

https://dspace.org/

Investigating the prevalence, evolution and phenotypic characteristics associated with Geobacillus and Parageobacillus plasmids

Bhoolia, Hena

2020
https://hdl.handle.net/10539/31593

Downloaded from a DSpace repository. https://dspace.org/



Investigating the prevalence, evolution and phenotypic characteristics associated with *Geobacillus* and *Parageobacillus* plasmids

by

Hena Bhoolia

(862322)

Dissertation

submitted in fulfilment of the requirements for the degree

Master of Science

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Dr. Pieter De Maayer

February 2020

Contents **Page No.**

Declaration iii

Abstract iv

Acknowledgements v

List of Figures vi

List of Tables vi

Chapter 1: Literature Review 1
1.1 Introduction 1
1.2 Characteristics of bacterial plasmids 2
1.3 Plasmid evolution 4
1.4 Plasmid prevalence in bacterial communities 6
1.5 Types of plasmids and their role in bacterial diversity 7
1.5.1 Resistance plasmids 7
1.5.2 Virulence plasmids 8
1.5.3 Col plasmids 8
1.5.4 Degradative plasmids 9
1.5.5 Fertility plasmids 10
1.5.6 Additional characteristics conferred by plasmids 10
1.5.7 Other classifications of plasmids 11
1.6 Plasmids in biotechnology 12
1.7 Aims and Objectives 15
1.8 References 16

Chapter 2: The prevalence and evolution of plasmids in the genera *Geobacillus* and *Parageobacillus* 21
Abstract 21
2.1 Introduction 21
2.2 Methods 24
2.2.1 Plasmid sequence extraction and assembly 24
2.2.2 Plasmid grouping 24
2.2.3 Phylogenetic analyses 25
2.3 Results and Discussio 26
2.3.1 Plasmids are prevalent features among the geobacilli 26
2.3.2 The geobacilli belong to nine distinct plasmid groups 29
2.3.3 Characteristics of the pGeo1 plasmids 32
2.3.4 Characteristics of the pGeo2 plasmids 37
2.3.5 Characteristics of the pGeo3 plasmids 43

2.3.6 Characteristics of the pGeo4 plasmid 45
2.3.7 Characteristics of plasmids in the groups pGeo5 - pGeo9 46
2.4 Conclusions 50
2.5 References 52

Chapter 3: Cargo genes of *Geobacillus* and *Parageobacillus* plasmids and their biotechnological potential 59
Abstract 59
3.1 Introduction 59
3.2 Materials and Methods 62
3.2.1 Annotation of plasmid protein coding sequences 62
3.2.2 Plasmid isolation and curing 62
3.2.2.1 Bacterial cultures 62
3.2.2.2 DNA extraction and electrophoresis 63
3.2.2.3 16S rRNA PCR amplification 64
3.2.2.4 Plasmid curing 64
3.3 Results and Discussion 65
3.3.1 Overview of the functional roles of the geobacilli plasmids 65
3.3.2 The geobacilli plasmids contribute to antibiotic resistance 68
3.3.3 The geobacilli plasmids contribute to antibiosis and interbacterial competition 70
3.3.4 The geobacilli plasmids contribute towards metabolic versatility 72
3.3.5 The geobacilli plasmids codes for five distinct restriction-modification systems and a recombinase of biotechnological value 77
3.3.6 Isolation and curing of plasmid DNA 84
3.4 Conclusions 89
3.5. References 90

4 General Discussion 99

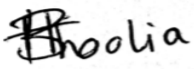
Declaration

I, Hena Bhoolia (862322), am a student registered for the degree of Masters in Science (MSc) in the academic year 2019. I hereby declare the following:

I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.

I confirm that the research proposal submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.

I have followed the required conventions in referencing the thoughts and ideas of others.

Candidate's Signature:  17/02 day of 2020 YEAR

Abstract

Plasmids are extrachromosomal elements that are commonly harboured by bacteria. They are a source of unique genes that influence the diversity and phenotypic characteristics of bacteria and are also widely utilized in a vast number of biotechnological applications. The genomes of several *Geobacillus* and *Parageobacillus* strains are observed to incorporate plasmids, however, only a few preliminary studies have highlighted the characteristics of some *Geobacillus* plasmids. This study aimed to investigate the prevalence and diversity of plasmids among the geobacilli, analyse their evolutionary dynamics and identify plasmid-encoded phenotypes that may be of biotechnological value. Here an exhaustive set of *Geobacillus* and *Parageobacillus* genomes were analysed to identify plasmid DNA and used comparative genomics to gain insight into their evolutionary dynamics. 56 plasmids were identified and grouped in nine distinct plasmid groups, indicating extensive genetic diversification of the plasmids of these genera. Furthermore, plasmids were functionally annotated to assess their biological roles among the geobacilli. From there is evidence of a range of potential phenotypes and possible thermostable proteins/enzymes that make these plasmids valuable candidates for biotechnological purposes. Lastly, plasmid DNA extraction and curing of selected *Geobacillus* and *Parageobacillus* strains was carried out to determine phenotypic differences between plasmid-cured and wild-type strains, this was unsuccessful attempt and may be as a result of possible low copy number plasmids as well as ineffective lysis of bacterial cells.

Acknowledgements

I would like to express my deepest appreciation to Dr. Pieter De Maayer, my supervisor throughout my postgraduate studies, for allowing me the opportunity to undertake both Honours and Masters under his guidance. I wish to thank him for his patience and enthusiasm when explaining new techniques and concepts to me and for his dedication to helping me achieve my goals of postgraduate qualifications. The skills I have gained under him are of immense value and will undoubtedly be useful in furthering a career in Science.

I would like to thank the National Research Foundation (NRF) for the financial support of the Block Grant Bursary for the years 2018 and 2019.

Lastly, a huge thank you to my family, friends and lab-mates in the Microbiology and Biotechnology discipline of Molecular and Cell Biology, who provided support and encouragement throughout my postgraduate studies.

List of Figures **Page No.**

Chapter 2

Figure 1: Plasmid prevalence among the genera *Geobacilus* and *Parageobacillus* 28
Figure 2: Circular map of the pGeo1 plasmids 33
Figure 3: Maximum likelihood phylogenies of the pGeo1 plasmids (left – 6 protein orthologues) and the core genomes (right - 1,441 protein orthologues) 34
Figure 4: Circular map of the pGeo2 plasmids 38
Figure 5: Maximum likelihood phylogenies of the pGeo2 plasmids (left – 23 protein orthologues) and the core genomes (right – 1,616 protein orthologues) 39
Figure 6: Circular map of the pGeo3 plasmids 43
Figure 7: Circular map of the pGeo5 plasmids 47

Chapter 3

Figure 1: Bar graph representing the 310 protein orthogroups assigned to COG functional categories 66
Figure 2: antiSMASH v 5.0 server (Blin *et al,* 2019) results depicting the coding regions of plasmid *G. thermodenitrificans* JSC_T9a involved in Geobacillin II synthesis and transport 77
Figure 3: Agarose gel of DNA samples extracted from the ABIOpure extraction kit (left) and plasmid extraction approach of Kado and Liu (1981) (right) 85
Figure 4: Agarose gel 16S rRNA PCR amplicon samples 86
Figure 5: Agarose gel of DNA samples extracted from cured and wild-type strain T2 *P. toebii* DSM 14590T using the ABIOpure extraction kit 88

List of Tables **Page No.**

Chapter 2

Table 1: General characteristics of the geobacilli plasmid groups 31
Table 2: Functionally annotated proteins encoded on the pGeo5 to pGeo9 plasmids 46

Chapter 3

Table 1: Plasmid loci of ABC- 2 type transporter subunits 69
Table 2: Plasmid loci of nucleobase-cation symporter and associated hydantoinase enzymes 74
Table 3: Type I R-M loci of five plasmids 78
Table 4: Plasmid loci of AlwI-dam Type II R-M system 80
Table 5: Plasmid loci of AlwI (BbvI/AlwXI) C5-Mtase R-M system 81
Table 6: Loci of the tyrosine recombinase in 41 plasmids 82
Table 7: Selected strains for plasmid DNA extraction 84

Chapter 1: Literature Review

1.1 Introduction

Plasmids are extrachromosomal double-stranded DNA molecules that are present within the cell. They are physically distinct from and smaller than the chromosomal DNA of a cell (Kroll *et al.*, 2010). Plasmids are found in a broad range of bacterial taxa, such as in members of the phyla Proteobacteria, Firmicutes, Spirochaetes, Actinobacteria, and Cyanobacteria, but have also been found in some Archaea and Eukarya (Shintani *et al.*, 2015). In bacteria, plasmids are generally circular, dsDNA molecules. However, they have been observed to be linearized in some members of the genera *Borellia, Mycobacterium, Streptomyces,* and *Rhodococcus* (Chen, 1996; Ventura *et al.*, 2007).

Plasmids are self-replicating elements and carry out their replication within a cell using the host cell replication machinery (del Solar *et al.*, 1998). All plasmids contain sites of around 50 to 100 bases making up a replicon region, being the origin of replication or *ori* site. This site enables a plasmid to replicate itself within the cell where host-cell enzymes bind to *ori* site, initiating replication (del Solar *et al.*, 1998; Lodish *et al.*, 2000). In addition to the *ori* site, plasmids encode *rep* genes that are involved in the initiation and maintenance of plasmid replication as well as *tra* genes encoding fertility factors that allow for their conjugative transfer between bacteria (del Solar *et al.*, 1998). Aside from proteins involved in replication, maintenance and dissemination, plasmids often contain additional genes that encode proteins that are able to confer unique and advantageous phenotypic traits to the organisms (Kroll *et al.*, 2010). Some of these traits include toxin production that inhibits other bacteria, antibiotic resistance, temperature resistance and the use of different carbohydrate, lipid and protein substrates (Kroll *et al.*, 2010; Shintani *et al.*, 2015; Smalla *et al.*, 2015).

Apart from conferring additional traits to bacteria, plasmids have also been widely utilized as vectors in genetic engineering (Lodish *et al.*, 2000). The presence of unique or valuable genes on plasmids as well as their structures and size has seen to their use as vectors in various applications, whereby the plasmids are manipulated to carry out mass production of specific genes and proteins (Lodish *et al.*, 2000).

Discussed in this chapter is the pertinent literature relating to the overall characteristics of plasmids. Given the topic at hand, which relates to the plasmids in the Gram-positive bacterial taxa *Geobacillus* and *Parageobacillus*, bacterial plasmids form the central focus of this review.

1.2 Characteristics of bacterial plasmids

As of 2015, there are over 4,600 complete plasmid sequences available in the NCBI Plasmid Genome database, 4,418 of which are from bacteria (Shintani *et al.*, 2015). Plasmids are a common genomic feature of many microbial communities and have been observed across all major bacterial phyla including the Proteobacteria, Firmicutes, Spirochaetes and Cyanobacteria (Shintani *et al.*, 2015). Bacterial plasmid DNA is typically smaller than chromosomal DNA, with chromosomal DNA being observed to range from 130 kilobases to over 14 megabases in size, while the majority of plasmids average around 80 kilobases in size (Shintani *et al.*, 2015). However, many bacterial taxa contain megaplasmids, which are far larger than the chromosome of some bacteria. For example, *Burkholderiae lata* 383 contains a 1.4 mb plasmid (Agnoli *et al.*, 2012), while a 1.8 mb linear plasmid (pSCL4) has been observed in *Streptomyces clavuligerus* (Medema *et al.*, 2010).

With plasmids being extrachromosomal elements, they are able to replicate independently of their host cell genome by utilizing the host cell replication machinery (del Solar *et al.*, 1998). Plasmid replication proceeds from one or more origin of replication (*ori*) sites located on the plasmid DNA (del Solar *et al.*, 1998; Lodish *et al.*, 2000). In addition to *ori* sites many plasmids incorporate genes coding for Rep proteins, which are generally localized in the first four kilobase region of the plasmid sequence, which along with other essential genes or sites assist in the replication of and regulation of replication of the plasmid (del Solar *et al.*, 1998; Lodish *et al.*, 2000).

Plasmids have various modes of replication such as rolling-circle-type, strand displacement type, and theta-type replication (del Solar *et al.*, 1998). In rolling-circle replication a nick is generated at the *ori* site, forming free 3'-OH and 5'-phosphate ends. The 3'-OH end forms the leading strand which is extended by replication enzymes, such as DNA polymerase, which continuously pushes the 5' end off the template strand until the reconstituted double-stranded *ori* site is reached (del Solar *et al.*, 1998). This generates one dsDNA (- parental strand and a new strand) plasmid and one single, displaced strand (+ parental strand). The ssDNA is then converted to dsDNA by host proteins (del Solar *et al.*, 1998). Rolling-circle replication is commonly observed in plasmids carried by Gram-positive bacteria as well as some plasmids that are smaller than 10 kilobases in size (Fire and Xu, 1995).

In strand-displacement and theta-type replication, the double-stranded DNA is denatured to form single-stranded DNA, which then allows for elongation of the plasmid DNA by DNA

polymerase. Strand displacement-type replication is initiated when the origin sites are exposed as single-stranded regions, which is carried out by melting of the dsDNA strands by plasmid Rep proteins RepA and RepC, which recognize AT-rich iterons (del Solar *et al.*, 1998). The DNA sequence is primed at either one of the single-stranded *ori* sites by plasmid-specific primases such as RepB. Extension of the primed strand occurs in the 5'-3' direction, which results in the displacement of one of the plasmid strands (del Solar *et al.*, 1998). Thus, like rolling circle replication, only one strand is replicated at once and is then copied into doublestranded DNA (dsDNA). This type of replication can result in the formation of the dsDNA plasmids, displaced single-stranded circular plasmids or plasmids that are partial doublestranded circles (del Solar *et al.*, 1998).

In theta-type replication, both single strands are used as templates to synthesize new complimentary strands in order to replicate the plasmid (del Solar *et al.*, 1998). The theta-type replication mechanism typically involves the melting of the parental strands, primer RNA (pRNA) synthesis and then DNA synthesis which is carried out by covalent extension of the pRNA generating both leading and lagging strands (del Solar *et al.*, 1998). Theta-type plasmid replication can start from one or several *ori* sites and, has also been observed to require plasmid-encoded Rep initiator proteins in most cases of replication.

Apart from extrachromosomal plasmids, some plasmids are able to insert or integrate themselves into the host cell's chromosome (del Solar *et al.*, 1998). These plasmids are termed episomes and their integration allows for their replication to be regulated as the chromosome itself undergoes replication (del Solar *et al.*, 1998).

Plasmid replication influences the plasmid copy number of a cell, which refers to the average number of copies of a plasmid per cell (Providenti *et al.*, 2006). In order to co-exist stably with their hosts, the replication-control mechanisms such as, the *ori* site and Rep proteins of a plasmid are responsible for maintaining a constant or fixed number of plasmid copies per cell (del Solar and Espinosa, 2000). Replication of plasmids can be relaxed, whereby a high plasmid copy number is established, or replication can be stringent, which produces a low plasmid copy number (Providenti *et al.*, 2006). It is essential for plasmids to maintain and regulate their copy number, as too high a number can result in a metabolic burden for the cell and therefore the plasmids cannot be replicated or maintained within the cell, while too low a number can result in plasmid loss upon cell division (del Solar and Espinosa, 2000; Providenti *et al.*, 2006). It is therefore also necessary for plasmids to ensure that there are enough plasmids present per chromosome during the division of the bacterial cells so that each daughter cell

receives copies of the plasmids (Providenti *et al.*, 2006). The plasmid copy numbers in bacteria have been observed to range from as little as 1 to 700 copies a cell (Providenti *et al.*, 2006).

Plasmid genomes consist of "core" genes, which include the *ori* site and the several other genes that are essential for plasmid replication, maintenance, and dissemination (Gogarten and Townsend, 2005). Apart from these core genes, plasmids are also observed to acquire additional genes that can increase the fitness of bacteria under different ecological conditions (Gogarten and Townsend, 2005). The acquisition of these genes can occur via several different processes such as transposition, transformation, phage-mediated transduction or during bacterial conjugation (Johnson and Grossman, 2015; Stevens, 2015). These genes are known as "cargo genes" and are not necessary for the functioning of a plasmid. However, due to plasmids being highly mobile elements as well as being easily manipulated by restriction enzymes, they represent "hotspots" for the integration of cargo genes and thus add to the characteristics and composition of plasmids (Smillie *et al.*, 2010). Cargo genes are observed to confer a range of phenotypes to the bacteria such as antibiotic resistance, resistance to heavy metals and/or the ability to metabolize different carbon sources (Johnson and Grossman, 2015; Stevens, 2015). Plasmid acquisition of cargo genes generally occurs at random (Johnson and Grossman, 2015). Therefore, these cargo genes can greatly influence plasmid evolution and diversification upon their integration, which in turn plays a role in the evolution and diversification of the host bacteria.

1.3 Plasmid Evolution

As is typical in Eukarya, bacterial chromosomal DNA is generally transmitted vertically from mother to daughter cells. Therefore, the extent of evolution of the genotypic and phenotypic characteristics of bacteria is limited to mutations and transposon activity within the bacterial chromosome itself (Carattoli, 2008; MacLean and San Millan, 2015). By contrast, plasmids are able to be transmitted both vertically and horizontally in bacteria, between both closely and distantly related bacteria, and therefore can largely influence the evolution of bacteria (MacLean and San Millan, 2015). Plasmids provide bacteria with access to a vast reservoir of genes that enables bacteria to display a plethora of new or unique characteristics, such as increased virulence or metabolic fitness (Shintani *et al.*, 2015). Over time the retention or loss of these plasmids directly impacts the evolving nature of different genera and species of bacteria by a gain, loss or alteration to the characteristics or phenotypes that the bacteria display

(MacLean and San Millan, 2015; Shintani *et al.*, 2015). Furthermore, plasmids themselves are observed to undergo evolution via changes to their gene sequence, which in turn influences the evolution and diversification of bacteria (Carattoli, 2008; MacLean and San Millan, 2015).

Plasmids are thought to have originated from primordial oligoribonucleotides which evolved into simple, primitive replicons, that largely resembled viral nucleic acid molecules such as RNA ribozymes (Kado, 1998). This similarity was established from the results of comparative analysis of the replication origins of virions and plasmids (Kado, 1998). As plasmid survival is dependent on the survival of the host cell, the relationship between plasmid and bacteria can be considered parasitic in nature (MacLean and San Millan, 2015). Therefore, the similarity of early plasmids to viral-based replicons could be as a result of the evolution of nucleic acids from a viral source or vice versa (Kado, 1998). Over time the eventual association of protein-coding genes with these viral-like replicons provided added efficiency and protective advantages for these nucleic acids to persevere as individual genetic elements, which led to the propagation of these replicons into the more complex plasmid structure known today (Kado, 1998).

Like other genetic materials, plasmids are observed to evolve as a result of various changes to their gene sequences. Changes to a plasmid sequence can occur via mutations that present themselves within the DNA during events such as, plasmid DNA replication and repair (MacLean and San Millan, 2015). Mutations are a strong driving force in evolution as it creates variability within populations establishing evolutionary change (Loewe and Hill, 2010). Mutations that occur within DNA sequences can be advantageous, disadvantageous or indifferent in nature and where they can either increase cell fitness, decrease cell fitness or result in no changes in a cell, respectively (Loewe and Hill, 2010). In the case of plasmids, advantageous mutations can result in changes to the gene sequence which may alter a gene's normal functions by encoding for specific metabolic activities, increased virulence or increased resistance to temperature or antibiotics (MacLean and San Millan, 2015). These evolved plasmid functions, in turn, increase the fitness of the host bacteria by providing them with said characteristics. Plasmids that acquire disadvantageous mutations are observed to lose stability and cannot be maintained within the bacterial population and are therefore lost (MacLean and San Millan, 2015). As a result of plasmid loss, the molecular evolution of plasmids is seen to select for beneficial mutations and thus it is expected that common evolutionary paths are followed to ensure plasmid stability (MacLean and San Millan, 2015).

Another driving factor of plasmid evolution is intracellular mobility, which refers to the exchange or movement of genetic materials between bacteria (Carattoli, 2008). Intracellular mobility introduces foreign DNA or genes into bacterial cells and can occur via conjugation (transfer of plasmids or genetic material between bacteria through a pilus), transformation (bacteria picking up DNA from their surrounding environments) and transduction (bacteriophage/viruses bringing DNA into their host bacterial cell) (Johnson and Grossman, 2015; Stevens, 2015). Once foreign DNA is present within a cell, recombination events occur which enables the insertion of these foreign genes into a plasmid sequence, thereby establishing new cargo genes encoding new functions within the plasmid or creating chimeric plasmids (Carattoli, 2008).

Intracellular mobility can also occur via the process of transposition, where transposons present within one plasmid, can insert themselves into the DNA sequence of another plasmid (Carattoli, 2008). Transposition can also occur via transposons sourced from chromosomal DNA (Carattoli, 2008). For example, antibiotic resistance in many pathogenic bacteria has evolved by the acquisition of plasmids carrying antibiotic-resistance genes, these genes were derived from the genome of environmental bacteria, which includes many bacteria that produce antibiotics themselves (Martínez, 2008). Intracellular mobility is, therefore, a key factor in plasmid evolution as the insertion of foreign cargo genes into a plasmid change or increases the overall fitness of the plasmid and in turn the bacteria.

1.4 Plasmid prevalence in bacterial communities

Bacterial plasmids were first identified as independent strands of DNA in the 1940s during research on antibiotic resistance (Clark and Pazdernik, 2013). Since then plasmids have been observed to be highly prevalent in many microbial communities, where they have been found to be a common genetic constituent of bacterial cells (Clark and Pazdernik, 2013). Plasmids can be introduced into bacterial cells via the processes of conjugation, transformation or transduction (Kroll *et al.,* 2010; Shintani *et al.,* 2015). However, bacteria can also readily lose their plasmids in cases when a bacterium divides into two, one of the daughter cells might miss out on getting a plasmid (Kroll *et al.,* 2010; Shintani *et al.,* 2015). The acquisition or loss of a plasmid thereby plays a role in the prevalence of plasmids within different microbial taxa.

Plasmids have been identified in and isolated from many bacteria, such as those belonging to the genera *Escherichia, Bacillus, Lactobacillus, Pseudomonas, Borellia, Mycobacterium, Staphylococcus, Streptomyces* and *Rhodococcus* (Bernet-Camard *et al.*, 1994; Chen, 1996;

Meletzus *et al.,* 1993; Ventura *et al.,* 2007). Within genera, the distribution of plasmids is dependent on the genus being studied, where in some cases all species and strains of a genus may contain plasmids, while in other genera, a more sporadic distribution of plasmid at both species and strain level may be observed (Clark and Pazdernik, 2013).

The prevalence of different plasmid types is also found to vary, whereby some types are more common in certain genera and species than others. For example, resistance type plasmids coding for drug resistance against antibiotics and other antibacterial drugs such as sulphonamide were found to be the most common plasmid type, existing within *Escherichia coli* strains isolated from humans and pig carcasses (Wu *et al.,* 2010). In other cases, the same plasmid type or more than one plasmid type may exist within different species belonging to the same genus (Clark and Pazdernik, 2013).

1.5 Types of plasmids and their role in bacterial diversity

As highlighted above, aside from core genes required for replication, dissemination, and maintenance of the plasmid, these extrachromosomal elements have also been observed to carry a variety of "cargo genes". Cargo genes differ from plasmid to plasmid and are able to confer distinct genotypic and phenotypic functions on the plasmids. As such, plasmids have been classified into five distinct plasmid types, namely, resistance plasmids, virulence plasmids, Col plasmids, degradative plasmids, and fertility plasmids (Shintani *et al.,* 2015). These different types of plasmids, if acquired by bacteria, can greatly influence bacterial phenotypes and their various capabilities by providing bacteria with a plethora of traits that under new or adverse environmental conditions, may be advantageous to these bacteria (Smalla *et al.,* 2015).

1.5.1 Resistance plasmids

Drug resistance in bacteria occurs following exposure to different compounds providing selective pressure for low permeability cell walls or the acquisition/evolution of, and expression of, drug deactivating enzymes or efflux pumps (San Millan 2019). In most cases of drug resistance in bacteria these characteristics are found to be plasmid-encoded (Stone, 1975). Plasmids coding for drug resistance are known as Resistance plasmids and typically include at least one or two genes involved in conferring drug resistance (Stone, 1975). These genes are termed resistance transfer factors or R-factors, which confer resistance against specific antibiotics and in some cases, they are able to impart resistance towards multiple, unrelated antibiotics (Stone, 1975). The most common drug resistance is encoded against antibiotic drug

classes including, beta-lactam-, fluoroquinolone- and aminoglycoside- based antibiotics, such as penicillin, ciprofloxacin, levofloxacin, gentamycin, neomycin and kanamycin (Jacoby, 2005; Garneau-Tsodikova and Labby, 2015). (Munita, and Arias, 2016). The presence of these genes are beneficial for bacteria as it enables them to colonize a variety of environments without being hindered. Furthermore, drug resistance adds to the virulent nature of pathogenic bacteria (Munita, and Arias, 2016). Several plasmids of the group IncP-1 are described as natural carries of antibiotic resistance and thus are typical examples of Resistance plasmids (Popowska and Krawczyk-Balska, 2013). IncP-1 plasmids show resistance towards a broad spectrum of antibiotics by encoding for efflux pumps and enzymatic modification or hydrolysis of the antibiotic molecule to mediate this resistance (Popowska and Krawczyk-Balska, 2013).

1.5.2 Virulence plasmids

Apart from plasmid-encoded drug resistance adding to the virulence of bacteria, some plasmids are seen to carry genes specifically encoding virulence factors that can make bacteria pathogenic in nature or increase their already pathogenic nature (Peterson, 1996). Plasmids that carry genes that code for these factors are known as "virulence plasmids" (Peterson, 1996). Virulence factors are pathogenic in nature and confer invasive and disease-causing properties to the plasmids that encode them (Peterson, 1996). Some of the most common plasmid virulence factors include; adherence factors, invasion factors, mobility, biofilm formation, capsule formation and the production of endotoxins and exotoxins (Meletzus *et al.,* 1993; Peterson, 1996). The presence of virulence plasmids allows for bacteria to aggressively invade, colonize and damage plant, animal and human host cells resulting in disease or death of the host (Meletzus *et al.,* 1993). Studies with regards to pathogenic *Escherichia coli* have found that their enteropathogenic nature, which includes enterotoxin production, enteroinvasion, enterohemorrhage, and enteroaggregative abilities, are as a result of virulence plasmids carried that code for multiple virulence factors (Johnson and Nolan, 2009). The *Agrobacterium* plasmid pTi can be considered a virulence plasmid as it encodes the tumour inducing T-DNA as well as the virulence operon region *virABCDEFG* (Zambryski *et al.,* 1983). The genes of the *vir* operon code for enzymes that enable members of *Agrobacterium tumefaciens* to the transfer T-DNA into host plant cells which results in the formation of tumour-like, crown gall disease within the plant (Zambryski *et al.,* 1983).

1.5.3 Col plasmids

In addition to virulence plasmids, another plasmid type providing a level of virulence to the bacteria that carry them are Col plasmids. Col plasmids carry genes that encode bacteriocins

such as colicins. Bacteriocins are protein toxins that have antimicrobial properties ranging from inhibition of cell growth to cell death of bacteria, in the cases of competitive symbiotic relationships (Riley and Gordo, 1992). Bacteriocins are produced by one bacterial taxon that is active against closely related strains (Shintani *et al.*, 2015). In the case of Col plasmids, colicins are mostly produced by and active against *E. coli* strains and related bacteria. Furthermore, colicins are almost always carried on plasmids, with over 20 distinct plasmid-encoded colicins having been described (Riley and Gordo, 1992). Plasmid-encoded bacteriocins have been observed in many taxa including *E. coli, Streptococcus* spp., *Lactobacillus* spp. and *Pseudomonas* spp. (Bernet-Camard *et al.*, 1994). The most common Col type plasmids are the ColE1 and ColE1-like plasmids (Rijavec *et al.*, 2007; Smajs *et al.*, 2010). These plasmids are common amongst *E. coli* and enable the production of colicins such as E1 or K which are toxic to and inhibits other strains of *E. coli* (Rijavec *et al.*, 2007; Smajs *et al.*, 2010).

1.5.4 Degradative plasmids

Plasmids are also seen to play a major role in bacterial metabolism. Degradative plasmids are plasmids that are observed to carry catabolic genes that confer metabolic capabilities to the host bacteria, enabling them to metabolize a range of substrates or compounds (Pathak and Navneet, 2017; Urbanek *et al.*, 2018). The proteins encoded by these genes are able to degrade a variety of different compounds such as aromatic- or aliphatic-containing and xenobiotic compounds. In addition to this, they have also been involved in metabolizing crude oil, soil pollutants and herbicides (Pemberton, 1983). Bacteria from the taxa *Alcaligenes* spp., *Rhodococcus* spp., *Bacillus* spp. and *Pseudomonas* spp. have been observed to degrade a variety of both natural and man-made substrates such as lignin, cellulose, terpenes, salicylates, toluene, xylene, naphthalene and phenols due to the presence of degradative plasmids (Segura *et al.*, 2014; Urbanek *et al.*, 2018). Plasmid-linked metabolic abilities of bacteria allow for these bacteria to thrive in different environments, including environments that contain pollutants as a result of many of the aforementioned compounds (Segura *et al.*, 2014). While most plasmid mediated metabolism involves catabolic activity resulting in the degradation of substrates, there is evidence of plasmid-mediated anabolic pathways in bacteria. Examples of plasmid encoded anabolic metabolism include some strains of *Rhizobium* that are able to carry out nitrogen fixation of atmospheric nitrogen to synthesize nitrogenous compounds like ammonia, as well as the ability of some bacteria to produce pigments or antibiotics (Hynes and McGregor, 1990).

1.5.5 Fertility plasmids

The fifth plasmid type that is commonly found is Fertility plasmids. These plasmids carry genes, namely *tra* genes, that encode fertility or F-factors conjugative machinery (Shintani *et al.,* 2015). These F-factors allow the plasmid DNA to be transferred between cells by initiating the process of conjugation (Shintani *et al*, 2015). Cells that carry Fertility plasmids are denoted F^+ and those without as F. The F^+ cells have the ability to form a conjugative pilus that makes contact with F^- cells, which then allows for the transfer of plasmid DNA between the two cells (Shintani *et al.,* 2015). Fertility plasmids are generally low copy number plasmids, occurring as single or two copies, however they play a major role in the spread of genetic material amongst bacteria and can be considered one of the most important means for bacteria to acquire new and advantageous traits (Shintani *et al.,* 2015). Amongst Fertility types, some examples include the plasmids of the IncP-1 group as well as plasmids of the IncW and IncU groups (Brown *et al.,* 2013; Popowska and Krawczyk-Balska, 2013). These plasmids encode for conjugative machinery that enable extensive plasmid transfer between bacteria and as a result, are often considered to be broad-host range plasmids (Brown *et al.,* 2013; Popowska and Krawczyk-Balska, 2013).

1.5.6 Additional characteristics conferred by plasmids

Some unique characteristics of bacteria that may be plasmid inherited include temperature resistance, heavy metal resistance, and spore formation. Temperature resistance is a typical characteristic amongst extremophilic bacteria, whereby they can withstand significantly higher or lower temperatures in environments such as deserts, volcanoes or ice glaciers (Vásquez *et al.,* 1984). Plasmids coding for temperature resistance factors has been observed in *Streptococcus thermophilus* and a variety of *Bacillus* species that were isolated from hot springs in Jordan (Khalil *et al.,* 2003).

Heavy metal resistance of bacteria has been observed against metals such as silver (Ag^+), arsenic (As), cadmium (Cd^{2+}), cobalt, (Co^{2+}), copper (Cu^{2+}), mercury (Hg^+) and zinc (Zn^{2+}) (Ji and Sliver, 1995). This characteristic has been observed to be plasmid-encoded in the case of *Bacillus thuringiensis, Bacillus subtilis* and *Staphylococcus aureus*, whereby plasmids isolated from these strains were found to carry genes enabling resistance towards the heavy metals' arsenic and cadmium (Guo and Mahillon, 2013). Several studies have also highlighted the plasmid-encoded mercury resistance of several plasmid carrying Gram-positive and Gram-negative bacterial species such as *S. aureus, E. coli, Enterobacter* spp., *Clostridium* spp., *Pseudomonas* spp., *Shigella* spp. and *Serretia* spp. (Furukawa and Tonomura, 1972;

Nascimento and Chartone-Souza, 2003; Nucifora *et al.,* 1989; Osborn *et al.,* 1997). This resistance is as a result of the *mer* operon. There are over ten different *mer* operons that have been identified and categorized into two groups; those that code for broad-spectrum enzymes that confer resistance to both organic and inorganic mercury and those that encode mercuric reductase which is active against the inorganic mercuric anion Hg^{2+} (Osborn *et al.,* 1997). These enzymes catalyse the reduction of mercury to a volatile Hg^{0} form, leaving the surrounding environment free of mercury (Furukawa and Tonomura, 1972; Nascimento and Chartone-Souza, 2003; Nucifora *et al.,* 1989; Osborn *et al.,* 1997).

Some bacteria are observed to undergo spore formation under adverse conditions (Adams *et al.,* 2014). These spores are known as endospores and allow for bacteria to become the dormant form of its vegetative state, making it highly resistant to physical or chemical influence (Adams *et al.,* 2014). Some bacteria, including strains belonging to genera *Bacillus* and *Clostridium,* have a plasmid-inherited ability to undergo sporulation (Adams *et al.,* 2014; Manetsberger *et al.,* 2015). Plasmids encoding for spore formation as well as temperature or heavy metal resistance are highly beneficial to their host bacteria, as it allows these bacteria to thrive in environments that are unfavorable to themselves and other organisms. Furthermore, these acquired characteristics also provide these bacteria with novel or increased virulence capabilities, as they are able to maintain themselves within adverse environments thereby allowing them to dominate within these environments.

1.5.7 Other classifications of plasmids

Aside from a classification system based on function, plasmids can be further subdivided into two categories on the basis of their ability to be transferred from one bacterium to another (conjugative plasmids) or not (non-conjugative plasmids). (Shintani et al., 2015). Plasmids that carry transfer (*tra)* genes, which are typically fertility plasmids, are categorized as conjugative plasmids and these genes enable the process of conjugation between the bacteria to occur (Shintani *et al.,* 2015). Conjugation is the direct contact between bacterial cells which allows for plasmids and other genetic material to be transferred from one bacterium to another via sex pili (Lodish *et al.,* 2000). By contrast, non-conjugative plasmids lack their own *tra* genes and can only be transferred between bacteria with the assistance of conjugative/fertility plasmids if they contain genes that are compatible with the genes encoding the *tra* system of the "helper" conjugative/fertility plasmid (Shintani *et al.,* 2015).

Plasmids can also be further categorized on the basis of their compatibility; which is dependent on the replicon type of the plasmid and is largely governed by the *ori* site region of the replicon (Thomas, 2014). If different plasmids share the same *ori* sites then these plasmids are incompatible, as they both have the same specificity for replication (Rep) protein or controlling elements and therefore, cannot coexist in the same cell (Thomas, 2014). Plasmids with different *ori* sites tend to be compatible and can coexist in the same cell, thus some bacteria can carry more than one kind of plasmid (Thomas, 2014). A study by Novick and Ruby (1975) found that six *S. aureus* plasmids, all carrying resistance to different antibiotics, were found to be compatible and coexist within the same cell.

1.6 Plasmids in biotechnology

Given the features of plasmids described above, it is not surprising that they have found wide ranging application in biotechnology (Lodish *et al.,* 2000; Shintani *et al.,* 2015). Their small size and simplistic structure allow for them to be separated from chromosomal DNA and readily modified by restriction enzymes and polymerase chain reaction.

Plasmids are most commonly used as cloning vectors in order to make many copies of genes of interest (Lodish *et al.,* 2000). This involves the plasmid being manipulated by restriction and recombination type enzymes in order for a gene of interest to be inserted into the multiple cloning site region of a vector, creating a recombinant plasmid (Lodish *et al.,* 2000). A recombinant plasmid is typically created by using a vector that carries genes for antibiotic resistance, as these resistance genes are used as selectable markers (Lodish *et al.,* 2000; Ochman *et al.,* 2000). Bacteria are transformed with the recombinant plasmid and are then grown in the presence of antibiotics to confirm if the recombination and transformation procedures were successful. The bacteria carrying the recombinant plasmid can then replicate and clone significant amounts of the genes of interest (Lodish *et al.,* 2000; Ochman *et al.,* 2000). The *ori* site feature of a plasmid is imperative for the amplification of insert genes (Lodish *et al.,* 2000; Ochman *et al.,* 2000). When host-cell enzymes bind to *ori,* initiating replication of the plasmid, replication continues around the plasmid regardless of the nucleotide sequence (Lodish *et al.,* 2000; Ochman *et al.,* 2000). Thus, any DNA sequence inserted into a plasmid is replicated along with the rest of the plasmid, and so valuable cargo genes can be artificially inserted into plasmids when designing vectors, in order to clone genes of interest (Tran and Boedicker, 2017). This has received interest in practices such as gene therapy and plant transformation, whereby plasmids are used as vectors for gene transfer into

human cells as potential treatment or into plant cells to create transgenic plant species that may be resistant to drought or pesticides (Lodish *et al.,* 2000; Baban *et al.* 2010)

Another use of plasmids is as expression vectors. In this case bacteria transformed with recombinant plasmids can be induced to carry out mass production of proteins encoded by the inserted genes of interest (Lodish *et al.,* 2000). Expression vectors are commonly used in pharmaceutical industries, where they are involved in the manufacturing of therapeutic agents such as enzymes, antibodies, insulin, antibiotics and growth hormones (Sousa *et al.,* 2009). Plasmid based therapeutic agents are advantageous over the traditionally produced agents as they are found to be more stable in different temperatures, are easier to store and transport, can be manufactured on a large-scale, can be produced at a low cost and in the case of vaccines, limit the need to use and inject infectious agents (Lodish *et al.,* 2000; Shintani *et al.,* 2015; Sousa *et al.,* 2009).

Studies have also highlighted the use of bacteriocin-producing plasmids as expression vectors in the food and beverage industry (Mortvedt and Nes, 1990). Increased production of bacteriocins can allow for a more reliable fermentation process as the growth of spoilage bacteria can be prevented by the high concentration of bacteriocins being present (Mortvedt and Nes, 1990). Bacteriocin production or resistance to bacteriocins could also be used to selectable markers in recombinant plasmid construction, which can aid in the future production of food preservatives or additives from microorganisms (Mortvedt and Nes, 1990). Furthermore, due to their antimicrobial properties, bacteriocins may be highly useful in drug design research and can be a potential drug candidate, replacing antibiotics in order to treat drug resistance pathogens (Yang *et al.,* 2014). Plasmids carrying temperature resistance factors have also gained increased biotechnological interest. Due to being able to withstand increased temperatures, the use of these plasmids can make high-temperature industrial processes more effective, as the probability of DNA or protein denaturation at high temperatures is lowered (Vásquez *et al.*, 1984).

While plasmids are often used as cloning or expression vectors, plasmids themselves have found extensive use due to the inherent cargo genes they carry. In particular, bacteria carrying degradative plasmids or recombinant plasmids constructed with genes of interest are used to carry out the metabolism of rare substrates (Lodish *et al.,* 2000). These substrates include aromatic and aliphatic containing compounds, herbicides, pesticides, chlorinated solvents, cyanide as well as polyethylene, polypropylene, and polystyrene, all of which are common environmental pollutants (Pathak and Navneet, 2017). Heavy metals are also common

constituents of the environment and overall these environmental pollutants can pose a threat of increased toxicity and adverse conditions to microbes and other organisms present (Ji and Silver, 1995).

The presence of these natural plasmids or recombinant DNA technology enables the host bacteria to degrade these environmental pollutants into carbon dioxide, water, biomass or into other harmless intermediates (Pathak and Navneet, 2017). Several studies have highlighted the use of natural or recombinant degradative plasmids to carry out the metabolism of environmental pollutants. A study by Demnerova *et al.* (1994), found that the natural degradative plasmids of *Pseudomonas putida* were able to carry out the degradation of camphor, ~10 carbons n-alkanes and phenol when introduced into selected Gram-negative bacterial strains that were indigenous to different polluted environments within the Czech Republic. Another study by Hada and Sizemore (1981) found that *Vibrio* spp. carrying plasmids encoding hydrocarbon digestion were capable of clearing crude oil from spills in an oil field in the Gulf of Mexico.

There are also several studies that highlight the use of the plasmid-encoded *mer* operon in the metabolism and decontamination of mercury in soil and aquatic environments (Ji and Silver, 1995; Nascimento and Chartone-Souza, 2003; Wagner-Döbler 2013). Bacteria with plasmid encoded resistance to environmental pollutants can be of increased value as biotechnological tools that are involved within bioremediation. These natural degradative plasmid-encoded systems or recombinant plasmids carrying the degradative genes can be employed to clear up polluted environments, as well as be used as preventative measures to limit pollution build-up (Ji and Silver, 1995).

In addition to decontaminating the environment, these plasmids, and the bacteria that carry them can be used as biomarkers to indicate environments that have a low- or high-concentration of heavy metals (Ji and Silver, 1995). This is observed in a study by Raja and Selvam (2011), where the pEGFP-1 plasmid vector was manipulated to carry the *cadR* gene (cadmium resistance operon) and the reporter gene green fluorescent protein (GFP), to construct a GFP-based biosensor sensitive to the heavy metal cadmium. This fluorescent protein was active in the presence of cadmium and therefore provides an alternative means for detecting toxic heavy metals, such as cadmium, within the environment (Raja and Selvam, 2011).

1.7 Aims and Objectives

This study consists of two main aims. Aim one focuses on determining the prevalence, evolutionary histories and functional characterization of the plasmids of the genera Geobacillus and Parageobacillus. This was carried out by identifying plasmid-related sequences in the available genome sequences of Geobacillus and Parageobacillus strains to determine the prevalence of plasmids within these genera. The sequence assembly of identified plasmids were improved to allow for structural and functional annotation, to gain insight into their encoded functions. Comparative genomics was also used to determine the evolutionary relationships of the plasmids among the genera. The second aim focuses on extracting plasmid DNA from a subset of Geobacillus and Parageobacillus strains to potentially functionally characterize plasmid-borne phenotypes. This was carried out by attempting plasmid elimination from selected Geobacillus and Parageobacillus strains using DNA extraction protocols and further curing strains of their plasmids to potentially determine phenotypic differences between wild-type and plasmid-cured strains.

1.8 References

Adams, V., Li, J., Wisniewski, J. A., Uzal, F. A., Moore, R. J., McClane, B. A. and Rood, J. I. (2014) Virulence Plasmids of Spore-Forming Bacteria. *Microbiology Spectrum,* 4(6), 116-121

Agnoli, K., Schwager, S., Uehlinger, S., Vergunst, A., Viteri, D., F., Nguyen, D. T., Sokol, P. A., Carlier, A. and Ebed, L. (2012) Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. *Molecular Microbiology*, 83, 362–378

Baban, C. K., Cronin, M., O'Hanlon, D., O'Sullivan, G. C., and Tangney, M. (2010). Bacteria as vectors for gene therapy of cancer. *Bioengineered Bugs*, 1(6), 385–394

Bernet-Camard, M. F., Liévin, V., Brassart, D., Neeser, J. R., Servin, A. L. and Hudault, S. (1994) *Lactobacillus acidophilus* LA1 binds to cultured human intestinal cell lines and inhibits cell-attachment and cell-invasion by enterovirulent bacteria. *Gut,* 35, 483-489

Brown, C. J., Sen, D., Yano, H., Bauer, M. L., Rogers, L. M. and Van der Auwera, G. A. (2013) Diverse broad-host-range plasmids from freshwater carry few accessory genes. *Applied and Environmental Microbiology*, 79, 7684–7695

Carattoli, A. (2008) Chapter 15: Evolution of Plasmids and Evolution of Virulence and Antibiotic-Resistance Plasmids. *American Society for Microbiology*, 1-2

Chen, C. W. (1996) Complications and implications of linear bacterial chromosomes. *Trends in Genetics*, 12, 192-196

Clark, D. P. and Pazdernik, N. J. (2013) Plasmids, *Molecular biology*, 20(2), 473-478 del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M., and Diaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiology and Molecular Biology Reviews*. 62, 434–464

del Solar, G. and Espinosa, M. (2000) Plasmid copy number control: an ever-growing story. *Molecular Microbiology*, 37(3), 492-500

Demnerova, K., Kralova, B. and McBeth, D. L. (1994) Introduction of Biodegradative Plasmids to Indigenous. *International Biodeterioration & Biodegradation, Elsevier science limited,* 27, 311-318

Fire, A. and Xu, S. Q. (1995) Rolling replication of short DNA circles. *Proceedings of the National Academy of Sciences of the United States of America*, 92: 4641-4645

Furukawa, K. and Tonomura, K. (1972) Mettalic mercury releasing enzyme in mercury resistant *Pseudomonas*. *Agricultural and Biological Chemistry*. 36, 217-226

Garneau-Tsodikova, S. and Labby, K. J. (2015) Mechanisms of Resistance to Aminoglycoside Antibiotics: Overview and Perspectives. *MedChemComm,* 7(1), 11–27

Gogarten, J. P. and Townsend, J. P. (2005) Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology*, 3, 679-687

Guo, S. and Mahillon, J. (2013) pGIAK1, a heavy metal resistant plasmid from an obligate alkaliphilic and halotolerant bacterium isolated from the Antarctic Concordia station confined environment. *PloS one*, 8(8), e72461

Hada, H. S. and Sizemore, R. K. (1981) Incidence of plasmids in marine Vibrio spp. isolated from an oil field in the northwestern Gulf of Mexico. *Applied Environmental Microbiology*, 41, 199-202

Hoshino, T., Ikeda, T., Narushima, H. and Tomizuka, N (1985) Isolation and characterization of antibiotic-resistance plasmids in thermophilic bacilli. *Canadian Journal of Microbiology*, 31(4), 339-345

Hynes, M. F. and McGregor, N. F. (1990) Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen-fixing nodules by *Rhizobium leguminosarum*. *Molecular Microbiology*, 4(4), 567-574

Jacoby, G. A. (2005) Mechanisms of resistance to quinolones. *Clinical Infectious Disease*, 41(15), 2-6

Ji, E. G. and Silver, S. (1995) Heavy Metal Resistance Plasmids and Use in Bioremediation. *Environmental Biotechnology*, 20, 47-62

Johnson, C. M. and Grossman, A. D. (2015). Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annual Review of Genetics*, 49, 577–601.

Johnson, T. J. and Nolan, L. K. (2009) Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiology and Molecular Biology Reviews*, 73(4), 750-774

Kado, C. I. (1998) Origin and evolution of plasmids. *Antonie Van Leeuwenhoek*, 73(1), 117126

Khalil, A. B., Anfoka, G. H. and Bdour, S. (2003) Isolation of plasmids present in thermophilic strains from hot springs in Jordan. *World Journal of Microbiology and Biotechnology*, 19(3), 239-241

Kroll, J., Klinter, S., Schneider, C., Voss, I., and Steinbuchel, A. (2010) Plasmid addiction systems: perspectives and applications in biotechnology. *Microbial Biotechnology*, 3(6), 634657

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D. and Darnell, J. (2000) DNA Cloning with Plasmid Vectors, *Molecular Cell Biology*, (4)

Loewe, L., and Hill, W. G. (2010) The population genetics of mutations: good, bad and indifferent. *Biological Sciences*, 365(1544), 1153–1167

MacLean, R. C. and San Millan, A. (2015) Microbial Evolution: Towards Resolving the Plasmid Paradox, *Cell press*, 25(17), 764-767

Manetsberger, J., Hall, E. A. and Christie, G. (2015) Plasmid-encoded genes influence exosporium assembly and morphology in *Bacillus megaterium* QM B1551 spores. *FEMS Microbiology Letters*, 362(18), 10

Martínez, J. L. (2008) Antibiotics and antibiotic resistance genes in natural environments. *Science*, 321, 365-367

Medema, M. H., Trefzer, A., Kovalchuk, A., Van den Berg, M., Muller, U., Heijne, W., Wu, L., Alam, M, T., Ronning, C. M., Nierman, W. C., Bovenberg, R. A. L., Breitling, R. and Takano, E. (2010) The Sequence of a 1.8-Mb Bacterial Linear Plasmid Reveals a Rich Evolutionary Reservoir of Secondary Metabolic Pathways. *Genome Biology and Evolution*, 2, 212-224

Meletzus, D., Bermphol, A., Dreier, J. and Eichenlaub, R. (1993) Evidence for plasmid-encoded virulence factors in the phytopathogenic bacterium *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382. *Journal of Bacteriology*, 175(7), 2131–2136

Mortvedt, C. I and Nes, I. F. (1990) Plasmid-associated bacteriocin production by a *Lactobacillus sake* strain. *Microbiology,* 136, 1601-1607

Munita, J. M. and Arias, C. A. (2016) Mechanisms of Antibiotic Resistance. *Microbiology Spectrum,* 4(2), 10

Nascimento, A. M. A. and Chartone-Souza, E. (2003) Operon *mer:* Bacterial resistance to mercury and potential for bioremediation of contaminated environments. *Genetics and Molecular Research,* 2(1), 92-101

Novick, R. P. and Ruby, C. (1975) Plasmid interactions in *Staphylococcus aureus*: Nonadditivity of compatible plasmid DNA pools. *Proceedings of the National Academy of Sciences of the United States of America,* 72(12), 5031-5035

Nucifora, G., Chu, L., Silver, S. and Misra, T. K. (1989) Mercury Operon Regulation by the merR Gene of the Organomercurial Resistance System of Plasmid pDU1358. *Journal of Bacteriology,* 171(8), 4241-4247

Ochman, H., Lawrence, J. G. and Groisman, E. A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature,* 405, 299-304

Osborn, M. A., Bruce, K. D., Strike, P. and Ritchie, D. A. (1997) Ditribution, diversity and evolution of the bacterial mercury resistance (mer) operon. *FEMS Microbiology reviews,* 19, 239-262

Pathak, V. and Navneet, M. (2017) Review on the current status of polymer degradation: a microbial approach. *Bioresour Bioprocess* 4, 15

Pemberton, M. (1983) Degradative Plasmids. *International Review of Cytology.* 84, 155-183

Peterson, J.W. (1996) Chapter 7 Bacterial Pathogenesis, *Medical Microbiology*. 4(1), 17

Popowska, M. and Krawczyk-Balska, A. (2013) Broad-host-range IncP-1 plasmids and their resistance potential. *Frontiers in Microbiology*, 4, 44

Providenti, M.A., O'Brien, J.M., Ewing, R.J., Paterson, E.S. and Smith, M.L. (2006) The copynumber of plasmids and other genetic elements can be determined by SYBR-Green-based quantitative real-time PCR. *Journal of Microbiology Methods,* 65, 476–487

Raja, E. C. and Selvam, G. S. (2011) Construction of green fluorescent protein based bacterial biosensor for heavy metal remediation. *International Journal of Environmental Science and Technology,* 8 (4), 793-799

Rijavec, M., Budic, M., Mrak, P., Muller-Premru, M., Podlesek, Z. and Zgur-Bertok, D. (2007) Prevalence of ColE1-Like Plasmids and Colicin K Production among Uropathogenic *Escherichia coli* Strains and Quantification of Inhibitory Activity of Colicin K, *Applied and Environmental Microbiology,* 73(3), 1029-1032

Riley, M.A. and Gordo, D.M. (1992) A survey of Col plasmids in natural isolates of *Escherichia coli* and an investigation into the stability of Col-plasmid lineages. *Journal of General Microbiology*, 138, 1345-1352

San Millan, A. (2018) Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. *Trends in Microbiology*, 26(12), 978-985

Segura, A., Molina, L. and Ramos, J. L. (2014) Plasmid-Mediated Tolerance Toward Environmental Pollutants. *Microbiology Spectrum*, 2(6), 4

Shintani, M., Sanchez, Z. K. and Kimbara, K., 2015) Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in Microbiology*, 6, 242

Smalla, K., Jechalke, S., and Top, E. M. (2015) Plasmid detection, characterization and ecology. *Microbiology Spectrum*, 3(1)

Smajs, D., Micenkova, L., Smarda, J., Vrba, M., Sevcikova, A., Valisova, Z. and Woznicova, V. (2010) Bacteriocin synthesis in uropathogenic and commensal *Escherichia coli*: colicin E1 is a potential virulence factor, *BMC Microbiology*, 288(10), 7488

Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., and de la Cruz, F. (2010) Mobility of Plasmids. *Microbiology and Molecular Biology Reviews: MMBR*, 74(3), 434–452

Sousa, F., Passarinha, L. and Queiroz, J. A. (2009) Biomedical application of plasmid DNA in gene therapy: A new challenge for chromatography. *Biotechnology and Genetic Engineering Reviews*, 26(1), 83-116

Stevens, R. H. (2015) Transduction-mediated horizontal gene transfer in the oral microbiome. *Frontiers in Cellular and Infection Microbiology*, 5(12), 1-2

Stone, A. B. (1975) R factors: plasmids conferring resistance to antibacterial agents. *Science Progress*, 62(245), 89-101

Thomas, C. M. (2014) Plasmid Incompatibility, In: Bell E. (eds) *Molecular Life Sciences*, 8, 130

Tran, F., & Boedicker, J. Q. (2017). Genetic cargo and bacterial species set the rate of vesicle mediated horizontal gene transfer. *Scientific Reports*, 7, 8813

Urbanek, A. K., Rymowicz, W., and Mirończuk, A. M. (2018) Degradation of plastics and plastic-degrading bacteria in cold marine habitats. *Applied Microbiology and Biotechnology*, 102(18), 7669–7678

Vásquez, C., González, B. and Vicuña, R. (1984) Plasmids from thermophilic bacteria. *Comparative Biochemistry*, 78(3), 507-514

Ventura M., Canchaya C., Tauch A., Chandra G., Fitzgerald G. F. and Chater K. F. (2007) Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. *Microbiology and Molecular Biology Reviews.* 71, 495–548

Wagner-Döbler, I. (2013) Bioremediation of Mercury: Current Research and Industrial Applications. *Scientific Press*, 62, 124-133

Wu, S., Dalsgaard, A., Hammerum, A. M., Porsbo, L. J. and Jensen, L. B. (2010) Prevalence and characterization of plasmids carrying sulfonamide resistance genes among *Escherichia coli* from pigs, pig carcasses and human. *Acta Veterinaria Scandinavia*¸50(3), 47-52

Yang, S. C., Lin, C. H., Sung, C. T. and Fang, J. Y. (2014) Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Frontiers in Microbiology*, 5, 241

Zambryski, P., Joos, H., Genetello, C., Leemans, J., Montagu, M. V. and Schell, J. (1983) Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. *The EMBO Journal*, 2(12), 2143–2150

Chapter 2: The prevalence and evolution of plasmids in the genera *Geobacillus* and *Parageobacillus*

Abstract

A few isolated studies have identified plasmids in several members of the genera *Geobacillus* and *Parageobacillus*. However, little is known about the prevalence and diversity of plasmids within these genera. Here plasmids have been identified in a comprehensive set of *Geobacillus* and *Parageobacillus* strains for which genome sequences are available through comparative genomic approaches. There is a broad range of plasmids amongst taxa of both genera, that can be classified into nine distinct plasmid groups (pGeo1 to pGeo9) on the basis of protein orthology. These plasmids are found to be highly structurally diverse and are predicted to be replicated and maintained by a range of distinct mechanisms. Furthermore, the low number of conserved proteins amongst these plasmids suggest these genetic elements contribute to the genetic diversity of individual taxa in these genera as a whole.

1.1 Introduction

The genera *Geobacillus* and *Parageobacillus* are comprised of Gram-positive, rod shaped, spore-forming, aerobic or facultatively anaerobic bacteria (Hussein *et al.*, 2015; Zeigler, 2005). Members of these genera are further characterized by their thermophilicity, growing optimally at temperatures between 45 to 90 °C (Zeigler, 2005; Nazina *et al.*, 2001). As such, they are frequently isolated from high temperate environments such as geothermal pools, hydrothermal vents and oil wells (Zeigler, 2014). The genus *Geobacillus* was first proposed in 2001, when members of the previously categorized genetic group V within the genus *Bacillus*, were transferred to the new genus *Geobacillus* on the basis of their distinct phenotypic features and 16S rRNA gene sequences (Nazina *et al.*, 2001). Subsequently, some species within the genus *Geobacillus* were moved to a novel genus, *Parageobacillus*, on the basis of genome-wide phylogenetic analysis and their distinct genomic G+C content (Aliyu *et al.*, 2016). Currently, sixteen and five species have been validly described for the genus *Geobacillus* and *Parageobacillus*, respectively.

Due to their thermophilic nature, members of the genera *Geobacillus* and *Parageobacillus* are able to produce a range of thermostable enzymes. As such, these taxa and their enzymes have garnered intense research interest, as thermostable enzymes can greatly benefit many different industrial applications where high operating temperatures are often used (Balan *et al.*, 2012). Thermostable enzymes are able to withstand these temperatures which also increases their catalytic activity and thus increases product turnover (Hussein *et al.*, 2015). Furthermore, these

enzymes are often resistant to organic solvents and can have unique enantio specificity and stereo specificity which suggests they may be of use in a wide variety of industrial applications (Hussein *et al.,* 2015). Enzymes from the geobacilli which have been studied and employed include thermostable α-amylase which cleaves the a-1,4-glucosidic linkages of starch to generate glucose monomers, proteases used in leather processing and the manufacturing of food, detergents and pharmaceuticals, as well as lipases and carboxylases which catalyse the hydrolysis of carboxyl esters and are important in the cosmetic and food industries (AbdelFattah and Gaballa, 2008; Ewis *et al.,* 2004; De Souza and de Oliveira Magalhaes, 2010; Hussein *et al.,* 2015; Khanniri *et al.,* 2015; Montoro-García *et al.,* 2009).

In addition to the thermostable enzymes produced by members of these genera, their thermophilic nature and their metabolic versatility also imply that they can be used themselves in a wide range of whole cell applications. These include the biorefinement of linen fibres and the bioremediation of environmental pollutants such as polyaromatic hydrocarbons (PAH), alkanes, phenols and organophosphate-containing pesticides (Hussein *et al.,* 2015; Obojska *et al.,* 2002).

The genome sequence of several *Geobacillus* and *Parageobacillus* species have indicated the presence of plasmids, that range in size from small (<5kb) to mega-plasmids (Hussein *et al.*, 2015). The genome of *Geobacillus thermodenitrificans* NG80-2 includes a 57.7 kilobase plasmid (pLW1071) (Feng *et al.*, 2006), while two plasmids, 58.8 and 57.0 kb in size, are present in *G. thermodenitrificans* T12 (Daas *et al.*, 2018). Larger plasmids have also been observed in the genomes of other *Geobacillus* species, including *G. stearothermophilus* 18 (pGS18) and *G. kaustophilus* HTA426 (pHTA426) (Stuknyte *et al.,* 2008; Takami *et al.,* 2014) Similarly, the genomes of *P. thermoglucosidasius* NCIMB 11955 (pNIC001 – 84.0 kb and pNIC002 – 47.9 kb) and C56-YS93 (pGeoTH01 – 80.8 kb and pGEOTH02 – 19.6 kb) incorporate two large plasmids each (Brumm *et al.,* 2015; Sheng *et al.,* 2016). Smaller plasmids have also been identified and characterized. For example, *Geobacillus* sp. 610 and 1121 incorporate the plasmids, pGTD7 and pGTG5, 3 279 bp and 1 540 bp in size, respectively (Kananavičiūtė *et al.,* 2014).

Plasmids are capable of carrying genes, different to the host cell chromosome, which can confer various phenotypes to their host bacterium (De Maayer *et al.,* 2012; Gull and El-Baz, 2018). Thus far, several plasmids in *G. thermodentrificans* and *G. stearothermophilus* have been shown to be involved in the production of thermostable α-amylase, encode resistance against the antibiotics streptomycin, tetracycline and kanamycin, the production of bacteriocins and

encode genes that enable the degradation of polyaromatic hydrocarbon (PAH) and alkanes (Mielenz, 1983; Hoshino *et al.*, 1985; Stahl, 1991; Pokusaeva *et al.*, 2009; Kumar *et al.*, 2012; Feng *et al.*, 2006; Hussein *et al.*, 2015). However, while some preliminary research has focused on a few *Geobacillus* plasmids, little is known about the prevalence and diversity of plasmids in members of the genera *Geobacillus* and *Parageobacillus*. Here the genomes of an exhaustive set of geobacilli were analyzed and a large number of plasmids across the broad taxonomic spectrum of the two genera were identified. Comparative genomic and phylogenetic analyses revealed that these plasmids can be organized into nine distinct plasmid groups that have been derived through distinct evolutionary events, thus suggesting that extensive diversity exists among the geobacilli plasmids.

2.2 Methods

2.2.1 Plasmid sequence extraction and assembly

The genome sequences of 63 *Geobacillus* and 22 *Parageobacillus* spp. were obtained from the National Center for Biotechnology Information (NCBI) genome database (Pruitt *et al.,* 2005) and the Integrated Microbial Genome (IMG) Database (Chen *et al.,* 2018) (Accession date: 05/03/2018-07/03/2018). The assemblies of draft genome sequences were improved by scaffolding using the multi-draft-based scaffolder MeDuSa (Bosi *et al.,* 2015) and Mauve v. 2.3.0 (Darling *et al.,* 2004). The complete and resultant high-quality draft genomes were structurally and functionally annotated using the Rapid Annotation using Subsystems Technology (RAST) server (Overbeek *et al.,* 2014). The functional annotations were then searched for plasmid-related key words, including "plasmid", "replication" and "origin of replication". Furthermore, plasmid replication (Rep) and maintenance proteins from previously completed *Geobacillus/Parageobacillus* plasmids were used in local tBLASTn against the draft genome to identify contigs encoding orthologues of these proteins. Plasmid contigs were extracted from the draft genomes and the ends were subjected to BLASTn searches against the NCBI non-redundant (nr) nucleotide database to identify potential overlaps across the sequence ends. The final plasmid sequences were again subjected to structural and functional annotation using the RAST server (Overbeek *et al.,* 2014). The G+C contents of plasmids were determined using Bioedit v. 7.2.5 (Hall, 1999). To identify the origin of replication (*ori*) on the geobacilli plasmids, the plasmid sequences were submitted to the DoriC database and *ori* regions were identified by nucleotide BLAST analysis (Luo and Gao, 2018).

2.2.2 Plasmid grouping

The protein complements encoded on each of the plasmids were used to determine the presence of orthologous proteins amongst the plasmids. This was done using OrthoFinder v 1.1.4, which performs pairwise BLASTp analysis between each individual plasmid protein set and assigned orthologous proteins (shared between two or more of the plasmid datasets) to distinct orthogroups (Emms and Kelly, 2015). On the basis of shared orthogroups, the plasmids were assigned to different groups, where a plasmid group was considered for those plasmids which share >6 orthologues between all plasmids in the group, however, the plasmids of group pGeo5 were grouped together on the basis of two shared orthologues given the few protein coding sequences identified on each plasmid. Partial plasmids were not included in this analysis. However, local BLASTp analyses in Bioedit v 7.2.5 with the core orthologues from each group were used to classify these partial plasmids according to their plasmid group (Hall, 1999).

Orthologues were considered as those proteins sharing >50% amino acid identity over 50% of the alignment length. The conserved orthologues in each plasmid group were functionally annotated by BLASTp comparison against the NCBI non-redundant (nr) protein database and the Conserved Domain Database using the Batch Conserved Domain (CD) Search function (Marchler-Bauer *et al.,* 2010, Marchler-Bauer *et al.,* 2014).

2.2.3 Phylogenetic analyses

A core genome phylogeny was constructed on the basis of the 85 geobacilli included in this study. Single-copy orthologues (SCOs) were identified using Orthofinder v. 1.1.4 (Emms and Kelly, 2015). The resultant 1,245 proteins were aligned using the m-Coffee implementation of the T-Coffee multiple sequence alignment algorithm v 11.00.8 (Notredame *et al.,* 2000). The alignments were concatenated, and poorly aligned regions were trimmed using Gblocks v. 0.91b (Talavera and Castresana, 2007). The resultant alignment comprised of 321,025 amino acids and was used to construct a maximum likelihood phylogeny using Phyml-SMS 3.0 (Guindon *et al.,* 2010; Lefort *et al.,* 2017) with the predicted best-fit Le and Gasquel (LG) amino acid substitution model and bootstrap support (n = 100 replicates).

Similarly, phylogenies were constructed on the basis of orthologous proteins conserved on the plasmids belonging to each potential plasmid group (pGeo1 (6 SCOs) and pGeo2 (23 SCOs) plasmids). The trimmed and concatenated alignments included 1,195 and 6,676 amino acid sites for pGeo1 and pGeo2, respectively. For comparative purposes core genome phylogenies were constructed on the basis of 1,441 SCOs and 1,616 SCOs conserved among the strains containing the pGeo1 and pGeo2 plasmids, respectively. These trimmed alignments were 379,449 and 436,053 amino acids in length, respectively and the phylogenies were constructed as described above. Plasmid phylogenies for groups pGeo3 and pGeo5 could not be constructed because of the lack of orthologous proteins and the low number of strains (<3 strains) that contain each plasmid.

The CGViewer server applet (Stothard and Wishart, 2008) was subsequently used to draw circular maps for the plasmid groups pGeo1, pGeo2, pGeo3 and pGeo5. This server incorporates genetic sequences and uses BLASTp or BLASTn (depending on input sequences) to compare the sequences, providing a graphical output showing the regions of conserved genes or proteins amongst all the input sequences (Grant and Stothard, 2008).

2.3 Results and Discussion

2.3.1 Plasmids are prevalent features among the geobacilli

A total of 63 plasmids were identified among the geobacilli. Of these, seven plasmids were either embedded within chromosomal contigs and therefore could not be extracted or they comprised of too many contigs to be assembled into complete plasmids. The remaining 56 plasmids were either complete and circularized (48 plasmids) or assembled to near-complete status (eight plasmids). The plasmids showed a broad distribution across the geobacilli, with 30/63 (47.62%) *Geobacillus* and 16/22 (72.73%) *Parageobacillus* strains harbouring plasmid elements. At the species level, plasmids show a random distribution, with no species harbouring plasmids in all strains incorporated in the study. When considering those species where more than one genome is available, the highest prevalence for the genus *Geobacillus* was observed for *G. thermodenitrificans* (6/9 strains; 66.67%), *G. stearothermophilus* (8/13 strains; 61.54%), *G. thermoleovorans* (9/19 strains; 47.42%), *G. icigianus* (2/4 strains; 50%), *Geobacillus* genomosp. 4 (1/5 strains; 20%) and *Geobacillus* genomosp. 1 (1/2 strains 50%), while no plasmids were observed in *G. thermocatenulatus* (0/3 strains) or *G. subterraneus* (0/3 strains). Among the *Parageobacillus* spp., plasmids were prevalent among *P. caldoxylosilyticus* (2/3 strains; 66.67%), *P. thermoglucosidasius* (8/11 strains; 72.73%) and *P. toebii* (4/5 strains; 80%) (Figure 1).

The majority of plasmid-harbouring strains incorporate only a single plasmid (30/46; 65.22%), while fourteen strains (30.43%) harbour two plasmids (Figure 1). Particularly in *P. thermoglucosidasius*, the occurrence of two plasmids is a frequent observation, with 6/8 plasmid-bearing strains incorporating two distinct plasmids. Two taxa, namely *G. thermoleovorans* N7 and *P. caldoxylolyticus* B4119 include three plasmids, but two of these are partial and could not be assembled further in the latter strain (Figure 1). The geobacilli plasmids show great diversity both in plasmid size and the number of proteins encoded on these extrachromosomal elements. The geobacilli plasmids range in size from 1.4 (*G. thermoleovorans* A8) to 95.4 (*P. thermoglucosidasius* G08C001) kilobases and encode between two (*G. thermoleovorans* A8) and 134 (*P. thermoglucosidasius* G08C001) proteins (Table 1). As such, plasmid-encoded proteins contribute between 0.05% (*G. thermoleovorans* A8) and 3.2% (*P. thermoglucosidasius* G08C001) of the total protein content in the geobacilli [This was calculated by the number of plasmid encoded proteins divided by the total number of genome encoded proteins and then multiplied by 100].

Analysis of the protein complement showed that no single protein is perfectly conserved among all the geobacilli plasmids. However, some plasmids contain a number of conserved proteins, and on the basis of this, the plasmids were grouped according to distinct plasmid types.

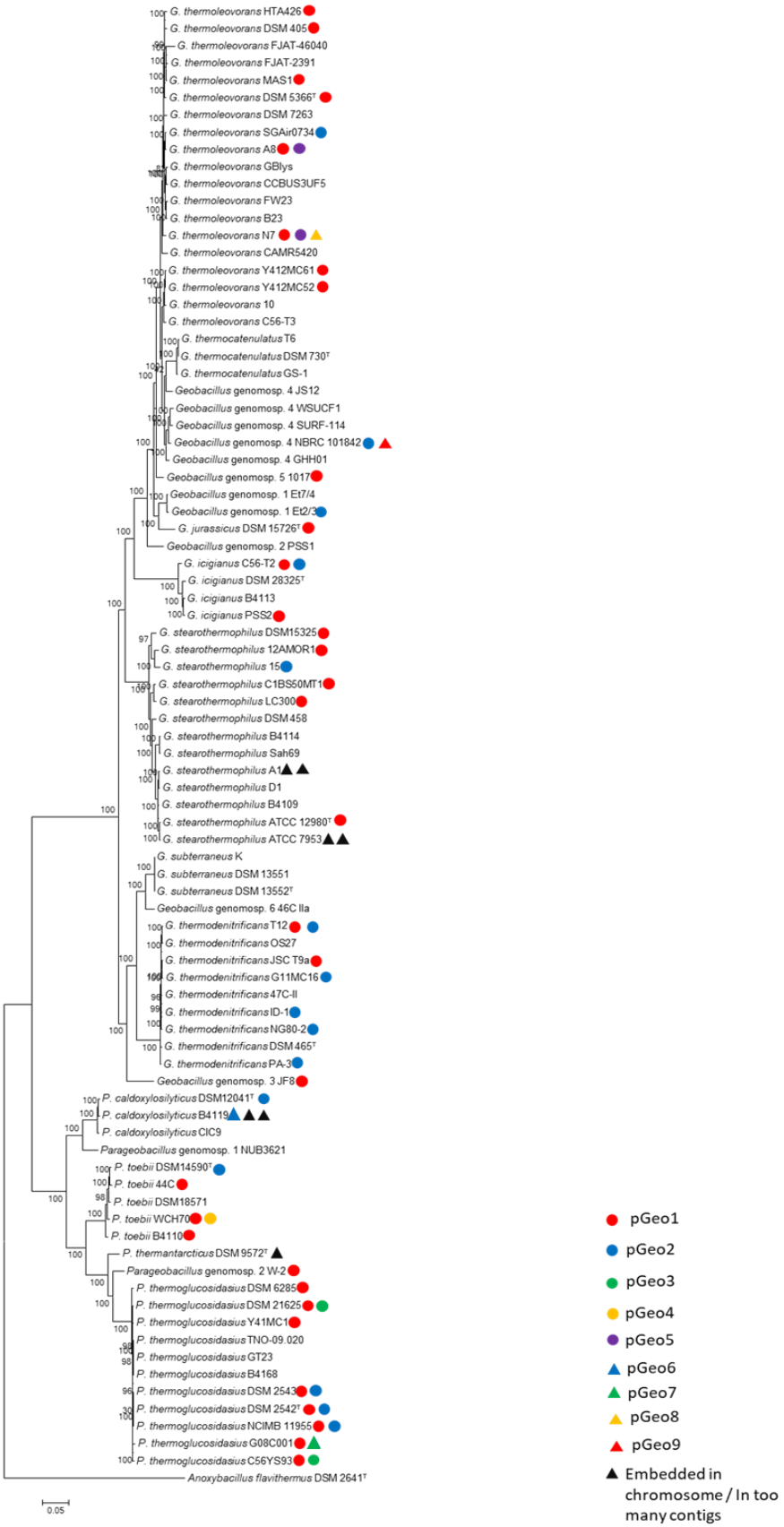


Figure 1: Plasmid prevalence among the genera *Geobacillus* and *Parageobacillus*. A core genome ML phylogeny constructed on the basis of the 85 strains analyzed in this study (*Anoxybacillus flavithermus* DSM 2641^T was used as outgroup to root the tree). Each plasmid group is denoted by a different colour as well as by either a circle or triangle. The seven plasmids that were unable to be further analyzed in this investigation are indicated by black triangles.

2.3.2 The geobacilli plasmids belong to nine distinct plasmid groups

The protein complements of the 56 complete or near-complete geobacilli plasmids were used to identify potential orthologues shared among the plasmids using Orthofinder v 1.1.4 (Emms and Kelly, 2015). Given the cutoff criteria, no proteins were found to be universally conserved among all the plasmids suggesting that the plasmids, or the groups identified either have completely distinct evolutionary histories or diverged long ago. In general, bacterial plasmids are far more flexible in their DNA sequence structure than the chromosome (Dutta and Pan, 2002). This is due to the abundant presence of transposable elements that facilitate inter- or intra-molecular genetic recombination, the frequency with which they gain or lose genes during conjugation or replication and, no requirement to encode functions vital for cell survival (e.g. rDNA) (Dutta and Pan, 2002; Fernández-López *et al.*, 2006). As a result, the evolutionary pathways of plasmids are difficult to understand, as there is a lack of extensive similarities between them and this therefore limits the evolutionary dynamics of plasmids to the elements that are shared between them (Dutta and Pan, 2002; Fernández-López *et al.*, 2006)

Of the 56 complete or near-complete plasmids, a total of 50 plasmids are > 10 kilobases in size. Orthology analysis identified thirty-two plasmids which share six orthologous proteins, while fifteen plasmids shared twenty-two orthologues and a further two plasmids shared six orthologues. These were subsequently grouped as pGeo1, pGeo2 and pGeo3 plasmids, respectively. No SCOs are retained when comparing the pGeo1, pGeo2 and pGeo3 plasmids, suggesting they represent evolutionarily independent elements. A single plasmid, at 10.3 kilobases and sharing no orthologues to any of the other plasmids was termed pGeo4 (Table 1). Lastly, of the remaining plasmids < 10 kb in size, two plasmids shared one ortholog and these were classified as pGeo5. Finally, the remaining four plasmids were found not to share any orthologues with other plasmids and were termed pGeo6, pGeo7, pGeo8 and pGeo9 (Table 1).

The pGeo1 plasmids are the most abundant constituting 50.79% of the identified plasmids and occur in species of both genera. They are observed to be particularly concentrated in the species *G. thermoleovorans* (eight strains), *G. stearothermophilus* (five strains) and *P. thermoglucosidasius* (eight strains), while the remaining eleven pGeo1 plasmids are dispersed amongst other species of the two genera. The pGeo2 plasmids make up 23.81% of the identified plasmids and are also dispersed amongst different species. Only eight of pGeo2 plasmids are found in the same species, with plasmids conserved among four *G. thermodenitrificans* and three *P. thermoglucosidasius* strains, respectively (Figure 1).

The remaining plasmid groups embody ~ 13% of the plasmids of these genera, with four harboured by *Geobacillus* spp. (pGeo5, pGeo8, pGeo9) and five by *Parageobacillus* spp. (pGeo3, pGeo4 pGeo6 and pGeo7), respectively. Furthermore, most of these plasmids are found in strains that harbour either pGeo1 or pGeo2 plasmids, with the exception of the pGeo7 plasmid of *P. caldoxylosilyticus* B4119 which is observed to coexist with two plasmids which are fragmented and were thus not included in this investigation (Figure 1). The co-occurrence of these smaller plasmids with pGeo1 and pGeo2 suggests that, due to their smaller sizes, they may lack self-transmission genes, thus requiring the aid of larger plasmids to assist in their dissemination and maintenance in the geobacilli (Smillie *et al.*, 2010). The distinct plasmid groups were subsequently assessed individually to gain a better insight into the core elements of each plasmid and their evolutionary histories.

Table 1: General characteristics of the geobacilli plasmid groups

Plasmid group	# plasmids	Species (# of strains)	Size range (kbp)	Average G+C content (%)	# protein coding sequences	# shared orthologues
pGeo1	32	*Geobacillus* genomosp. 3 (1) *Geobacillus* genomosp. 5 (1) *G. icigianus* (2) *G. jurassicus* (1) *G. stearothermophilus* (5) *G. thermodenitrificans* (2) *G. thermoleovorans* (8) *Parageobacillus* genomosp. 2 (1) *P. thermoglucosidasius* (8) *P. toebii* (3)	16.2 – 95.4	43.8	20 – 102	6
pGeo2	15	*Geobacillus* genomosp. 1 (1) *Geobacillus* genomosp. 4 (1) *P. caldoxylosilyticus* (1) *G. icigianus* (1) *G. staerothermophilus* (1) *G. thermodenitrificans* (5) *G. thermoleovorans* (1) *P. thermoglucosidasius* (3) *P. toebii* (1)	43.4 – 70.1	39.9	48 – 75	23
pGeo3	2	*P. thermoglucosidasius* C56YS93 *P. thermoglucosidasius* DSM 21625	19.9 and 30.6	40.4	21 and 36	6
pGeo4	1	*P. toebii* WCH70	10.3	39.6	12	0
pGeo5	2	*G. thermoleovorans* A8 *G. thermoleovorans* N7	1.4 and 3	47.6	2 and 5	2
pGeo6	1	*P. caldoxylosilyticus* B4119	1.8	41.64	2	0
pGeo7	1	*P. thermoglucosidasius* G08C001	1.6	42.4	3	0
pGeo8	1	*G. thermoleovorans* N7	6.2	41.5	14	0
pGeo9	1	Geobacillus genomosp. 4 NBRC101842	1.9	47.6	4	0

2.3.3 Characteristics of the pGeo1 plasmids

The pGeo1 plasmid group comprises thirty-two plasmids and occurs in both *Geobacillus* and *Parageobacillus* spp. These plasmids vary significantly in size, ranging between 16,2 kb (*G. thermoleovorans* MAS1 [also classified as *G. thermopakistaniens* (Siddiqui *et al.,* 2014)]) and 95,4 kb (*P. thermoglucosidasius* G08C001). Furthermore, they encode between 20 (*G. thermoleovorans* MAS1) and 102 (*P. thermoglucosidasius* G08C001) proteins (Table 1). The G+C contents of these plasmids range from 38.91% (*G. thermodenitrificans* T12) to 47.4% (*G. thermoleovorans* MAS1). The overall genomic G+C content of the *Geobacillus* and *Parageobacillus* strains that harbour pGeo1 plasmids ranges from 42.14% (*P. toebii* 44C) to 53.07% (*G. stearothermophilus* ATCC 12980). The plasmid DNA shows an average G+C deviation of -4.96% (range: -0.64% {*P. toebii* 44C} to -13.27% {*G. stearothermophilus* ATCC 12980}) from the chromosome.

It is predicted that when a plasmid is maintained within a cell, the same factors that homogenize the nucleotide signatures of the chromosomal DNA will also influence the plasmid nucleotide signature to drift toward that of the host by amelioration (Campbell *et al.,* 1999; Ochman *et al.,* 2000; Rocha and Danchin, 2002). Most horizontally transferred DNA elements are found to have a lower G+C content than that of the host chromosomal DNA and the extent to which the values differ can suggest if these elements have recently been acquired or established within the host cell or is long-term resident of a cell (Campbell *et al.,* 1999; Ochman *et al.,* 2000; Rocha and Danchin, 2002).

Six protein orthologues were identified amongst the thirty-two pGeo1 plasmids. This constitutes between 5.88% (*P. thermoglucosidasius* G08C001) and 30.00% (*G. thermoleovorans* MAS1) of the total proteins encoded on each pGeo1 plasmids. This is depicted by the circular map which represents an overview of the protein coding sequence of the pGeo1 plasmids. An extensive mosaic structure in conserved protein coding sequences is observed from the few shared protein coding regions amongst the plasmids, thus suggesting increased diversification of these plasmids among the geobacilli. (Figure 2). Furthermore, the circular map depicts regions of genomic islands, represented as large increases or decreases from average G+C content of the plasmids used in the map. These genomic islands can suggest that these regions are prone to the insertion of new genes and could thus be indicative of horizontal gene transfer (Guo *et al.,* 2014).

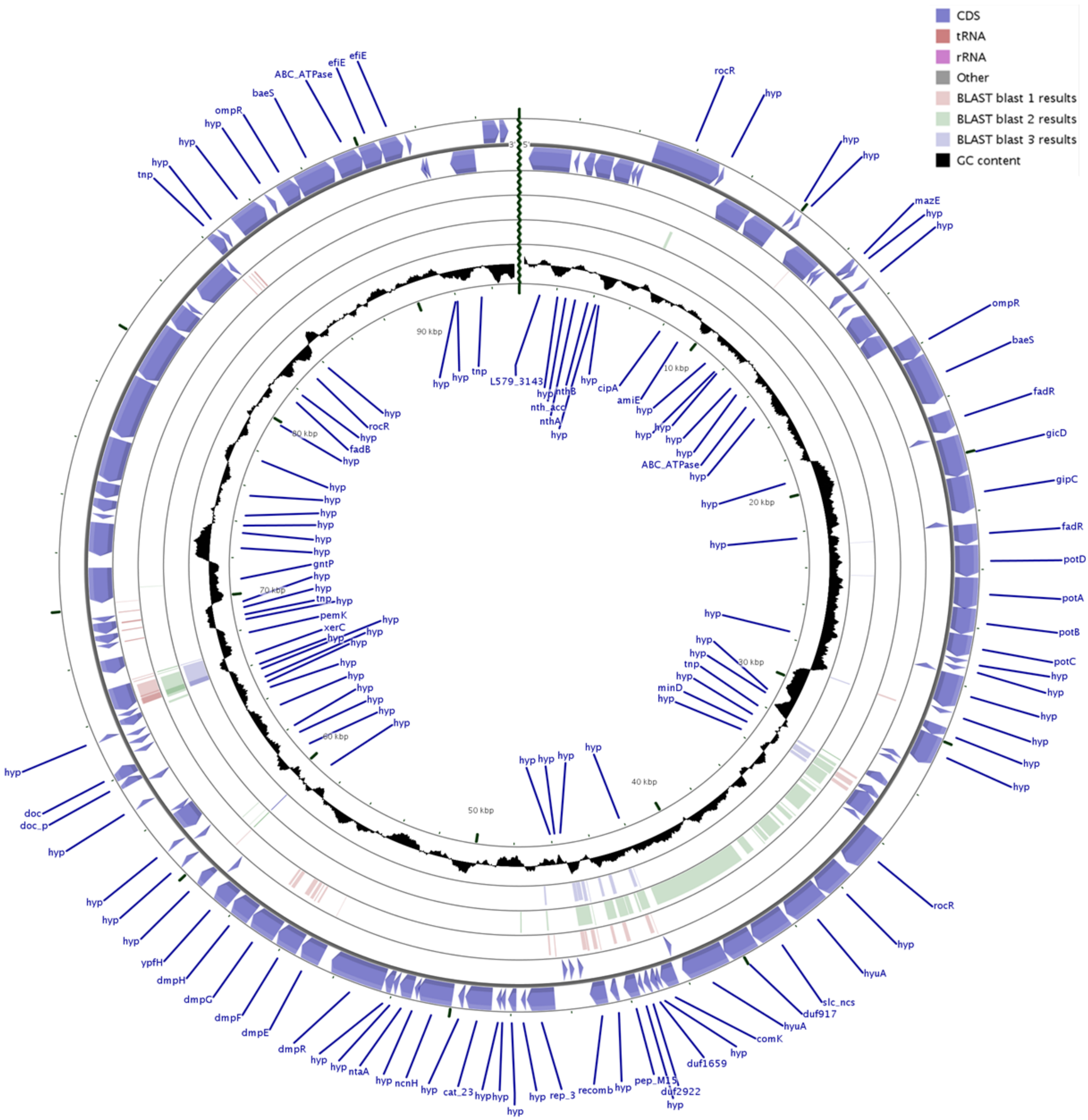


Figure 2: Circular map of the pGeo1 plasmids. The largest pGeo1 plasmid (*P. thermoglucosidasius* G08C001) was used as reference. The plasmid nucleotides sequences of pGeo1 plasmids *Geobacillus* genomospecies 3 JF8 (third circle from outside; BLAST 1 results), *Parageobacillus toebii* 44C (fourth circle from outside, BLAST 2 results) and *G. thermoleovorans* HTA426 (fifth circle from outside, BLAST 3 results) were mapped against the reference plasmid using BLASTn in CGView (Grant and Stothard, 2008). Protein coding sequences annotated as "hyp" and "tnp" refer to hypothetical proteins and transposons, respectively

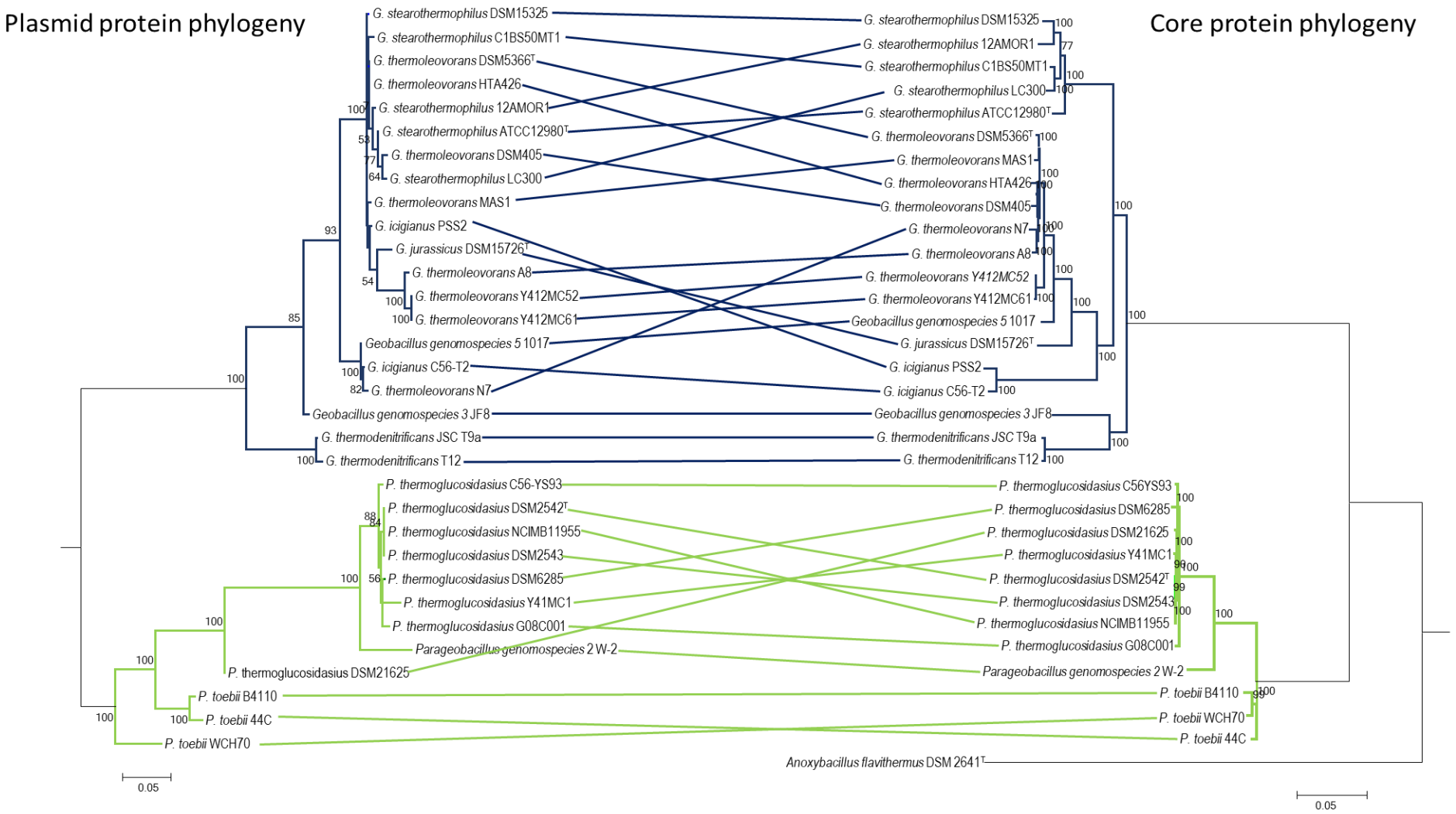


Figure 3: Maximum likelihood phylogenies of the pGeo1 plasmids (left – 6 protein orthologues) and the core genomes (right - 1,441 protein orthologues). Bootstrap values (n = 100 replicates) > 50% are shown and the bar length indicates 0.05 substitutions per site. The core protein phylogeny was rooted with *A. flavithermus* DSM 2641^T and the plasmid protein phylogeny was rooted at the mid-point.

A ML phylogeny was constructed on the basis of the six conserved pGeo1 protein orthologues. Comparison against an extensive core protein phylogeny of the pGeo1-harbouring strains showed similar clustering between the two phylogenies, with two distinct genus clades for the *Geobacillus* and *Parageobacillus* strains (Figure 3), suggesting that these plasmids have evolved from a common ancestral plasmid, prior to splitting of the genus and subsequent speciation events. Within the genus clades, however, some differences could be observed, with the pGeo1 plasmid proteins of *G. thermoleovorans* clustering together with those of *G. stearothermophilus* and *G. iciganus*, and the *P. thermoglucosidasius* DSM 21625 pGeo1 plasmid clustering distinctly from the other *P. thermoglucosidasius* strains (Figure 3). These changes in branching patterns may be due to the limited number of proteins compared for the pGeo1 plasmid phylogeny in contrast to the 1,441 proteins incorporated to the core protein phylogeny. Alternatively, there may have been exchange of this plasmid between different species within each genus clade which coincides with some of the larger differences in G+C contents that were observed between the plasmid and chromosomal DNAs.

Functional annotation of the six conserved pGeo1 proteins showed that they code for a plasmid replication initiator (Rep_3), a competence protein ComK, plasmid partitioning protein MinD, peptidase M15_4, as well as two proteins containing domains of unknown function, DUF2922 and DUF1659.

The Rep_3 proteins of the geobacilli contain the conserved domain pfam01051 (Average Bitscore; 206.73; Average E-value: $1.80e^{-52}$), are on average 385 amino acids in length and share approximately 75.48% average amino acid identity (AAI) amongst the pGeo1 plasmids. Rep_3 plasmid replication proteins are site-specific, possess topoisomerase activity, interact with both the plasmid and host-encoded replication proteins and are found to initiate plasmid replication via the theta-replicating mechanism. In addition to this they have also been observed to provide stringent control of plasmid copy number, thereby maintaining low plasmid copy numbers in host bacteria (Khan, 2003; Bertini *et al.,* 2010; Fueki and Yamaguchi, 2001; Thompson *et al.,* 2018; Kwong *et al.,* 2017; Salto *et al.,* 2018). Rep_3 bind to a specific site within the plasmid *ori*, which generally varies in position within theta replicating plasmids, but is typically found to consist of A+T rich regions, direct repeats or iterons (del Solar *et al.,* 1998; Lilly and Camps, 2015). The *ori* site of the pGeo1 plasmids were found to share an average of 66.86% A+T nucleotide composition. The DoriC database identified the *ori* sites of only 23/32 of the pGeo1 plasmids, all of which matched the same reference sequence, *G. thermoleovorans*

Y412MC52 (DoriC acc: pORI00000492) at an average of 91.07%, indicating that the *ori* regions are nearly identical among a subset of the pGeo1 plasmids (Luo and Gao, 2018).

The pGeo1 MinD orthologues are on average 265 amino acid residues in length, share 89.41% average amino acid identity and incorporate a conserved ParAB_family domain (CD02042; Bitscore: 162.06; E-value: 3.78e^{-47}). Members of this family assist with the stable vertical inheritance of the chromosome or plasmids upon bacterial cell division and are typically associated with low copy number plasmids as there is a higher chance of plasmid loss upon cell division (Lutkenhaus, 2012). In addition, MinD/ParA family proteins have been observed to play a role in the segregation of cytoplasmic proteins and the polar localization of type IV pili, chemotaxis proteins, conjugative transfer machinery, as well as cellulose synthesis (Lutkenhaus, 2012; Römling and Galperin, 2015; Krasteva *et al.*, 2017).

The plasmid ComK proteins orthologues (pfam06338; Bitscore: 150.62; E-value: 1.59e^{-03}) are approximately 176 aa in size and share 81.41% average amino acid identity. These proteins incorporate the conserved domain ComK has been observed to be plasmid-encoded in a number of *Bacillus subtilis* strains, where it is predicted to be a master regulator of bacterial competence (Singh *et al.*, 2012; Singh *et al.*, 2013; Miguel-Arribas *et al.*, 2017). This protein functions as a transcription factor that upregulates genes that make bacterial cells competent (Blokesch, 2016). This is typically observed by the formation of pili structures that allow for the competent bacterial cell to take up or absorb exogenous DNA from the environment and carry out the recombination of that DNA within their chromosomal or plasmid sequences (Blokesch, 2016). Thus, competence can play a major role in the evolution of both chromosomal and plasmid DNA as it facilitates horizontal gene transfer (HGT) between cells, both closely or distantly related, and the recombination of exogenous DNA into another chromosome or plasmid sequence can result in the establishment of new genes and ultimately phenotypic characteristics (Blokesch, 2016). The presence of ComK proteins on the pGeo1 plasmids suggests they have the ability to evolve and acquire accessory genes at a higher rate than plasmids that do not encode for such proteins, as the ComK proteins allow for an increased gene flow within host cells (Eberhard, 1990).

The Peptidase_M15_4 enzyme (pfam13539; Bitscore: 81.47; E-value: 4.12e^{-21}) possesses D-alanyl D-alanine carboxypeptidase activity and belongs to the VanY protein superfamily (Lee *et al.*, 2001). These proteins orthologues are 156 amino acid residues in length and share 85.11% average amino acid identity among the 32 plasmids. This superfamily is made up of

enzymes that are involved in the peptidoglycan cross-linking stage of bacterial cell wall synthesis, where they are observed to be membrane bound and function by cleaving the D-alanine residue at the carboxy-terminal of peptidoglycan pentapeptides (Lee *et al.,* 2001). Carboxypeptidase enzymes furthermore provide resistance towards antibiotics that interfere with bacterial cell wall synthesis, such as penicillin, vancomycin or β-lactam antibiotics (Spratt and Strominger, 1976; Dudley, 1995; Lee *et al.,* 2001). While antibiotic resistance genes and proteins are not essential elements of a plasmid, they are found to be common constituents of plasmid architecture (Condit and Levin, 1988).

Lastly, two proteins of unknown function are conserved among all pGeo1 plasmids. These contain conserved domains DUF1659 (pfam07872; Bitscore: 43.04; E-value: 3.03e^{-}06; 73 residues long) and DUF2922 (pfam11148; Bitscore: 75.05; E-value: 1.10e^{-}19; 70 residues long) and share average amino acid identities of 76.29% and 84.83% amongst the plasmids, respectively. No known function can be ascribed to the latter proteins.

2.3.4 Characteristics of the pGeo2 plasmids

The plasmid group pGeo2 is comprised of fifteen plasmids. Orthologous replicons are found in ten *Geobacillus* (six species) and five *Parageobacillus* (three species) strains. Only two species, namely *G. thermodenitrificans* (five strains) and *P. thermoglucosidasius* (three strains) include multiple strains harbouring this type of plasmid (Figure 1). The pGeo2 plasmids range in size between 43,4 kb (*P. caldoxylosilyticus* DSM 12041) and 70,1 kb (*G. thermodenitrificans* ID-1) and code between 48 (*P. caldoxylosilyticus* DSM 12041) and 75 (*G. thermodenitrificans* ID-1) proteins (Table 1). The G+C contents of these plasmids range from 38.9% (*P. thermoglucosidasius* DSM 2542) to 41.1% (*Geobacillus* genomospecies Et2/3). The chromosomal G+C content of pGeo2 harboring strains ranges from 42.14% (*P. toebii* DSM 14590) to 52.39% (*G. icigianus* C56-T2). On average, the G+C content of the plasmids is –8.25% below that of the chromosome (range: -2.56% {*P. toebii* DSM14590} to -12.58% {*G. stearothermophilus* strain 15}).

Twenty-three SCOs are conserved among the pGeo2 plasmids, constituting approximately 29.33% (*G. thermodenitrificans* ID-1) to 45.83% (*P. caldoxylosilyticus* DSM12041) of the total plasmid protein complement. As such, the pGeo2 plasmids are far more structurally conserved than the pGeo1 plasmids, with large sections of shared coding regions between the pGeo2 plasmids, particularly in one half of the pGeo2 plasmid sequences (Figure 4). Several

genomic islands in the G+C content are also observed for these plasmids, indicative of horizontal gene transfer.

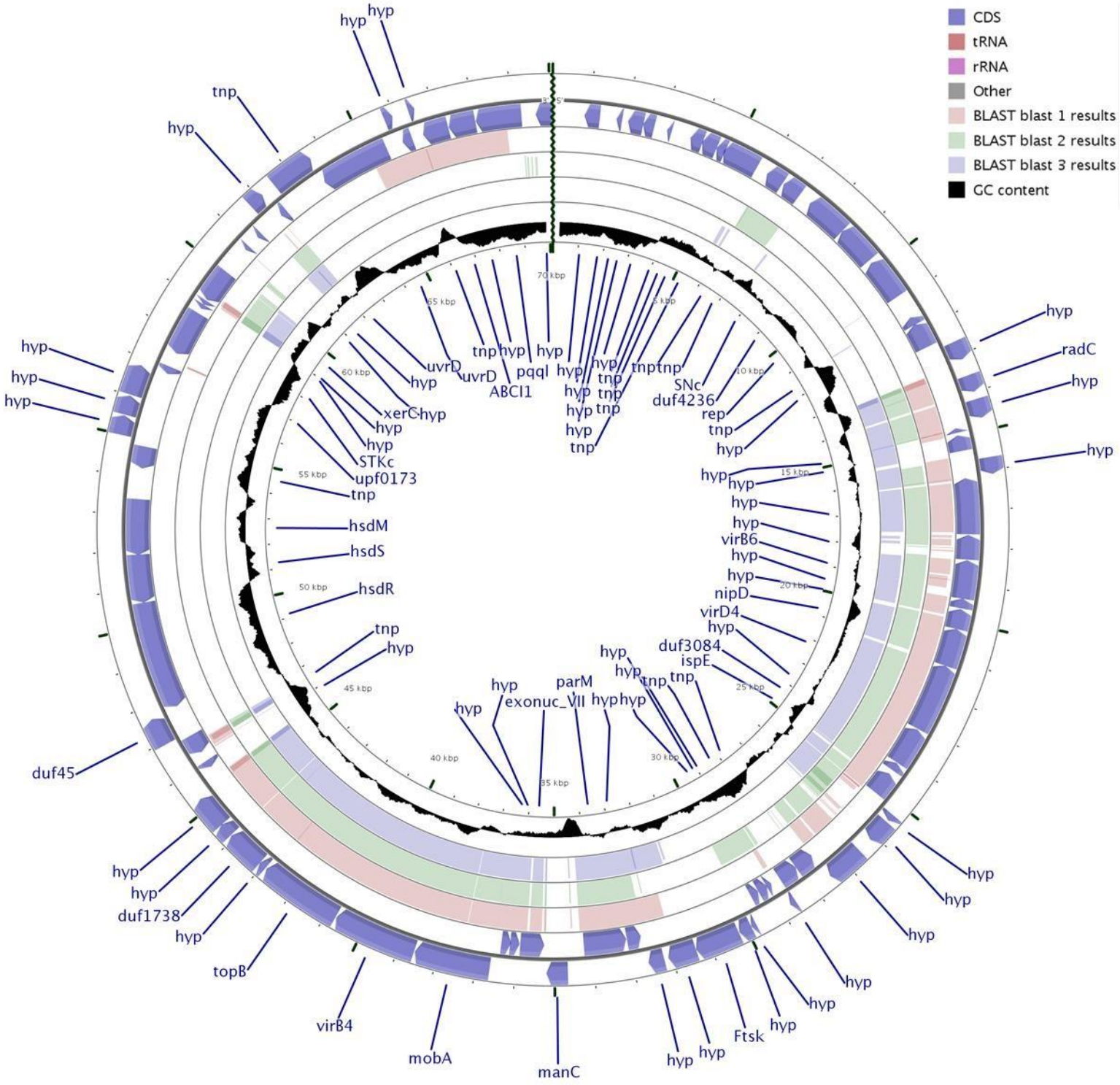


Figure 4: Circular map of the pGeo2 plasmids. The largest plasmid (*G. thermodenitrificans* ID-1) was used as a reference. The plasmid nucleotides sequences of pGeo2 plasmids *P. thermoglucosidasius* DSM 2542 (third circle from the outside, BLAST 1 result), *Geobacillus* genomospecies 1 Et2/3 (fourth circle from outside, BLAST 2 result) and *P. caldoxylosilyticus* DSM 12041 (fifth circle from outside, BLAST 3 result) were mapped against the reference plasmid using BLASTn in CGView (Grant and Stothard, 2008). Protein coding sequences annotated as "hyp" and "tnp" refer to hypothetical proteins and transposons, respectively.

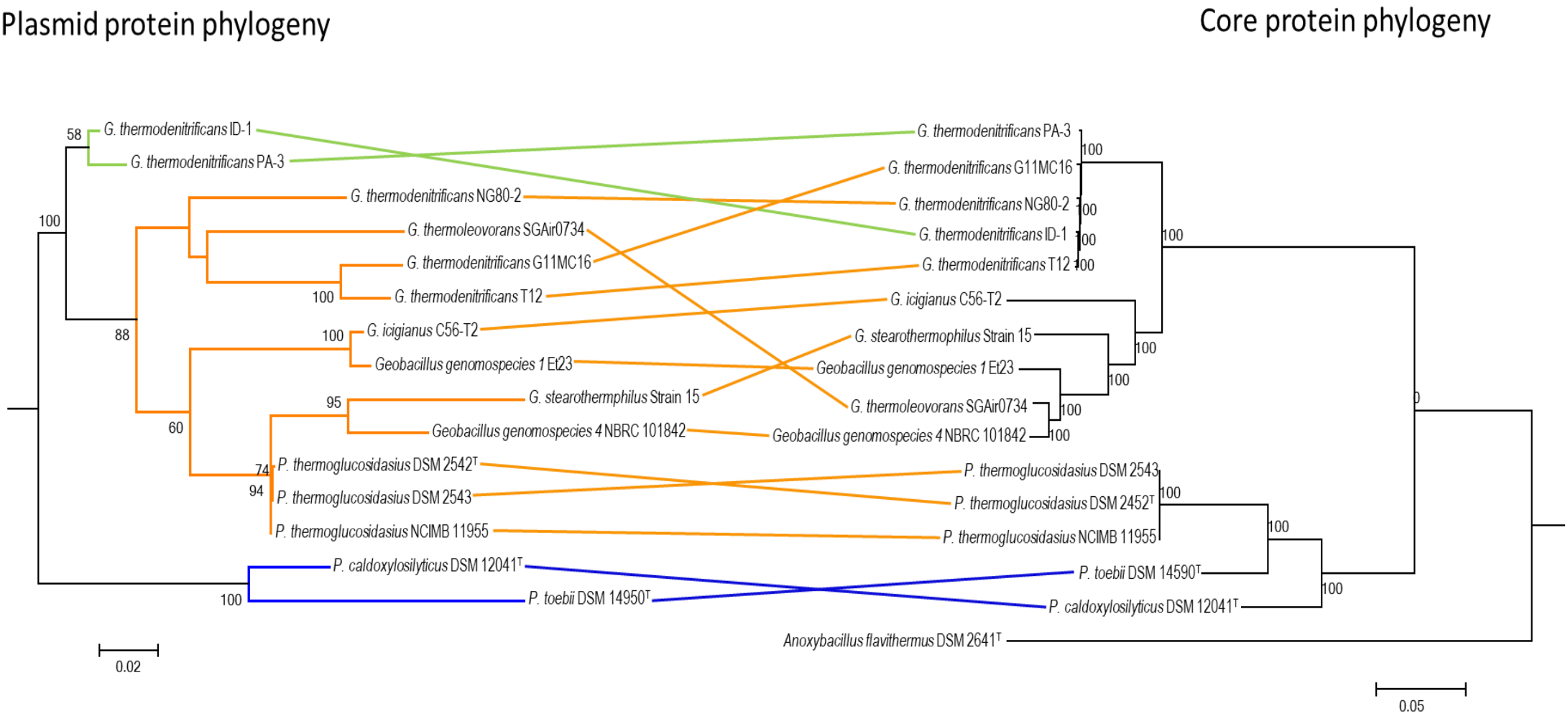


Figure 5: Maximum likelihood phylogenies of the pGeo2 plasmids (left – 23 protein orthologues) and the core genomes (right – 1,616 protein orthologues). Bootstrap values of (n = 100 replicates) > 50% are shown and the bar length indicates 0.02 and 0.05 substitutions per site. The core protein phylogeny was rooted with *A. flavithermus* DSM 2641^T and the plasmid protein phylogeny was rooted at the mid-point.

A ML phylogeny on the basis of the twenty-two conserved pGeo2 proteins showed disparate clading from the core protein phylogeny of the corresponding strains (Figure 5). The two genus clades in the core protein tree were not observed in the pGeo2 plasmid protein tree, instead, three main clades could be observed (Figure 5). The largest incorporates eight *Geobacillus* strains, as well as three strains of *P. thermoglucosidasius*. The other two clades were comprised of two *G. thermodenitrificans* and two *Parageobacillus* strains, respectively. The colocalization of *Parageobacillus* and *Geobacillus* spp. in the same branches in the pGeo2 plasmid phylogeny suggests a common ancestor plasmid prior to the divergence of the current pGeo2 plasmids and can further suggest that these plasmids have undergone intra-generic transmission, or HGT, subsequent to divergence of the genera and the speciation events inherent to each genus.

Of the twenty-three proteins conserved among the pGeo2 plasmids, twelve could be functionally annotated, while the remaining eleven represent hypothetical proteins without conserved functional domains. The eleven annotated proteins include, plasmid segregation protein ParM, an exonuclease (exonuclease VII), plasmid mobilisation protein MobA, Type IV secretion system DNA transfer proteins VirB4, VirB6 and VirD4, topoisomerase TopB, nuclease RadC, murine hydrolase protein NlpD, erythritol kinase IspE and two proteins contain a domain of unknown function (DUF1738 and DUF3084).

Of note are six proteins predicted to function in the conjugal transfer of plasmids via a Type IV secretion system (T4SS). These systems are one of several types of secretion systems that microbes employ to transport macromolecules across the cell envelope (Wallden *et al.,* 2010). The most common type of T4SSs found in bacteria are plasmid encoded and used for initiating conjugation for the transfer of plasmid DNA and other genetic material from cell to cell (Wallden *et al.,* 2010). T4SSs have been observed and detailed in plasmids such as the Ti plasmid from *Agrobacterium tumefaciens*, the pKM101 plasmid in *E. coli* and in several *Rickettsia* plasmids, where they are synthesised from multiprotein complexes (Wallden *et al.,* 2010). The T4SSs of these plasmids consist of up to twelve proteins (VirB1-VirB11 and VirD4) all of which play a specific role in the process of conjugation from DNA substrate processing and substrate recruitment to the translocation of substrate from cell to cell (Wallden *et al.,* 2010; Gillespie *et al.,* 2016).

The pGeo2 plasmids incorporate a 617 amino acid orthologue of VirB4 (TIGR0274; Bitscore: 103.59; E-value: $1.98e^{-22}$) sharing 98.20% average amino acid identity. VirB4 proteins

function in the assembly and maturation of a conjugative pilus. Pilus structures allows for cell to cell contact providing a bridge between two bacterial cells enabling the passage of plasmids and other genetic materials (Schandel *et al.,* 1992; Wallden *et al.,* 2010; Gillespie *et al.,* 2016).

The conserved protein orthologue of VirB6 (pfam04610; Bitscore: 38.55; E-value: 0.0034) is a 347 amino acid protein that incorporates the conserved domain and shares 90.12% average amino acid identity among the plasmids. VirB6 is an inner membrane protein that forms the membrane pore structure of the T4SS which allows for genetic material to be moved in and out of a cell (Wallden *et al.,* 2010; Gillespie *et al.,* 2016).

The MobA orthologue (pfam03389; Bitscore: 273.22; E-value: $9.07e^{-88}$) coded for by the pGeo2 plasmids is 634 amino acid resides in length, sharing 94.23% average amino acid identity. This protein functions as relaxase proteins in conjugation systems (Alvarez-Martinez and Christie, 2009; Garcillan-Barcia *et al.,* 2009; Wallden *et al.,* 2010; Ramsay *et al.,* 2016). MobA functions together with a topoisomerase (TopB), an orthologue of which is encoded adjacent of the *mobA* gene in the pGeo2 plasmids. The latter protein (PRK07726; Bitscore: 707.48; E-value: 0.0) shares 99.20% average amino acid identity (AAI) amongst the plasmids. MobA and TopB together prepare the DNA substrate for translocation through the pilus. TopB facilitates the unwinding of double-stranded plasmid DNA, by creating a nick at the origin of transfer (*oriT*) site of the plasmid. This leaves a 5′ end available on the now single-stranded DNA substrate for the relaxase to covalently bind to (Garcillan-Barcia *et al.,* 2009; Gadelle *et al.,* 2014). The relaxase protein remains linked at the 5' end of the donor strand, and once it is translocated to the recipient cell, it catalyzes the re-joining of the two ends of the strand, recircularizing the plasmid DNA into dsDNA in the new host cell (Scherzinger *et al.,* 1992).

In addition to their role in the T4SS, relaxase and topoisomerase activity is associated with the rolling circle replication mechanism of plasmids. As a result, it is observed that during conjugation the donor cell will retain the newly replicated dsDNA plasmid, while the recipient cell will acquire the ss-plasmid DNA, which is then re-circularized in the new host cell via the relaxase protein activity (Garcillan-Barcia *et al.,* 2009).

Acting in conjunction with the relaxase protein is the T4SS DNA transfer protein VirD4, an ATPase that assists in the translocation of the substrate DNA to the recipient cell through the conjugative pilus (Atmakuri *et al.,* 2003*;* Alvarez-Martinez and Christie, 2009; Wallden *et al.,* 2010; Ramsay *et al.,* 2016). The pGeo2 plasmids incorporate a 711 amino acid orthologue of VirD4 (pfam02534; Bitsore: 256.78; E-value: $3.99e^{-77}$; AAI: 99.32%).

Another pGeo2 orthologue is predicted to play a peripheral role in conjugative transfer is a murein hydrolase NlpD (COG0739; Bitscore: 115.99; E-value: 1.50e^{-}28; AAI: 93.94%). This protein is reported to be involved in the degradation of the cell wall, in order to prepare a region of the cell for the assembly of the conjugative pilus (Laverde-Gomez *et al.,* 2013).

Furthermore, all pGeo2 plasmids encode a 215 amino acid long orthologue of an exonuclease VII (pfam02601; Bitscore: 145.54; E-value: 2.85e^{-}05; AAI: 97.55%). This enzyme catalyses the exonucleolytic cleavage of DNA in both the 5'-3' and 3'-5' direction to yield 5' phosphomononucleotides (Chase and Richardson, 1974). Exonuclease VII has been identified in *E. coli* K12, where it is found to be specific for ssDNA and plays a role in the proofreading and digestion of UV damaged DNA (Chase and Richardson, 1974; Lovett, 2011).

While exonucleases are not typically encoded by plasmids it is likely that they are acquired over time as a result of conjugation and subsequent transposon activity (Lovett, 2011). Apart from its activity on damaged DNA, it may also play a role in degrading incoming single stranded plasmid DNA from neighbouring cells during conjugation, contributing to the entry exclusion characteristics associated with conjugative plasmids. Such activity has been observed in IncP-1 plasmids, where exonuclease activity reduced the frequency with which the plasmids where transferred to cells possessing the same plasmids (Norberg *et al.,* 2011). In addition to the exonuclease VII protein, the pGeo2 plasmids encode an orthologue of RadC (PRK00024; Bitscore: 175.26; E-value: 4.45e^{-}50; AAI: 93.89%), that is 146 amino acids in length. While the exact function of RadC is unknown, it is predicted to function as a nuclease and thus may also be involved in the degradation of foreign or damaged DNA (Iyer *et al.,* 2011).

A predicted ParM-like protein (CD10227; Bitscore: 173.96; E-value: 2.15e^{-}51; AAI: 99.24%) of 383 amino acid residues is also conserved amongst the pGeo2 plasmids. ParM proteins are actin-based plasmid segregation proteins that are involved in the control of plasmid partition during cell division and ensure that each daughter cell receives a copy of the plasmid (Orlova *et al.,* 2007; LeBrasseur, 2007).

2.3.5 Characteristics of the pGeo3 plasmids

The pGeo3 plasmid group comprises two plasmids found in *P. thermoglucosidasius* C56YS93 and DSM 21625. These plasmids are 19.9 and 30.6 kb in size and code for 21 and 36 proteins, respectively. They both have a G+C content of 40.4%, which differs by -3.48% from the host strain core genomes.

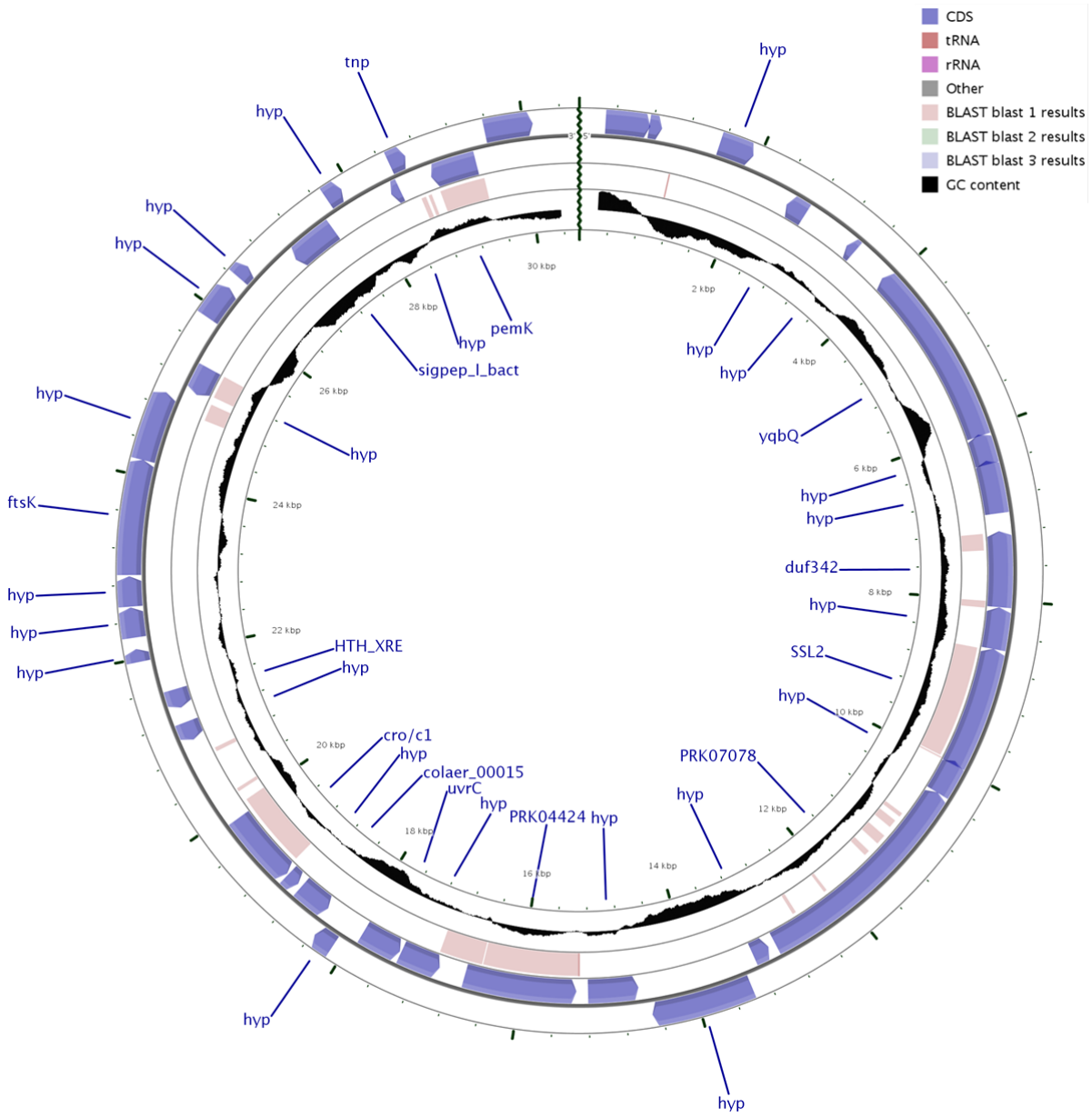


Figure 6: Circular map of the pGeo3 plasmids. The largest plasmid *P. thermoglucosidasius* DSM 21625 was used as a reference. The plasmid nucleotide sequence of *P. thermoglucosidasius* C56YS93 (third circle from the outside, BLAST 1 result) was mapped against the reference plasmid using BLASTn in CGView (Grant and Stothard, 2008). Protein coding sequences annotated as "hyp" and "tnp" refer to hypothetical proteins and transposons, respectively.

A total of six orthologous proteins are shared between the two pGeo3 plasmids. The circular map (Figure 6) indicates the conserved nature of the two plasmids, whereby regions of sequence

similarity are depicted in both halves of the reference plasmid. This may be suggestive of a common ancestral plasmid from which these two plasmids have since diversified. Four of the orthologous proteins could be functionally annotated, while the remaining two represent a hypothetical protein (PRK07078; Bitscore: 473.87; E value: 2.17e154; AAI: 62.72%) and protein with domain of unknown function DUF342 (pfam03961; Bitscore: 34.16; E value: 8.76e^{-}03; AAI: 61.35%).

Among the functionally characterized proteins is a superfamily II DNA and RNA helicase (COG1061; Bitscore: 198.05; E value: 3.91e^{-}58; AAI 88.64%). Helicases carry out the unwinding dsDNA or RNA to form ss-nucleic acid regions substrates for numerous nucleic acid metabolic pathways. Plasmid-encoded helicases are mostly found to be involved in plasmid replication, recombination and repair events as well as transcription and translation (Byrd and Raney, 2012). Superfamily II (SF2) is the largest and most diverse of all helicase families (Byrd and Raney, 2012). SF2 helicases are specifically involved in transcription, DNA repair, chromosomal and plasmid sequence rearrangement as well as many aspects of RNA metabolism (Byrd and Raney, 2012).

Two potential transcription factors are also shared by the pGeo3 plasmids. The first incorporates a conserved domain (PRK04424; Bitscore: 38.46; E value: 2.52e^{-}03; AAI: 96.52%) which has been provisionally annotated as a transcriptional regulator involved in fatty acid biosynthesis. The second potential transcription factor contains an HTH_26 motif (pfam13443; Bitscore: 75.86-bitscore; E value: 7.59e^{-}20; AAI: 91.03%) which has been described as a part of a Cro/C1-like transcription factors (Anderson *et al.*, 1981). Cro/C1 proteins are transcription repressors commonly associated with bacteriophages but are also found in both prokaryotes and eukaryotes, where the HTH_26 (helix-turn-helix) motif makes up part of the N-terminal region of the transcription factor and is responsible for binding the target gene site, inhibiting its expression (Anderson *et al.*, 1981).

The remaining orthologue represents a PemK toxin of a type II Toxin-Antitoxin (TA) system (pfam02452; Bitscore: 74.77; E value: 6.69e^{-}14; AAI: 94.36%) Genes encoding for TA systems are generally found on plasmids but are also common within bacterial chromosomal DNA (Tsuchimoto and Ohtsubo, 1988; Tsuchimoto *et al.*, 1993). These systems typically consist of a downstream toxin gene and an upstream antitoxin gene, (Tsuchimoto and Ohtsubo, 1988; Tsuchimoto *et al.*, 1993). Plasmid-encoded TA systems are involved in the stabilization of plasmids within its host bacterial population, by mediating the post-segregational killing of

daughter cells in the event of plasmid loss upon cell division (Gerdes *et al.*, 1986; Díaz-Orejas *et al.*, 2017). The PemK toxin has been identified in *E. coli* plasmids where it is has endoribonuclease activity on cellular mRNA, inhibiting protein synthesis (Díaz-Orejas *et al.*, 2017).

2.3.6 Characteristics of the pGeo4 plasmid

The pGeo4 plasmid is unique to *P. toebii* WCH70, has a total size of 10. 3 kb and codes for twelve proteins, four of which have been functionally annotated. The G+C content of this plasmid is 39,6% which differs by -3.2% to the 42.8% G+C content of the host cell. On the basis of the Conserved Domain Database annotation, no replication initiation protein could be identified. However, the RAST annotation identified a 448 amino acid protein on the pGeo4 plasmid as a putative candidate. BLASTp analysis against the NCBI protein database showed that the pGeo8 protein shares 33% sequence identity with the protein AAV98_19225 in *Bacillus* sp. CHD6a. In the latter organism, this protein is predicted to incorporate a pfam16793 RepB DNA-primase domain. However, due to the limited information available and lack of conserved domains, the role of this protein cannot be deduced. The remaining three proteins have been annotated as an orthologue of the restriction enzyme TdeIII (pfam09520; 256.05; E value: $6.78e^{-}86$), a DNA methylase Dcm (COG0270; Bitscore: 255.82; E value: $3.63e^{-}83$) and a GInR-like transcription factor (CD01105; Bitscore: 35.67; E-value: 0.0022).

2.3.7 Characteristics of the plasmids in the groups pGeo5 to pGeo9

The remaining five plasmid groups are found singly in individual strains, with the exception of pGeo5, which is made up of two plasmids. These plasmids range in size between 1.4 kb (pGeo3 *G. thermoleovorans* A8) to 6.2 kb (pGeo7 in *G. thermoleovorans* N7), and code for between two (pGeo3 in *G. thermoleovorans* A8) and fourteen (pGeo8 in *G. thermoleovorans* N7) proteins (Table 1). Only a few of these proteins could be functionally annotated. The majority of these comprise plasmid replication initiator proteins, plasmid maintenance proteins, transcription factors and enzymes (Table 2).

Table 2: Functionally annotated proteins encoded on the pGeo5pGeo9 plasmids

Plasmid Group	Strains	Number of plasmids	Functionally annotated proteins
pGeo5	*G. thermoleovorans* A8, N7	2	Rep_1
pGeo6	*P. caldoxylosyliticus* B4119	1	Rep_1
pGeo7	*P. thermoglucosidasius* G08C001	1	-
pGeo8	*G. thermoleovorans* N7	1	Rep_trans
pGeo9	*Geobacillus* genomosp. 4 NBRC 101842	1	Rep protein RelE RelB

pGeo5 group plasmids are found in *G. thermoleovorans* A8 and N7. They are 1,4 kb and 3 kb in size and comprise two and five coding regions, respectively. The G+C content of these two plasmids is 47,54% (A8) and 47.71% (N7), while the G+C content of the host strains is 52.42%. Two protein orthologues are shared between these two plasmids.

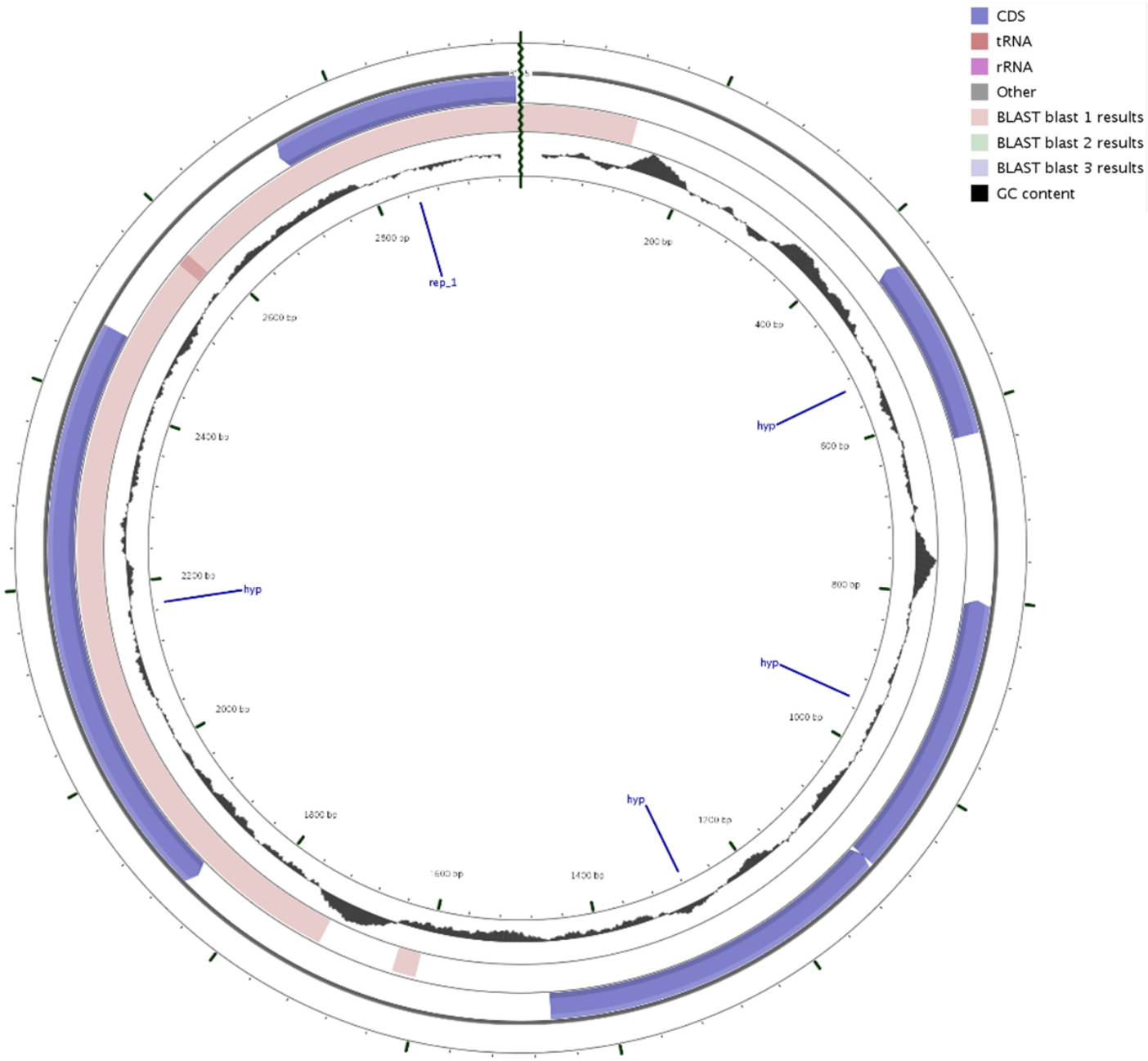


Figure 7: Circular map of the pGeo5 plasmids. The largest plasmid *G. thermoleovorans* N7 was used as a reference and the nucleotide sequence of the *G. thermoleovorans* A8 plasmid (third circle from the outside; BLAST 1 result) was mapped against the reference plasmid using BLASTn in CGView (Grant and Stothard, 2008). Protein coding sequences annotated as "hyp" and "tnp" refer to hypothetical proteins and transposons, respectively.

The circular map indicates that the plasmid of *G. thermoleovorans* A8 share sequence similarity with nearly half of the nucleotide sequence of the N7 plasmid, suggesting a highly conserved nature amongst these two plasmids (Figure 7). The shared orthologues amongst the pGeo5 plasmid are annotated as a hypothetical protein and a plasmid replication protein belonging to the Rep_1 superfamily (pfam01446; Bitscore: 83.1; E value: 0.0014; AAI: 99.89%). These proteins have been implicated in initiating rolling circle replication of plasmids (Chandler *et al*, 2013; Kwong *et al*, 2017). Rep_1 proteins contain a catalytic tyrosine residue and a HTH (helix turn helix) motif that is involved in site-specific DNA binding and uses topoisomerase activity to introduce a nick in only one of the strands of the *ori* site region of the plasmid (Chandler *et al.,* 2013; Kwong *et al.,* 2017).

Group pGeo6 is made up of the *P. caldoxlosilyticus* B4119 plasmid at 1.6 kb in size. This plasmid has a G+C content of 41.64%, -2.38% below that of the host chromosome. Two proteins are encoded by this plasmid of which only one was functionally annotated as a protein of the Rep_1 superfamily (pfam01446; Bitscore: 161.843; E-value: 4.09789e-48) domain. This Rep_1 protein is observed to share 26.75% amino acid identity with that of the pGeo5 Rep_1 proteins, which could suggest that these plasmids may have been derived from a common ancestor and subsequently undergone extensive diversification.

The plasmid group pGeo7 comprises a 1.8kb sized plasmid of *P. thermoglucosidasius* G08C001. This plasmid has a G+C content of 42.4% which is -1.83% below the host chromosome, suggesting that this plasmid may be a long-term resident within this particular strain. Three proteins are encoded by this plasmid all of which were annotated as hypothetical proteins.

The 6.2 kb pGeo8 plasmid from *G. thermoleovorans* N7 codes for twelve distinct proteins which share no orthology with proteins encoded on any of the other plasmids. The chromosomal G+C content of N7 is 52.42 % while the plasmid has a G+C content of 41.5% (10.92% difference), suggesting this plasmid has been acquired through a recent HGT event.

Only one protein could be functionally annotated as a replication initiation protein belonging to the Rep_trans superfamily (pfam02486; Bitscore: 165.17; E value: $2.96e^{-}50$) domain. These proteins are also associated with rolling circle replication, using topoisomerase activity to generate a nick in one of the strands at the *ori* site of the plasmid (Balson and Shaw 1990; Kwong *et al.,* 2017; Wawrzyniak *et al.,* 2017).

The final plasmid group, pGeo9 is harboured by *Geobacillus* genomospecies 4 NBRC101842, is 1,9 kb in size and comprises four protein coding sequences. The G+C content of the plasmid is -4.34% below that of the chromosome. Of the four pGeo8 proteins, three could be functionally annotated. One belongs to the PHA00330 replication initiation protein superfamily (PHA00330; Bitscore: 38.69; E value: 0.00229). This superfamily includes relaxases that have been associated with rolling circle replication of both plasmid and viral DNA (Lorenzo-Diaz *et al*, 2014). The other two encoded proteins are annotated as the RelE toxin (COG3657; Bitscore: 35.1; E value: 0.00096) and the RelB (COG3077; Bitscore: 33.1; E value: 0.00055) antitoxin. Much like PemI and PemK, RelB is the antitoxin protein that neutralizes the activity of the RelE toxin, forming part of the RelEB TA system. The RelEB TA system is described as the most abundant of all TA systems and is also involved in plasmid maintenance within a host cell population, ensuring each daughter cell inherits copies of the plasmid during cell division (Gerdes *et al.*, 1986; Díaz-Orejas *et al.*, 2017). In the event of plasmid loss, the RelB toxins' mode of action is to bind to ribosomes, preventing translation from occurring where the inhibition of protein production eventually results in cell death (Gerdes *et al.*, 1986; Díaz-Orejas *et al.*, 2017).

2.4 Conclusions

Plasmids are observed as prevalent features among the geobacilli, with 47.62% and 72.73% of all evaluated *Geobacillus* and *Parageobacillus* species found to harbour plasmids, respectively. Plasmids are particularly concentrated in the species *G. thermodenitrificans, G. stearothermophilus, G. thermoleovorans, P. thermoglucosidasius* and *P. toebii*. A total of 56 plasmids at a wide range of sizes were incorporated into this study and upon orthology analysis were found to be split into nine plasmid groups (pGeo1-pGeo9) on the basis of orthologous proteins shared between them.

The pGeo1 group (32 plasmids) shared six orthologous proteins between them, displaying significant diversification in their plasmid structure given their larger sizes. Phylogenetic analysis of these plasmids suggest that this plasmid has been vertically derived from the common ancestor of both genera. By contrast, the second most prevalent plasmid group (pGeo2, fifteen plasmids) is more structurally conserved but phylogenetic analysis showed a distinct evolutionary history of the pGeo2 plasmids from the core genome, suggesting these plasmids are prone to horizontal transfer between strains of these genera. The conserved proteins on the pGeo2 plasmids include six with predicted roles in the synthesis of a conjugative T4SS pilus and for plasmid DNA transfer, which may underlie the predicted horizontal plasmid transmission between geobacilli taxa.

On the basis of the annotated plasmid proteins encoded, inferences could be made on the mode of replication of each plasmid. Plasmids of pGeo1 are deduced to use theta replication, while those of pGeo2 are possible rolling circle replicating plasmids, as a result of the relaxase and topoisomerase components of the T4SS. Plasmids of groups pGeo5, pGeo6, pGeo8 and pGeo9 are also deduced to replicate via rolling circle type replication as a result of their encoded rep proteins. The mode of replication for plasmids of the groups pGeo3, pGeo4 and pGeo7 remains unclear, as no replication proteins could be identified.

While there are some commonalities among plasmids in terms of replication mechanisms and plasmid maintenance proteins, it is noteworthy that there are very few orthologous proteins shared among the plasmids, with no protein conserved in all plasmids and at most only twenty-three proteins shared among the pGeo2 group plasmids. This hints to extensive diversification of the geobacilli plasmids. In total these 56 plasmids encode approximately 2,691 proteins, of which a total of 37 contribute to conserved orthologues among the plasmids. As up to 98.62% of the proteins encoded on these plasmids are non-conserved, they are likely to serve as potent

sources of cargo genes with a wide range of phenotypic and genotypic effects on the organisms that harbor them. This was the subject of further analysis in Chapter 3.

2.5 References

Abdel-Fattah, Y. R., and Gaballa, A. A. (2008). Identification and over-expression of a thermostable lipase from *Geobacillus thermoleovorans* Toshki in *Escherichia coli*. *Microbiological Research*, 163, 13-20

Aliyu, H., Lebre, P., Blom, J., Cowan, D. and De Maayer, P (2016) Phylogenomic ReAssessment of the Thermophilic Genus *Geobacillus*, *Systematic and Applied Microbiology*, 39(8), 527-533Hussein, A. H., Lisowska, B. K. and Leak, D. J. (2015) The genus *Geobacillus* and their biotechnological potential. *Advances in Applied Microbiology*, 92, 1-48

Alvarez-Martinez, C. E., and Christie, P. J. (2009) Biological diversity of prokaryotic type IV secretion systems. *Microbiology and Molecular Biology Reviews: MMBR*, 73(4), 775–808

Anderson, W. F., Ohlendorf, D. H., Takeda, Y. and Matthews, B. W. (April 1981) Structure of the *cro* repressor from bacteriophage lambda and its interaction with DNA. *Nature*, 290(5809), 754–758

Atmakuri, K., Ding, Z. and Christie, P. J. (2003) VirE2, a type IV secretion substrate, interacts with the VirD4 transfer protein at cell poles of *Agrobacterium tumefaciens. Molecular Microbiology*, 49, 1699–1713

Balan, A., Ibrahim, D., Rahim, R. A. and Rashid, F. A. A. (2012) Purification and Characterization of a Thermostable Lipase from *Geobacillus thermodenitrificans* IBRL-nra. *Enzyme research*, vol. 2012, Article ID 987523, 1-7

Balson, D. F. and Shaw, W. V. (1990) Nucleotide sequence of the rep gene of staphylococcal plasmid pCW7. *Plasmid*, 24(1), 74-80

Bertini, A., Poirel, L., Mugnier, P. D., Villa, L., Nordmann, P. and Carattoli, A. (2010) Characterization and PCR-based replicon typing of resistance plasmids in *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*, 54(10), 4168-4177

Blokesch, M. (2016) Natural competence for transformation, *Current biology*, 26(23), 3255

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M. F., Lió, P., Crescenzi, P., Fani R. and Fondi M. (2015) MeDuSa: a multi-draftbased scaffolder. *Bioinformatics*, 31(15), 24432451

Brumm, P., Land, M. L., Hauser, L. J., Jeffries, C. D., Chang, Y. J. and Mead, D. A. (2015) Complete genome sequences of Geobacillus sp. Y412MC52, a xylan-degrading strain isolated from obsidian hot spring in Yellowstone National Park. *Standards in Genomic Sciences*, 10, 81

Byrd, A. K., and Raney, K. D. (2012) Superfamily 2 helicases. *Frontiers in Bioscience (Landmark edition)*, 17, 2070–2088

Campbell, A., Mrázek, J. and Karlin, S. (1999). Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16), 9184–9189

Chandler, M., de la Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G. and Ton-Hoang, B. (2013) Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nature Reviews Microbiology*, 11(8), 525-538

Chase, J. W. and Richardson, C. C. (1974) Exonuclease VII of *Escherichia coli*. *Journal of Biological Chemistry*, 49(14), 4553-4561

Chen, I, A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J. R., Seshadri, R., Smirnova, T., Kirton, E., Jungbluth, S. P., Woyke, T., Eloe-Fadrosh, E. A., Ivanova, N. N. and Kyrpides, N. C. (2018) IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acid Research,* 8(47), D666-677

Condit, R. and Levin, B. R. (1988) The evolution of plasmids carrying multiple resistance genes: The role of segregation, transposition and homologous recombination. *The American Naturalist,* 135(4), 573-596

Daas, M. J. A., Vriesendorp, B., van de Weijer, A H. P., van der Oost, J. and van Kranenburg, R. (2018) Complete Genome Sequence of *Geobacillus thermodenitrificans* T12. A Potential Host for Biotechnological Applications, *Current Microbiology,* 75(1), 49-56

Darling, A. C., Mau, B., Blattner, F. R. and Perna, N. T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research,* 14(7), 1394-1403

De Maayer, P., Chan, W-Y., Blom, J., Venter, S. N., Duffy, B., Smits, T. H. M. and Coutinho, T. A. (2012) The large universal Pantoea plasmid LPP-1 plays a major role in biological and ecological diversification. *BMC Genomics,* 13(625), 1471-2164

De Souza, P. M., and de Oliveira Magalhães P. (2010). Application of microbial α-amylase in industry: A review. *Brazilian Journal of Microbiology*, 41(4), 850-861

del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M., and Diaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiology and Molecular Biology Reviews*. 62, 434–464

Díaz-Orejas, R., Espinosa, M. and Yeo, C. C. (2017) The Importance of the Expendable: Toxin-Antitoxin Genes in Plasmids and Chromosomes. *Frontiers in Microbiology*, 8, 1479

Dudley, M. (1995) Bacterial resistance mechanisms to beta-lactam antibiotics: assessment of management strategies, *Pharmacotherapy,* (1), 9-14

Dutta, C. and Pan, A. (2002) Horizontal gene transfer and bacterial diversity. *Journal of Biosciences,* 27(1), 27-33

Eberhard, W. G. (1990) Evolution in bacterial plasmids and levels of selection. *The Quarterly Review of Biology,* 65(1), 3-22

Emms, D. M. and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 157(16), 1

Ewis, H. E., Abdelal, A. T., and Lu, C.-D. (2004). Molecular cloning and characterization of two thermostable carboxyl esterases from *Geobacillus stearothermophilus*. *Gene,* 329, 187195

Fernández-López, R., Garcillán-Barcia, M. P., Revilla, C., Lázaro, M. and Vielva, L. (2006) dlC F: Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiology Reviews,* 30(6), 942-966

Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., Tang, Y., Liu, X., Han, W., Peng, X., Liu, R. and Wang, L. (2006) Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proceedings of the National Academy of Sciences of the United States of America*, 104(13), 5602-5607

Fueki, T. and Yamaguchi, K. (2001) The structure and function of the replication initiator protein (Rep) of pSC101: an analysis based on a novel positive-selection system for the replication-deficient mutants. *Journal of Biochemistry,* 130(3), 399-405

Gadelle, D., Krupovic, M., Raymann, K., Mayer, C., and Forterre, P. (2014) DNA topoisomerase VIII: a novel subfamily of type IIB topoisomerases encoded by free or integrated plasmids in Archaea and Bacteria, *Nucleic Acids Research,* 42(13), 8578–8591

Garcillan-Barcia, M. P., Francia, M. V. and De La Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiology Revisions*, 33, 657– 687.

Gerdes, K., Rasmussen, P. B. and Molin, S. (1986) Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proceedings of the National Academy of Sciences of the United States of America*, 83, 3116–3120

Gillespie, J. J., Phan, I. Q., Driscoll, T. P., Guillotte, M. L., Lehman, S. S., Rennoll-Bankert, K. E., Subramanian, S., Beier-Sexton, M., Myler, P. J., Rahman, M. S. and Azad, A.F. (2016) The *Rickettsia* type IV secretion system: unrealized complexity mired by gene family expansion. *Pathogens and Disease*, 74(6), 58

Gull, M. and El-Baz, S. (2018) Introductory Chapter: Preface to Plasmids. *InTechOpen* Peerreviewed chapter, 130(3), 399-405

Guo, F. B., Xia, Z. K., Wei, W. and Zhao, H. L. (2014) Statistical analyses of conserved features of genomic islands in bacteria. *Genetic and Molecular Research,* 13(1), 1782-1793

Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3), 307-21

Hall, T. A. (1999) BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series,* 41, 95-98

Hoshino, T., Ikeda, T., Narushima, H. and Tomizuka, N (1985) Isolation and characterization of antibiotic-resistance plasmids in thermophilic bacilli. *Canadian Journal of Microbiology*, 31(4), 339-345

Iyer, L. M., Zhang, D., Rogozin, I. B., and Aravind, L. (2011) Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Research,* 39(22), 9473-9497

Kananavičiūtė, R., Butaitė, E. and Citavičius, D. (2014) Characterization of two novel plasmids from *Geobacillus* sp. 610 and 1121 strains. *Plasmid,* 71, 23-31

Khanniri, E., Bagheripoor-Fallah, N., Sohrabvandi, S., Mortazavian, A. M., KhosraviDarani, K., and Mohammadi, R. (2015). Application of liposomes in some dairy products. *Critical Reviews in Food Science and Nutrition,* 20

Krasteva, P. V., Bernal-Bayard, J., Travier, L., Martin, F. A., Kaminski, P-A., Karimova, G., Fronzes, R. and Ghigo, J-M. (2017) Insights into the structure and assembly of a bacterial cellulose secretion system. *Nature Communications,* 8, 2065

Kumar A, Munjal A, Khan A, Singh RN (2012) Mega plasmid mediated anthracene degradation by novel strain *Geobacillus stearothermophilus* AAP7919 isolated from oil contaminated soil. *Journal of Pharmacy Research*. 5, 381-384

Kwong, S. M., Ramsay, J. P., Jensen, S. O. and Firth, N. (2017) Replication of Staphylococcal Resistance Plasmids, *Frontiers in Microbiology*, 8, 2279

Laverde-Gomez, J. A., Bhatty, M. and Christie, P. J. (2013) PrgK, a multidomain peptidoglycan hydrolase, is essential for conjugative transfer of the pheromone-responsive plasmid pCF10, *Journal of Bacteriology*, 196(3), 527-539

LeBrasseur N. (2007) ParM pushes plasmids apart. *The Journal of Cell Biology*, 179(5), 806

Lee, W., McDonough, M. A., Kotra, L. P., Li, Z-H., Silvaggi, N. R., Takeda, Y., Kelly, J. A. and Mobashery, S. (2001) A 1.2-Å snapshot of the final step of bacterial cell wall biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 98(4), 1427,1431

Lefort, V., Longueville, J.E. and Gascuel, O (2017) "SMS: Smart Model Selection in PhyML.", *Molecular Biology and Evolution,* 59(3), 307-21

Lilly, J. and Camps, M. (2015) Mechanisms of Theta Plasmid Replication. *Microbial Spectrum,* 3(1), 228-235

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D. and Darnell, J. (2000) DNA Cloning with Plasmid Vectors., *Molecular Cell Biology*, (4)

Lorenzo-Diaz, F., Fernandez-Lopez, C., Garcillan-Barcia, M. P., and Espinosa, M. (2014) Bringing them together: plasmid pMV158 rolling circle replication and conjugation under an evolutionary perspective. *Plasmid*, 74, 15–31

Lovett S. T. (2011) The DNA Exonucleases of Escherichia coli. *EcoSal Plus*, 4(2), 22

Luo, H. and Gao, F. (2018) DoriC 10.0: an updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acid Research,* 47(D1), D74-D77

Lutkenhaus, J. (2012) The ParA/MinD family puts things in their place. *Trends in Microbiology,* 20(9), 411-418

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchler, G. H., Mullokandov, M., Omelchenko, M. V., Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N., Zheng, C. and Bryant, S. H. (2010) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research,* 39, D225-9.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., and Bryant, S. H. (2014) CDD: NCBI's conserved domain database. *Nucleic Acids Research*, 43, D222-6.

De Souza, P. M., and de Oliveira Magalhães P. (2010). Application of microbial α-amylase in industry: A review. *Brazilian Journal of Microbiology*, 41(4), 850-861

Miguel-Arribas, A., Hao, J-A., Luque-Ortega, J. R., Ramachandran, G., Val-Calvo, J., GagoCórdoba, C., González-Álvarez, D., Abia, D., Alfonso, C., Wu, L. J. and Meijer, W. J. J. (2017) The Bacillus subtilis Conjugative Plasmid pLS20 Encodes Two Ribbon-Helix-Helix Type Auxiliary Relaxosome Proteins That Are Essential for Conjugation, *Frontiers in microbiology*, 8, 70

Montoro-García, S., Martínez-Martínez, I., Navarro-Fernandez, J., Takami, H., GarcíaCarmona, F., & Sanchez-Ferrer, A. (2009). Characterization of a novel thermostable carboxylesterase from *Geobacillus kaustophilus* HTA426 shows the existence of a new carboxylesterase family. *Journal of bacteriology*, 191, 3076-3085

Nazina, T. N., Tourova, T. P., Poltaraus, A. B., Novikova, E. V., Grigoryan, A. A., Ivanova, A. E., Lysenko, A. M., Petrunyaka, V. V., Osipov, G. A. & other authors (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* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G. kaustophilus, G. thermoglucosidasius* and *G. thermodenitrificans*. *International journal of systematic and evolutionary microbiology*, 51, 433-446

Norberg, P., Bergstrom, M., Jethava, V., Dubhashi, D. and Hermansson, M. (2011) The IncP1 plasmid backbone adapts to different host bacterial species and evolves through homologous recombination, *Nature communications*, 268(2), 1-11

Notredame, Higgins and Heringa, (2000) T-Coffee: A novel method for multiple sequence alignments, *JMB*, 302, 205-217

Obojska A, Ternana NG, Lejczak B, Kafarski P & McMullan P (2002) Organophosphate utilization by the thermophile *Geobacillus caldoxylosilyticus T20*. *Applied Environmental Microbiology*, 68, 2081- 2084

Ochman, H., Lawrence, J. G. and Groisman, E. A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature*, 405, 299-304

Orlova, A., Garner, E. C., Galkin, V. E., Heuser, J., Mullins, R. D. and Egelman, E. H. (2007) The structure of bacterial ParM filaments. *Nature structural & molecular biology*, 14(10), 921–926

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwrads, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F. and Stevens, R. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic acids research*, 42(Database issue), D206–D214

Pokusaeva, K., Kuisiene, N., Jasinskyte, D., Rutiene, K., Saleikiene, J. and Chitavichius, D. (2009) Novel bacteriocins produced by *Geobacillus stearothermophilus*, *Central European journal of biology*, 4(2), 196-203

Pruitt, K. D., Tatusova, T. and Maglott, D. R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, *Nucleic Acids Research. 33 (Database* issue): D501–504

Ramsay, J. P., Kwong, S. M., Murphy, R. J., Yui Eto, K., Price, K. J., Nguyen, Q. T., O'Brien, F. G., Grubb, W, B., Coombs, G. W. and Firth, N. (2016) An updated view of plasmid conjugation and mobilization in Staphylococcus. *Mobile Genetic Elements*, 6(4), e1208317

Rocha, E. P. and Danchin, A. (2002) Base composition bias might result from competition for metabolic resources. *Trends in Genetics*, 18(6), 291–294

Römling, U. and Galperin, M. Y. (2015) Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. *Trends in Microbiology*. 23, 545–557

Salto, I. P., Tjereizo, G. T., Wibberg, D., Puhler, A., Schluter, A. and Pistorio, M. (2018) Comparative genomic analysis of *Acinetobacter* spp. plasmids originating from clinical settings and environmental habitats. *Nature Scientific Reports,* 8, 2287

Schandel, K. A., Muller, M. M. and Webster, R. E. (1992) Localization of TraC, a protein involved in assembly of the F conjugative pilus, *Journal of Bacteriology,* 174(11), 3800-3806

Sheng, L., Zhang, Y. and Minton, N. (2016) Complete Genome Sequence of Geobacillus thermoglucosidasius NCIMB 11955, the Progenitor of a Bioethanol Production Strain. *Genome Announcements*, 4(5), 1065-1071

Scherzinger, E., Lurz, R., Otto, S. and Dobrinski B. (1992) In vitro cleavage of double- and single-stranded DNA by plasmid RSF1010-encoded mobilization proteins, *Nucleic Acids Research,* 20, 41-48 Singh, P. K., Ramachandran, G., Durán-Alcalde, L., Alonso, C., Wu, L. J. and Meijer, W. J. J. (2012) Inhibition of *Bacillus subtilis* natural competence by a native, conjugative plasmidencoded comK repressor protein. *Environmental Microbiology,* 14(10), 1462-1468

Siddiqui, M. A., Rashid, N., Ayyampalayam, S. and Whitman, W. B. (2014). Draft Genome Sequence of *Geobacillus thermopakistaniensis* Strain MAS1. *Genome Announcements*, 2(3), e00559-14

Singh, P. K., Ramachandran, G., Ramos-Ruiz, R., Peiró-Pastor, R., Abia, D., Wu, L. J. and Meijer, W. J. J. (2013) Mobility of the Native *Bacillus subtilis* Conjugative Plasmid pLS20 Is Regulated by Intercellular Signaling. *PLoS Genetics*, 9, 10

Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., and de la Cruz, F. (2010) Mobility of Plasmids, *Microbiology and molecular biology reviews: MMBR*, 74(3), 434–452

Spratt, B. G. and Strominger, J. L. (1976) Identification of the major penicillin-binding proteins of Escherichia coli as D-alanine carboxypeptidase IA. *Journal of Bacteriology*, 127(1), 660663

Stahl, S. (1991) Plasmids *in Bacillus stearothermophilus* coding for bacteriocinogeny and temperature resistance. *Plasmid*, 26, 94-107

Stothard, P. and Wishart, D. S. (2008) Circular genome visualization and exploration using CGView. *Bioinformatic*s, 21, 537-539

Stuknyte, M., Guglielmetti, S., Mora, D., Kuisiene, N., Parini, C. and Citavicius, D. (2008) Complete nucleotide sequence of pGS18, a 62.8-kb plasmid from *Geobacillus stearothermophilus* strain 18. *Extremophiles*, 2 415-2 429

Talavera, G., and Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* 56, 564-577

Takami, H., Takaki, Y., Chee, G. J., Nishi, S., Shimamura, S., Suzuki, H., Matsui, S. and Uchiyama, I. (2004) Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*. *Nucleic Acids Research*, 32(21), 6292–6303

Thompson, M. G., Sedaghatian, N., Barajas, J. F., Wehrs, M., Bailey, C. B.,Kaplan, N., Hillson, N. J., Mukhopadyay, A. and Keasling, J. D. (2018) Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. *Scientific Reports*, 8, 1590

Tsuchimoto, S., Ohtsubo, H., and Ohtsubo, E. (1988) Two genes, pemK and pemI, responsible for stable maintenance of resistance plasmid R100. *Journal of Bacteriology*, 170, 1461-1466

Tsuchimoto, S. and Ohtsubo, H. (1993) Autoregulation by cooperative binding of the PemI and PemK proteins to the promoter region of the pem operon. *Molecular and General Genetics MGG*, 237, 81-88

Wallden, K., Rivera-Calzada, A., and Waksman, G. (2010) Type IV secretion systems: versatility and diversity in function. *Cellular Microbiology*, 12(9), 1203–1212

Wawrzyniak, P., Plucienniczak, G. and Bartosik, D. (2017) The Different Faces of RollingCircle Replication and Its Multifunctional Initiator Proteins. *Frontiers in Microbiology*, 23, 1664

Zeigler, D. R. (2005) Application of a recN sequence similarity analysis to the identification of species within the bacterial genus *Geobacillus. Intenational Journal of Systematic and Evolutionary Microbiology*, 55, 1171-1179

Zeigler, D. R. (2014) The Geobacillus paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet, *Microbiology*, 160, 1-11

Chapter 3: Cargo genes of *Geobacillus* and *Parageobacillus* plasmids and their biotechnological potential

Abstract

Plasmids are prevalent features among members of the genera *Geobacillus* and *Parageobacillus*. To date there little is known about the functional role of these plasmids. Here 56 geobacilli plasmids were analyzed to gain insight into the cargo genes they carry as well as their associated phenotypes. The geobacilli plasmids show extensive versatility in their cargo genes, coding for a total of 2,961 proteins belonging to 618 distinct orthogroups. Functional annotation of the proteins reveal that the plasmids contribute to a wide range of phenotypes, including antibiotic resistance, antimicrobial compounds production and metabolism of a broad array of substrates. Furthermore, these plasmids code for a range of enzymes that could be of use in biotechnological applications. Finally, to further characterize the functional potential of the plasmids, plasmid isolation and curing of a subset of strains was attempted.

3.1 Introduction

Plasmids typically comprise core elements such as the *ori* site of replication as well as genes that code for proteins involved in replication, maintenance and transmission (del Solar *et al.*, 1998; Lodish *et al.*, 2000). In addition to these essential elements, plasmids also harbour a variety of cargo genes, that are not necessary for the functioning of a plasmid but encode proteins that confer a panoply of phenotypic functions to plasmid-host cells (Smillie *et al.*, 2010; Johnson and Grossman, 2015). These accessory genes can be acquired throughout a plasmid's existence, where they may be maintained for multiple generations but can also be lost if the surrounding environment does not provide a selective pressure to ensure their maintenance (Burrus and Waldor, 2004). Thus, the phenotypic characteristics and capabilities of bacteria can vary depending on the types of genes present within their plasmid sequences (Burrus and Waldor, 2004).

As discussed earlier (section 1.5), phenotypic traits contributed by plasmids includes heightened resistance to temperature or antibiotics as well as, the production of metabolic enzymes, toxins or virulence factors all of which contribute to increased metabolic and pathogenic characteristics of bacteria and can thus provide them with competitive advantages over other organisms (Johnson and Grossman, 2015). Furthermore, some of these factors can also act as defense mechanisms against other bacteria, invading bacteriophages or adverse environmental conditions, enabling the colonization of different environments (De Maayer *et*

al., 2012; Gull and El-Baz, 2018). Plasmids are thus able to make a significant contribution to the survival of a bacterium in a particular environment and also equip bacteria with a variety of characteristics to carry out the colonization of novel niches (De Maayer *et al.*, 2012; Gull and El-Baz, 2018). In addition to being highly beneficial to the bacteria that harbour them, some plasmids may represent a source of genes and proteins that are useful in a range of different biotechnological applications and are thus widely utilized to optimize production processes, product quality or costs within different industries and applications (Lodish *et al.*, 2000).

A number of studies have focused on select cargo proteins encoded on plasmids of members of the genera *Geobacillus* and *Parageobacillus*. These cargo proteins have been shown to play several distinct roles, including in metabolism, antibiotic resistance and antibiosis. One study identified four plasmids ranging in size from 12kb to 108kb, in a strain of *G. stearothermophilus* isolated from a compost pile in Japan (Mielenz, 1983). Of these plasmids, a 26 kilobase plasmid incorporates a gene coding for a thermostable α-amylase enzyme (Mielenz, 1983). α-Amylases catalyse the cleavage of the α-1,4-glucosidic linkages between glucose molecules in starch (De Souza and de Oliveira Magalhaes, 2010). A plasmid in *G. stearothermophilus* AAP7919 harbours genes that facilitate degradation of anthracene, a polycyclic aromatic hydrocarbon (PAH) generally used in the production of dyes and pesticides (Kumar *et al.*, 2012). PAHs are organic compounds that are formed during combustion of organic molecules such as coal, and are found to be pollutants of the air, soil, foodstuffs and water and have been described as toxic and potentially carcinogenic (Bisht *et al*, 2015). The plasmid pLW107 of *G. thermodenitrificans* NG80-2 encodes a monooxygenase (LadA) which facilitates the degradation of alkanes that naturally occur in crude oil and which are common environmental pollutants (Feng *et al.*, 2007). *G. stearothermophilu*s was reported to carry plasmids that confer resistance against several antibiotics, including streptomycin, tetracycline and kanamycin (Hoshino *et al.*, 1985). Another study highlighted the production of antimicrobial bacteriocins by plasmids in *G. stearothermophilus* 32A, 17, 30 and 20 (Pokusaeva *et al.*, 2009). These bacteriocins were effective against closely related *Geobacillus* species and the pathogens *Bacillus cereus* and *Staphylococcus haemolyticus* (Pokusaeva *et al.*, 2009).

Much of the interest in the plasmids of *Geobacillus* and *Parageobacillus* species can be linked to the fact that they carry genes of biotechnological interest. However, to date, studies have solely focused on isolated plasmids or plasmid-encoded proteins. As such, there is a need to

obtain a global picture of the phenotypes encoded on the plasmids of these taxa and to further exploit their biotechnological potential. Here, a holistic analysis of 56 plasmids in *Geobacillus* (34 plasmids) and *Parageobacillus* (22 plasmids) was undertaken to identify and characterize plasmid-encoded phenotypes. Plasmid isolation and curing protocols were also attempted, for further validation of plasmid-encoded phenotypes and biotechnological application.

3.2 Materials and Methods

3.2.1 Annotation of plasmid protein coding sequences

The plasmid nucleotide sequences of 56 complete or near-complete plasmids from *Geobacillus* and *Parageobacillus* were structurally and functionally annotated using the RAST server (Overbeek *et al.*, 2014) as discussed in Chapter 2 section 2.2.1. The resultant protein datasets were compared and clustered into orthologous groups using Orthofinder v 1.1.4 (Emms and Kelly, 2015). The protein datasets were then uploaded onto the eggNOG-mapper 4.5 server (Huerta-Cepas *et al.*, 2016) and classified according to their Conserved Orthologous Groups (COG) categories (Tatusov *et al.*, 2000). The parameters HMMER (Hidden Markov Models) mapping mode and "Bacteria database" were selected for the analysis.

The plasmid amino acid files of each individual plasmid were also uploaded onto the NCBI Batch Conserved Domain (CD) Search Server. This server uses RPS-BLAST to scan a set of pre-calculated position-specific scoring matrices (PSSMs) of a protein query (Marchler-Bauer *et al.*, 2010, Marchler-Bauer *et al.*, 2014). Furthermore, the protein datasets were annotated by comparison against the NCBI non-redundant protein database (Using BLASTp), InterPro and KEGG databases (Pruitt *et al.*, 2005; Bauer *et al.*, 2015; Hunter *et al.*, 2009; Selengut *et al.*, 2009; Kanehisa and Goto, 2000). Finally, secondary metabolite biosynthetic gene clusters were identified by searching the plasmid nucleotide sequences against the antiSMASH v 5.0 server (Blin *et al.*, 2019).

3.2.2 Plasmid extraction and curing

3.2.2.1 Bacterial cultures

Bacterial cultures of the type strains of three *Geobacillus* (*G. thermoleovorans* DSM 5366^T, *G. kaustophilus* DSM 7263^T, *G. stearothermophilus* ATCC 12980^T), and three *Parageobacillus* species (*P. thermoglucosidasius* DSM 2542^T, *P. caldoxylosilyticus* DSM 12041^T and *P. toebii* DSM 14590^T) were used in this study. The cultures were maintained in glycerol stocks at the School of Molecular and Cell Biology (University of Witwaterstrand, South Africa). The cultures were reconstituted on modified Luria Bertani (mLB) (originally called Lysogeny Broth) agar: 10 g tryptone, 5 g yeast extract, 5 g sodium chloride (NaCl), 995 ml Water, 1.25 ml sodium hydroxide (NaOH 10% w/v), 15 g bacteriological agar powder and 1 ml each of filter sterilized stocks of nitrilotriacetic acid (1.05 M NTA), manganese sulphate (0.59 M

$MnSO_4.7H_2O$), calcium chloride (0.91 M $CaCl_2.2H_2O$) and iron sulphate (0.04 M $FeSO_4.7H_2O$). The cultures were incubated at 55°C.

3.2.2.2 DNA extractions and gel electrophoresis

DNA extraction was performed using two different approaches. Firstly, genomic DNA was extracted using the ABIOpure™ Total DNA extraction kit (AllianceBio, U.S.A). The recommended protocol for Gram-positive was used to perform the extraction. Isolated bacterial colonies were picked from mLB agar plates and suspended in 2,000 µL distilled water, prior to pelleting by centrifugation (Spectrafuge 24D, Labnet International, Inc.) at 1,000 rpm for five minutes. The supernatant was discarded, and the pellet was resuspended in 180 µL of lysis buffer (30mg/ml lysozyme in buffer GP from the kit) and incubated at 37˚C for one hour. Proteinase K was added to the lysed cells and incubated at 56˚C and 70˚C for 30 minutes at each temperature. A volume of 400 ul of 100% ethanol was added to the lysed cells and the mixture was transferred to a mini G column and centrifuged at 10,000 rpm for one minute. Wash buffer BW was added, and the column was centrifuged at 10,000 rpm for one minute. Extracted DNA was dissolved and eluted from the column in 50 µL of buffer AE (10,000 rpm, one minute) and stored in the freezer.

Secondly, genomic DNA was extracted using the protocol of Kado and Liu (1981) for the specific isolation of plasmid DNA. Bacterial cultures were cultured in 50 ml mLB broth, for 48 hours on a shaker at 55˚C. Optical density readings of bacterial cultures were taken in triplicates and were between 0.4 and 0.6 (OD600 nm) amongst all cultures. Cells were pelleted at 4˚C at 5,700 rpm for 7 minutes in a Heraeus X1R centrifuge (Thermo fischer Scientific; U.S.A). The cell pellets were resuspended in 5 ml E buffer (40 mM Tris-HCl, 2mM Sodium EDTA (adjusted to pH 7,9 using NaOH pellets)). A volume of 10 ml lysis solution (3% SDS, 50 mM Tris adjusted to pH 12.6 using 2M NaOH) was added and the lysate was incubated in a waterbath at 60˚C for 30 minutes. 30 ml of phenol-chloroform (1:1 v/v) was added, and the solution was centrifuged at 4˚C at 6,000 rpm for 15 minutes. The aqueous layer was transferred to a new tube and placed on ice. A volume of 100 µl sodium acetate (3 M NaOAc) was added to 5ml of the aqueous DNA layer collected from each sample and then 800 µL of 100% ethanol was added to precipitate the DNA. This solution was centrifuged at 6,000 rpm for 30 minutes, with the supernatant being discarded. Samples were washed twice with 70% ethanol and centrifuged at 6,000 rpm for 5 minutes. The supernatant was discarded, and samples were left to dry over a heating block for 2 minutes. The DNA extracts were stored in E buffer at -20˚C.

Gel electrophoresis was carried out using a low melting point agarose gel (0.4% w/v in 1x TBE buffer). The DNA extracts were mixed with Novel Juice DNA Staining reagent (GeneDireX; USA) in a 1:5 dilution and were loaded onto the gel, along with 3 µL of the O'GeneRuler 1 kb DNA ladder (Thermo Fischer Scientific; USA). Electrophoresis was performed at 4˚C for ~48 hours at 25 V. A 0.7% (w/v in 1x E buffer) agarose gel was used to ascertain plasmid isolation from the DNA extraction described by Kado and Liu (1981). The gel was run at 10 V/cm in 1x E buffer for ~3 hours. All agarose gels were viewed using a UV transilluminator.

3.2.2.3 16S rRNA PCR amplification

To determine whether the extracted DNA represented plasmid DNA or genomic DNA, PCR amplification of the chromosomal 16S rRNA gene was performed. PCR amplification was performed according to the instructions of the MyTaq™ Polymerase Kit (Bioline; USA). The universal bacterial 16S rRNA primers 8F (5'-AGAGTTTGATCCTGGCTA-3') forward and 1492R (3'- GGTTACCTTGTTACGACTT -5') reverse primers (Turner *et al,* 1999; Lane, 1991) were used in preparation of the PCR master mix. PCR amplifications was performed using a BioRad™ MyCycler thermal cycler (BioRad; USA) using the cycle program: denaturation at 95˚C for five minutes; 30 cycles of 30 seconds at 95˚C, 30 seconds at 55˚C, 90 seconds at 72˚C and final extension at 72˚C for ten minutes. Amplicons were run on a 0.7% agarose gel, as described above.

3.2.2.4 Plasmid curing

P. toebii DSM 14590^{T} was cultured in mLB broth media containing 50 µL of 20M acridine orange as described by Mizuno *et al* (2017). The cultures were incubated at 55˚C, on a rotary shaker at 160 rpm for one week. Serial dilutions (10^{-1}, 10^{-2} and 10^{-3}) of each culture were prepared and spread plated onto mLB agar and incubated at 55℃ overnight. To ascertain plasmid curing efficacy, DNA extraction of the cured strain and wild-type was carried out using the ABIOpure™ Total DNA extraction kit and provided protocol (AllianceBio, U.S.A) to identify any changes in banding pattern.

3.3 Results and Discussion

3.3.1 Overview of the functional roles of the geobacilli plasmids

The 56 geobacilli plasmid incorporated into this study code for a total of 2,691 proteins, with the larger pGeo1 plasmids contributing 1,717 proteins (63.73%), the pGeo2 plasmids contributing 875 proteins (32.48%) and, the pGeo3 to pGeo9 plasmids collectively contributing 99 proteins (3.67%). The protein complements of all 56 plasmids were organized into their orthogroups (groups of proteins sharing orthology) using Orthofinder v 1.1.4 (Emms and Kelly, 2015). A total of 618 distinct protein orthogroups were identified amongst all 56 plasmids. The pGeo1, pGeo2 and pGeo3-pGeo9 plasmids, incorporate 429, 225 and 91 distinct protein orthogroups amongst each group, respectively. This highlights the extensive diversity of the geobacilli plasmids even within plasmid groups.

The protein complements of each plasmid were annotated according to their COG functional categories with eggNOG mapper (Marchler-Bauer *et al.,* 2010*;* Marchler-Bauer *et al.,* 2014; Tatusov *et al.,* 2000), with 1,573 proteins (58.45% of the total geobacilli plasmid protein complement) being assigned to COG categories, while the remaining proteins could not be assigned. A total of 1,111 (41.24%), 433 (16.07%) and 29 (1.84%) proteins encoded by the pGeo1, pGeo2 and pGeo3-pGeo9, respectively, were identified by eggNOG mapper and assigned to COG groups. Of the categorised proteins, 572 are involved in COG super-category "Information storage and Processing", 283 belonged to the COG super category "Cellular processes and Signalling", 299 in "Metabolism", while 419 belong to the super category of "General or Unknown functions" proteins (Galperin *et al.*, 2015).

When considering the 618 protein orthogroups, 310 were classified according to their COG functional categories (Figure 1). Most of the protein orthogroups (32.26%) are placed in the COG super-functional category metabolism, suggesting a major role for the geobacilli plasmids in this cellular function. Of the remaining proteins, 23.87%, 13.55% and 30.32% are involved in information storage and processing, cellular processes and signalling and general function prediction only/unknown function, respectively (Figure 1).

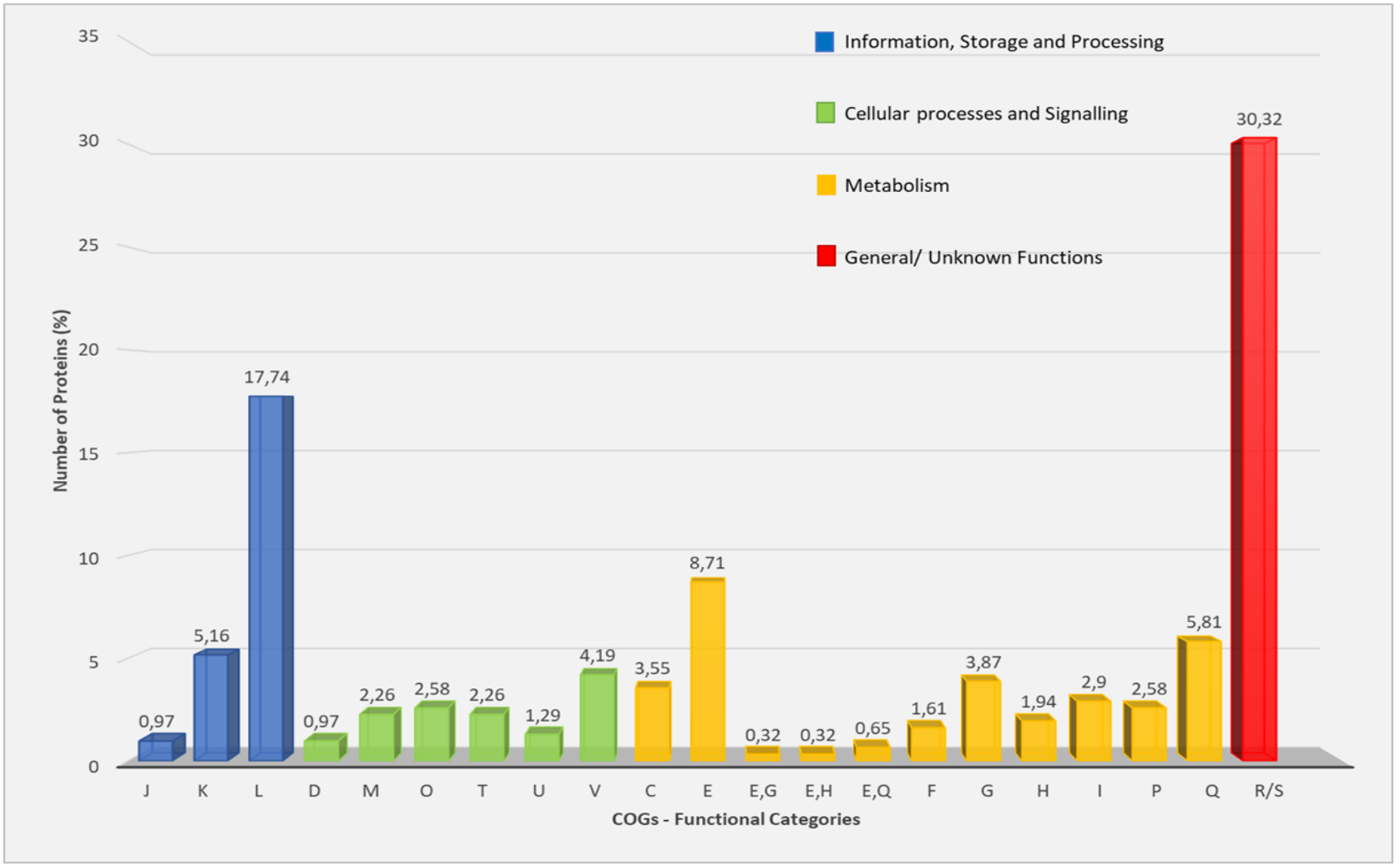


Figure 1: Bar graph representing the 310 protein orthogroups assigned to COG functional categories. Sub-categories of each COG supercategory is characterized by blue (information storage and processing), green (cellular processes and signalling), orange (metabolism and red (general/unknown functions). The 308 orthogroups not found in COGS are not presented here.

Proteins categorised in information storage and processing are predominated by those involved in DNA replication, recombination and repair (COG category L; 55 distinct orthogroups; 17.74% of total annotated proteins) (Figure 1). This is typical for plasmids, as proteins in this category include replication proteins, helicase enzymes, DNA repair proteins, recombinase enzymes, DNA methylases and numerous transposition-related proteins (Fondi *et al.*, 2010). In addition, sixteen and three distinct orthogroups are involved in transcription (COG category K) and translation (COG category J), respectively.

Proteins involved in cellular processes and signalling are mostly associated with the COG functional categories cell cycle control, cell division and chromosome partitioning (COG category D; three orthogroups), cell wall/membrane/envelope biogenesis (COG category M; seven orthogroups) and intracellular trafficking and secretion (COG category U; four orthogroups) (Figure 1). COG category D is dominated by plasmid partitioning proteins, with orthologues of the plasmid partitioning proteins MinD (CD02042; Bitscore: 162.06; E-value: $3.78e^{-}47$) and ParM (CD10227; Bitscore: 173.96; E-value: $2.15e^{-}51$) encoded on the pGeo1 and pGeo2 plasmids, respectively. Type 4 secretion system proteins, which are predicted to contribute to the conjugative transmission of the pGeo2 plasmids represent the majority of the proteins in category U. A total of twelve distinct orthogroups (3.87% of the total proteins are involved in cellular defence (category V; 13 orthogroups) and includes restriction endonuclease enzymes and proteins involved in resistance to antibiotics.

The geobacilli plasmids play a substantive role in metabolism, with amino acid transport and metabolism (category E; 31 distinct orthogroups) and secondary metabolite biosynthesis, transport and metabolism (category Q; 20 distinct orthogroups) representing the predominant metabolic functional categories (Figure 1). Overall, approximately 12.96% of the total proteins involved in metabolism are contributed by the pGeo1 plasmids, while ~ 2% is contributed by the pGeo2 plasmids, suggesting a role for the pGeo1 plasmids in the metabolic versatility of the *Geobacillus* and *Parageobacillus* spp. that harbour them.

The protein complement of the geobacilli plasmids was further functionally characterized on the basis of the RAST annotation, BLASTp analysis against the NCBI non-redundant protein database and using Batch CD-search (Huerta-Cepas *et al.*, 2016). This provides a more comprehensive overview of the potential biological roles of the geobacilli plasmids and their cargo genes.

3.3.2 The geobacilli plasmids contribute to antibiotic resistance

Antibiotic resistance proteins are frequently encoded by plasmids, with these extrachromosomal elements being one of the main drivers of the spread of antibiotic resistance amongst bacteria (Condit and Levin, 1988). The geobacilli plasmids are observed to encode a variety of proteins potentially involved in resistance to antibiotics.

The pGeo1 plasmid of *G. stearothermophilus* DSM 15325 incorporates a cluster of genes coding for proteins that share orthology with the MacAB-TolC-like ABC (ATP binding cassette) transporter found in a wide range of bacteria, including *Escherichia coli* and *Bacillus subtilis* (Davidson *et al.,* 2008; Yamanaka *et al.,* 2008; Fitzpatrick *et al.,* 2017). These proteins include the ATPase binding permease protein MacB (DSM15325pGeo1_CDS4; pfam12704; Bitscore: 45.96; E-value: 6.96e^{-}08; 59 aa), accessory protein MacA (DSM15325pGeo1_CDS6; CD03255; Bitscore: 279.84; E-value: 2.29e^{-}95; 66 aa) and transmembrane pore forming protein TolC (DSM15325pGeo1_CDS7; PRK10535; Bitscore: 50.11; E-value: 3.76e^{-}09; 216 aa) (Yamanaka *et al*, 2008). This type of transporter is reportedly involved in the resistance against and efflux of 14- and 15-membered macrolide antibiotics, such as erythromycin and azithromycin as well as polypeptide virulence factors (Davidson *et al*, 2008; Yamanaka *et al*, 2008; Fitzpatrick *et al*, 2017).

Two pGeo1 (*G. icigianus* C56-T2pGeo1_CDS10 and *P. thermoglucosidasius* C56YS93pGeo1_CDS38) plasmids and four pGeo2 (*G. icigianus* C56-T2pGeo2_CDS39, *Geobacillus* genomosp. 1 Et2/3pGeo2_CDS43, *G. stearothermophilus* 15pGeo2_CDS42 and *G. thermodenitrificans* NG80-2pGeo2_CDS66) plasmids code for a MefA-like transmembrane protein of a major facilitator superfamily (MFS) transporter (TIGR00900; Bitscore: 137,95; Evalue: 1.16e^{-}23; 419 aa; AAI: 52.64%). MefA transporters were originally identified in *Streptococcus pyogenes* as an efflux protein specific to macrolide resistance (Clancy *et al.,* 1996). The geobacilli MefA-like proteins are found to share an average amino acid identity of 20.61% with that of *S. pyogenes* 36 (WP_063853602) (Sangvik *et al.,* 2005).

The pGeo1 plasmids of six *P. thermoglucosidasius* strains (DSM 2542^{T}, DSM2543, NCIMB 11955, DSM6285, DSM21625, G08C001) were observed to encode an ATPase component (CD03268; Bitscore: 290.65; E-value: 1.66e-99; 303 aa; AAI 78.56%) of an ABC-type transporter. Four of the six plasmids were also observed to incorporate two co-localised genes encoding transmembrane proteins (COG4200; Bitscore: 116.67; E-value: 1.37e-31; 243 aa; AAI: 99.55%) belonging to the family of ABC-2 type transporters (Table 1).

Table 1: Plasmid loci of ABC- 2 type transporter subunits

Genus and species	Strains/plasmids (pGeo1)	Locus tag	Locus tag
		ATPase	Transmembrane proteins
P. thermoglucosidasius	DSM 2542	pGeo1_CDS107	-
	NCIMB 11955	pGeo1_CDS110	-
	G0C8001	pGeo1_CDS126	pGeo1_CDS127-128
	DSM 6285	pGeo1_CDS9	pGeo1_CDS10-11
	DSM 21625	pGeo1_CDS10	pGeo1_CDS11-12
	Y41MC1	pGeo1_CDS105	pGeo1_CDS106-107

These ABC-2 type transporters are reported to play a role in bacitracin resistance in *Bacillus licheniformis* and *B. subtilis* and are also predicted to be involved in lantibiotic resistance in *Lactococcus lactis* (Reizer *et al.*, 1992; Podlesek *et al.*, 2003; Polgar and Bates, 2005; Mirkovic *et al.,* 2016). The geobacilli ATPases and transmembrane proteins are found to share 55.47% and 35.65% average amino acid identity with orthologues in *B. licheniformis* LMG 7559 (KUL13544; KUL13543) (Podlesek *et al.,* 2003).

The pGeo2 plasmid of *G. thermodenitrificans* T12 encodes an orthologue of the phosphotransferase enzyme APH (T12pGeo2_CDS45; pfam01636; Bitscore: 67.92; E-value: 2.90e^{-}13; 317 aa), belonging to the aminoglycoside phosphotransferase family (APHs) (Sarwar and Akhtar, 1990). Members of this family are reported to confer resistance against aminoglycoside-based antibiotics such as neomycin, kanamycin, streptomycin and gentamicin in a range of bacteria (Sarwar and Akhtar, 1990). These enzymes inactivate the antibiotic by catalysing the phosphorylation of the binding site, reducing the antibiotics affinity for its binding target (Sarwar and Akhtar, 1990). The geobacilli enzyme is observed to share low amino acid identity with the Aph enzymes of *Bacillus circulans* NCTC2610 (SPT80732; 14.82% AAI) and *Legionella pneumophila* 130b (AAB58447; 17.95% AAI) (Scheeff and Bourne, 2005; Fong *et al.,* 2010). It has, however, been reported that these enzymes are highly diverse in structure and substrate specificity with representatives of the APH family sharing at most 35% amino acid identity amongst each other (Wright and Thompson, 1999).

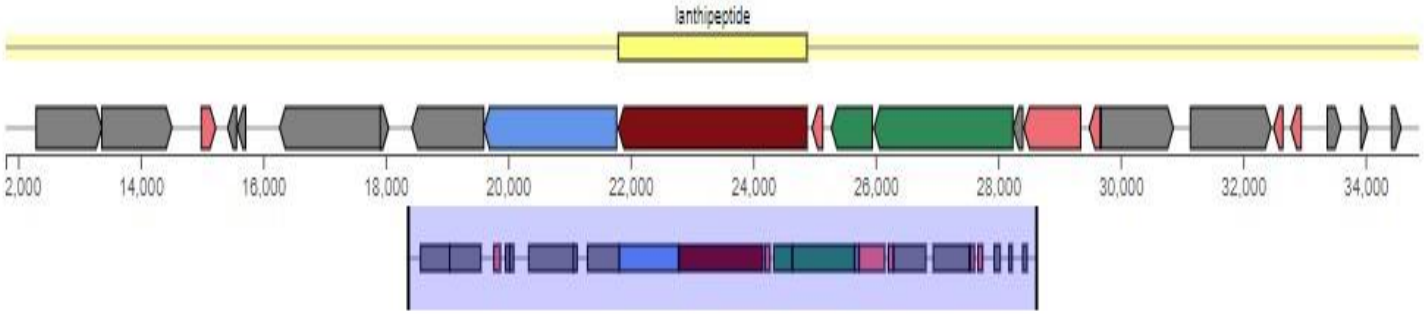
3.3.3 The geobacilli plasmids contribute to antibiosis and interbacterial competition

Apart from potentially conferring resistance to antibiotics and bacteriocins, analysis of the geobacilli plasmid sequences provides evidence that they are involved in the production of secondary metabolites with antimicrobial properties. The production of these antibiotic compounds may contribute to the ability of *Geobacillus* and *Parageobacillus* to outcompete other bacterial taxa for nutrients as well as enable them to colonize novel ecological niches (Jackson *et al.*, 2011).

Analysis of the pGeo1 plasmid of *G. thermodenitrificans* JSC_T9a using antiSMASH v 5.0 server identified a gene cluster predicted to be involved in the synthesis of the lanthipeptide Geobacillin II (Figure 2A). This gene cluster includes a 3,080 nucleotide long gene coding for an orthologue of the lanthipeptide synthetase, LanM (JSC_T9apGeo1_CDS26; TIGR03897; Bitscore: 1110.79: E-value: 0; 1,026). LanM is a bifunctional class II enzyme that carries out the phosphorylation of serine or threonine residues of the peptide substrate and subsequently eliminates the phosphate to generate the dehydroamino acid/thioether crosslinks of a typical lanthipeptide (Yu *et al.,* 2013). This protein shares 99.90% (Bitscore: 2094; E-value: 0.0) amino acid identity with the lanthipeptide synthase GeoM in *G. thermodenitrificans* NG80-2 (ABO67414) (Figure 2B). In the latter organism, *geoM* is flanked at the 5′ end by a predicted lantibiotic ABC transporter (GeoTII) and at the 3′ end by the *geoAII* gene coding for the 57 aa precursor peptide (Figure 2B) (Garg *et al.*, 2012). These genes are similarly found flanking the *lanM* gene on the pGeo1 plasmid of *G. thermodenitrificans* JSC_T9a with their protein products sharing 99.86% and 100% amino acid identity with the orthologues in *G. thermodenitrificans* NG80-2, respectively (Figure 2B).

Geobacillin II is one of two reported geobacilli lantibiotics. The first, Geobacillin I, was also identified and characterized in *G. thermodenitrificans* NG80-2 and is observed to share similarities with nisin, a *Lactococcus lactis* lantibiotic commonly used in food preservation (Garg *et al.,* 2012). Nisin and Geobacillin I are predicted to be involved in lipid binding and pore formation within the cellular membrane of competing bacteria (Garg *et al.*, 2012; Garg *et al.,* 2014). Geobacillin I is described to have activity against a wide range of Gram-positive bacteria, while Geobacillin II is reported to be specific for members of the genus *Bacillus* (Garg *et al.,* 2012; Garg *et al.,* 2014).

A.



B.

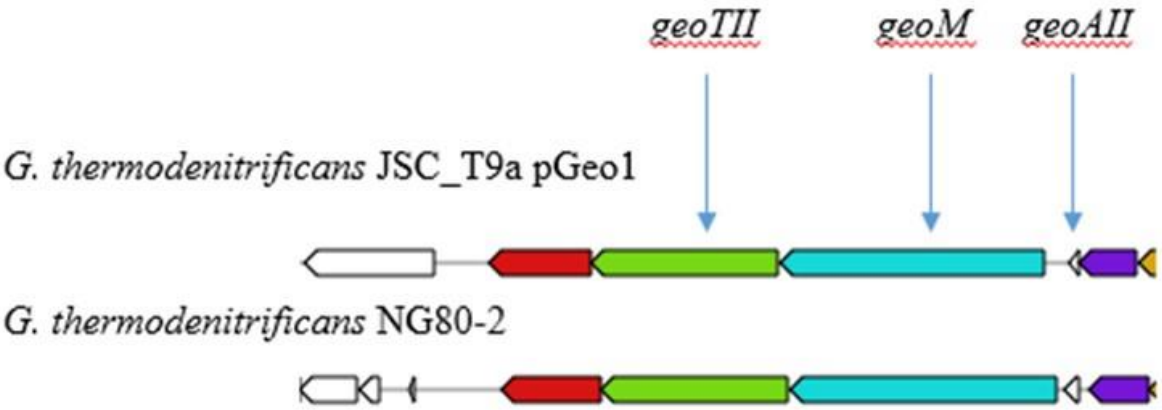


Figure 2: antiSMASH v 5.0 server (Blin *et al,* 2019) results depicting the coding regions of plasmid *G. thermodenitrificans* JSC_T9a involved in Geobacillin II synthesis and transport. (A) Lanthipeptide synthase LanM is depicted as the red bar and the ABC type transporter as the blue bar. **(B)** antiSMASH ClusterBlast output depicting the similar gene clusters between the pGeo1 plasmid *G. thermodenitrificans* JSC_T9a and *G. thermodenitrificans* NG80-2. The lanthipeptide synthetase genes *lanM/geoM* are presented as a blue-bars, the ABC-transporter genes *geoTII* as the green bars and the peptide precursor as purple bars.

The pGeo1 plasmids of five *P. thermoglucosidasius* strains (DSM2542pGeo1_CDS66; DSM2543pGeo1_CDS48; NCIMB11955pGeo1_CDS71; DSM6285pGeo1_CDS72 and G08C001pGeo1_CDS74) encode orthologues of the enzyme napthocyclinone hydroxylase NcnH (CD01159; Bitscore: 453.27: E-value: 1.97e^{-}159; 400 aa; AAI: 99.36%). This enzyme is reported to be involved in the synthesis of the secondary metabolite naphthocyclinone, an aromatic polyketide (Brunke *et al.,* 2001). The latter polyketide is produced by *Streptomyces arenae* DSM 40737, and is mostly active against Gram-positive bacteria (Brunke *et al.,* 2001). The geobacilli NcnH is found to share 25.52% amino acid identity with that of *S. arenae* (AF218066).

The pGeo1 plasmid of *G. stearothermophilus* ATCC 12980^{T} (ATCC12980pGeo1_CDS37-39) encodes three enzymes of the SagBCD complex, that is involved in the production of heterocyclic bacteriocins (Lee *et al.,* 2008). The proteins SagC (TIGR04424: Bitscore: 248.01; E-value: 9.95e^{-}82; 297 aa) and SagB (CD02142; Bitscore: 76.68; E-value: 4.14e^{-}17; 270 aa) function as a cyclohydratase and dehydrogenase enzyme, respectively, and are involved in the cyclizing of the thiazole ring and maturation of the bacteriocin peptide, while the docking protein SagD (TIGR03604; Bitscore: 63.59; E-value: 1.75e^{-}10; 713 aa) acts as a scaffolding protein during thiazole formation (Lee *et al.,* 2008). These proteins play a role in the synthesis and modification of bacteriocins such as microcin B17 and streptolysin S in a range of Bacillaceae (Pierrat and Maxwell, 2003; Molloy *et al.,* 2011).

3.3.4 The geobacilli plasmids contribute towards metabolic versatility

Plasmids frequently incorporate genes that code for pathways that enhance the metabolic versatility of the host bacterium and allow them to utilise a broad range of substrates as sources of nutrients (Pathak and Navneet, 2017; Urbanek *et al.,* 2018).

The pGeo1 and pGeo2 plasmid of *G. thermodenitrificans* T12 (T12pGeo1_CDS25-27) and PA-3 (PA-3pGeo2_CDS8-10), respectively, encode three co-localized proteins that form part of a potential ABC-type transporter potentially involved in the uptake of high molecular weight alginate, a polysaccharide produced by brown algae and some bacteria (Momma *et al.,* 2000; Mishima *et al.,* 2001). These transporter systems are typically made up of five components, two extracellular substrate binding proteins AlgQ1 and AlgQ2, two permease transport proteins AlgM1 and AlgM2 and an ATP-binding protein AlgS (Momma *et al.,* 2000; Mishima *et al.,* 2001). The two geobacilli plasmids encode orthologues of permease proteins AlgM1 (COG4209; Bitscore: 334,68; E-value: 3,58e^{-}115; 315 aa; AAI: 99.37%) and AlgM2

(COG0395; Bitscore: 193.91, E-value: 3.11e^{-}61; 272 aa; AAI: 87.73%), as well as the binding protein AlgQ1 (CD13581; Bitscore: 571.52; E-value: 0, 534 aa; AAI: 98.59%). These proteins share 58.55%, 61.27% and 61.51% average amino acid identity with the corresponding components of an ABC transporter in *Sphingomonas paucimobilis* NCTC11032 (VDG77686688). In the latter organism, this transporter has been shown to play a role in the uptake and metabolism of alginate, maltose and glycerol-3-phosphate as a carbon source (Momma *et al.,* 2000; Mishima *et al.,* 2001).

Three pGeo2 plasmids of *G. thermodenitrificans* (G11MC16pGeo2_CDS9, T12pGeo2_CDS4 and ID-1pGeo2_CDS49) encode orthologues of the isomerase enzyme ManC (COG0662; Bitscore: 106.72; E-value: 3.88e^{-}30; 181 aa; AAI: 100%). This enzyme catalyses the conversion of mannose-6-phosphate into fructose-6-phosphate (Sigdel *et al.,* 2015). Mannose6-phosphate isomerases are involved in the metabolism of glucose as well as the biosynthesis of the exopolysaccharide colanic acid in numerous bacteria (Hanna *et al.,* 2013; Sigdel *et al.,* 2015). This enzyme is also of biotechnological value, as it can be used in the production of rare sugars such as L-ribose, L-ribulose, L-tagatose and L-xylulose, which are used as sweeteners in the food industry (Sigdel *et al.,* 2015).

Fourteen of the pGeo1 plasmids (Table 2) encode a protein that may be involved in the uptake and metabolism of heterocyclic compounds. These proteins share orthology with the nucleobase-cation Symporter NCS1 (CD11484; Bitscore: 294.52, E-value: 3.32e-96; 438 aa; AAI: 96.94%). NCS1 is a membrane transporter that is involved in the specific uptake of purines, pyrimidines, hydantoins and other related metabolites (Suzuki and Henderson, 2006). In addition to the symporter protein these plasmids encode two orthologues of a hydantoinase enzyme HyuA (COG0145; Bitscore: 350.45; E-value: 1.92e-113; 519 aa) that are split into two orthogroups (a and b; Table 2). These enzymes share 93.25% and 97.18% amino acid identity within each orthogroup, respectively, but share only 35.21% amino acid identity with each other.

Hydantoinase catalyses the hydrolysis of hydantoin substrates to form N-carbamoyl-α-amino acids which are then converted to corresponding free, optically pure D- or L-amino acids. Bacteria are likely to have evolved such systems for the scavenging specific nucleobase and/or hydantoin compounds from their environments, which are then metabolized and enter salvage pathways for carbon- and nitrogen-containing nutrients (Suzuki and Henderson, 2006). HyuA, in particular, is specific for D-5 substituted hydantoins, where it hydrolyses the cyclic amide bond of D-hydantoin derivatives with aromatic side chains at the 5'-position. (Suzuki and

Henderson, 2006). D- and L-amino acids are often used as intermediates in several industries with D-amino acids being utilised in the synthesis of pesticides, sweeteners, antibiotics and other biologically active peptides (Las Heras-Vazquez *et al*, 2008).

Table 2: Plasmid loci of nucleobase-cation symporter and associated Hydantoinase enzymes

Genus species	Plasmids	Locus tag	Locus tag	Locus tag
		NCS1	HyuA (a)	HyuA (b)
G. thermoleovorans	A8	pGeo1CDS21	pGeo1CDS20	pGeo1CDS23
	DSM 405	pGeo1CDS14	pGeo1CDS13	pGeo1CDS16
	Y412MC52	pGeo1CDS8	pGeo1CDS9	pGeo1CDS6
	Y412MC61	pGeo1CDS27	pGeo1CDS26	pGeo1CDS29
Parageobacillus genomosp. 2	W-2	pGeo1CDS106	pGeo1CDS107	pGeo1CDS104
P. thermoglucosidasius	C56-YS93	pGeo1CDS57	pGeo1CDS58	pGeo1CDS55
	DSM 2542	pGeo1CDS45	pGeo1CDS44	pGeo1CDS47
	DSM 2543	pGeo1CDS70	pGeo1CDS71	pGeo1CDS68
	NCIMB 11955	pGeo1CDS49	pGeo1CDS48	pGeo1CDS51
	DSM 6285	pGeo1CDS53	pGeo1CDS52	pGeo1CDS55
	G08C001	pGeo1CDS52	pGeo1CDS51	pGeo1CDS54
	DSM 21625	pGeo1CDS55	pGeo1CDS54	pGeo1CDS57
	Y41MC1	pGeo1CDS42	pGeo1CDS41	pGeo1CDS44
P. toebii	44C	pGeo1CDS10	pGeo1CDS9	pGeo1CDS12

Ten pGeo1 plasmids of *Parageobacillus* genomosp. 2 W-2pGeo1_CDS19-20, *P. thermoglucosidasius* C56-YS93pGeo1_CDS76-79, DSM2542pGeo1CDS_20-23, DSM2543pGeo1_CDS92-95, NCIMB11955pGeo1_CDS24-27, DSM6285pGeo1_CDS31-34, DSM21625pGeo1_CDS35-38, G08C001pGeo_CDS32-35, Y41MC1pGeo1_CDS18-20 and *P. toebii* B4110pGeo1CDS7-10, encode four proteins that embody a spermidine/putrescine polyamine ABC transport system (Shah and Swiatlo, 2008). These proteins are orthologues of the ATPase domain protein PotA (COG3842; Bitscore: 438,64; E-value: 1,32e-154; 365; AAI: 97.98%), the transmembrane polypeptides PotB (COG1176; Bitscore: 218.25; E-value: 5.11e70; 310 aa; AAI: 96.26%) and PotC (COG1177; Bitscore:

221.68; E-value: 4.11e-72; 264 aa; AAI: 97.88%) and the periplasmic substrate binding protein PotD (CD13589; Bitscore: 250.22; E-value: 8.38e-82; 367 aa; AAI: 97.36%) (Shah and Swiatlo, 2008).

Polyamines are polycationic molecules consisting of a hydrocarbon backbone and multiple amino groups, with spermidine and putrescine being the most common types found in bacteria (Shah and Swiatlo, 2008). Many different microorganisms employ these transporter systems to regulate polyamine levels within the cell, where they participate in gene expression, cell growth, survival, bacteriocin production, stress response, proliferation and biofilm formation (Chattopadhyay *et al.,* 2003; Shah and Swiatlo, 2008; Gevrekci, 2017). Polyamines have been implicated in the thermophilic characteristics of *G, kaustophilus*, where they are thought to play a role in stabilizing DNA, RNA, phospholipids or proteins by binding them and increasing their melting temperatures (Takami *et al.,* 2004). Additionally, *G. thermodenitrificans* NG802 is predicted to metabolize polyamines, from which it obtains carbon and nitrogen nutrients (Feng *et al.,* 2007).

Plasmid pGeo1 of *P. thermoglucosidasius* G08C001 encodes an amidase (G08C001pGeo1_CDS1; PRK07235; Bitscore: 846.6; E-value: 0) clustered with a cobalt-containing nitrile hydratase enzyme NHase (Alpha chain: G08C001pGeo1_CDS4; pfam02979; Bitscore: 314.14; E-value:1.30e^{-}110; 219 aa) (Beta chain: G08C001pGeo1_CDS5; pfam02211; Bitscore: 279.49; E-value: 4.38e^{-}96; 229 aa) with an adjacent cobalt containing transmembrane accessory protein (G08C001pGeo1_CDS3; TIGR03889; Bitscore: 81.70; Evalue: 8.05e^{-}22; 122 aa) that is responsible for the maturation and activation of the enzyme (Mylerova and Martinkova 2003). NHase is responsible for catalyzing the hydration of nitriles to amides which are further metabolized by an amidase, to their corresponding acids and ammonia, that is used as a source of nitrogen and carbon by a wide range of bacterial taxa (Mylerova and Martinkova 2003). This reaction is of interest in various industries, where the use of this enzyme has been reported in the production of acrylamide from acrylonitrile and the production of pharmaceuticals, agrochemicals and herbicides (Mylerova and Martinkova 2003; Kamble *et al.,* 2013). Furthermore, bacterial NHase-mediated nitrile metabolism can be used to replace chemically driven industrial processes for the bioremediation of nitrile-polluted wastewater (Brady *et al.,* 2004; Kamble *et al.,* 2013).

The pGeo1 plasmid of *Geobacillus* genomospecies 3 JF8 harbours three co-localized genes coding for a ring hydroxylating dioxygenase enzyme BphA (JF8pGeo1_CDS3-4; CD08881; Bitscore: 270.27; E-value: 1.81e^{-}89; 616 aa), a dihydrodiol dehydrogenase BphB (JF8pGeo1_CDS5-6; CD05348: Bitscore: 383.24; E-value: 9.17e^{-}136; 347 aa) and a 2,3catechol dioxygenase (JF8pGeo1_CDS7; TIGR03211; Bitscore: 267.03; E-value: 1.29e^{-}54; 315 aa). These enzymes play a role in the degradation of aromatic containing compounds such as Polycyclic Aromatic Hydrocarbon (PAHs) and PolyChlorinated biphenyls (PCBs). PAHs exist naturally in petroleum, coal, tar or as gases but are also are widely used in the production of plastics, detergents, pesticides or dyes and thus are common environmental pollutants (Wallace, 1989; van Schie and Young, 2007). PCBs are not natural compounds and contribute to environmental pollution from their use in different industries (Holoubek *et al.,* 2001).

Several microorganisms have demonstrated the capacity to degrade these compounds and concomitantly obtain carbon, nitrogen and energy (Wallace, 1989; van Schie and Young, 2007). The *Geobacillus* genomospecies 3 JF8, plasmid-encoded BphA and BphB enzymes are predicted to be involved in the degradation of PCBs and share 74.10% and 84.44% amino acid identity with respective orthologous enzymes in *Paenibacillus validus* NBRC 15382 (WP_127610294 and WP_127610296) (Abraham, 2002; Uhlik *et al,* 2009). The 2,3-catechol dioxygenase enzyme is mostly associated with degradation of phenols and PAHs such as napthelene, toluene or xylene, however, this enzyme shares 83.82% amino acid identity with the 2,3- dihydroxybiphenyl dioxygenase BphC- like enzyme of *P. validus* NBRC 15382 (Locus: WP_127610298) (Abraham, 2002; Uhlik *et al.,* 2009). During the degradation of PCB's BphA introduces oxygen at the 2,3-position of a PCB ring structure to form a dihydrodiol, which is then dehydrogenated by BphB to 2,3-dihydroxybiphenyl. BphC then cleaves the 2,3-dihydroxybiphenyl at the 1,2position forming a 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate which is further hydrolysed to its corresponding chlorobenzoic acid by BphD a hydrolase enzyme (Furukawa and Miyazaki, 1986). No BphD-like enzyme was identified on the pGeo1 plasmid of *Geobacillus* genomosp. 3 JF8, however the presence of enzymes BphA, BphB and a possible BphC can suggest the role of this plasmid in PCB degradation.

While no orthologues of the enzymes BphA and BphB were found to be encoded on any of the other geobacilli pGeo1 plasmids, five pGeo1 plasmids of *P. thermoglucosidasius* (DSM2542pGeo1_CDS63, DSM2543pGeo1_CDS51, NCIMB11955pGeo1_CDS68, DSM6285pGeo1_CDS70 and G08C001pGeo1_CDS72) encode a 2,3-catechol dioxygenase

enzyme sharing 99.82% average amino acid identity. These enzymes share only 28.87% amino acid identity with the BphC-like enzyme encoded by the plasmid of *Geobacillus* genomosp. 3 JF8, thus they may not be involved in the degradation of PCB's specifically, however, they are found to share 48.09% amino acid identity with the 2,3-catechol dioxygenase encoded by a plasmid of *Pseudomonas putida* GJ31 (AAD05250) (Mars *et al.,* 1999). This enzyme plays a role in the degradation of a variety of aromatic hydrocarbon compounds by carrying out the meta-cleavage of the aromatic ring structure to catechols and its derivatives (He *et al.,* 2016). The presence of this enzyme amongst the geobacilli plasmids can therefore infer that pathways involved in the degradation of aromatic compounds may be encoded by these plasmids.

3.3.5 The geobacilli plasmids codes for five distinct restriction-modification systems and a recombinase of biotechnological value

Restriction modification (R-M) systems are enzyme-based DNA cleaving mechanisms found in a wide range of unicellular organisms (Roberts *et al.,* 2003; Pingoud *et al.,* 2005; Vasu and Nagaraja, 2013). These systems typically comprise a restriction endonuclease and corresponding methyltransferase and are classified into four main types on the basis of their subunit composition, sequence recognition, cleavage position or substrate specificity (Roberts *et al.,* 2003; Pingoud *et al.,* 2005; Vasu and Nagaraja, 2013). Bacteria have evolved R-M systems as defence mechanisms for the cell, protecting them against invading viral or foreign DNA that may pose a threat to the cell (Roberts *et al.,* 2003; Pingoud *et al.,* 2005; Vasu and Nagaraja, 2013). In addition to defence, R-M systems have been associated with plasmid stability within a cell, mediating the post-segregational cell killing of daughter cells that lose a plasmid upon division (Vasu and Nagaraja, 2013). The wide distribution of R-M systems amongst bacteria is largely influenced by HGT, thus the majority of these systems are found to be plasmid encoded (Pingoud *et al.,* 2005; Vasu and Nagaraja, 2013).

Analysis of the protein complement of the geobacilli plasmids showed that six distinct R-M systems are encoded on the plasmids. Two pGeo1 plasmids of *Geobacillus* genomospecies 3 JF8 and *G. jurassicus* DSM 15726 and the pGeo2 plasmid of *G. thermodenitrificans* ID-1 encode proteins that make up a potential Type I R-M system (Murray, 2000; Roberts *et al.,* 2003) (Table 4). Type I R–M systems are dependent on methylase activity and are encoded by three genes, *hsdM* encodes the methyltransferase modification M subunit, *hsdS* codes for the specificity recognition S subunit *hsdR* encodes for the RE subunit (Murray, 2000; Roberts *et al*, 2003). Together the M and S subunits form a methyltransferase, that modifies the adenine residues within the target DNA sequence, which are generally bipartite sequences such as 5'-

AAC(N6)GTGC–3' (Kan *et al.,* 1979; Roberts *et al.,* 2003). The R subunit recognizes the methylated site and subsequently initiates endonuclease activity cutting one of the DNA strands at a variable site that is usually distant from the methylase target site (Murray, 2000; Roberts *et al.,* 2003).

Table 3: Type I R-M loci of five plasmids

Genus species	Locus tag	Locus tag	Locus tag
	HsdM	HsdS	HsdR
Geobacillus geneomosp 3 JF8	pGeo1_CDS18	pGeo1_CDS19	pGeo1_CDS20
G. jurassicus DSM 15726	pGeo1_CDS30	pGeo1_CDS31	pGeo1_CDS32
G. thermodenitrificans ID-1	pGeo2_CDS65	pGeo2_CDS64	pGeo2_CDS63
P. toebii DSM 14590	pGeo2_CDS21	-	-
P. caldoxylosilyticus DSM 12041	pGeo2_CDS43	-	-

The genes coding for HsdM (COG0286; Bitscore: 185.31; E-value: 3.54e^{-}56; 522 aa; AAI: 97.75%), HsdS (COG0732; Bitscore: 117.07; E-value: 2.42e^{-}29; 425 aa; AAI: 58.44%) and HsdR (COG0610; Bitscore: 574.74; E-value: 0; 985 aa; AAI: 95.62%) are found co-localised within the geobacilli plasmid sequences (Table 3). The Type I enzymes studied in most detail are those of the EcoKI system of *E. coli* K-12 (Murray, 2000; Loenen *et al.,* 2014). These geobacilli proteins share average amino acid identities of 23.33%, 19.33% and 17.67% with the HsdM, HsdS and HsdR proteins of the EcoKI Type I system of *E. coli* K-12 substr MG1655 (P08957, P05719, P08956) respectively (Hayashi *et al.,* 2006). Apart from the three plasmids encoding all three subunits of the R-M system, two pGeo2 plasmids (*P. caldoxylosilyticus* DSM 12041 and *P. toebii* DSM 14590) are observed to solely encode the HsdM subunit (Table 3).

The pGeo1 plasmid of *G. thermodenitrificans* T12 encodes an orthologue of the restriction enzyme BamHI (CDS67; pfam02923; Bitscore: 126.46; E-value: 5.34e^{-}37; 221 aa), which was first isolated from *Bacillus amyloliquefaciens* strain H (Brooks *et al.,* 1991). This pGeo1 protein shares 69.86% amino acid identity with that of *B. amyloliquefaciens* KCP2 (WP_013353547) (Newman *et al.,* 1994). This enzyme recognizes the target site 5'-GGATCC-3' and carries out endonuclease activity after the first 5'-guanine residue of each DNA strand generating sticky ends. The pGeo1 plasmid is also observed to encode a putative companion methyltransferase enzyme, N4_Mtase (CDS65; pfam01555; Bitscore: 179.93; E-value:

1.64e54; 414 aa). This enzyme is cytosine specific and is reported to carry out the methylation of the internal cytosine residue of the target site which may inhibit the activity of BamHI (Brooks *et al.,* 1991).

The pGeo1 plasmid of *G. stearothermophilus* C1BS50MT1 encodes an orthologue of the *Bacillus pumilus* RE BpuJI (CDS10; pfam11564; 304.77; E-value: 1.32e^{-}101; 468 aa) and a companion methyltransferase Dcm (CDS11; COG0270; Bitscore:265.06; E-value: 3.19e^{-}87; 334 aa). The pGeo1 enzyme shares 28.48% amino acid identity with that of *B. pumilus* RFL1458 (ABN54439) (Sukackaite *et al.,* 2007). This enzyme recognizes the asymmetric sequence 5′-CCCGT-3′ and carries out endonucleolytic cleavage of both DNA strands at multiple sites downstream of the target sequence (Sukackaite *et al.,* 2007). This enzyme is also reported to show a preference for generating blunt DNA ends (Sukackaite *et al.,* 2007). The methyltransferase Dcm, is a cytosine specific enzyme, that recognizes the double-stranded sequence 5'-NGNCC-3', and methylates the internal cytosine residues of both DNA strands (Gomez-Eichelmann *et al.,* 1991; Kumar *et al.,* 1994).

The *P. toebii* WCH70 pGeo4 plasmid encodes an orthologue of RE TdeIII (CDS5; pfam09520; Bitscore: 256,05; E-value: 6,78e^{-}86; 268 aa) and an associated methyltransferase enzyme Dcm (CDS4; COG0270; Bitscore: 265.06; E-value: 3.19e^{-}87; 268 aa). This enzyme was originally isolated from *Treponema denticola* ATCC 35405 (NP_971521) and shares 22.62% amino acid identity with that of the pGeo4 enzyme (Bian and Li, 2011). TdeIII recognizes the DNA site 5'-GGNCC-3', cleaving both strands of DNA within the target site, however the ends generated are not described (Bian and Li, 2011). As described above, the Dcm enzyme encoded in this plasmid is most likely responsible for methylating the cytosine residues within the target sequence of the host DNA, protecting it from cleavage by TdeIII (Gomez-Eichelmann *et al.,* 1991).

A R-M system found on seven geobacilli plasmids (two pGeo1 and five pGeo2 plasmids) (Table 4) comprises the RE AlwI (pfam09491; Bitscore: 546,76; E-value: 1,67e-84; 546 aa; AAI: 100%) and the companion adenine methyltransferase Dam (TIGR00571; Bitscore: 265,78; E-value: 1,45e—84; 317 aa; AAI: 100%). AlwI was originally isolated from *Acinetobacter iwoffii* SH145 (AAL25126) which shares 40.15% amino acid identity with the geobacilli enzymes (Steczkiewicz et al., 2012; Suzuki and Yoshida, 2012).

AlwI is a Type II nicking enzyme that recognizes asymmetric target sequences and carries out asymmetric cleavage of DNA (Xu et al., 2001). It recognizes the target sequence 5′-GGATC-

3′ and cleaves four (5’-3’of the 3’ end) and five (3’-5’ of the 5’ end) bases in the same direction outside the target sequence and thus most likely generates sticky ends (Xu et al., 2001). Endonucleolytic activity of AlwI has is observed to cease in the presence of an adenine methyltransferase, thus this enzyme most likely methylates the adenine residue within the RE’s target site (Xu et al, 2001; Chen et al, 2013).

Table 4: Plasmid loci of AlwI-dam Type II R-M system

Genus species	Strains	Locus tag	Locus tag
		AlwI	Dam
Geobacillus genomosp 1	Et2/3 (pGeo2)	pGeo2_CDS51	pGeo2_CDS52
Geobacillus genomosp 4	NBRC 10184 (pGeo2)	pGeo2_CDS2	pGeo2_CDS3
G. icigianus	C56-T2 (pGeo2)	pGeo2_CDS32	pGeo2_CDS31
G. stearothermophilus	Strain 15 (pGeo2)	pGeo2_CDS59	pGeo2_CDS60
G. thermoleovorans	HTA426 (pGeo1)	pGeo1_CDS13	pGeo1_CDS12
	N7 (pGeo1)	pGeo1_CDS25	pGeo1_CDS24
	SGAIR0734 (pGeo2)	pGeo2_CDS45	pGeo2_CDS44

A further eleven pGeo1 plasmids are observed to encode an enzyme belonging to the AlwI enzyme superfamily (pfam09491; Bitscore: 58,39; E-value: 5,47e^{-}09; 543 aa; AAI: 98.56%) and a co-localised methyltransferase 5-cytosine Mtase (pfam00145; Bitscore: 270.68; E-value: 1.58e^{-}76; 370 aa; AAI: 100%) (Table 5). These enzymes share only 19.21% sequence identity with that of *A. iwoffi* and the seven geobacilli AlwI-dam R-M systems but are observed to share 60.11% sequence identity with the BbvI RE enzyme of *Brevibacillus brevis* (ADQ20502). BbvI is described as an isoschizomer of the AlwXI enzyme isolated from *A. iwoffi* X (Roberts, 1990). Isoschizomers are restriction endonuclease enzymes that recognize the same target site and cleave them in the similar manner (Roberts, 1990). BbvI/AlwXI are also Type II enzymes and are reported to recognize the target sequence 5’-GCACG-3’, cleaving six bases from the target site in the same direction in both the 5’-3’and 3’-5’ strands, generating blunt ends (Roberts, 1990; Castel *et al.,* 2001). The companion methyltransferase is cytosine specific and found in numerous R-M systems (Lauster *et al.,* 1989). The BbvI enzyme has been found to display sensitivity to this methyltransferase, thus it is most likely responsible for methylating the cytosine residues of the target site, inhibiting the RE activity (Castel *et al.,* 2001).

Table 5: Plasmid loci of AlwI (BbvI/AlwXI) C5-Mtase R-M system

Genus species	Strains	Locus tag	Locus tag
		AlwXI/BbvI	5C- Mtase
Geobacillus genomosp. 5	1017	pGeo1_CDS11	pGeo1_CDS12
G. icigianus	PSS2	pGeo1_CDS22	pGeo1_CDS21
	C56-T2	pGeo1_CDS55	pGeo1_CDS54
G. jurassicus	DSM 15726	pGeo1_CDS4	pGeo1_CDS5
G. stearothermophilus	12AMOR1	pGeo1_CDS5	pGeo1_CDS4
	DSM 15325	pGeo1_CDS31	pGeo1_CDS32
	LC300	pGeo1_CDS26	pGeo1_CDS27
G. thermoleovorans	A8	pGeo1_CSD28	pGeo1_CDS29
	MAS1	pGeo1_CDS10	pGeo1_CDS11
	Y412MC52	pGeo1_CDS22	pGeo1_CDS23
	Y412MC61	pGeo1_CDS13	pGeo1_CDS12

A prevalent feature among the geobacilli plasmids is a gene that encodes an orthologue of the XerC tyrosine recombinase (TIGR02224; Bitscore: 186.27; E-value: 2.03e^{-}57; 297 aa; AAI: 92.48%) (Table 6), with orthologues observed in 41 of the plasmids (Table 6). XerC was originally identified in *E. coli* (WP_017383588) with which the geobacilli recombinases share 28.10% average amino acid identity (Yates *et al.,* 2003). In *E. coli*, XerC is found to imitate topoisomerase activity, where it relaxes the supercoiled DNA structure and generates a nick at the target DNA sequence, carrying out recombination events at the *dif* region of the chromosome and a range of plasmid-borne sites (Cornet et al., 1997; Yates et al., 2003).

In addition to the XerC-like recombinase, twenty-six of the plasmids (Table 6) encode an uncharacterized recombinase (CD01192; Bitscore: 195.20; E-value: 1.70e^{-}64; 189 aa; av AAI: 91.98%) sharing 56.84% sequence identity to a lambda (Int-like), tyrosine recombinase of the phage integrase family from *Domibacillus enclensis* (SIR57214) (Nunes-Düby *et al.,* 1998).

Table 6: Loci of the Tyrosine recombinase in 41 plasmids

Genus species	Strains	Locus tag - XerC	Locus tag- Uncharacterized tyrosine recombinase
Geobacillus genomosp. 1	Et2/3	pGeo2_CDS47	-
Geobacillus genomosp. 3	JF8 (pGeo1)	pGeo1_CDS15	pGeo1_CDS31
Geobacillus genomosp. 4	NBRC101842 (pGeo2)	pGeo2_CDS56	-
Geobacillus genomosp. 5	1017 (pGeo1)	pGeo1_CDS6	pGeo1_CDS21
G. icigianus	PSS2 (pGeo1)	pGeo1_CDS29	pGeo1_CDS12
	C56-T2 (pGeo1)	pGeo1_CDS29	pGeo2_CDS45
	C56-T2 (pGeo2)	pGeo2_CDS36	-
G. stearothermophilus	C1BS50MT1 (pGeo1)	pGeo1_CDS8	pGeo1_CDS23
	12AMOR1 (pGeo1)	pGeo1_CDS11	pGeo1_CDS38
	DSM 15325 (pGeo1)	pGeo1_CDS54	pGeo1_CDS41
	LC300 (pGeo1)	pGeo1_CDS18	pGeo1_CDS41
	Strain 15 (pGeo2)	pGeo2_CDS55	-
G. thermodenitrificans	JSC_T9a (pGeo1)	pGeo1_CDS3	pGeo1_CDS5
	A8 (pGeo1)	pGeo1_CDS24	pGeo1_CDS6
	T12 (pGeo1)	pGeo1_CDS30	pGeo1_CDS16
	T12 (pGeo2)	pGeo2_CDS57	-
	NG80-2 (pGeo2)	pGeo2_CDS59	-
	G11MC16 (pGeo2)	pGep2_CDS51	-
	ID-1 (pGeo2)	pGeo2_CDS74	-
	PA-3 (pGeo2)	pGeo2_CDS13	-
G. thermoleovorans	DSM 405 (pGeo1)	pGeo1_CDS18	pGeo1_CDS35
	DSM 5366 (pGeo1)	pGeo1_CDS34	pGeo1_CDS15
	HTA426 (pGeo1)	pGeo1_CDS18	pGeo1_CDS57
	Y412MC52 (pGeo1)	pGeo1_CDS5	pGeo1_CDS49
	Y412MC61 (pGeo1)	pGeo1_CDS30	pGeo1_CDS38
	N7 (pGeo1)	pGeo1_CDS29	pGeo1_CDS16
	SGAir0734 (pGeo2)	pGeo2_CDS49	-
Parageobacillus genomosp. 2	W-2 (pGeo1)	pGeo1_CDS65	-
P. thermoglucosidasius	C56_YS93 (pGeo1)	pGeo1_CDS103	pGeo1_CDS45
	DSM 6285 (pGeo1)	pGeo1_CDS102	pGeo1_CDS63
	G08C001 (pGeo1)	pGeo1_CDS102	pGeo1_CDS63
	DSM 21625 (pGeo1)	pGeo1_CDS88	pGeo1_CDS66

	DSM 2543 (pGeo1)	pGeo1_CDS1	pGeo1_CDS59
	DSM 2543 (pGeo2)	pGeo2_CDS40	-
	NCIMB 11955 (pGeo1)	pGeo1_CDS117	pGeo1_CDS60
	NCIMB 11955 (pGeo2)	pGeo2_CDS40	-
	DSM 2542 (pGeo2)	pGeo2_CDS40	-
P. toebii	B4110 (pGeo1)	pGeo1_CDS29	pGeo1_CDS64
	44C (pGeo1)	pGeo1_CDS45	pGeo1_CDS20
	WCH70 (pGeo1)	pGeo1_CDS11	pGeo1_CDS25
	DSM 14590 (pGeo2)	pGeo2_CDS11	-

In general tyrosine recombinases use a conserved nucleophilic, tyrosine amino acid residue to carry out a site-specific attack of target DNA, in order to perform rearrangements of DNA segments. The enzyme recognizes and binds to short DNA sequences which are cleaved from the DNA backbone, and then exchanged with another piece of excised DNA with the helices being re-joined (Gaj *et al.,* 2014). These enzymes can control and coordinate a number of diverse functions, which include the integration or excision of plasmid or viral genomes from chromosomal DNA, gene expression, chromosome/plasmid partitioning, control of plasmid copy number as well as the transposition of mobile genetic elements (Cornet *et al.,* 1997; Nunes-Düby *et al.,* 1998; Yates *et al.,* 2003).

Both REs and recombinases have been utilized extensively in recombinant technology, vector design, DNA mapping and gene editing and are therefore of significant biotechnological value (Roberts, 2005; Gaj *et al.,* 2014). In particular, there has been extensive interest in both types of enzymes from thermophilic bacteria as they are able to withstand temperatures that typically denature those enzymes of mesophilic bacteria (Buchholz *et al.,* 1996; Sharma *et al.,* 2013). Previously characterized geobacilli restriction enzymes such as GeoICI and BstVI, have been reported to withstand and have increased activity at temperatures ranging from 55°C to 75°C with optimum activity at 65°C, which were greater than temperature ranges of enzymes from other thermophilic bacteria such as members of the genera *Lactobacillus* and *Thermobifida* (Runswick and Harris, 1978; Zebrowska *et al.,* 2016). In addition to their inferred thermostability, several of the RE's encoded on the geobacilli plasmids were found to share low, restricted orthology to characterized restriction enzymes, as such, these plasmids may serve as a new source of restriction modification systems for use in recombinant technology.

3.3.6 Isolation and curing of plasmid DNA

Plasmid curing of bacteria refers to the use of chemical reagents that interfere with plasmid replication or procedures such as increased temperatures to remove plasmids from their host cells (Trevors, 1986). This is often desirable when determining the functional roles of a plasmid within bacteria as it allows for comparisons to be made between cells that contain their plasmids and cells that have had their plasmids removed, which therefore allows for plasmid encoded characteristics to be observed on a phenotypic level (Trevors, 1986). The last portion of this investigation aimed at extracting plasmid DNA from five selected *Geobacillus* and *Parageobacillus* strains (Table 7), with the main aim of further curing these strains of their plasmids to elucidate differences in phenotypic characteristics between wild-type strains and plasmid cured strains.

Table 7: Selected strains for plasmid DNA extraction

Sample names	Type Strains	Number of plasmids	Plasmid type
T1	*P. thermoglucosidasius* DSM 2542^T	2	pGeo1; pGeo2
T2	*P. toebii* DSM 14590^T	1	pGeo2
T3	*G. thermoleovorans* DSM 5366^T	1	pGeo1
T4	*G. stearothermophilus* ATCC 12980^T	1	pGeo1
T5	*G. kaustophilus* DSM 7263^T	0	-
T6	*P. caldoxylosilyticus* DSM 12041^T	1	pGeo2

To ascertain the presence of plasmids amongst the five geobacilli strains two DNA extraction procedures were performed to isolate plasmid DNA. DNA extraction with the ABIOpure kit yielded single bands for all samples, which is expected to represent chromosomal DNA (Figure 3). No smaller bands, representative of plasmid DNA could be observed. The second extraction carried out using the protocol of Kado and Liu (1981), observed bands for two strains, namely *P. toebii* DSM 14590^T and *G. kaustophilus* DSM 7263^T (Figure 3). The latter strain is suggested not to contain any plasmids as previous literature/studies have not associated plasmids with this strain, however, it cannot be excluded that a plasmid may not be integrated into the chromosomal assembly for the latter strain.

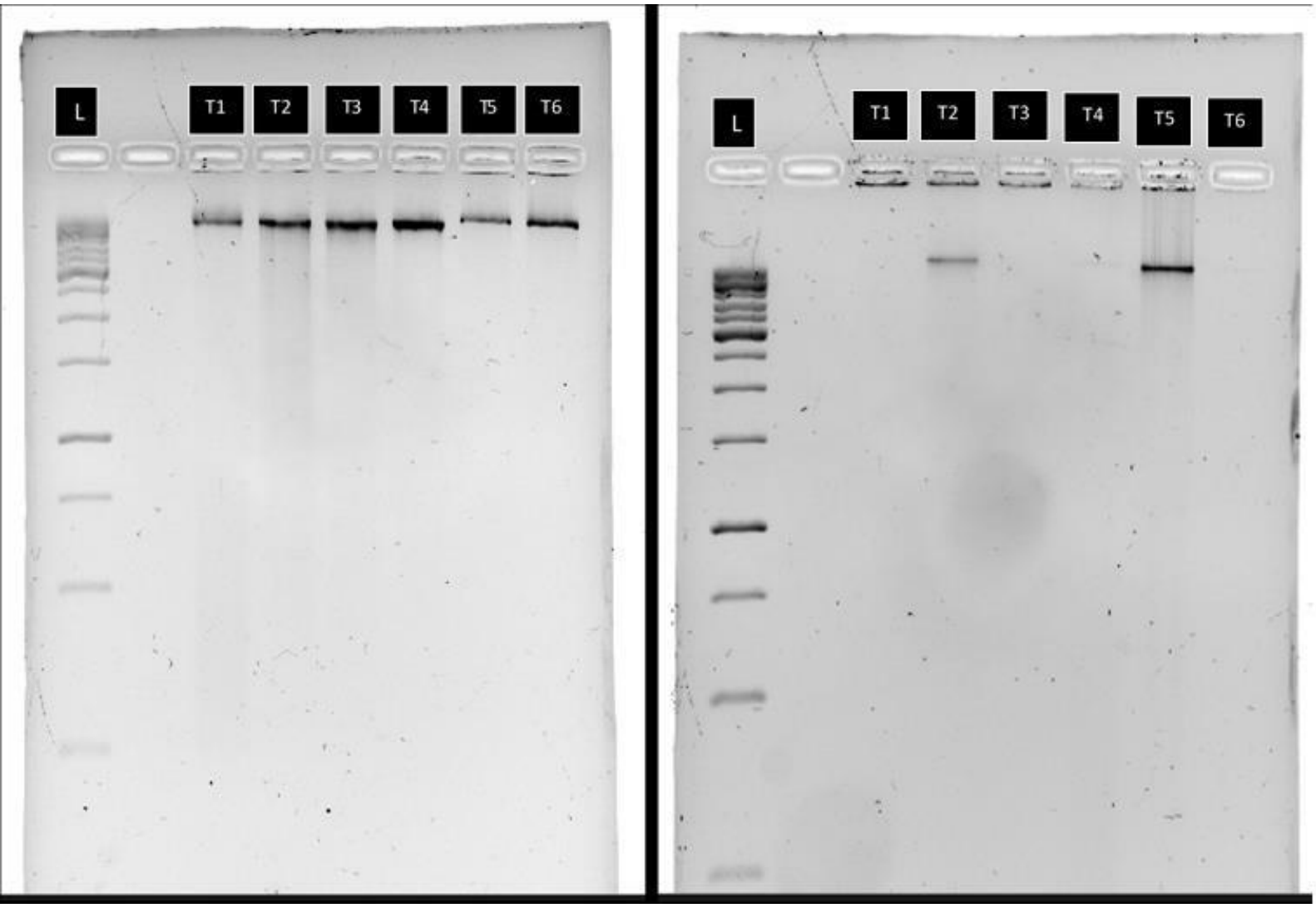


Figure 3: Agarose gel of DNA samples extracted from the ABIOpure extraction kit (left) and plasmid extraction approach of Kado and Liu (1981) (right). The strains in order are, T1 *P. thermoglucosidasius* DSM 2542^{T}, T2 *P. toebii* DSM 14590^{T}, T3 *G. thermoleovorans* DSM 5366^{T}, T4 *G. stearothermophilus* ATCC 12980^{T}, T5 *G. kaustophilus* DSM 7263^{T} and T6 *P. caldoxylosilyticus* DSM 12041^{T}. L denotes the 1kb Generuler molecular weight marker.

To validate whether the two bands represented plasmid DNA, PCR amplification of the 16S rRNA gene was performed. The 16S rRNA genes are expected to occur solely on the chromosome, so no amplicons were expected. However, amplicons of the expected size (~1,500 bp) were obtained for both samples (Figure 4), suggesting that, if plasmid DNA is present, it is contaminated with chromosomal DNA.

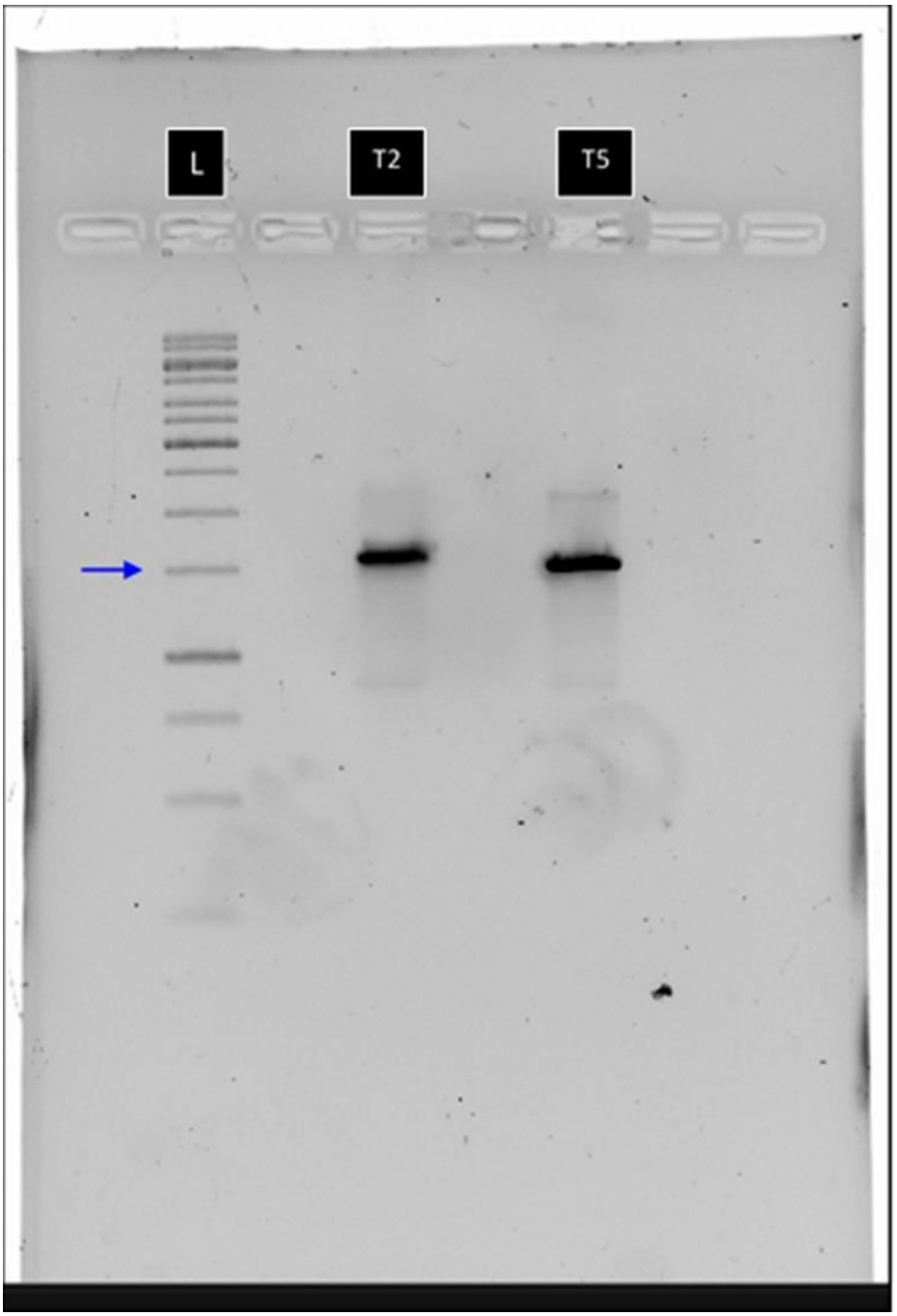


Figure 4: Agarose gel 16S rRNA PCR amplicon samples. The strains are in the order T2 *P. toebii* DSM 14590^{T} and T5 *G. kaustophilus* DSM 7263^{T}. L denotes the molecular weight marker (1kb ruler), and the blue arrow indicates the 1 500 base pair marker.

Potential reasons why plasmid DNA could not be extracted include the possible low copy number nature of the plasmids, as well as inefficacy of the plasmid DNA extraction protocols. The rep_3 proteins, identified in the pGeo1 plasmids in Chapter 2, are associated with a low plasmid copy number in the cases of the plasmids pSC101 (*E. coli*), pD36-3, pD36-4 (*A. baumannii* spp.) and pSK639 (*S. epidermidis*). In these plasmids the rep_3 proteins were observed to have stringent control of the copy number maintaining copies of only 3-10 plasmids per cell (Kwong *et al.*, 2017; Bertini *et al.*, 2010; Fueki and Yamaguchi, 2001; Thompson *et al.*, 2018). Thus, *P. thermoglucosidasius* DSM 2542^{T}, *G. thermoleovorans* DSM 5366^{T} and *G. stearothermophilus* ATCC 12980^{T} may harbour too few pGeo1 plasmids for effective extraction. Similarly, the pGeo2 plasmids (*P. thermoglucosidasius* DSM 2542^{T}, *P. toebii* DSM 14590^{T} and *P. caldoxylosilyticus* DSM 12041^{T}), encode a ParM-like plasmid partitioning protein. This is a common feature of low copy number plasmids and ensures that their few copies are equally distributed amongst all daughter cells, to avoid plasmid loss (LeBrasseur, 2007).

In the original protocol paper, Kado and Liu (1981) evaluated plasmid extraction on members of the genera *Agrobacterium, Rhizobium, Escherichia, Salmonella, Erwinia, Pseudomonas* and *Xanthomonas*, all of which are Gram-negative bacteria (Kado and Liu, 1981). Alkaline lysis extractions work well for Gram-negative bacteria, whereas in some cases Gram-positive cells are inadequately lysed due to their thicker cell wall structures yielding low concentrations of plasmid DNA (Wada *et al.*, 2012). As such, this particular protocol may not be suited for the extraction of plasmid DNA from *Geobacillus* and *Parageobacillus* spp.

Although a suitable plasmid DNA extraction protocol was not available, we attempted plasmid curing on *P. toebii* DSM 14590^{T} using the protocol of Mizuno *et al.* (2017). This approach makes use of acridine orange, an intercalating agent that inhibits plasmid replication, ultimately preventing the vertical transmission of plasmids to daughter cells during cell division (Spengler *et al.*, 2006).

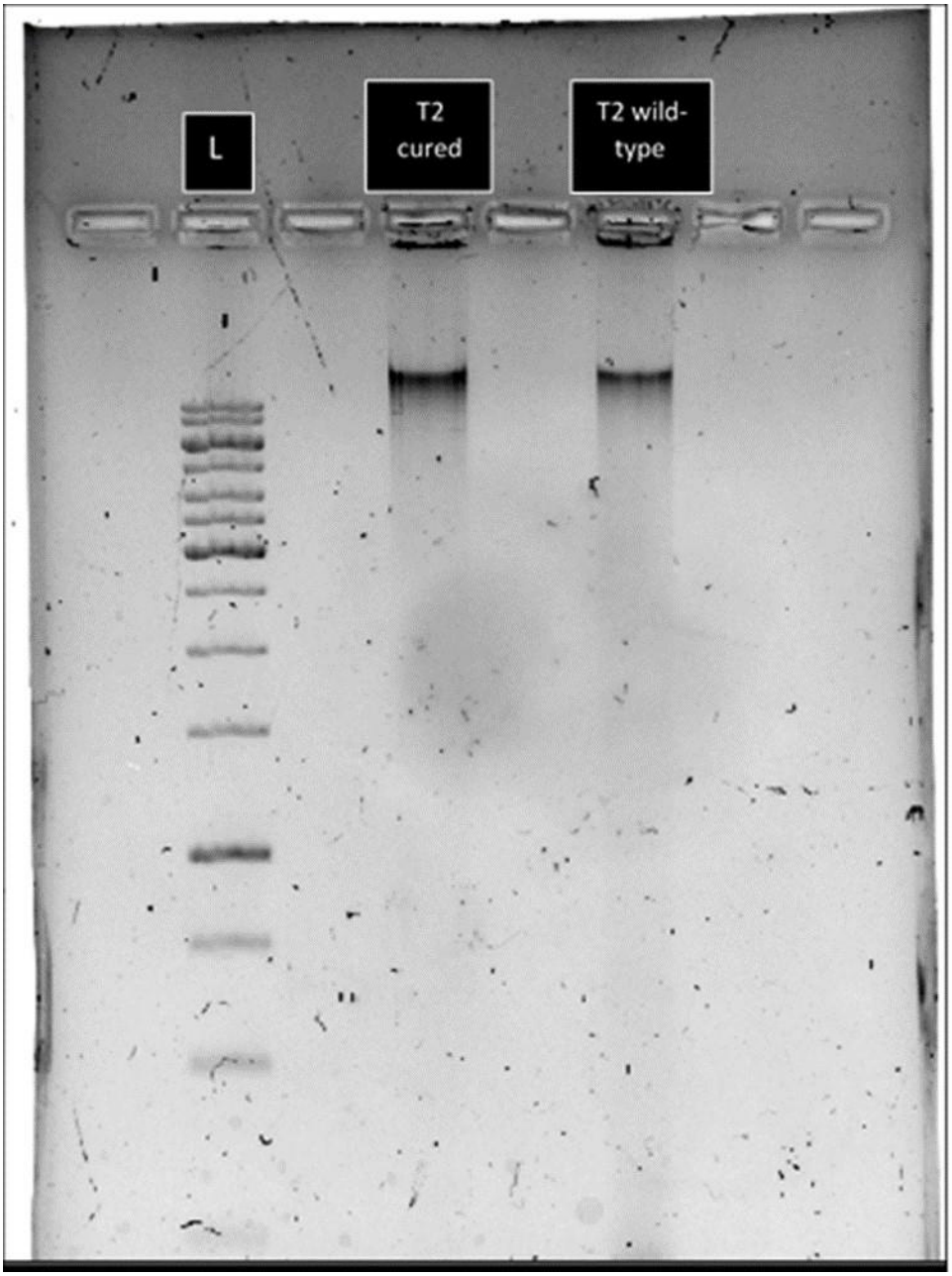


Figure 5: Agarose gel of DNA samples extracted from cured and wild-type strain T2 *P. toebii* DSM 14590^{T} using the ABIOpure extraction kit. L denotes the molecular weight marker (1kb ruler).

Agarose gel electrophoresis of DNA extracts (as per Kado and Liu, 1981) showed a single band occurring in both the "plasmid cured" and wild-type *P. toebii* DSM 14590^{T}, with no distinct plasmid band observed for the wild-type strain (Figure 5). As such, the efficacy of plasmid curing could not be determined. Future work should consider optimizing plasmid curing protocol as well as evaluating different curing protocols to obtain plasmid-cured geobacilli strains.

3.4 Conclusions

The geobacilli plasmids are highly versatile entities, with few proteins conserved among the plasmids of same plasmid group and no proteins conserved amongst all 56 plasmids. Furthermore, the geobacilli plasmids serve as a backbone for an extensive set of cargo genes. Functional annotation of approximately 1,573 of the plasmid cargo proteins shows that these plasmids play diverse roles in the metabolic capabilities, cellular processes, signalling and information storage and processing of their host cells. The plasmids encode a range of phenotypes that could provide members of the genera with a selective advantage over other organisms allowing them to establish themselves in a wide range of ecological niches and make use of a broad range of nutrients. Furthermore, they encode antibiotics and antibiotic resistance mechanisms that could allow them to outcompete other microorganisms inhabiting the same ecological niche.

The geobacilli plasmids encode a number of proteins which could be employed in a broad range of biotechnological applications. Plasmid-encoded antimicrobial compounds, such as bacteriocins or Geobacillin II, may be of value in the production of foods which employ high temperatures during preparation. Furthermore, the diverse range of DNA modifying enzymes (restriction endonucleases and recombinases) encoded on the geobacilli plasmids can be considered as thermostable alternatives for use in genetic engineering and recombinant technology. There is also evidence of enzymes involved in catabolic pathways that could potentially be utilized in the development of genetic tools for the bioremediation of PAH, PCB and nitrile contaminated environments. As members of the genera *Geobacillus* and *Parageobacillus* represent thermophiles, these proteins have the added advantage of being thermostable entities that may have different substrate specificities, are able withstand increased bioprocessing temperatures and have higher catalytic capacities or product yield compared to their mesophilic counterparts. Thus, overall the geobacilli plasmids represent a rich source of biotechnological potential.

Finally, plasmid DNA isolation from *Geobacillus* and *Parageobacillus* strains was attempted, but this was unsuccessful and the curing efficacy of acridine orange could not be assessed. Further testing of protocols which are specific to low copy number plasmid and work effectively in Gram-positive bacteria such as the geobacilli should be undertaken. In addition to this several plasmid curing approaches should also be considered to obtain plasmid-cured strains.

3.5 References

Abraham, W-R. (2002) Microbial degradation of polychlorinated biphenyls (PCBs) in the environment, *Progress in Industrial Microbiology*, 36, 29-67

Bauer, A. M., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D. and Zheng, C. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Research*,43 (D1), D222-D226

Bertini, A., Poirel, L., Mugnier, P. D., Villa, L., Nordmann, P. and Carattoli, A. (2010) Characterization and PCR-based replicon typing of resistance plasmids in *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*, 54(10), 4168-4177

Bian, J. and Li, C. (2011) Disruption of a type II endonuclease (TDE0911) enables Treponema denticola ATCC 35405 to accept an unmethylated shuttle vector. *Applied Environmental Microbiology*, 77, 4573–4578

Bisht, S., Pandey, P., Bhargava, B., Sharma, S., Kumar, V., & Sharma, K. D. (2015). Bioremediation of polyaromatic hydrocarbons (PAHs) using rhizosphere technology. *Brazilian Journal of Microbiology*, 46(1)

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H. and Weber, T. (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, 47(W1), W81-W87

Brady, D., Beeton, A., Zeevaart, J., Kgaje, C., van Rantwijk, F. and Sheldon, R. A. (2004) Characterization of nitrilase and nitrile hydratase biocatalytic systems. *Applied Microbiological Biotechnology*, 4, 76–85

Brooks, J. E., Blumenthal, R. M., and Gingeras, T. R. (1983) The isolation and characterization of the Escherichia coli DNA adenine methylase (dam) gene. *Nucleic Acids Research*, 11(3), 837–851

Brooks, J. E., Brenner, J. S., Heiter, D. F., Siber, K. R., Sznyter, L. A., Jager-Quinton, T., Moran, L. S., Slatko, B. E., Wilson, G. G. and Nwankwo, D. O. (1989) Cloning the BamHI restriction modification system. *Nucleic Acids Research*, 17(3), 979-997

Brooks, J. E., Nathan, P. D., Landry, D., Sznyter, L. A., Waite-Rees, P., Ives, C. L, Moran, L. S., Slatko, B. E. and Benner, J. S. (1991) Characterization of the cloned BamHI restriction modification system: its nucleotide sequence, properties of the methylase, and expression in heterologous hosts. *Nucleic Acids Research*, 19(4), 841–850

Bruins, M. E., Janssen, A. E. M. and Boom, M. (2001) Thermozymes and their applications. *Applied Biochemistry and Biotechnology*, 9, 155-186

Brunke, P., Stener, O., Bailey, J. E. and Minas, W. (2001) Heterologous expression of the naphthocyclinone hydroxylase gene from Streptomyces arenae for production of novel hybrid polyketides. *Antonie Van Leeuwenhoek*, 79(3), 235-245

Burrus, V. and Waldor, M. K. (2004) Shaping bacterial genomes with integrative and conjugative elements., *Research in Microbiology*, 155, 376–386

Cameron, D. E., Bashor, C. J. and Collins, J. J. (2014) A brief history of synthetic biology. *Nature Reviews Microbiology*, 12, 381-390

López Castel, A., Nakamori, M., Thornton, C. A. and Pearson, C. E. (2011) Identification of restriction endonucleases sensitive to 5-cytosine methylation at non-CpG sites, including expanded (CAG)n/(CTG)n repeats. *Epigenetics*, 6(4), 416–420

Chattopadhyay, M.K., Tabor, C.W. and Tabor, H. (2003) Polyamines protect Escherichia coli cells from the toxic effect of oxygen. *Proceedings of the National Academy of Sciences of the United States of America* USA, 100, 2261-2265

Chen, F. and Zhao, Y. (2013) Methylation-blocked enzymatic recycling amplification for highly sensitive fluorescence sensing of DNA methyltransferase activity. *The Analyst*, 7(138) 284-289

Clancy, J., Petitpas, J., Dib-Hajj, F., Yuan, W., Cronan, M., Kamath, A.V., Bergeron J. and Retsema, J.A. (1996) Molecular cloning and functional analysis of a novel macrolideresistance determinant, mefA, from *Streptococcus pyogenes. Molecular Microbiology*, 22, 867879

Condit, R. and Levin, B. R. (1988) The evolution of plasmids carrying multiple resistance genes: The role of segregation, transposition and homologous recombination. *The American Naturalist*, 135(4), 573-596

Cornet, F., Hallet, B. and Sherratt, D. J. (1997) Xer Recombination in *Escherichia coli*. *Journal of Biological Chemistry*, 272. 21927-21931

Davidson, A. L., Dassa, E., Orelle, C., and Chen, J. (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews: MMBR*, 72(2), 317–364

De Maayer, P., Chan, W-Y., Blom, J., Venter, S. N., Duffy, B., Smits, T. H. M. and Coutinho, T. A. (2012) The large universal Pantoea plasmid LPP-1 plays a major role in biological and ecological diversification, *BMC Genomics*, 13(625), 1471-2164 del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M., and Diaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiology and Molecular Biology Reviews*. 62, 434–464

De Souza, P. M., and de Oliveira Magalhães P. (2010). Application of microbial α-amylase in industry: A review. *Brazilian Journal of Microbiology*, 41(4), 850-861.

Emms, D. M. and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 157(16), 1

Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., Tang, Y., Liu, X., Han, W., Peng, X., Liu, R. and Wang, L. (2007) Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proceedings of the National Academy of Sciences of the United States of America*, 104(13), 5602-5607

Fitzpatrick, A., Llabrés, S., Neuberger, A., Blaza, J. N., Bai, X. C., Okada, U., Murakami, S., van Veen, H. W., Zachariae, U., Scheres, S., Luisi, B. F. and Du, D. (2017) Structure of the MacAB-TolC ABC-type tripartite multidrug efflux pump. *Nature Microbiology*, 2, 17070

Fong, D. H., Lemke, C. T., Hwang, J., Xiong, B. and Berghuis, A. M. (2010) Structure of the Antibiotic Resistance Factor Spectinomycin Phosphotransferase from *Legionella pneumophila. The Journal of Biological Chemistry,* 285, 9545-9555

Fueki, T. and Yamaguchi, K. (2001) The structure and function of the replication initiator protein (Rep) of pSC101: an analysis based on a novel positive-selection system for the replication-deficient mutants. *Journal of Biochemistry,* 130(3), 399-405

Furukawa, K. and Miyazaki, T. (1986) Cloning of a Gene Cluster Encoding Biphenyl and Chlorobiphenyl Degradation in *Pseudomonas pseudoalcaligenes. Journal of Bacteriology*, 166(2), 392-398

Garg, N., Tang, W., Goto, Y., Nair, S. K. and van der Dork, W. A. (2012) Lantibiotics from *Geobacillus thermodenitrificans. Proceedings of the National Academy of Sciences of the United States of America*, 109(14), 5241-5246

Garg, N., Oman, T. J., Andrew Wang, T. S., De Gonzalo, C. V., Walker, S. and van der Donk, W. A. (2014). Mode of action and structure-activity relationship studies of geobacillin I. *The Journal of Antibiotics*, 67(1), 133–136

Garg, N., Goto, Y., Chen, T., and van der Donk, W. A. (2016) Characterization of the stereochemical configuration of lanthionines formed by the lanthipeptide synthetase GeoM. *Biopolymers*, 106(6), 834–842

Gevrekci, A. O. (2017) The roles of polyamines in microorganisms. *World Journal of Microbiology and Biotechnology*, 33(11), 204

Gull, M. and El-Baz, S. (2018) Introductory Chapter: Preface to Plasmids. *InTechOpen* Peerreviewed chapter, 130(3), 399-405

Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B. L., Mori, H. and Horiuchi, T. (2006) Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Molecular Systems Biology*, 2006(2), 0007

Hanna, A., Berg, M., Stout, V., and Razatos, A. (2003) Role of capsular colanic acid in adhesion of uropathogenic *Escherichia coli*. *Applied and Environmental Microbiology*, 69(8), 4474–4481

He, P., Li., L., Liu, J., Bai, Y. and Fang, X. (2016) Diversity and distribution of catechol 2, 3dioxygenase genes in surface sediments of the Bohai Sea. *FEMS Microbiology Letters*, 363(10), fnw086

Huerta-Cepas J., Szklarczyk D., Forslund K., Cook H., Heller D., Walter M. C., Rattei T., Mende D. R., Sunagawa S., Kuhn M., Jensen L. J., von Mering C., and Bork, P. (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research,* 69(8), 4474–4481

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, A. F., Selengut, J.D., Sigrist, C.J., Thimma, M., Thomas, P.D., Valentin, F., Wilson, D., Wu,C.H. and Yeats, C. (2009) InterPro: the integrative protein signature database. *Nucleic Acid Research*, 37(Database issue), D211-215

Jackson, R. W., Vinatzer, B., Arnold, D. L., Dorus, S. and Murillo, J. (2011) The influence of the accessory genome on bacterial pathogen evolution. *Mobile Genetic Elements*, 1(1), 55–65

Johnson, C. M. and Grossman, A. D. (2015) Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annual Review of Genetics*, 49, 577–601

Kado, C. I. (1989) Origin and evolution of plasmids. *Antonie Van Leeuwenhoek*, 73(1), 117126

Kamble, A. L., Banoth, L., Meena, V. S., Singh, A., Chisti, Y., & Banerjee, U. C. (2013). Nitrile hydratase of Rhodococcus erythropolis: characterization of the enzyme and the use of whole cells for biotransformation of nitriles. *3 Biotech*, 3(4), 319–330

Kan, N. C., Lautenberger, J. A., Edgell, M. N. and Hutchison, C. A. (1979) The nucleotide sequence recognized by the *Escherichia coli K12* restriction and modification enzymes. *Journal of Bacteriology*, 30, 191-209

Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27-30

Kristjansson, J. K. (1989) Thermophilic organisms as sources of thermostable enzymes. *Trends in Biotechnology*, 7, 349-353

Kumar A, Munjal A, Khan A, Singh RN (2012) Mega plasmid mediated anthracene degradation by novel strain *Geobacillus stearothermophilus* AAP7919 isolated from oil contaminated soil. *Journal of Pharmacy Research*. 5, 381-384

Kumar, S., Mukherjee, M. M. and Varela, M. F. (2013) Modulation of Bacterial Multidrug Resistance Efflux Pumps of the Major Facilitator Superfamily. *International Journal of Bacteriology*, 2013, 204141

Kwong, S.M., Ramsay J.P., Jensen, S.O. and Firth, N. (2017) Replication of Staphylococcal Resistance Plasmids. *Frontiers in Microbiology*, 8, 2279

Lane, D. J. (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, editors. Nucleic acid techniques in bacterial systematics. New York: Wiley. 115–175.

Las Hera-Vazquez, F. J., Clemente-Jimenez, J. M., Martinez-Rodriguez, S. and Rodriguez-Vico, F. (2008) Optically Pure α-Amino Acids Production by the "Hydantoinase Process. *Recent Patents on Biotechnology*, 2(1) 35-46

LeBrasseur N. (2007) ParM pushes plasmids apart. *The Journal of Cell Biology*, 179(5), 806

Lee, S. W., Mitchell, D. A., Markley, A. L., Hensler, M. E., Gonzalez, D., Wohlrab, A., Dorrestein, P. C., Nizet, V. and Dixon, J. E. (2008) Discovery of a widely distributed toxin biosynthetic gene cluster. *Proceedings of the National Academy of Sciences of the United States of America*, 105(15), 5879-5884

Lee, S., Ward, T. J., Siletzky, R. M. and Kathariou, S. (2012) Two novel type II restriction modification systems occupying genomically equivalent locations on the chromosomes of Listeria monocytogenes strains. *Applied and Environmental Microbiology*, 78(8), 2623–2630

Loenen, W. A., Dryden, D. T., Raleigh, E. A., and Wilson, G. G. (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Research*, 42(1), 20–44

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D. and Darnell, J. (2000) DNA Cloning with Plasmid Vectors. *Molecular Cell Biology*, (4), 1000

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott,

C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchler, G. H., Mullokandov, M., Omelchenko, M. V., Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N., Zheng, C. and Bryant, S. H. (2010) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, 39, D225-9

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Mandal, M., mandal, S. and Das, S. N. (2016) Plasmid Mediated Antibiotic and Heavy Metal Co-Resistance in Bacterial Isolates from Mahananda River Water (Malda, India). *Translational Medicine*, 6(4), 185

Mars, A. E., Kingma, J., Kaschabek, S. R., Reineke, W. and Janssen, D. B. (1999) Conversion of 3-chlorocatechol by various catechol 2,3-dioxygenases and sequence analysis of the chlorocatechol dioxygenase region of *Pseudomonas putida* GJ31. *Journal of Bacteriology*, 181(4), 1309-1318

Mirkovic, N., Polovic, N., Vukotic, G., Jovcic, B., Miljkovic, M., Radulovic, Z., Diep, D. B. and Kojic, M. (2016) *Lactococcus lactis* LMG2081 Produces Two Bacteriocins, a Nonlantibiotic and a Novel Lantibiotic. *Applied and Environmental Microbiology*, 82(8), 2555–2562

Mishima, Y., Momma, K., Hashimoti, Q., Mikami, B. and Murata, K. (2001) Super-channel in bacteria: function and structure of the macromolecule import system mediated by a pitdependent ABC transporter. *FEMS Microbiology Letters*, 204(2), 215-221

Mizuno, T., Ohshiro, T. and Suzuki, H. (2017) Plasmid Curing is a Promising Approach to Improve Thermophiles for Biotechnological Applications: Perspectives in Archaea. *InTechOpen*, 5, 84-99

Molloy, E. M., Cotter, P. D., Hill, C., Mitchell, D. A., and Ross, R. P. (2011) Streptolysin Slike virulence factors: the continuing sagA. *Nature reviews. Microbiology*, 9(9), 670–681

Momma, K., Okamoto, M., Mishima, Y., Mori, S., Hashimoto, W. and Murata, K. (2000) A Novel Bacterial ATP-Binding Cassette Transporter System That Allows Uptake of Macromolecules. *Journal of Bacteriology*, 182(14), 3998-4004

Mylerova, V. and Martinkova, L. (2003) Synthetic applications of nitrile-converting enzymes. *Current Organic Chemistry*, 7, 1–17

Murray, N. E. (2000) Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). *Microbiology and Molecular Biology Reviews: MMBR*, 64(2), 412–434

Newman, M., Strzelecka, T., Dorner, L. F., Schiildkraut, I. and Aggarwal, A. K. (1994) Structure of restriction endonuclease BamHI and its relationship to EcoRI. *Nature*, 368(6472), 660-664

Nunes-Düby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T., & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Research*, 26(2), 391–406

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwrads, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F. and Stevens, R. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, 42(Database issue), D206–D214

Pathak, V. and Navneet, M. (2017) Review on the current status of polymer degradation: a microbial approach. *Bioresour Bioprocess* 4, 15

Pierrat, O. A. and Maxwell, A. (2003) The Action of the Bacterial Toxin Microcin B17. *Journal of Biological Chemistry*, 278, 35016-35023

Pingoud, A., Fuxreiter, M., Pingoud, V. and Wende, W. (2005) Type II restriction endonucleases: structure and mechanism. *Cellular and Molecular Life Sciences*, 62, 685-707

Podlesek, Z., Comino, A., Herzog-Velikonja, B., Zgur-Bertok, D., Komel, R. and Grabnar, M. (1995) *Bacillus licheniformis* bacitracin-resistance ABC transporter: relationship to mammalian multidrug resistance. *Molecular Microbiology*, 16(5), 696-676

Podlesek, Z., Comino, A., Herzog-Velikonja, B. and Grabnar, M. (2000) The role of the bacitracin ABC transporter in bacitracin resistance and collateral detergent sensitivity. *FEMS Microbiology Letters*, 188(1), 130-106

Pokusaeva, K., Kuisiene, N., Jasinskyte, D., Rutiene, K., Saleikiene, J. and Chitavichius, D. (2009) Novel bacteriocins produced by *Geobacillus stearothermophilus*. *Central European Journal of Biology*, 4(2), 196-203

Pruitt, K. D., Tatusova, T. and Maglott, D. R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 33 (Database issue): D501–504

Rao, M. B., Tanksale, A. M., Ghatge, M. S. and Deshpande, V. V. (1998) Molecular and biotechnological aspects of microbial. *Microbiology and Molecular Biology Reviews: MMBR*, 62(3), 597–635.

Reizer, J., Reizer, A. and Saier, M. H. Jr. (1992) A new subfamily of bacterial ABC-type transport systems catalyzing export of drugs and carbohydrates. *Protein Science*, 1(10), 13261332

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinate, J., Blumenthal, R. M., Degtyarev, S. K., Dryden, D. T. F., Dybvig, K., Firman, K., Gromova, E. S., Gumport, R. I., Halford, S. E., Hattman, S., Heitman, J., Hornby, D. P., Janulaitis, A., Jetsch, A., Josephsen, J., Kiss, A., Klaenhammer, T. R., Kobayashi, I., Kong, H., Krüger, D. H,, Lacks, S., Marinus, M. G., Miyahara, M., Morgan, R. D., Murray, N. E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao, D. N., Reich, N., Repin, V. E., Selker, E. U., Shaw, P-C., Stein, D. C., Stoddard, B. L., Szybalski, W., Trautner, T. A., Van Etten, J .L., Vitor, J. M. B., Wilson, G. G. and Xu, S. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research*, 31, 1805–1812

Roberts, R. (1990) Restriction enzymes and their isoschizomers. *Nucleic Acids Research*, 18, 2331

Runswick, M. J. and Harris, J. I. (1978) A rapid procedure for the isolation of endonucleases from two thermophilic bacteria. *FEMS Microbiology Letters*, 94(2), 6-12

Sangvik, M., Littauer, P., Simonsen, G. S., Sundsfjord, A. and Dahl, K. H. (2005) *mef*(A), mef(E) and a new *mef* allele in macrolide-resistant *Streptococcus* spp. isolates from Norway, *Journal of antimicrobial chemotherapy,* 56(5), 841-846

Sarwar M. and Akhtar, M. (1990) Cloning of aminoglycoside phosphotransferase (APH) gene from antibiotic-producing strain of *Bacillus circulans* into a high-expression vector, pKK2233; Purification, properties and location of the enzyme. *Journal of Biochemistry,* 268, 671-677

Scheeff, E. D. and Bourne, P. E. (2005) Structural evolution of the protein kinase-like superfamily. *PLOS One,* 1(5), e49

Selengut, J. D., Sigrist, C. J., Thimma, M., Thomas, P. D., Valentin, F., Wilson, D., Wu, C. H. and Yeats, C. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Research,* (Database issue): D211-215

Shah, P. and Swiatlo, E. (2008) A multifaceted role for polyamines in bacterial pathogens. *Molecular Microbiology,* 68(1), 4-16

Sigdel, S., Singh, R., Kim, T. S., Li, J., Kim., S. Y., Kim, I. W., Jung, W. S., Pan, C. H., Kang, Y. C. and Lee, J. K. (2015) Characterization of a Mannose-6-Phosphate Isomerase from *Bacillus amyloliquefaciens* and Its Application in Fructose-6-Phosphate Production, *PLOS One,* 10(7), e0131585

Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., and de la Cruz, F. (2010) Mobility of Plasmids. *Microbiology and Molecular Biology reviews: MMBR*, 74(3), 434–452

Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., and Bryant, S. H. (2014) CDD: NCBI's conserved domain database. *Nucleic acids research*, 43, D222226Spengler, G., Molnar, A., Schelz, Z., Amaral, L., Sharples, D., and Molnar, J. (2006). The mechanism of plasmid curing in bacteria. *Current Drug Targets*, 7, 1–19

Steczkiewicz, K., Muszewska, A., Knizewski, L., Rychlewski, L. and Ginalski, K. (2012) Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Research,* 40(15), 7016-7045

Stahl, S. (1991) Plasmids *in Bacillus stearothermophilus* coding for bacteriocinogeny and temperature resistance. *Plasmid*, 26, 94-107

Sukackaite, R., Lagunavicius, A., Stankevicius, K., Urbanke, C., Venclovas, C. and Siksnys, V. (2007) Restriction endonuclease BpuJI specific for the 5'-CCCGT sequence is related to the archaeal Holliday junction resolvase family. *Nucleic Acids Research,* 35(7), 2377–2389

Suzuki, S. and Henderson, P. J. F. (2006) The Hydantoin Transport Protein from *Microbacterium liquefaciens*, *Journal of Bacteriology,* 188(9), 3329-3336

Suzuki, H. and Yoshida, K. (2012) Genetic Transformation of *Geobacillus kaustophilus* HTA426 by Conjugative Transfer of Host-Mimicking Plasmids. *Journal of Microbiology and Biotechnology,* 22(9), 1279-1287

Takami, H., Takaki, Y., Chee, G. J., Nishi, S., Shimamura, S., Suzuki, H., Matsui, S. and Uchiyama, I. (2004) Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*. *Nucleic Acids Research,* 32(21), 6292–6303

Tatusov, R. L., Galperin, M. Y., Natale, D. A. and Koonin, E. V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research,* 28(1), 33–36

Thompson, M. G., Sedaghatian, N., Barajas, J. F., Wehrs, M., Bailey, C. B.,Kaplan, N., Hillson, N. J., Mukhopadyay, A. and Keasling, J. D. (2018) Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. *Scientific Reports,* 8, 1590

Trevors, J. T. (1986) Plasmid curing in bacteria, *FEM Microbiology Letters,* 32, 3-4

Turner, S., Pryer, K. M., Miao, V. P. and Palmer, J. D. (1999) Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *Journal of Eukaryotic Microbiology,* 46(6), 327-338

Uhlik, O., Jecna, K., Mackova, M., Vlcek, C., Hroudova, M., Demnerova, K., Paces, V. and Macek, T. (2009) Biphenyl-metabolizing bacteria in the rhizosphere of horseradish and bulk soil contaminated by polychlorinated biphenyls as revealed by stable isotope probing. *Applied and Environmental Microbiology*, 75(20), 6471–6477

Urbanek, A. K., Rymowicz, W., and Mirończuk, A. M. (2018) Degradation of plastics and plastic-degrading bacteria in cold marine habitats. *Applied Microbiology and Biotechnology*, 102(18), 7669–7678.

Van Schie, P. M. and Young, L. Y. (2007) Biodegradation of Phenol: Mechanisms and Applications. *Bioremediation Journal,* 4(1), 1-18

Vasu, K. and Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiology and Molecular Biology Reviews: MMBR*, 77(1), 53–72

Wada, A., Kono, M., Kawauchi, S., Takagi, Y., Morikawa, T. and Funakoshi, K. (2012). Rapid discrimination of Gram-positive and Gram-negative bacteria in liquid samples by using NaOHsodium dodecyl sulfate solution and flow cytometry. *PloS one*, 7(10), e47093

Wallace, L. A. (1989) Major sources of benzene exposure. *Enironmental Health Perspectives,* 82, 165-169

Wright, G. D. and Thompson, P. R. (1999) Aminoglycoside phosphotransferases: proteins, structure, and mechanism. *Frontiers in Bioscience*, 4, D9-D21

Xu, Y., Lunnen, K. D. and Kong, H. (2001) Engineering a nicking endonuclease N.AlwI by domain swapping. *Proceedings of the National Academy of Sciences of the United States of America USA*, 98(23), 12990–12995

Yamanaka, H., Kobayashi, H., Takahashi, E and Okamoto, K. (2008) MacAB Is Involved in the Secretion of Escherichia coli Heat-Stable Enterotoxin II. *Journal of Bacteriology,* 190(23), 7693-7698

Yang, S. C., Lin, C. H., Sung, C. T. and Fang, J. Y. (2014) Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Frontiers in Microbiology*, 5, 241

Yates, J., Aroyo, M., Sherratt, D. J., and Barre, F. X. (2003) Species specificity in the activation of Xer recombination at dif by FtsK. *Molcular Microbiology,* 49(1), 241-249

Yu, Y., Zhang, Q. and van der Donk, W. A. (2013) Insights into the evolution of lanthipeptide biosynthesis. *Protein science: a publication of the Protein Society*, 22(11), 1478–1489

Zebrowska, J., Zolnierkiewicz, O., Skowron, M. A., Zylicz-Stachula, A., JezewskaFrackowiak, J. and Skowron, P. M. (2016) A putative Type IIS restriction endonuclease GeoICI from *Geobacillus* sp.--A robust, thermostable alternative to mezophilic prototype BbvI. *Journal of Bioscience*, 41(1), 27-28

Zhao, X. and Kuipers, O. P. (2016) Identification and classification of known and putative antimicrobial compounds produced by a wide variety of *Bacillales* species. *BMC Genomics*, 17(1), 882

Zinoviev, V. V., Yakishchik, S. I., Evdokimov, A. A., Malygin, E. G. and Hattman, S. (2004) Symmetry elements in DNA structure important for recognition/methylation by DNA [amino]methyltransferases. *Nucleic Acids Research*, 32(13), 3930-3934

Zwick, M. E., Joseph, S. J., Didelot, X., Chen, P. E., Bishop-Lilly, K. A., Stewart, A. C., Willner, K., Nolan, N., Lentz, S., Thomason, M. K., Sozhamannan, S., Mateczun, A. J., Du, L., and Read, T. D. (2012) Genomic characterization of the Bacillus cereus sensu lato species: backdrop to the evolution of Bacillus anthracis. *Genome Research*, 22(8), 1512-1524

4. General Discussion

The genera *Geobacillus* and *Parageobacillus* are made up of Gram-positive, aerobic or facultative anaerobic bacteria. Members of these genera are further characterized by their thermophilicity where they are able to grow in temperatures ranging from 45 - 75 ˚C. Plasmids represent extrachromosomal DNA molecules that frequently occur within bacteria communities and encode a multitude of genes that can confer additional characteristics to the bacteria that harbour them. The genomes of several *Geobacillus* and *Parageobacillus* strains are observed to incorporate plasmids of which only a few have been studied. These genera are of biotechnological interest as a result of their increased thermophilic nature and thus their plasmid are also an area of interest given the extent to which plasmids contribute to the advantageous or unique traits bacteria possess. This study aimed at a holistic characterization of the plasmids among members of the genera to elucidate their prevalence, evolutionary relationships and functionalities within the *Geobacillus* and *Parageobacillus.*

In the first portion of this investigation a collective total of 85 geobacilli genomes were screened, identifying 63 plasmids. This suggests that plasmids are highly prevalent genomic features among the two genera with 54.12% (46/85) of all geobacilli members harbouring at least one plasmid. Of the 63 plasmids, the sequences of 56 were assembled enough for further analyses. These 56 plasmids were grouped on the basis of their shared orthologues which resulted in the formation of nine plasmid groups, pGeo1 to pGeo9, highlighting the extent of genotypic diversity amongst plasmids of these genera, with no orthologue shared amongst all 56 plasmids incorporated into this study.

Phylogenomic analyses of the largest group, pGeo1 (32 plasmids, sharing six orthologues), revealed increased structural diversification amongst these plasmids suggesting that they were derived from a common ancestor and have subsequently undergone extensive diversification. Among their six shared orthologues, a replication initiator protein was functionally annotated as a Rep_3 protein suggesting that this group of plasmids may replicate via the theta-mechanism.

The second largest group, pGeo2 (fifteen plasmids, sharing twenty-three), is comprised of plasmids and are found to be far more structurally conserved, which also suggests their derivation from a common plasmid ancestor. Furthermore, phylogenomic analyses of the pGeo2 plasmids further suggests that these plasmids may also be prone to horizontal gene

transfer between members of each genera, which could be as a result of a potential Type 4 Secretion System encoded by all fifteen plasmids. The remaining nine plasmids were grouped as pGeo3 to pGeo9, with very few plasmid proteins were identified and annotated amongst each group. However, several possible replication initiator proteins were identified, which allowed for inferences to be made on their mode of replication as being rolling-circle type (pGeo5, pGeo6, pGeo8 and pGeo9).

The next portion of this study aimed at further characterization of the geobacilli plasmids in order to identify the general roles that these plasmids play within the two genera as well as any phenotypes that may be of biotechnological value. The geobacilli plasmids were found to code for a total of 2,691 proteins belonging to 618 distinct protein orthogroups. This places further emphasis on the extent of plasmid diversification amongst these genera. Functional annotation of the plasmid proteins revealed that these plasmids carry out a variety of functional roles and may contribute to antibiotic resistance, the production of antimicrobial metabolites, and metabolic versatility amongst members of these genera. As such, these plasmids may contribute to members of these genera inhabiting new ecological niches or eliminating other bacteria during competitive relationships and are thus highly beneficial to the bacteria that carry them. Future studies can aim at characterizing these proteins and their associated pathways to contribute to the knowledge pertaining to the functional roles of these plasmids.

In addition to this, these plasmids are also a potential source of valuable thermostable proteins/enzymes, including a range of restriction modification enzyme systems and phenol degrading enzymes, that may be better suited to biotechnological applications due to their ability to withstand higher processing temperatures over mesophilic proteins/enzymes.

Lastly, the extraction of plasmid DNA from a subset of *Geobacillus* and *Parageobacillus* was carried out in an attempt to elucidate their plasmid-encoded phenotypes. This was unsuccessful as a result of possible low copy number plasmids as well as ineffective lysis of bacterial cells. As such the curing efficacy of acridine orange curing procedure could not be assessed. Future work should focus on optimizing protocols used, as well as developing and using plasmid isolation and curing protocols that are suitable for both low copy number and Gram-positive bacteria. This may assist in elucidating some of the geobacilli plasmid-encoded characteristics at a phenotypic level for the subsequent use of these plasmids and their proteins in biotechnology.

Overall it can be concluded that plasmids are a prevalent feature of the genera *Geobacillus* and *Parageobacillus.* The formation of nine plasmid groups and the range of potential phenotypes identified reveal the extensive diversity amongst the plasmids of these genera and suggest that these plasmids are rich in biotechnological value. Future work in further characterizing and isolating these plasmids can be considered in order to see these plasmids employed in various biotechnological applications.