-405

Taylor MP, van Zyl L, Tuffin IM, Leak DJ, Cowan DA (2010). Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts. Microb Biotechnology. 4(4):438-448

van Zyl LJ, Sunda F, Taylor MP, Cowan DA, Trindade MI (2015) Identification and characterization of a novel *Geobacillus thermoglucosidasius* bacteriophage, GVE3. Arch of virology. 160(9):2269-82

Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Hönigschmid P, Schafferhans A, Roos M, Bernhofer M, et al. (2014) PredictProtein-an open resource

for online prediction of protein structural and functional features. Nucleic Acids research. Gku366

Zeigler DR (2014) The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? Microbiology 160 (1), 1–11

CHAPTER 4

Development of an integrating suicide vector and counterselection system for *Geobacillus* spp.

Marks, Timothy J.[AB] and Hamilton, Paul T.[B]

[A]Department of Pharmaceutical Sciences, Campbell University, Buies Creek, NC 27506

[B] Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695

Abstract

Members of the genus *Geobacillus* are thermophilic organisms of great biotechnological potential because of their metabolic diversity and their production of industrially relevant thermostable enzymes and products. However, a lack of genetic engineering tools has hindered efforts to use these organisms to their full potential. Bacteriophages that infect *Geobacillus* spp. represent an untapped resource for the development of genetic manipulation tools in these organisms. We have isolated and sequenced the genome for the lysogenic phage GBK1 from *Geobacillus kaustophilus* ATCC 8005 (the type strain). BLAST searches of GBK1 revealed an integrase/recombinase and a thymidylate synthase gene that were potential targets for development of new tools for *Geobacillus* spp. In this paper, we describe a novel integrating suicide vector, pTM2, for *Geobacillus thermoglucosidasius* NCIMB 11955 as well as an antibiotic-free counterselection system based around complementation of a *thyA* mutant of *Geobacillus thermoleovorans* DSM 5366 using replicating vector pTM5 and integrating suicide vector pTM6.

4.1 Introduction

Members of the genus *Geobacillus* are Gram-positive, thermophilic, aerobic or facultatively anaerobic microbes that were reclassified from the genus *Bacillus* in 2001 (Suzuki et al., 2012; Nazina et al., 2001). *Geobacillus* spp. are industrially relevant due to their ability to metabolize C5 and C6 sugars (Coorevits et al., 2012; Nazina et al., 2001; Omokoko et al., 2008; Zeigler, 2014), namely xylose, cellobiose, and lignocellulose. They have also shown arsenate resistance (Cuebas et al., 2011), ethanol tolerance (Fong et al., 2006), and the ability to degrade long-chain alkanes (Wang et al., 2006). There are numerous advantages to manufacturing products at high temperatures, including the easy removal of volatile products (such as ethanol), reduced risk of contamination by microbial competitors, and reduced energy expenditure for agitation and cooling (Weigel et al., 1986). *Geobacillus* spp. can be used to produce thermostable enzymes, such as amylases, lipases (McMullan et al., 2004), DNA polymerase (Mead et al., 1991) and proteases. There has also been research published on the production of ethanol (Cripps et al., 2009) and isobutanol (Lin et al., 2014), as *Geobacillus* species are mixed acid fermenters. Unfortunately, there is a lack of gene transfer and protein expression tools in *Geobacillus* spp. due in part to difficulties with transformation (Kananavičiūtė and Čitavičius 2015).

Transformation into thermophilic Gram-positive organisms is problematic due to low permeability of the plasma membrane and the thick peptidoglycan layer (Lin and Xu, 2013). Additionally, restriction-modification systems present in the host have to be considered in order to attain effective transformation, as some *Geobacillus* spp. degrade DNA that is not properly methylated (Suzuki and Yoshida, 2012). The size of the plasmid, the replicon, and the selection marker used all have effects on transformation efficiency (Liao and Kanikula,

1990; Narumi et al., 1992; Turgeon et al., 2006). Lack of reliability of antibiotic selection at higher temperatures is another issue that can pose problems when using thermophilic organisms like *Geobaicllus* spp. For example, Peteranderl et al. (1990) found that ampicillin and streptomycin had greatly reduced ability to inhibit growth of *Clostridium thermohydrosulfuricum* JW 102 at 50 ˚C within 18 h as opposed to kanamycin, which was still inhibitory under the same conditions. Additionally, there are limited thermostable antibiotic resistance markers which further exacerbate the problem of antibiotic selection in these thermophilic organisms.

The genetic diversity among phages represents a source of potential tools that can be developed to exchange, alter and express genes and gene products from an organism of interest. We have isolated and sequenced a lysogenic phage, GBK1 (Accession number MN727882), that infects *Geobacillus kaustophilus* ATCC 8005. GBK1 BLAST results revealed an integrase/recombinase (along with an *attP*/*attB* site) and a *thyA* gene encoding for thymidylate synthase (unpublished results). Thymidylate synthase is the enzyme responsible for the reductive methylation of deoxyuridylic acid (dUMP) to deoxythymidylic acid (dTMP). Therefore, *thyA* deficiency confers thymine auxotrophy, permitting counterselection using a *thyA* marker in *thyA*$^{-}$ mutant strains. Antibiotic-free systems depend on the complementation of auxotrophic marker genes that are essential in a metabolic pathway, but are missing in a mutant host strain (Kananavičiūtė and Čitavičius, 2015). Common complementation systems used in various microorganisms are involved in the synthesis of essential amino acids or nucleotides, such as thymidine, tryptophan, and uracil (Inoue and Sako, 2013; Taylor et al., 2011). Both the integrase and thymidylate synthase enzymes represent potential tools for genetic engineering in *Geobacillus* spp. Here we

present the development of an integrating suicide vector for *Geobacillus thermoglucosidasius* NCIMB 11955 and an antibiotic-free counterselection system based around thymidylate synthase in a *thyA*$^-$ mutant of *Geobacillus thermoleovorans* DSM 5366. Each of these tools addresses the need for an expanded repertoire of applications in the *Geobacillus* genus for further genetic manipulation efforts in these useful organisms.

4.2 Methods

4.2.1 Bacterial strains, culture conditions, plasmids, primers. Table 4-1 lists the strains and plasmids used in this study. Table 4-2 lists the primers used. *E. coli* strains were grown at 37˚C in Luria-Bertani (LB) medium supplemented with 100 µg/mL ampicillin when necessary. *G. thermoglucosidasius* and *G. thermoleovorans* were grown at 55 ˚C in TGP broth or minimal glucose + casamino acids (MgCas) media supplemented with 10 µg/mL kanamycin, 10 µg/mL trimethoprim, or 50 µg/mL thymine when necessary. TGP contained (per liter deionized water) 17 g tryptone (Difco Bacto), 3 g soy peptone (Sigma), 5 g NaCl, and 2.5 g K_2HPO_4. The pH was adjusted to 7.3 pre-autoclaving, and then 4 g sodium pyruvate and 4 mL glycerol were added post-autoclaving as a filter-sterilized 10× concentrate (Cripps et al., 2009). MgCas media contained (per liter deionized water) 20 mL of minimal salts and 680 µL of 10% (w/v) KOH. Minimal salts, composed of (per liter of deionized water) 26.1 g K_2HPO_4, 11.3 g KH_2PO_4, and 25 g NH_4NO_3, were filter sterilized and stored refrigerated. After autoclaving, 1 mL of each of the following stocks were added to the Mg media: 1.05M nitrilotriacetic acid, 0.59M $MgSO4{\cdot}7H_2O$, 0.91M $CaCl_2{\cdot}2H_2O$, and 0.04M $FeSO_4{\cdot}7H_2O$. 25 mL of 20% (w/v) filter-sterilized glucose and 0.3% (w/v) of 10% filter-sterilized casamino acids (Difco) were then added. Agar was added pre-autoclaving to 15 g/L for solid media of LB, TGP, and MgCas.

Table 4-1

Bacterial strains and plasmids used in this study.

Strain or Plasmid	Relevant Description	Reference or Source
Strain		
G. thermoglucosidasius		
NCIMB 11955	Wild type isolate (also known as DSM 2542)	BGSC Accession No. 95A1
TM2	Wild type that integrates pTM2 at *attB* locus	This study
G. thermoleovorans		
DSM 5366	Wild type isolate	BGSC Accession No. 96A1
thyA$^-$	*thyA*$^-$ variant of *G. thermoleovorans* DSM 5366	This study
thyA$^-$ TM6	Derivative of *thyA*$^-$ variant: integrates pTM6 at *attB* locus for *thyA* complementation	This study
E. coli		
S17-1	Donor strain for conjugative plasmid transfer; recA pro hsdR RP4-2-Tc::Mu-Km::Tn7 integrated into the chromosome	Wan 2013
Plasmids		
pJZ04e	Source of *mob*	Zhou et al., 2016
pG1AK-sfGFP	KanR, AmpR; *E. coli-Geobacillus* shuttle vector containing repBSTI from pBST22, KanR from pUB110 and sfGFP	Reeve et al., 2016
pTM1	pG1AK-sfGFP containing *mob* from pJZ04e	This study
pTM2	Integrating suicide vector for *G. thermoglucosidasius*; pTM1 with repBST1 replaced with *attP*/*int* from GBK1 bacteriophage	This study
pTM5	pTM1 containing *thyA* from GBK1 bacteriophage	This study
pTM6	Integrating suicide plasmid for *G. thermoleovorans*; pTM2 containing *thyA* from GBK1 bacteriophage	This study

Table 4-2

Primers used in this study.

Primer	Sequence 5' to 3'	Target
G1AK Mob Cloning XhoI Fwd	AAAACTCGAGAGATTGTACTGAGAGTGCACCA (underlined bases indicate introduced XhoI site)	1447 bp from pJZ04e that corresponds to *mob* region
G1AK Mob Cloning SalI Rev	AAAAGTCGACAATTCTTGATGGGGCAAGG (underlined bases indicate introduced SalI site)	
pG1AK attP/INT Fwd	CGCATAGCAGGGAGAGCTAC	1371 bp from GBK1 bacteriophage that corresponds to *attP/int* region
pG1AK attP/INT Rev	TTTTGGCTTTTGTTGACTACAC	
HiFi Primer FP1	AGGGAATGGCTAACATTAACGGAAA (underlined bases indicate annealing to insert)	774 bp from GBK1 bacteriophage that corresponds to *thyA* region
HiFi Primer RP1	AGGCGCGCCTTATCATTGTTCGCCCTC (underlined bases indicate annealing to insert)	
HiFi Primer FP2	AACAATGATAAGGCGCGCCTCTA (underlined bases indicate annealing to vector)	6323 bp from pTM1/6541 bp from pTM2 that corresponds to the entire plasmid with the exception of the TK101 Kan^R marker region
HiFi Primer RP2	CCGTTAATGTTAGCCATTCCCTTTCAGATAATT TTAGATTTG (underlined bases indicate annealing to vector)	
attB Flank Fwd	GAAACAGAAAGCGCTCCAAT	1046 bp fragment if no insert is present at *attB* in *G. thermoglucosidasius/ G. thermoleovorans*
attB Flank Rev	ATTCGTCAACCATCCGTCA	
Geo ThyA Fwd	GGATGGCTCCCGTGATTTT	965 bp fragment corresponding to *thyA* in *G. thermoglucosidasius* and *G. thermoleovorans*
Geo ThyA Rev	AGCCGGTTATCTTTGCCAAT	

4.2.2 Plasmid introduction into *Geobacillus* spp. Plasmids were introduced into *G. thermoglucosidasius* and *G. thermoleovorans* by conjugative transfer from *E. coli* S17-1 as described by Suzuki et al. (2012) with the following modifications. *E. coli* S17-1 donor cells containing the plasmid of interest were cultured (100 µL of frozen stock into 10 mL fresh TGP media with 100 µg/mL Amp) overnight at 37˚C, shaking at 250 rpm. *Geobacillus* spp. were subcultured in TGP and grown until an OD of 0.5 was reached. Donor *E. coli* cultures

were centrifuged for 10 min at 4,000 rpm, supernatant was discarded, and cultures were suspended in 10 mL TGP. Once recipient *Geobacillus* spp. cultures reached 0.5 OD, 5 mL of donor and 5 mL of recipient were mixed and centrifuged for 10 min at 4,000 rpm. The supernatant was discarded and pellets were suspended in residual TGP media. 100 µL spots were plated on TGP plates and incubated overnight at 37˚C. For selection of recipient cells containing the plasmid of interest using kanamycin, 750 µL of TGP was added to plates and spots (mixture of *E. coli* donors and *Geobacillus* recipients) were scraped using a sterilized glass L-rod. The media containing cells was aspirated with a pipette and spread on TGP plates with 10 µg/mL Kan in 100 µL aliquots to isolate transconjugants. For antibiotic-free selection, 750 µL of MgCas was added to plates and spots (mixture of *E. coli* donors and *Geobacillus* recipients) were scraped using a sterilized glass L-rod. The media containing cells was aspirated with a pipette and spread on MgCas plates. Plates were incubated at 55˚C overnight. Conjugation efficiencies were determined for each plasmid by taking the number of transconjugants and dividing by total number of cells in the absence of selection.

4.2.3 Construction of plasmid pTM1. The *E. coli-Geobacillus* spp. shuttle vector pTM1 was constructed with parts from previously described vector pG1AK-sfGFP (Reeve et al., 2016). pG1AK-sfGFP was a gift from Tom Ellis (Addgene plasmid # 71739 ; http://n2t.net/addgene:71739 ; RRID:Addgene_71739). Primers G1AK Mob Cloning XhoI Fwd and G1AK Mob Cloning SalI Rev were used to amplify a 1447 bp fragment corresponding to the *mob* region in pJZ04e (Zhou et al., 2016) with XhoI and SalI sites added for cloning into pG1AK-sfGFP. Digestion with XhoI and SalI allowed insertion of this fragment into similarly digested pG1AK-sfGFP resulting in pTM1.

4.2.4 Construction of plasmid pTM2. The integrating suicide vector pTM2 was constructed with parts from previously described pTM1. pTM1 was digested with HpaI and PmeI, which both leave blunt ends. Primers pG1AK attP/INT Fwd and pG1AK attP/INT Rev (primers ordered with 5' phosphorylation) were used to amplify a 1371 bp fragment encompassing the *attP* and *int* region from the GBK1 bacteriophage. The PCR fragment was cloned into the HpaI and PmeI sites in pTM1 resulting in pTM2, a plasmid capable of replication in *E. coli* but lacking an origin of replication for *Geobacillus* spp. The presence of the insert and correct orientation was confirmed by PCR and DNA sequencing (Eton Biosciences, USA).

4.2.5 Construction of plasmid pTM5. The *E. coli-Geobacillus* spp. shuttle vector pTM5 was constructed with parts from previously described vector pTM1. Primers HiFi Primer FP1 and HiFi Primer RP1 were used to amplify a 774 bp fragment from the GBK1 bacteriophage corresponding to the *thyA* gene. Primers HiFi Primer FP2 and HiFi Primer RP2 were used to amplify a 6323 bp fragment from pTM1 that included all components of that plasmid (*bla*, *E. coli* origin, pBST1 origin, *sfGFP*, *mob*) except the TK101 Kan^R marker. The NEBuilder HiFi Assembly Cloning Kit (NEB, USA) was used to assemble the *thyA* insert with the pTM1 fragment.

4.2.6 Construction of plasmid pTM6. The integrating suicide vector pTM6 was constructed with parts from previously described vector pTM2. Primers HiFi Primer FP1 and HiFi Primer RP1 were used to amplify a 774 bp fragment from the GBK1 bacteriophage corresponding to the *thyA* gene. Primers HiFi Primer FP2 and HiFi Primer RP2 were used to amplify a 6541 bp fragment from pTM2 that included all components of that plasmid (*bla*, *E. coli* origin, GBK1 *attP/int*, *sfGFP*, *mob*) except the TK101 Kan^R marker. The NEBuilder

HiFi Assembly Cloning Kit (NEB, USA) was used to assemble the *thyA* insert with the pTM2 fragment.

4.2.7 Integration confirmation of pTM2 in *G. thermoglucosidasius*. Confirmation of successful integration of pTM2 into the host *G. thermoglucosidasius* strain was achieved in two ways: (1) successful colony PCR on the *G. thermoglucosidasius* host strain using the pG1AK attP/INT Fwd and attB Flank Rev or pG1AK attP/INT Rev and attB Flank Fwd primer pairs and (2) digestion of the host genomic DNA flanking the integration site with religation and subsequent transformation of *E. coli* 10β cells. This involved genomic DNA prep of transconjugants using the Gram-positive protocol of the GenElute Bacterial Genomic Extraction Kit (Millipore Sigma, USA), digestion with MfeI (a site that is not present in pTM2 but will digest host genomic DNA 1506 bp upstream and 161 bp downstream of the *attB* site), ligation of MfeI digested genomic DNA, electroporation of *E. coli* 10β, plating at 37 ˚C for 16 hrs on LB plates (100 µg/mL Amp), aseptic transfer of colonies into LB liquid media (100 µg/mL Amp) at 37 ˚C for 16 hrs, shaking at 250 rpm, plasmid minipreps (Monarch Miniprep kits, NEB, USA) of transformants, PCR using pG1AK attP/INT Fwd and pG1AK attP/INT Rev primer pairs, and DNA sequencing (Eton Biosciences, USA) of PCR products.

4.2.8 Selection and confirmation of a *thyA*$^{-}$ mutant of *G. thermoleovorans*. *G. thermoleovorans* DSM 5366 were grown in 10 mL of TGP (10 µg/mL trimethoprim and 50 µg/mL thymine added) at 55 ˚C for 16 hrs, shaking at 250 rpm. A 100 µL aliquot was spread on MgCas (10 µg/mL trimethoprim and 50 µg/mL thymine added) plates and grown 16 hrs at 55 ˚C. Colonies were aseptically transferred into 10 mL of MgCas (50 µg/mL thymine added) and grown 16 hrs at 55 ˚C, shaking at 250 rpm. All cultures showing growth were

subcultured into MgCas (with and without 50 µg/mL thymine) and grown for 16 hrs at 55 ˚C, shaking at 250 rpm. Any cultures that had growth in MgCas with thymine added but lacked growth in MgCas were screened for mutations in the *thyA* gene by sequencing of products resulting from colony PCR using the Geo ThyA Fwd and Geo ThyA Rev primer pair.

4.3 Results

4.3.1 Construction of pTM1, pTM2, pTM5, and pTM6. Shuttle plasmid pTM1 (Fig. 4-1a) was successfully constructed by inserting the *mob* region from pJZ04e (Zhou et al., 2016) into pG1AK-sfGFP (Reeve et al., 2016). The *mob* region made pTM1 able to be conjugated from *E. coli* s17-1 into *Geobacillus* spp. Suicide integrating plasmid pTM2 was constructed by replacing the pBST1 origin in pTM1 with the *attP/int* region from bacteriophage GBK1 (Fig. 4-1b; accession number MN727882, GBK1_44). As a result, pTM2 would only be able to replicate in *Geobacillus* strains that have a homologous *attB* site. Shuttle plasmid pTM5 (Fig. 4-2a) was successfully constructed as a vector capable of antibiotic-free selection in a *G. thermoleovorans thyA*$^{-}$ mutant strain. This was achieved by replacing the TK101 Kan^{R} marker in pTM1 with the *thyA* gene from bacteriophage GBK1 (accession number MN727882, GBK1_37) using the NEBuilder HiFi Assembly Kit. A suicide integration vector named pTM6 (Fig 4-2b) was similarly constructed using pTM2 as the backbone. Conjugation efficiencies of each of these constructed plasmids are shown in Table 4-3.

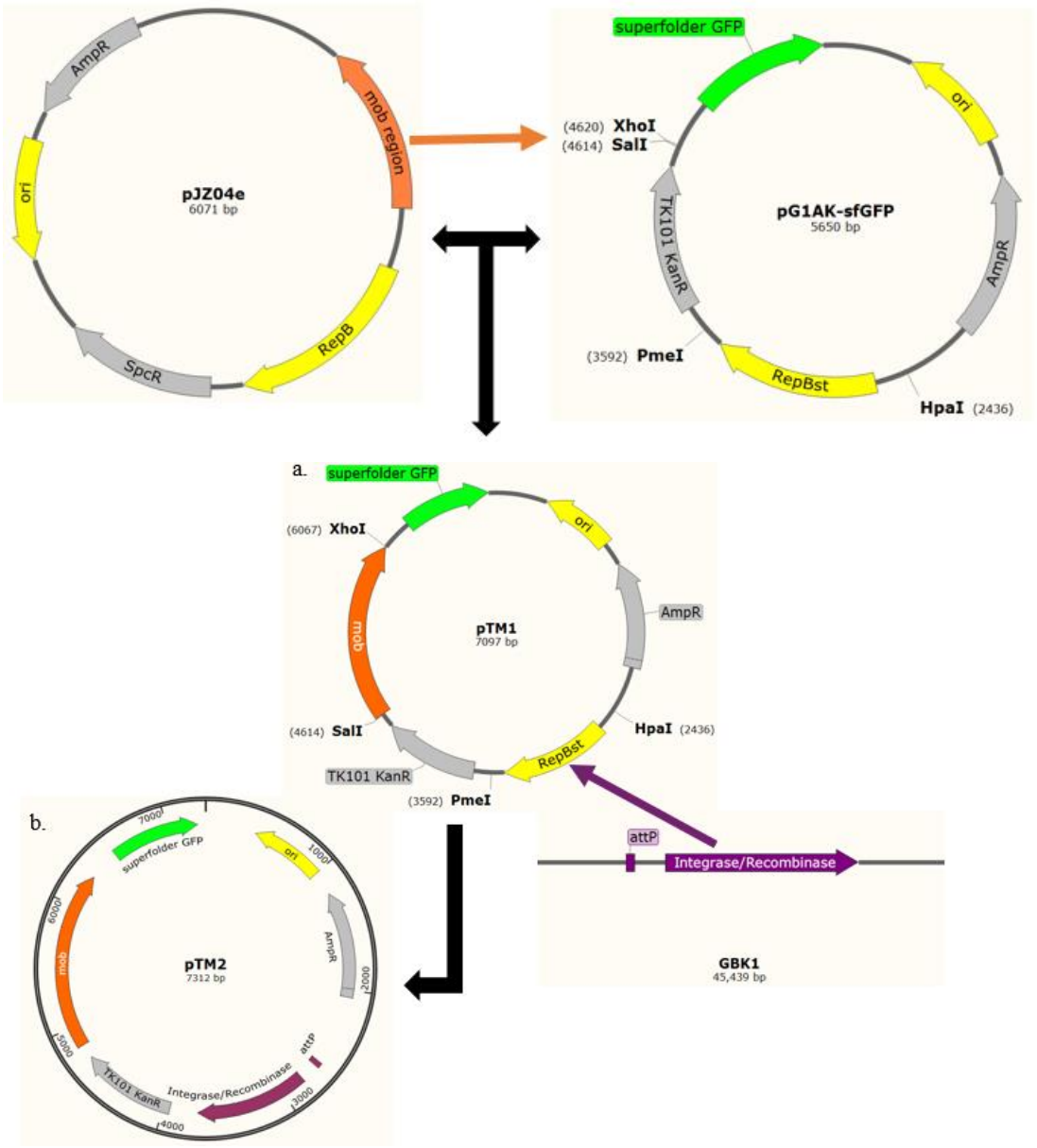

Figure 4-1 Construction of pTM1 and pTM2 plasmids. (a) pTM1 was constructed by adding XhoI and SalI sites via PCR to *mob* fragment and cloning into those sites in pG1AK-sfGFP. (b) Integrating suicide vector pTM2 was constructed by PCR amplifying *attP/int* from bacteriophage GBK1 and cloning into blunt sites PmeI and HpaI in pTM1.

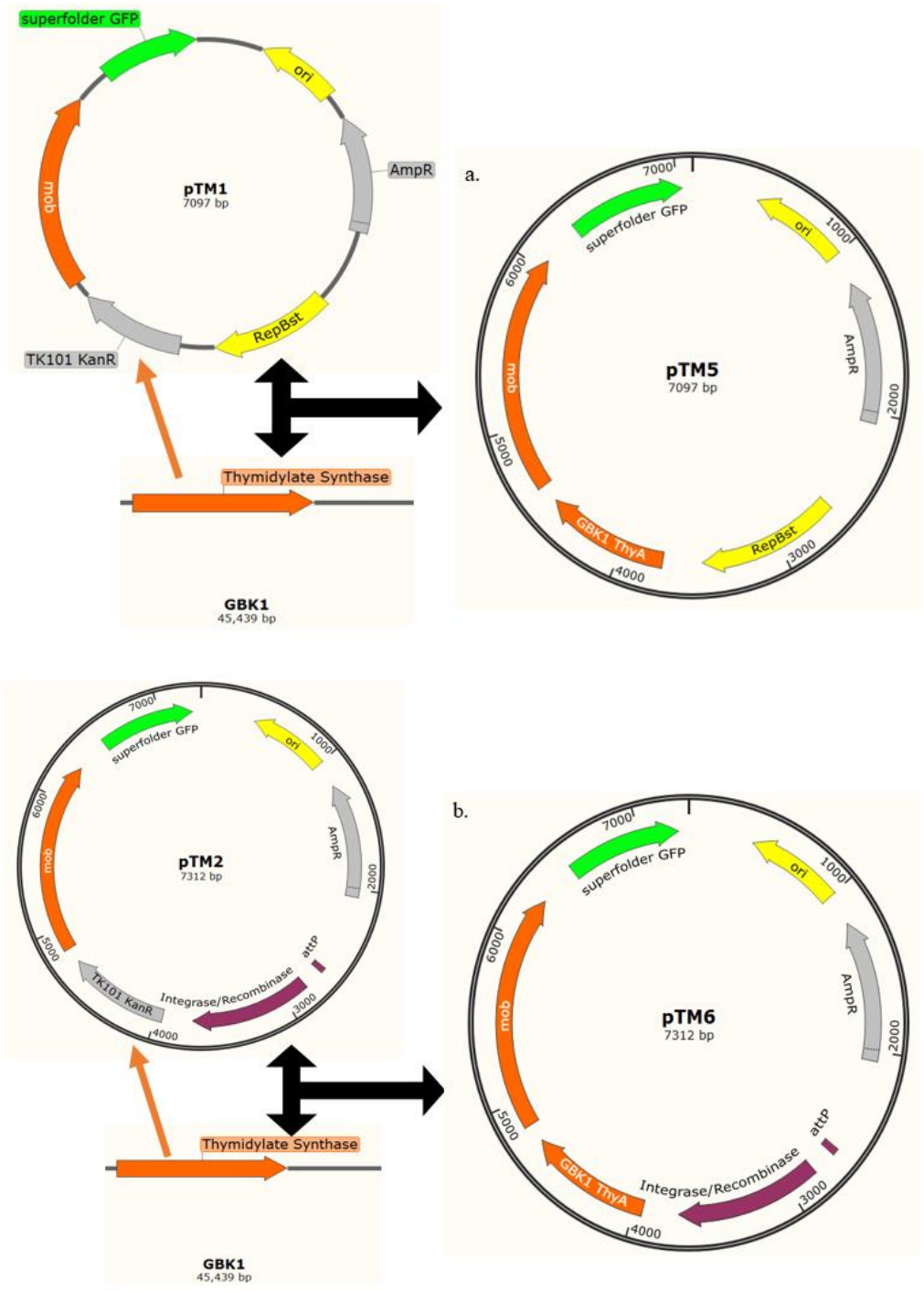

Figure 4-2 Construction of pTM5 and pTM6 plasmids. (a) pTM5 was constructed by replacing the Kan^R marker in pTM1 with the *thyA* gene from bacteriophage GBK1. (b) Integrating suicide vector pTM6 was constructed by replacing the Kan^R marker in pTM2 with the *thyA* gene from bacteriophage GBK1.

Table 4-3

Conjugation efficiencies of pTM1/pTM2 in *G. thermoglucosidasius* and pTM5/pTM6 in *G. thermoleovorans.* The conjugation efficiency is presented as the mean (+/- SD) of the ratio of exconjugant cells to the total number of recipient cells (N=3).

Plasmid	Conjugation Efficiency
pTM1	3.54×10^{-5} (+/-0.42)
pTM2	3.75×10^{-6} (+/- 0.78)
pTM5	5.21×10^{-5} (+/- 0.77)
pTM6	1.59×10^{-8} (+/- 1.12)

4.3.2 Conjugation of pTM1 and pTM2 into *G. thermoglucosidasius* using kanamycin selection.

Growth on TGP (10 µg/mL Kan) plates and expression of sfGFP were used as indicators of successful conjugation in *G. thermoglucosidasius* NCIMB 11955. Conjugation using pTM1 was successful, as was integration of suicide vector pTM2 (Fig 4-3a), although expression levels of sfGFP were lower with pTM2 when compared with pTM1. The wild-type *G. thermoglucosidasius* served as a control and was only able to grow on TGP plates without antibiotic (Fig 4-3b). Integration of pTM2 into the host *G. thermoglucosidasius* genome at the *attB* site was confirmed by colony PCR amplification of fragments using the pG1AK attP/INT Fwd/attB flank Rev and attB flank Fwd/pG1AK attP/INT Rev primer pairs (results not shown). Additionally, host genomic DNA was isolated and digested with MfeI, which does not digest pTM2 but does digest host genomic DNA at sites flanking the *attB* site. The digested genomic DNA was ligated back together to form circular fragments, transformed into *E. coli* 10β cells, and plated on LB (100 µg/mL Amp) to select for circular fragments that contained the pTM2 *E. coli* origin and selection marker.

pG1AK attp/INT Fwd and pG1AK attp/INT Rev primer pairs were used to amplify fragments for sequencing. DNA sequencing confirmed the presence of pTM2 and *G. thermoglucosidasius* host genomic DNA, proving integration at the *attB* site (results not shown).

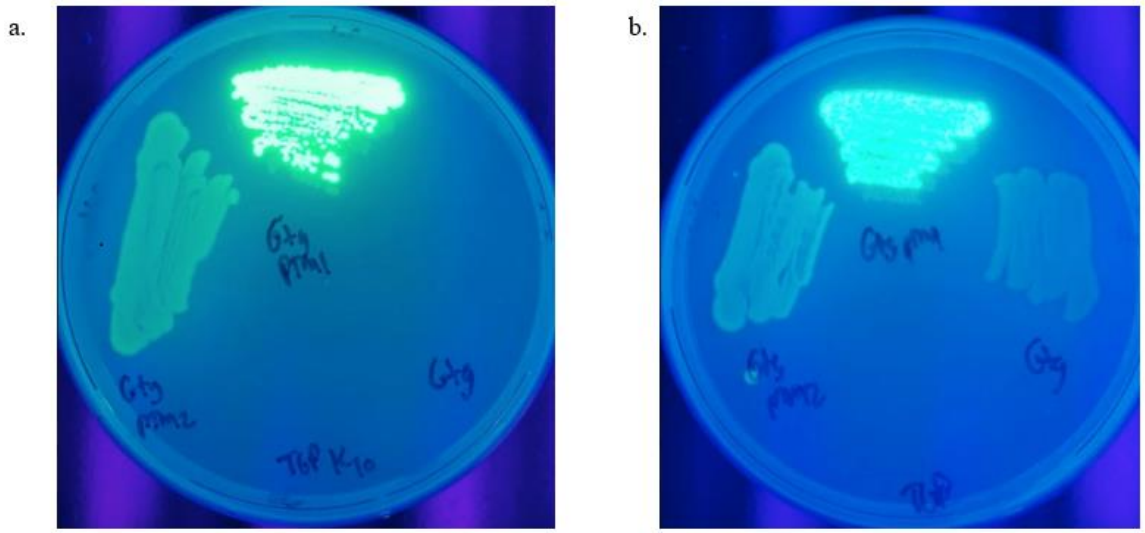

Figure 4-3 Plating of wild-type (right side of each plate) *G. thermoglucosidasius*, *G. thermoglucosidasius* harboring pTM1 (top of each plate), and *G. thermoglucosidasius* harboring pTM2 (left of each plate) on (a) TGP (10 µg/mL Kan) and (b) TGP plates.

4.3.2 Conjugation of pTM5 and pTM6 into *G. thermoleovorans thyA*$^-$ mutant

using thymine complementation. Growth on MgCas plates and expression of sfGFP were used as indicators of successful conjugation in a *G. thermoleovorans thyA*$^-$ mutant. Conjugation using pTM5 was successful, as was integration of suicide vector pTM6 (Fig 4-4a). The mutant *G. thermoleovorans thyA*$^-$ mutant served as a control and was only able to grow on MgCas (50 µg/mL thymine) plates (Fig 4-4b). Integration of pTM6 into the host genome of a *G. thermoleovorans thyA*$^-$ mutant at the *attB* site was confirmed by colony PCR amplification of fragments using the pG1AK attP/INT Fwd/attB flank Rev and attB flank Fwd/pG1AK attP/INT Rev primer pairs (results not shown). Additionally, host genomic DNA was isolated and digested with MfeI, which does not digest pTM6 but does digest host

genomic DNA at sites flanking the *attB* site. The digested genomic DNA was ligated back together to form circular fragments, transformed into *E. coli* 10β cells, and plated on LB (100 µg/mL Amp) to select for circular fragments that contained the pTM6 *E. coli* origin and selection marker.

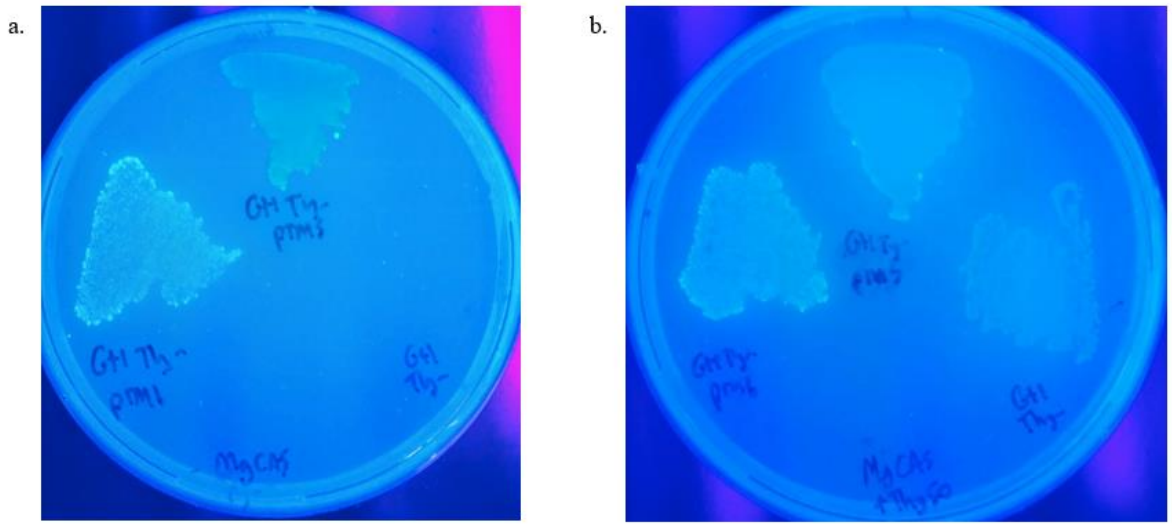

Figure 4-4 Plating of *G. thermoleovorans thyA*$^{-}$ mutant (right side of each plate), *G. thermoleovorans thyA*$^{-}$ mutant harboring pTM5 (top of each plate), and *G. thermoleovorans thyA*$^{-}$ mutant harboring pTM6 (left of each plate) on (a) MgCas and (b) MgCas (50 µg/mL thymine) plates.

4.4 Discussion

Use of *Geobacillus* spp. for industrial applications has been hindered by a shortage of tools for genetic manipulation. DNA uptake can be easily achieved in various *E. coli* (Aune and Aachmann, 2010), *Bacillus*, and *Lactobacillus* strains (Landete et al., 2014; Vojcic et al., 2012). In *Geobacillus* spp., there are three DNA uptake methods currently used: formation of protoplasts, electroporation, and conjugative transfer (Imanaka et al., 1982; Narumi et al., 1992; Suzuki and Yoshida, 2012). Protoplast methods are tedious and time-consuming, and kanamycin cannot be used as a selectable marker (Imanaka et al., 1982; Liao et al., 1986; Wu and Welker, 1989). DNA transformation by electroporation has been successful in some

Geobacillus species but not all and requires optimization for each species (Bartosiak-Jentys et al., 2013; Cripps et al., 2009; Studholme et al., 1999, Kananavičiūtė et al., 2014). Conjugative DNA transfer methods have been generally successful but the efficiency is limited (Suzuki and Yoshida, 2012). In this study, we developed new shuttle vectors (pTM1 and pTM5) that incorporate the *mob* region from pJZ04e (Zhou et al., 2016) with elements from pG1AK-sfGFP (Reeve et al., 2016). This allowed for conjugation of these 7 kb plasmids from *E. coli* s17-1 into *G. thermoglucosidasius* NCIMB 11955 as well as *G. thermoleovorans* DSM 5366. *E. coli* s17-1 has the RP4 elements incorporated into its genome, so it does not require any additional helper plasmids in order to be a conjugation donor. Additionally, we developed pTM2 and pTM6 which are suicide vectors that integrate into the host *Geobacillus* spp. strains at a specific *attB* site. This was accomplished by replacing the pBST1 *Geobacillus* spp. origin in each suicide vector with an *attP/int* segment from novel bacteriophage GBK1. There are a few examples of plasmids used in geobacilli that have replicons that are stable only under certain conditions or completely lack functional *Geobacillus* spp. replicons altogether. Vector pSTE12 needs selective pressure to be maintained and is unstable otherwise (Nakayami et al., 1993), while pTMO31 can only grow at temperatures below 55 ˚C (Cripps et al., 2009). The pGAM plasmids, pTM2, and pTM6 lack *Geobacillus* spp. replicons altogether, and are therefore called suicide plasmids (Suzuki et al., 2012; this study). These vectors are desirable because they allow use of other plasmids with functional *Geobacillus*-compatible replicons for further genetic manipulations. Replicative vectors work well for studies of gene expression and promoters, but for long-term genetic manipulation of *Geobacillus* spp., stable integration of genes (e.g. endoglucanases/cellobiohydrolases for crystalline cellulose degradation or glycoside

hydrolases for hydrolysis of noncrystalline polymeric substrates) into the host chromosome is preferable (Hussein et al., 2015).

Additionally, the ability to get functional integration of plasmid DNA in both *G. kaustophilus* and *G. thermoleovorans* indicates that the GBK1 integration system works in geobacilli beyond phage GBK1's host range of *G. kaustophilus* ATCC 8005. Plasmids pTM2 and pTM6 should also integrate in *G. thermodenitrificans, G. lituanicus, G. subterraneus, G. stearothermophilus, G. kaustophilus, Geobacillus* sp. JS12, *Geobacillus* sp. LC300, *Geobacillus* sp. 12AMOR1, *Geobacillus* sp. GHH01, *Geobacillus* sp. Y412MC52, *Geobacillus* sp. C56-T3, *Geobacillus* sp. Y412MC61, *Geobacillus* sp. Y4.1MC1, *Geobacillus* sp. WCH70, *Parageobacillus thermoglucosidasius, Aeribacillus pallidus, Anaerococcus prevotii, Bacillus caldolyticus,* and *Anoxybacillus amylolyticus* which have the *attB* site that matches the *attP* site of GBK1.

Using conjugative DNA transfer, shuttle vectors (replicative plasmids pTM1 and pTM5) had efficiencies of 3.54×10^{-5} (+/-0.42) and 5.21×10^{-5} (+/- 0.77) conjugated cells/total recipient cells in *G. thermoglucosidasius* and *G. thermoleovorans thyA*$^{-}$, respectively (Table 4-3). These efficiencies are similar to what has been previously reported with pUCG18T (10^{-5}-10^{-3}) and pSTE33T (10^{-7}-10^{-6}) in *G. kaustophilus* HTA426 (Suzuki and Yoshida, 2012). Conjugation efficiencies of the integrating suicide plasmids were lower, with pTM2 at 3.75×10^{-6} (+/- 0.78) into *G. thermoglucosidasius* and pTM6 at 1.59×10^{-8} (+/- 1.12) into *G. thermoleovorans thyA*$^{-}$. The lower efficiencies are most likely a result of the requirement of a successful integration event in addition to the expression of the necessary selection marker (TK101 in pTM2 and ThyA in pTM6).

For selection in DNA uptake experiments, there is a shortage of both antibiotics and antibiotic selection markers that are functional at temperatures above 55 ˚C. The two most common antibiotics used with geobacilli are kanamycin and chloramphenicol. Kanamycin is one of the most thermostable antibiotics (Peteranderl et al., 1990) and has a thermostable resistance determinant; TK101 is stable to 70 ˚C (Liao and Kanikula, 1990). Chloramphenicol is stable only up to 60 ˚C, (Liao and Kanikula, 1990; Narumi et al., 1992), and its resistance determinant, chloramphenicol acetyltransferase from *Staphylococcus aureus* plasmid pC194 (Horinouchi and Weisblum, 1982), is unreliable at higher temperatures because this is a mesophilic protein. The shortage in reliable antibiotic selection and antibiotic selection markers and the high costs of antibiotics at larger scales are some of the reasons why antibiotic-free selection systems are desirable. The only auxotrophic complementation system described in the literature for *Geobacillus* spp. is based on the complementation of a *pyrF* gene (*ura3* in eukaryotes) responsible for orotidine 5'-phosphate decarboxylase in a *G. kaustophilus* HTA426 *pyrF*$^{-}$ mutant strain (Suzuki et al., 2012). Thymidylate synthase (*thyA)* is essential for DNA synthesis as it is an enzyme that catalyzes the conversion of dUMP to dTMP. Several microbial auxotrophic *thyA* complementation systems have been described, including in *Streptococcus gordonii* (Lee et al., 2016), *Salmonell typhi* Ty21a (Bumann et al., 2001; Tacket et al., 1997), *Lactobacillus acidophilus* (Fu and Xu, 2000), *Vibrio cholera* (Liang et al., 2003), and *Streptococcus thermophiles* (Sasaki et al., 2004). In this study, we describe a successful *thyA*-based marker-free selection system in a *thyA*$^{-}$ mutant strain of *Geobacillus thermoleovorans.* Expression of *sfGFP* using a replicating vector (pTM5) and integrating suicide vector (pTM6) were successful. A higher level of GFP expression was seen in each replicative plasmid (pTM1 and pTM5) when

compared to their integrating counterparts (pTM 2 and pTM6), most likely due to the high copy number of the replicative plasmids (copy number of approximately 160 for repBST replicon) as opposed to the copy number of 1 when integrated into the chromosome. This is beneficial because kanamycin can be used as a selection method with other plasmids for further genetic manipulation in *Geobacillus* spp. Furthermore, pTM6 allows the integration of exogenous DNA into the host strain without the use of kanamycin or a *Geobacillus*-compatible replicon. Future research will involve optimizing a conjugation protocol for these plasmids in MgCas media to improve conjugation efficiency, especially in the case of the integrating pTM6 suicide vector into the *G. thermoleovorans thyA*$^{-}$ mutant.

In this study, we report an expansion of the genetic toolbox using elements found in bacteriophage that infect geobacilli. Future work will include developing a *thyA*$^{-}$ mutant for *G. thermoglucosidasius* that can be used for marker-free *thyA* selection with pTM5 and pTM6. Incorporation of an inducible promoter (like the maltose-inducible system mentioned previously) and the *recT* gene from bacteriophage GBK2 (Marks and Hamilton, 2014) could be used to develop a recombineering system in any *thyA*$^{-}$ mutant strain of *Geobacillus* spp. while still maintaining the availability of kanamycin and *Geobacillus*-compatible replicons for future genetic manipulations.

References

Aune, T., & Aachmann, F. (2010). Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. *Appl Microbiol Biotechnol, 85*(5), 1301-1313. doi:10.1007/s00253-009-2349-1

Bartosiak-Jentys, J., Hussein, A., Lewis, C., & Leak, D. (2013). Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. *Microbiology, 159*(Pt 7), 1267-1275. doi:10.1099/mic.0.066332-0

Bumann, D., Metzger, W., Mansouri, E., Palme, O., Wendland, M., Hurwitz, R., Meyer, T. (2001). Safety and immunogenicity of live recombinant *Salmonella enterica* serovar Typhi Ty21a expressing urease A and B from *Helicobacter pylori* in human volunteers. *Vaccine, 20*(5-6), 845-852. doi:10.1016/s0264-410x(01)00391-7

Coorevits, A., Dinsdale, A., Halket, G., Lebbe, L., De Vos, P., Van Landschoot, A., & Logan, N. (2012). Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus, G. stearothermophilus, G. jurassicus, G. toebii, G. thermodenitrificans* and *G. thermoglucosidans* (nom. corrig., formerly 'thermoglucosidasius'); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. *Int J Syst Evol Microbiol, 62*(Pt 7), 1470-1485. doi:10.1099/ijs.0.030346-0

Couñago, R., Chen, S., & Shamoo, Y. (2006). In vivo molecular evolution reveals biophysical origins of organismal fitness. *Mol Cell, 22*(4), 441-449. doi:10.1016/j.molcel.2006.04.012

Couñago, R., & Shamoo, Y. (2005). Gene replacement of adenylate kinase in the gram-positive thermophile *Geobacillus stearothermophilus* disrupts adenine nucleotide homeostasis and reduces cell viability. *Extremophiles, 9*(2), 135-144. doi:10.1007/s00792-004-0428-x

Cripps, R., Eley, K., Leak, D., Rudd, B., Taylor, M., Todd, M., Atkinson, T. (2009). Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metab Eng, 11*(6), 398-408. doi:10.1016/j.ymben.2009.08.005

Cuebas, M., Sannino, D., & Bini, E. (2011). Isolation and characterization of arsenic resistant *Geobacillus kaustophilus* strain from geothermal soils. *J Basic Microbiol, 51*(4), 364-371. doi:10.1002/jobm.201000314

Fong, J., Svenson, C., Nakasugi, K., Leong, C., Bowman, J., Chen, B., Rogers, P. (2006). Isolation and characterization of two novel ethanol-tolerant facultative-anaerobic thermophilic bacteria strains from waste compost. *Extremophiles, 10*(5), 363-372. doi:10.1007/s00792-006-0507-2

Fu, X., & Xu, J. (2000). Development of a chromosome-plasmid balanced lethal system for *Lactobacillus acidophilus* with thyA gene as selective marker. *Microbiol Immunol, 44*(7), 551-556. doi:10.1111/j.1348-0421.2000.tb02533.x

Hussein, A., Lisowska, B., & Leak, D. (2015). The genus *Geobacillus* and their biotechnological potential. *Adv Appl Microbiol*, 92, 1-48. doi:10.1016/bs.aambs.2015.03.001

Horinouchi, S., & Weisblum, B. (1982). Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J Bacteriol, 150*(2), 815-825.

Imanaka, T., Fujii, M., Aramori, I., & Aiba, S. (1982). Transformation of *Bacillus stearothermophilus* with plasmid DNA and characterization of shuttle vector plasmids between *Bacillus stearothermophilus* and *Bacillus subtilis*. *J Bacteriol, 149*(3), 824-830.

Inoue, T., & Sako, Y. (2013). *Host-Vector Systems in Thermophiles. In Thermophilic Microbes in Environmental and Industrial Biotechnology*: Springer, Dordrecht.

Kananavičiūtė, R., Butaitė, E., & Citavičius, D. (2014). Characterization of two novel plasmids from *Geobacillus* sp. 610 and 1121 strains. *Plasmid, 71*, 23-31. doi:10.1016/j.plasmid.2013.10.002

Kananavičiūtė, R., & Čitavičius, D. (2015). Genetic engineering of *Geobacillus* spp. *J Microbiol Methods, 111*, 31-39. doi:10.1016/j.mimet.2015.02.002

Landete, J., Arqués, J., Peirotén, A., Langa, S., & Medina, M. (2014). An improved method for the electrotransformation of lactic acid bacteria: A comparative survey. *J Microbiol Methods, 105*, 130-133. doi:10.1016/j.mimet.2014.07.022

Lee, S., Hulbah, M., & Halperin, S. (2016). Development of a gene delivery system in *Streptococcus gordonii* using thymidylate synthase as a selection marker. *J Microbiol Methods, 125*, 43-48. doi:10.1016/j.mimet.2016.04.003

Liang, W., Wang, S., Yu, F., Zhang, L., Qi, G., Liu, Y., Kan, B. (2003). Construction and evaluation of a safe, live, oral *Vibrio cholerae* vaccine candidate, IEM108. *Infect Immun, 71*(10), 5498-5504. doi:10.1128/iai.71.10.5498-5504.2003

Liao, H., & Kanikula, A. (1990). Increased efficiency of transformation of *Bacillus stearothermophilus b*y a plasmid carrying a thermostable kanamycin resistance marker. *Current Microbiol, 21*, 301-306.

Liao, H., McKenzie, T., & Hageman, R. (1986). Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. *Proc Natl Acad Sci U S A, 83*(3), 576-580. doi:10.1073/pnas.83.3.576

Lin, L., & Xu, J. (2013). Dissecting and engineering metabolic and regulatory networks of thermophilic bacteria for biofuel production. *Biotechnol Adv, 31*(6), 827-837. doi:10.1016/j.biotechadv.2013.03.003

Lin, P., Rabe, K., Takasumi, J., Kadisch, M., Arnold, F., & Liao, J. (2014). Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*. *Metab Eng, 24*, 1-8. doi:10.1016/j.ymben.2014.03.006

Marks, T., & Hamilton, P. (2014). Characterization of a thermophilic bacteriophage of *Geobacillus kaustophilus*. *Arch Virol*, 159(10), 2771-2775. doi:10.1007/s00705-014-2101-8

McMullan, G., Christie, J., Rahman, T., Banat, I., Ternan, N., & Marchant, R. (2004). Habitat, applications and genomics of the aerobic, thermophilic genus *Geobacillus*. *Biochem Soc Trans*, *32*(Pt 2), 214-217. doi:10.1042/bst0320214

Mead, D., McClary, J., Luckey, J., Kostichka, A., Witney, F., & Smith, L. (1991). Bst DNA polymerase permits rapid sequence analysis from nanogram amounts of template. *Biotechniques, 11*(1), 76-78, 80, 82-77.

Nakayama, N., Narumi, I., Nakamoto, S., & Kihara, H. (1993). Complete nucleotide sequence of pSTK1, a cryptic plasmid from *Bacillus stearothermophilus T*K015. *Biotechnol Lett, 15*, 1013-1016.

Narumi, I., Sawakami, K., Nakamoto, S., Nakayama, N., Yanagisawa, T., Takahashi, N., & Kihara, H. (1992). A newly isolated *Bacillus stearothermophilus K*1041 and its transformation by electroporation. *Biotech Techniques, 6*, 83-86.

Nazina, T., Tourova, T., Poltaraus, A., Novikova, E., Grigoryan, A., Ivanova, A., Ivanov, M. (2001). Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus, Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus,*

G. thermocatenulatus, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius and G. thermodenitrificans*. *Int J Syst Evol Microbiol, 51*(Pt 2), 433-446. doi:10.1099/00207713-51-2-433

Omokoko, B., Jäntges, U., Zimmermann, M., Reiss, M., & Hartmeier, W. (2008). Isolation of the phe-operon from *G. stearothermophilus* comprising the phenol degradative meta-pathway genes and a novel transcriptional regulator. *BMC Microbiol*, *8*, 197. doi:10.1186/1471-2180-8-197

Peteranderl, R., Shotts, E. B., & Wiegel, J. (1990). Stability of antibiotics under growth conditions for thermophilic anaerobes. *Appl Environ Microbiol*, *56*(6), 1981-1983.

Reeve, B., Martinez-Klimova, E., de Jonghe, J., Leak, D., & Ellis, T. (2016). The *Geobacillus* Plasmid Set: A Modular Toolkit for Thermophile Engineering. *ACS Synth Biol, 5*(12), 1342-1347. doi:10.1021/acssynbio.5b00298

Sasaki, Y., Ito, Y., & Sasaki, T. (2004). ThyA as a selection marker in construction of food-grade host-vector and integration systems for *Streptococcus thermophilus*. *Appl Environ Microbiol, 70*(3), 1858-1864. doi:10.1128/aem.70.3.1858-1864.2004

Sharma, A., Adhikari, S., & Satyanarayana, T. (2007). Alkali-thermostable and cellulase-free xylanase production by an extreme thermophile *Geobacillus thermoleovorans*. *World Journal of Microbiology and Biotechnology*, *23*, 483-490.

Studholme, D., Jackson, R., & Leak, D. (1999). Phylogenetic analysis of transformable strains of thermophilic *Bacillus* species. *FEMS Microbiol Lett, 172*(1), 85-90. doi:10.1111/j.1574-6968.1999.tb13454.x

Suzuki, H., Murakami, A., & Yoshida, K. (2012). Counterselection system for *Geobacillus kaustophilus* HTA426 through disruption of pyrF and pyrR. *Appl Environ Microbiol, 78*(20), 7376-7383. doi:10.1128/AEM.01669-12

Suzuki, H., & Yoshida, K. (2012). Genetic transformation of *Geobacillus kaustophilus* HTA426 by conjugative transfer of host-mimicking plasmids. *J Microbiol Biotechnol, 22*(9), 1279-1287. doi:10.4014/jmb.1203.03023

Suzuki, H., Yoshida, K., & Ohshima, T. (2013). Polysaccharide-degrading thermophiles generated by heterologous gene expression in *Geobacillus kaustophilus* HTA426. *Appl Environ Microbiol, 79*(17), 5151-5158. doi:10.1128/AEM.01506-13

Tacket, C., Kelly, S., Schödel, F., Losonsky, G., Nataro, J., Edelman, R., Curtiss, R. (1997). Safety and immunogenicity in humans of an attenuated *Salmonella typhi* vaccine vector strain expressing plasmid-encoded hepatitis B antigens stabilized by the Asd-balanced lethal vector system. *Infect Immun, 65*(8), 3381-3385.

Taylor, M. P., van Zyl, L., Tuffin, I. M., Leak, D. J., & Cowan, D. A. (2011). Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts. *Microb Biotechnol, 4*(4), 438-448. doi:10.1111/j.1751-7915.2010.00246.x

Turgeon, N., Laflamme, C., Ho, J., & Duchaine, C. (2006). Elaboration of an electroporation protocol for *Bacillus cereus* ATCC 14579. *J Microbiol Methods, 67*(3), 543-548. doi:10.1016/j.mimet.2006.05.005

Uma Maheswar Rao, J., & Satyanarayana, T. (2003). Enhanced secretion and low temperature stabilization of a hyperthermostable and Ca2+-independent alpha-amylase of *Geobacillus thermoleovorans* by surfactants. *Lett Appl Microbiol, 36*(4), 191-196. doi:10.1046/j.1472-765x.2003.01283.x

Vojcic, L., Despotovic, D., Martinez, R., Maurer, K., & Schwaneberg, U. (2012). An efficient transformation method for *Bacillus subtilis* DB104. *Appl Microbiol Biotechnol, 94*(2), 487-493. doi:10.1007/s00253-012-3987-2

Wan, H. (2013). *Transformation of the thermophilic bacterium, Geobacillus debilis , by conjugation with the mesophilic bacterium, Escherichia coli.* (Master of Science). University of Manitoba,

Wang, L., Tang, Y., Wang, S., Liu, R., Liu, M., Zhang, Y., Feng, L. (2006). Isolation and characterization of a novel thermophilic *Bacillus* strain degrading long-chain n-alkanes. *Extremophiles, 10*(4), 347-356. doi:10.1007/s00792-006-0505-4

Weigel, J., & Ljungdahl, L. (1986). The importance of thermophilic bacteria in biotechnology. *Crit. Rev. Biotechnol., 3*, 39-108.

Wu, L., & Welker, N. (1989). Protoplast transformation of *Bacillus stearothermophilus* NUB36 by plasmid DNA. *J Gen Microbiol, 135*(5), 1315-1324. doi:10.1099/00221287-135-5-1315

Zeigler, D. (2014). The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology, 160*(Pt 1), 1-11. doi:10.1099/mic.0.071696-0

Zhou, J., Wu, K., & Rao, C. (2016). Evolutionary engineering of *Geobacillus thermoglucosidasius* for improved ethanol production. *Biotechnol Bioeng, 113*(10), 2156-2167. doi:10.1002/bit.25983

CHAPTER 5

Concluding Remarks

Timothy J. Marks and Paul T. Hamilton

This work sought to isolate, sequence, annotate, and characterize two novel bacteriophage that infect *Geobacillus kaustophilus* ATCC 8005 (type strain). These phages were isolated from a backyard compost pile in Cary, NC and were named GBK1 and GBK2. Illumina sequencing of each phage and subsequent annotation of ORFs revealed genes that could potentially fill a need regarding expanding the available tools for genetic manipulation of *Geobacillus* spp. Chapter 2 describes GBK2 as a lytic, circularly permuted bacteriophage that has two ORFs of interest. ORF 39 has homology to *recT* and could be used to develop a recombineering system in *Geobacillus* spp. ORF 51 has homology to a thymidylate synthase and could be applicable as an antibiotic-free auxotrophic selection marker (similar to our work with GBK1 *thyA*). Chapter 3 describes GBK1 as a lysogenic, linear bacteriophage that has two ORFs of interest. Chapter 3 also examined the integration and sequencing of GBK1 into the *G. kaustophilus* type strain (ATCC 8005), which had not been sequenced previously. ORF 37 from GBK1 has homology to a thymidylate synthase which we used to create a functional antibiotic-free auxotrophic selection system in a *thyA*$^{-}$ mutant of *G. thermoleovorans*. ORF 43 has homology to a tyrosine recombinase which we used (along with the *attP* sequence) to construct an integrating suicide vector in *G. thermoglucosidasius* and *G. thermoleovorans*. Both of these tools are explained in Chapter 4.

Sequencing and annotation of both GBK1 and GBK2 led us to look at phage evolutionary relationships between these novel thermophilic bacteriophages and closely related mesophilic phages that infect *Bacillus* spp. This mirrors the proposed evolutionary relationship between the mesophilic *Bacillus* and thermophilic *Geobacillus* host bacteria of these phages. There were several ORFs from *Bacillus* phage vB_BpsS-36 that had homology to GBK1, including the terminase, head morphogenesis protein, scaffold protein, DNA

helicase, DNA primase, DNA polymerase, and several other hypothetical proteins. A similar situation was discovered when looking at GBK2, which had homology with the recombinase, RecT, single strand DNA binding protein (SSB), helicase loader, helicase, and several other hypothetical proteins of *Bacillus* phage SPP1. Commonalities between GBK1/GBK2 regions of homology with *Bacillus* phages mostly include the ORFs responsible for DNA replication.

Future work should seek to develop a recombineering system using recT from GBK2 similar to those that can be found in *E. coli*. Additionally, the development of other *thyA*-mutant strains of *GeoBacillus* spp. and the optimization of conjugative transfer of replicating vector pTM5 and integrating suicide vector pTM6 into those mutant strains should be a focus of future work. The addition of an inducible promoter and use of pTM6 to integrate genes (cellulases or endoglucanases for example) in *GeoBacillus* spp. like *G. thermoglucosidasius* would serve to expand the metabolic capabilities of these organisms. These tools could be invaluable for the genetic engineering of these industrially useful thermophilic microorganisms.