UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Molecular characterization of *Bacillus sporothermodurans* in Ultra High Temperature (UHT) milk

By

Rodney Nana Owusu-Darko

Submitted in partial fulfilment of the requirements for the degree

PhD (Food Science)

In the Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

May 2019

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

DECLARATION

I, Rodney Nana Owusu-Darko declare that the dissertation, which I hereby submit for the degree PhD (Food Science) at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature: …….

Date: …….

DEDICATION

I dedicate this work to the Almighty God for His love, and the gift of life and to my parents Ebenezer and Lucy Owusu-Darko and my siblings Sidney, Sharon and Stephanie.

To Mom Dee, Auntie Vi, Auntie Ste and Auntie Suzie; gone, but not forgotten.

ACKNOWLEDGEMENTS

My supervisor, Prof. Elna M. Buys for her patience, guidance and insights to my research, the enthusiasm, and the independence afforded me to explore and apply myself and helping me hone my research and teaching capabilities.

My co-supervisor, Prof. Sílvia Dias de Oliveira for her supervision, constructive criticisms and suggestions that enhanced the dissertation.

Prof. Carlos A. S. Ferreira for your tireless efforts and direction with the molecular aspects of my research.

The South African Department of Science and Technology (DST) – National Research Foundation (NRF), in the form of a scholarship disbursed through the Center of Excellence (COE) of the University of Pretoria, South Africa and a HESA/IBSA Research Cooperation administered through the NRF.

The staff of the Sequencing Core Facility of the National Institute of Communicable Diseases, Johannesburg, for assistance with the whole genome sequence wet lab analysis, and especially Dr. Mushal Allam for help with genome analysis.

Dr. Frederick Tabit for his support in the early stages of my work.

Staff and postgraduate students of the Department of Consumer and Food Sciences for the support and encouragement throughout my research stay.

My parents, Mom Lou and Dad Ray, and my siblings Sidney, Sharon, and Stephanie, thanks for your moral, financial and spiritual support. I couldn't have done it without you.

To auntie Sly, uncle Amo, my cousins, nephews and nieces, God bless you for your encouragement.

To Amogelang, friends Clement, Franklin, Naa, Joyce, Selorm, Matthew, Pastor Landman and the Eli church choir members for your companionship, love and support.

ABSTRACT

Molecular characterization of *Bacillus sporothermodurans* in Ultra High Temperature (UHT) milk

By

Rodney Nana Owusu-Darko

Supervisor: Professor Elna M. Buys

Co-supervisor: Professor Sílvia Dias de Oliveira

Degree: PhD Food Science

Bacillus sporothermodurans spores are known to withstand Ultra high temperature (UHT) processing. The surviving spores germinate and cause the technological problem of non-sterility in milk and milk products. This high heat resistance has been attributed to clonal effects, with the highly heat resistant spore (HRS) responsible. This study sought to use whole genome sequencing (WGS) to characterize strains of *B. sporothermodurans* isolated from UHT milk and to determine and compare heat inactivation dynamics of an HRS and non-HRS strain. Subsequent comparative genetic analysis was undertaken to compare *B. sporothermodurans* strains with other closely related *Bacillus* sp. and determine the pangenome of *B. sporothermodurans*.

The closest phylogenetic neighbour of *B. sporothermodurans* was identified as *B. oleronius* using WGS. Further, the closest relatives besides *B. oleronius* were identified as *B. acidicola* and *B. coagulans* rather than *B. lentus* and *B. firmus* as previously thought. The *B. sporothermodurans* strains had genome sizes ranging from 3.4 Mbp – 4.0 Mbp and an average

G + C content of 36% as compared to 5.1 Mbp and 35% respectively for *B. oleronius*. The number of mobile element proteins in *B. sporothermodurans* strains ranged from 61 – 141, with the HRS strains SA01 and BR12 generally having twice as much as compared to the non-HRS strains. This will probably make the HRS strains more prone to horizontal gene transfer (HGT). Additionally, the HRS strains exhibited the Tn552 and Tn554 transposons, responsible for coding for beta-lactamases and erythromycin and kanamycin resistance and possibly also affecting the control of heat resistance in *B. sporothermodurans*.

Heat inactivation dynamics determined for the two South African strains SA01 (HRS) and SAD (non-HRS) showed similar inactivation data. Data from the study suggests spores of *B. sporothermodurans* strains can survive UHT temperatures for over 6 s at 140 °C. The Weibull inactivation model gave a superior fit as compared to especially the linear inactivation model.

Comparative genetic analysis of *B. sporothermodurans* strains show no observable difference in the nucleotide sequence of the GrpE heat shock protein and other proteins in the DnaK cluster of heat resistance related proteins, suggesting the influence of mobile genetic elements (MGEs) in their varying heat resistance. Pangenome analysis of *B. sporothermodurans* strains exhibited a core genome covering 69.9 – 92.2% and the accessory genome covering 7.8 – 30.1%. *B. sporothermodurans* core/pangenome ratio of 55% suggests an open pangenome thereby making it more prone to HGT events.

Results from this study suggests that while HRS and non-HRS strains have a similar complement of heat resistant genes, their differing heat resistance may be as a result of MGEs, since *B. sporothermodurans* has an especially open pangenome.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AMR	Antimicrobial resistance
BHI	Brain heart infusion
BL	Beta lactamase
CDS	Coding sequences
COGs	Clusters of orthologous groups
CRISPR	Clustered regularly interspaced short palindromic repeats
DNA	Deoxyribonucleic acid
DNAP	Deoxyribonucleic acid polymerase
DPA	Dipicolinic acid
EC	European Council
ESL	Extended shelf life
FDA	Food & Drug Administration
GCMS	Gas chromatography mass spectroscopy
HGT	Horizontal gene transfer
HPLC	High performance liquid chromatography
HRS	Highly heat resistant spores
HSPs	Heat shock proteins
ISs	Insertion sequences
ITEP	Integrated toolkit for exploration of microbial Pan-genomes
ITM	Intestinal mucin of invertebrates
LGT	Lateral gene transfer
MALDI	Matrix assisted laser desorption ionisation

MGE	Mobile genetic elements
MITEs	Miniature inverted repeat transposable elements
MLST	Multilocus sequence typing
MSE	Mean square error
NCBI	National Centre for Biotechnology Information
NGS	Next generation sequencing
ONT	Oxford nanopore technologies
PacBio	Pacific biosciences
PATRIC	Pathosystems resource integration centre
PCR	Polymerase chain reaction
PGAP	Prokaryotic genome annotation pipeline
PGAT	Prokaryotic genome analysis tool
PHASTER	PHAge search tool enhanced release
PMA	Propidium Monoazide
R^2	Regression of coefficient
RAPDs	Random amplified polymorphic DNA
RAST	Rapid annotation using subsystem technology
REP	Repetitive extragenic palindromic
RGPs	Regions of genomic plasticity
RNA	Ribonucleic acid
RNAP	Ribonucleic acid polymerase
RSM	Response surface methodology
SAM	S-adenosyl-L-methionine

SDS PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
sHSPs	Small heat shock proteins
SMRT	Single molecule real time
SNPs	Single nucleotide polymorphism
TEM	Transmission electron Microscope
UHT	Ultra-high temperature
UHPH	Ultra-high pressure homogenization
USB	Universal serial bus
UV	Ultraviolet
WGS	Whole genome sequencing

Chapter 1: INTRODUCTION

1.1 Introduction

Spore-forming bacilli are an important group of contaminants affecting the dairy industry. Albeit generally non-pathogenic, their presence in dairy products is an indicator of poor hygiene and numbers generally above 10^4 CFU ml^{-1} (Burgess *et al.*, 2010; Flint *et al.*, 2016) contravenes good manufacturing practices. Owing to their ubiquity, and highly adaptable spores, these bacilli affect a range of food products including dairy products. Contamination with these spores may occur along the entire milk supply chain and during reprocessing of milk products. Spore-forming bacteria including *Bacillus sporothermodurans*, *B. thermoamylovorans*, *B. licheniformis*, *B. coagulans*, *B. pumilus* and *B. subtilis* are usually selected for by high temperature processing including pasteurisation and ultra-high temperature processing (UHT) during dairy manufacture (Pettersson *et al.*, 1996; Burgess *et al.*, 2014). UHT milk involves processing at temperatures of between 135 to 140 °C for 2 – 5 seconds (Kessler, 1981) and packaging under aseptic conditions. The processing of milk at such high temperatures is aimed at inactivating vegetative cells and endospores present in raw milk with the view to long term storage (6 – 9 months) and under room temperature (Kessler, 1981).

1.2 Problem statement

The dairy industry has many challenges in the production of commercially sterile products as a result of reduced shelf life caused by microbial contamination (Burgess *et al.*, 2010). The culprits of dairy contamination are usually spore forming bacilli, which are challenging to inactivate because of the production of thermo-resistant endospores coupled with the ability to

contaminate milk along several routes of the dairy processing chain. Their contamination and growth may result in several milk defects as a result of the production of acids or proteolytic and lipolytic enzymes (Gundogan and Arik, 2004; Murugan and Villi, 2009; Burgess *et al.*, 2010). These challenges result in unfavourable organoleptic properties as unpleasant colour, smell, taste and textural changes (Rawat, 2015). Milk is a suitable medium for bacterial growth, which makes it prone to bacterial contamination and subsequently aggravates the previously mentioned unfavourable organoleptic changes. The formation of bacilli spores, resistant to high heat, disinfection and desiccation enhances their survival during industrial cleaning regimes (Ryu and Beuchat, 2005). The apparent rise in the occurrence of these highly resistant endospore-forming bacilli in the dairy industry, could be as a result of environmental and processing adaptation and the selection for highly heat resistant spore-formers such as *B. sporothermodurans*, that are exposed to elevated heat processing regimes such as UHT of dairy products (Pettersson *et al.*, 1996; Postollec *et al.*, 2012). Alternatively, the increase in occurrence of highly heat resistant spore formers such as *B. sporothermodurans* could be as a result of transfer of bulk ingredients around the world (Flint *et al.*, 2016), improved detection methods and horizontal gene transfer of traits from previously uncharacterised environmental bacterial strains. Regardless of the cause, improved characterisation of *B. sporothermodurans* is critical in identifying the unique attributes of this important bacteria.

B. sporothermodurans, an emerging spore-forming high heat resistant *Bacillus* specie has been continuously identified in dairy plants around the world since the early 1980s. With the ability to form biofilms (Jindal *et al.* 2016), *B. sporothermodurans* is one of the most heat resistant endospore forming bacteria known to the food industry at present, easily able to survive UHT temperatures (Klijn *et al.*, 1997; Huemer *et al.*, 1998). Its heat resistance is known to be clonal,

with the highly heat resistant spore (HRS) clone particularly resistant against UHT temperatures.

In recent times, characterisation of *B. sporothermodurans* and its comparison to its closest phylogenetic neighbour, *B. oleronius* (Heyndrickx *et al.*, 2012), has mainly been carried out using 16S rDNA sequencing. However multiple copies of this gene are located on different sites on the genome, the mechanism for generating multiple ribotypes by probing with a 16S rDNA complimentary probe, hence misidentification especially at the species level (De Clerck and De Vos, 2004; Woo *et al.*, 2008). Protein encoding genes (including *atpD*,*gyrB*, *nifD*, *recA* and *rpoB*) though proposed as better alternatives as a result of the handicap of the 16S *rRNA* gene (Caamano-Antelo *et al*., 2015), lacks the discriminatory power of recent next generation sequencing technologies. Also, the use of DNA-DNA hybridisation, SDS-PAGE analysis of cellular proteins, cell wall (polar lipid) and menaquinone analysis (Heyndrickx *et al.*, 2012), though useful, has resulted in reclassification of some related strains that were initially thought to be *B. sporothermodurans*, and are not effective against identifying strain differences.

The use of whole genome sequencing has not been utilised in the study of *B. sporothermodurans* and most importantly to differentiate HRS and non-HRS clones. However, the use of whole genome sequencing (WGS) is especially effective at elucidating strain differences as a result of genomic variations (Kim *et al.*, 2017). Furthermore, downstream analysis including pangenome studies enhances the identification of core and strain specific genes that may infer phenotypic characteristics (Bazinet, 2017; Kim *et al.*, 2017). There is no pangenome study undertaken on *B. sporothermodurans* strains to date.

Additionally, limited information exists for the thermal inactivation studies of *B. sporothermodurans* spores, with past studies traditionally based on first order kinetics. The

determination of D-values using Weibullian models rather than the traditional first order kinetics model would offer more accurate heat inactivation data. This will provide valuable information on temperature time dynamics in *B. sporothermodurans* inactivation, prevent over-processing and save energy in the processing plant. This would maintain organoleptic properties and help preserve food nutrients (Claeys *et al.*, 2001). New evidence suggest that the microbial sterilization models based on first order kinetics is the exception rather than the rule (Van Boekel, 2002), especially in the case of bacterial spores. Furthermore, the heat inactivation dynamics of HRS and non-HRS clones of *B. sporothermodurans* will offer interesting perspectives on heat resistance.

Chapter 2: LITERATURE REVIEW

2.1 First cases of UHT milk contamination

In 1985, *Bacillus* species were detected in UHT processed milk in many European countries (Hammer *et al*., 1995). At that time, spoilage encountered in UHT milk products was thought to have been caused by contamination during filling. Members of the genus *Bacillus*, especially *B. badius*, *B. cereus*, *B. licheniformis*, *B. subtilis* and others outside the genus *Bacillus* including *Paenibacillus polymyxa* and *Geobacillus stearothermophilus*, were identified as responsible for the spoilage of UHT milk (Pettersson *et al.*, 1996). However, it was also found that other heat resistant spores could survive UHT treatment and subsequently grow in the stored products. The first detection of these heat resistant spores in UHT milk was in Germany in 1990, and subsequently spread to several other countries (Hammer *et al*., 1995).

In 1985, the European Council (EC) Milk Hygiene Directive 85/397 stipulated a maximum of 10 CFU/0.1 ml for UHT milk after an incubation period of 15 days at 30 °C and this led to a change in the quality control protocol for UHT treated milk. When the German government enforced the EC Milk Hygiene Directive 85/397 in their national legislation through the Hygiene Ordinance of June 1989 (Klijn *et al.*, 1997) the quality control for UHT milk was modified to include bacteriological analysis techniques instead of the usual physio-chemical methods like pH, sensory and stability tests, which lacked effectiveness to detect slow-growing spore forming bacteria (Klijn *et al.*, 1997).

The initial heat resistant bacteria was temporarily referred to as the highly heat resistant spore former (HHRS or HRS) as they were normally isolated from the bypass of an indirect UHT system and immediately after the heating section (Scheldeman *et al.*, 2006). The bacterium

responsible for the non-sterility of UHT milk was subsequently identified to be *B. sporothermodurans* (Pettersson *et al.*, 1996).

2.2 Milk contamination routes

There's a clear connection between soil-borne endospore forming bacteria and their subsequent presence in foods as contaminants. Spores of the genus *Bacillus*, which are formed at the end of the growth phase and within the bacterial vegetative cell are released into the environment as a means of survival during harsh conditions. The ubiquity of these spores, coupled with their heat resistance, spore adhesive characteristics and ability to regenerate during favourable conditions make them a food safety and quality concern (Heyndrickx 2011a). In the dairy industry, bacterial sporeformers contaminate processed milk through filler nozzles which may have been compromised by contaminated air or water. The contaminating spores can be long-time residents of the filling machine or in its immediate surroundings (Eneroth *et al*., 2000). Consequently, silage, factory equipment and packaging materials have been identified as main sources of contamination of raw milk by bacterial spores (Te Giffel *et al*., 2002). Feed concentrates for dairy cattle has also been implicated as a source of raw milk contamination by *B. sporothermodurans* and other bacterial sporeformers at dairy farms, with spores counts ranging from $4{\cdot}0 \times 10^3$ to $1{\cdot}1 \times 10^6$ spores g^{-1} of feed (Vaerewijck *et al.*, 2001). Studies suggest that bacteria associated with dairy contamination do not necessarily come from raw milk alone, thereby confirming the theory of contamination and propagation during the manufacturing process (Banykó and Vyletelová, 2008, Heyndrickx 2011, Kmiha *et al.* 2017, Mugadza 2018). Studies by Mugadza *et al.* (2018) albeit on extended shelf life (ESL) milk also showed the carryover of *Bacillus* sp. from raw milk through various section of the processing plant to the

final product. Research from Yuan *et al*. (2012) found the presence of thermophilic bacteria in processed milk powders including infant formula milk powders and whole milk powders manufactured in China. Thermophilic bacteria where found in eighteen of twenty-two milk powder producers.

2.3 UHT milk processing and systems

2.3.1 UHT milk processing

UHT processing involves exposure of liquids such as milk to brief, intense heating within the temperature range of 135 °C to 140 °C for 2 - 10 seconds, packaged under aseptic conditions and stored at room temperature for months (Pinto *et al.*, 2018). The processing of milk at high temperatures is aimed at destroying vegetative cells as well as endospores present in raw milk for long term storage without refrigeration (Elliott *et al.*, 2005; Ritota *et al.*, 2017).

UHT plants operate in a manner to maintain the temperature-time relationship at a given range. Two types of UHT processes exist: the direct UHT, which includes a heating and cooling phase, and indirect UHT. The heating and cooling phases increase with a higher temperature during indirect UHT processes. This implies a longer holding time and, therefore, an increase in the undesired heat-induced changes in milk (Elliott *et al.*, 2005; Ritota *et al.*, 2017). For the direct UHT process, which could be either injection or infusion, a high heating temperature can be achieved without the heating and cooling phases. The use of flash evaporation to cool milk rapidly in direct UHT results in heat loss as compared to indirect UHT processes. This is because during flash evaporation, the steam formed is approximately 80 °C and can only be used for milk preheating (Grijspeerdt *et al.*, 2004).

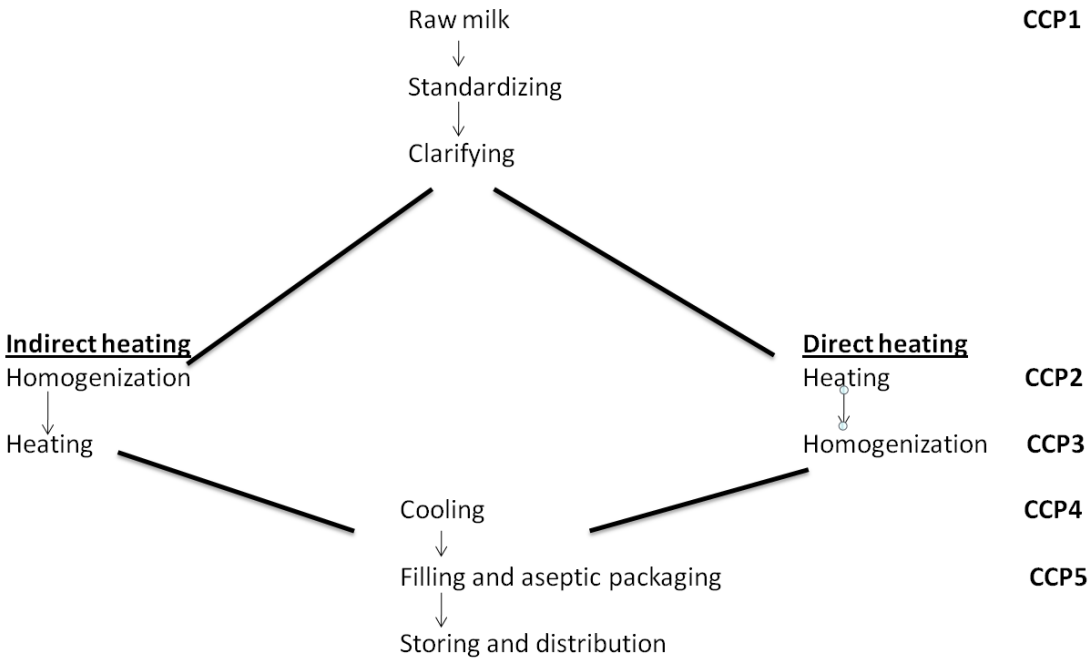


Figure 2.1: Flow diagram for production of UHT milk. (Sandrou and Arvanitoyannis, 2000)

2.4 UHT treatment systems

2.4.1 Direct UHT system

The direct UHT system involves heating by mixing the product and steam at high pressure of approximately 9 bars. The heat transfer is higher than in indirect UHT systems, and because the residence time is very short, it can result in less fouling. This process is difficult to control and therefore not often used (Grijspeerdt *et al.*, 2004). Two types of direct heating systems exist, direct heating by injection and direct heating by infusion as shown in Figure 2.2. With the injection system, the steam injected encounters the product through a specially designed nozzle. In the infusion system however, the product is distributed in strings across the centre of a chamber in which steam is distributed.

2.4.2 Indirect UHT system

The indirect UHT system consists of a solid barrier that separates the heat transfer medium (often water or steam) and the dairy product. The indirect system can further be subdivided into two types depending on the nature of the heat exchanger incorporated into it. These include the plate and tubular heat exchangers as shown in figures 2.3 and 2.4 (Grijspeerdt *et al.*, 2004). The quality of UHT milk in packs processed via indirect UHT differed from that produced by direct UHT when they were analysed after storage of 24 weeks. Thermally induced changes in lactulose and furosine, formed during heat treatment and acid-soluble whey proteins revealed that directly heated UHT milks suffer less heat damage than indirectly heated milk (Elliott *et al.*, 2005).

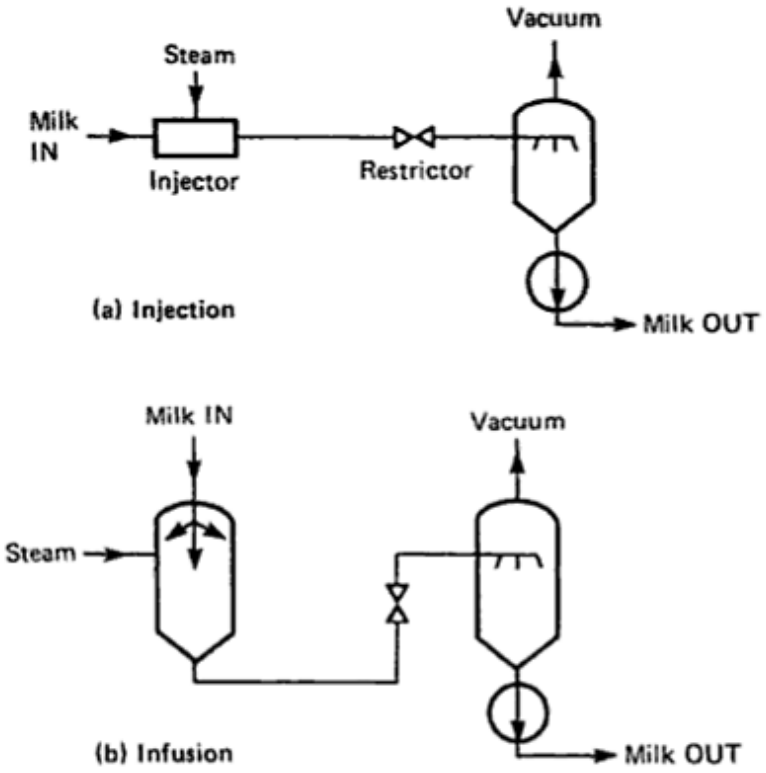


Figure 2.2: Diagram of UHT (a) injection and (b) infusion systems (Lewis and Heppel, 2000)

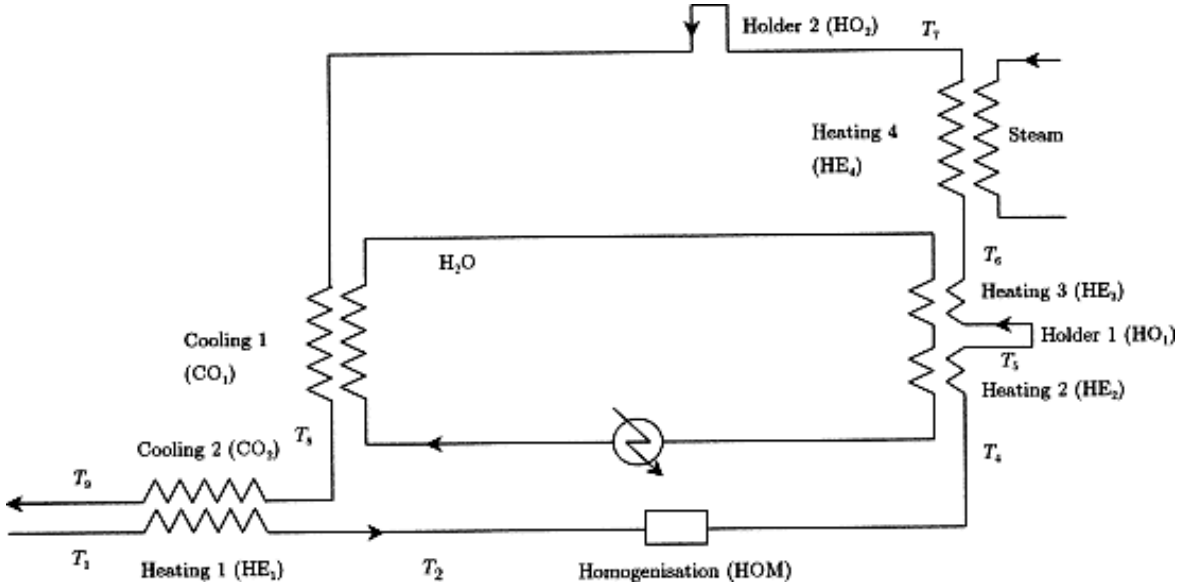


Figure 2.3: Schematic layout of a UHT system using a tubular heat exchanger (Grijspeerdt *et al.*, 2004).

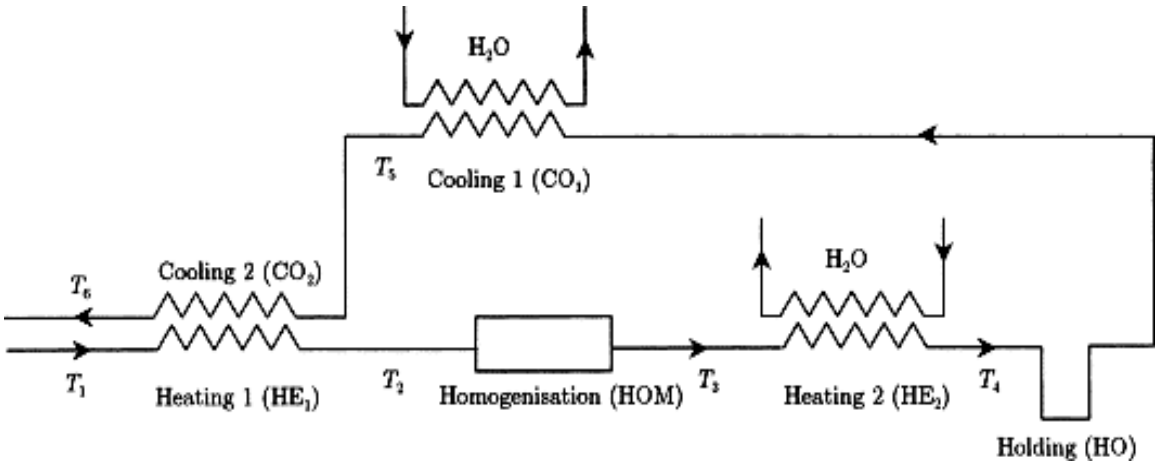


Figure 2.4: Schematic layout of a UHT system using a plate heat exchanger (Grijspeerdt *et al.*, 2004).

2.5 *B. sporothermodurans* in milk

B. sporothermodurans spores have been found to be more resistant than other heat resistant spores at temperatures above 130 °C with D_{140} ranging from 3.4 - 7.9 s and Z-values ranging from 13.1 - 14.2 °C (Huemer *et al.*, 1998). The D-value is the time required in a given medium at a given temperature to inactivate 90% of the microbial population, whiles the Z-value, a simplified version of the Arrhenius equation, is the calculated increases of temperature required to reduce the D-value by 90% (Xu *et al.*, 2006). Any increase in the processing temperature or the holding time to inactivate *B. sporothermodurans*, negatively impacts the quality of UHT milk (Claeys *et al.*, 2001). When unopened packages of UHT treated milk are incubated at 30 °C for 15 days, *B. sporothermodurans* counts rarely exceed a maximum of 10^5 CFU ml^{-1} (Tabit and Buys, 2011). Isolation of *B. sporothermodurans* from raw milk can be challenging because of its inability to grow and compete with other faster growing bacteria in raw milk. However, there is no competition from other bacteria in UHT processed milk hence it grows without hindrance, making isolation in this case easier (Huemer *et al.*, 1998).

Albeit having been discovered in the 1980's, its route of contamination is a matter of debate and of varied opinion, with diverse possible contamination routes including soil, fodder, digestive tract, dung, udder, teat, milking utensils, raw milk and feed mentioned by various studies (Vaerewijck *et al.*, 2001). Even though *B. sporothermodurans* is not known to be pathogenic (Pettersson *et al.*, 1996), its presence in UHT milk is against GMP with regards to the non-sterility of commercially sterile milk and milk products. This constitutes a quality problem, leading to possible considerable economic losses (Scheldeman *et al.*, 2006). Proteolysis in UHT milk by *B. sporothermodurans* is minor, with mild proteolytic changes experienced when the sell by date of UHT packages has been reached (Tabit, 2018). Aerobic

incubation with high concentration of *B. sporothermodurans* for 72 h caused the breakdown of casein and the presence of small peptides, however as a result of poor growth in the milk this proteolytic activity caused minor changes in the milk (Klijn *et al.*, 1997). Real spoilage defects are rare and evident by a slight pink colour change, with off flavours and coagulation but notably with packaging with a low oxygen barrier (Lembke, 1995). Significant proteolysis activity in UHT milk due to *B. sporothermodurans* has been reported to occur long after the exponential growth phase has been attained (Tabit, 2018).

2.6 Culturing and physiological traits

B. sporothermodurans react with Gram stain producing a granular appearance with uneven staining (Hammer *et al*., 1995). Brain heart infusion (BHI) agar plates that have been supplemented with Vitamin B_{12} are used for the isolation of *B. sporothermodurans* when incubated at 37 °C. In industry though, milk plate count agar is used to check for contamination, however growth is poor (Scheldeman *et al.*, 2006). Due to the competitive nature of the background flora, it is difficult to isolate *B. sporothermodurans* from raw milk or other farm sources. However, samples heated at 100 °C for 30 - 40 min and plated on BHI agar, makes isolation easier (Scheldeman *et al.*, 2006). Some phenotypic characteristics are depicted in Table 2.1.

Table 2.1: Phenotypic characteristics of *B. sporothermodurans*

Characteristic	Result	Characteristic	Result
Motility	v	API 50 CHB tests contd.	
Spores observed	Rarely	*myo*-Inositol	−
Spores spherical	v	Mannitol	v
Sporangia swollen	v	Sorbitol	−
Anaerobic growth	−	Methyl d-glucoside	v
Gram reaction	+	Methyl d-glucoside	v
Starch hydrolysis	−	Amygdalin	v
Casein hydrolysis	−	Arbutin	v
API 20E tests		Salicin	v
ONPG reaction	−	Cellobiose	v
Citrate utilization (Simmons')	v	Maltose	+
Urease production	−	Lactose	−
Voges–Proskauer test	v	Melibiose	−
Gelatin hydrolysis	+	Sucrose	(+)
Nitrate reduction	+	Trehalose	(+)
API 50 CHB tests		Inulin	−
Glycerol	v	Melezitose	v
l-Arabinose	−	Raffinose	−
Ribose	−	Glycogen	−
d-Xylose	−	Xylitol	v/w
Methyl xyloside	−	Gentiobiose	v
Galactose	−	Turanose	v
d-Mannose	v		

All characteristics were determined using tests in the API 20E and API 50CHB systems with the exception of microscope observations, and starch and casein hydrolysis. +, >85% positive; (+), 75–84% positive; v, variable (26–74% positive); −, 0–15% positive; w, weakly positive reaction; +/w, positive or weakly positive reaction; v/w, variable and when positive the reaction is weak (adapted from Heyndrickx *et al.*, 2012)

2.7 Characterisation of the *B. sporothermodurans* spore

2.7.1 Spore structure

The *B. sporothermodurans* spore has four conspicuous layers; cortex, spore coat, exosporium and the surface layer and the densities of the different layers are in the order; core > spore coat > cortex > exosporium. The core is presented on the transmission electron microscope (TEM) micrograph as a central compact structure of approximately 500 nm in diameter which is surrounded by the cortex of about 62 nm wide. The coat which is the second dense layer is relatively thin compared to the other layers but is surrounded by the exosporium, which is approximately 187 nm wide.

2.7.2 The induction of endospore germination before heat treatment

The induction of germination of the *B. sporothermodurans* spore prior to heat treatment in processed products has been studied. Induced germination of spores could lower the severity of heat treatment during milk processing. Naturally used amino acids, purine ribonucleotides (inosine), and sugars and the effect of different factors such as pH, temperature, and cations were used to optimize and analyse the nutrient-induced germination of up to 100% *B. sporothermodurans* spores. D-glucose, l-alanine, and inosine were the germinants specific to *B. sporothermodurans* LTIS27 with an optimal germination rate occurring after 60 min following incubation of spore at 35 °C in the presence of 9 mM of d-glucose and 60 mM of l-alanine (Aouadhi *et al.*, 2013).

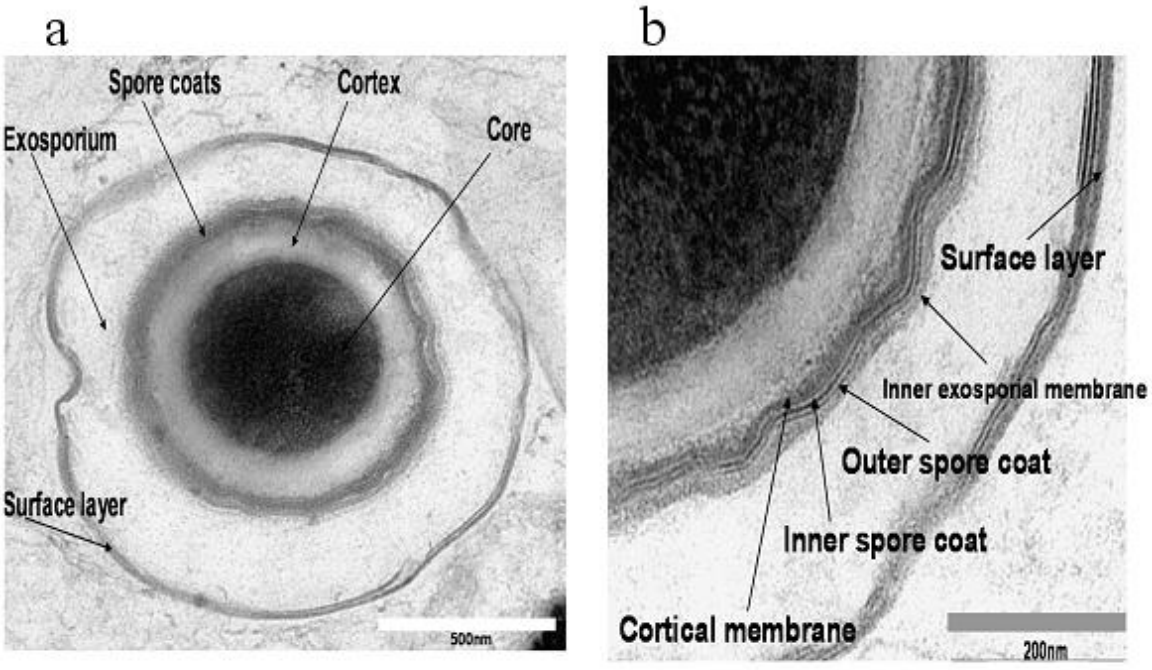


Figure 2.5: Micrographs generated from TEM analysis (a & b), showing the detail structures of unheated spores of *B. sporothermodurans* UP20A, harvested from 14 days old BHI agar plates (Tabit and Buys, 2010).

An alternative study conducted to investigate the potential use of pressure-induced germination of *B. sporothermodurans* spores in water and in milk, looked at the effects of the pressure, temperature, pressure-holding time and post-pressurization incubation time and temperature. The germination of *B. sporothermodurans* spores was induced at lower pressures with maximum germination achieved at 200 MPa and faster in milk than in distilled water (Aouadhi *et al.*, 2012). Given that temperature had no significant effect on germination, and complete germination was not realised, suggest it will be difficult to use this technique in solving the *B. sporothermodurans* contamination problem. On the other hand, nutrient-induced germination has potential for industrial applications provided 100% *B. sporothermodurans* spore germination can be attained.

2.7.3 Effects of wet heat treatment on bacterial spores

The ability of *B. sporothermodurans* to survive extreme temperatures has been attributed to strain specific heat resistant clones. The inactivation of *B. sporothermodurans* spores during wet heat treatment occurs when hot moisture penetrates the spore, moistens spore components and inactivates enzymes resulting in the release of spore components such as proteins and dipicolinic acid (DPA) due to high water pressure in the spores (Tabit and Buys, 2010). The release of DPA indicates the breach in the structural integrity of spores during wet heat treatment. Thus, any inactivation technique that compromises the structural integrity of spores with heat application can lead to complete inactivation of spores at relatively shorter time and lower temperature.

For example, Baril *et al*. (2011) discovered with *B. weihenstephanensis* KBAB4 spores that it wasn't so much about the conditions prior to sporulation that affected thermal resistance but rather the sporulation temperature. Thermal resistance in *B. weihenstephanensis* KBAB4 spores was found to be dependent on the temperature that initiated the sporulation process. This assertion however contradicts the previous theory on thermal resistance of bacterial spores with respect to thermal resistance being affected mainly by conditions prior to sporulation. In my opinion though, both situations could influence thermal resistance since the genetic response to environmental stress resulting in the production of proteins can occur at any stage of the introduction of an unfavourable condition. This could be an isolated case rather than the norm in *B. weihenstephanensis* KBAB4.

2.7.4 Effects of sporulation temperature

The sporulation temperature is the temperature at which sporulation initiates and has been directly found to affect the thermal resistance of *Bacillus* spores. This has practical importance in the food industry as it has a bearing on methods for the effective inactivation of bacterial spores in food products. Garcia *et al*. (2010) suggested that the resistance of bacterial spores to inactivation increases when the sporulation temperature is high. The reverse is the case with high pressure treatments where Olivier *et al*. (2012) realised that high sporulation temperatures correlated to a reduction in the resistance to high pressure processing.

The above results make for interesting deductions as manipulation of the sporulation temperature would affect various processing parameters, including high pressure or high temperature processing. Thus, combinations of these two processing techniques could provide important answers to bacterial spore inactivation. Aouadhi *et al*. (2012), combined the effects of pressure and temperature to firstly induce germination and spores and then apply heat treatment to inactivate the resulting vegetative cells. The study above showed the potential of the use of high hydrostatic pressure to induce germination of *B. sporothermodurans* spores before treatment with heat. As a result, the temperatures subsequently used for processing in this case would be lower, as inactivation of vegetative cells rather than spores would be the case. However, the above study did not induce a 100% germination rate with pressure treatment.

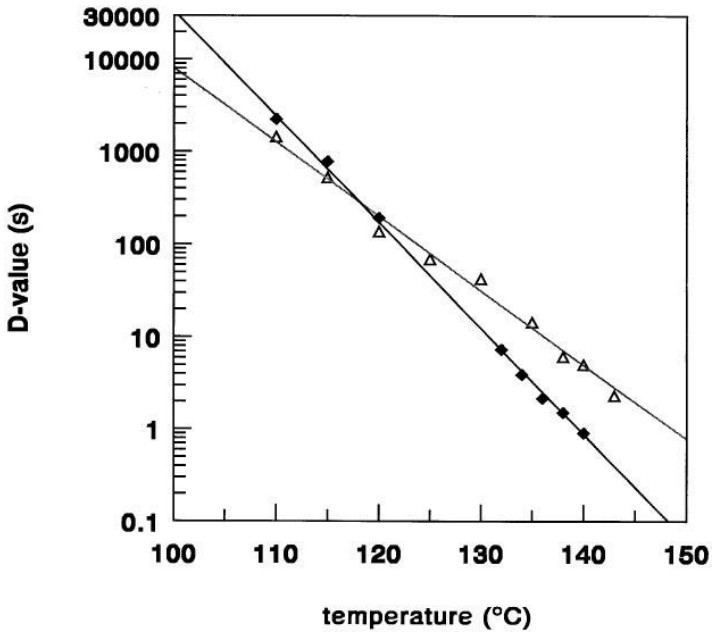


Figure 2.6: Thermal death time curves of *B. stearothermophilus* spores and *B. sporothermodurans* spores. *G. stearothermophilus* spores (▲) and *B. sporothermodurans* spores J16B (Δ); best plot lines through experimental data (Huemer *et al.*, 1998)

2.7.5 Thermal death kinetics of B. sporothermodurans spores

Spores of *B. sporothermodurans* strains isolated from non-sterile UHT milk are exceptionally heat resistant at higher temperatures (130 – 145 °C) under UHT conditions. And at higher temperature the D_{140} values of *B. sporothermodurans* range from 3.4 to7.9 s compared to 0.9 s for *G. stearothermophilus* (Figure 2.6) (Huemer *et al.*, 1998). In a similar study conducted by Tabit and Buys (2011), spores of two strains of *B. sporothermodurans*: DSM 10599^T and UP20A survived heat treatment at 130 °C for 4 minutes as shown in Figure 2.7. Molecular typing shows a heterogeneous group of farm isolates (non-HRS strains), but a clonal group of UHT isolates from diverse European countries and other continents (HRS-clone) suggesting a

common source (Scheldeman *et al.*, 2006). There are observed differences in heat resistance between HRS and non-HRS strains but as well spores isolated from UHT milk and from other sources. For example the D_{100} (D-value at 100 °C) of HRS strains was determined to be 800 mins as opposed to a range of 23 - 468 mins in non-HRS strains (Scheldeman *et al.* 2006). Even for spores isolated from various UHT based products, there was a wide variation in D_{100} of 800 mins, 160 mins and 77 mins of industrial spores in naturally contaminated UHT milk, spores from UHT milk and spores from UHT based vanilla drink respectively (Scheldeman *et al.*, 2006). Spores from feed concentrate on average had higher D_{100} (132 – 476 mins) as opposed to those from raw milk (23 mins). The properties of the medium of isolation selects for spores with varying heat resistance.

2.7.6 Effects of medium characteristics on isothermal and non-isothermal heat resistance

The effect of the medium characteristics (pH and buffer/food) on the thermal inactivation of *B. sporothermodurans* spores upon exposure to isothermal and non-isothermal heating and cooling treatments has been investigated (Esteban *et al.*, 2013). The survival curves showed the shoulder phenomena common with highly heat-resistant bacteria spores hence the Weibull and Geeraerd non lineal models used for analysing the survival curves. The heat resistance of *B. sporothermodurans* spores diminishes with low pH in both isothermal and non-isothermal heat treatments (Esteban *et al.*, 2013).

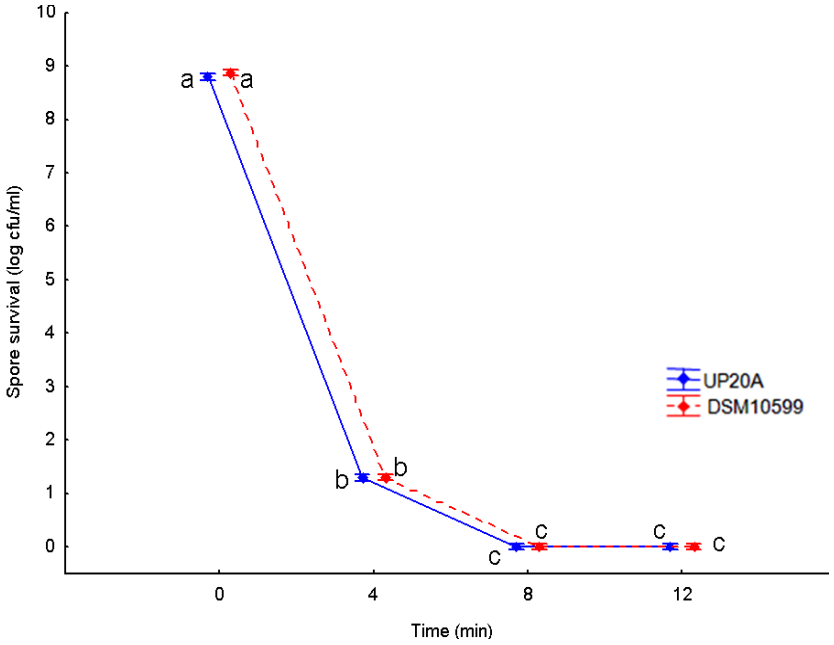


Figure 2.7: Inactivation trend of spores of two strains of *B. sporothermodurans*: DSM 10599^{T} and UP20A, following heat treatment at 130 °C at different times. Different letters a, b and c denote statistical differences at $p \leq 0.05$ (n = 3). Log CFU ml^{-1} of 0 represent counts ranging from 1 to 0 spore ml^{-1}(n = 3) (Tabit and Buys, 2011).

2.8 Modelling microbial behaviour

In pursuit of production of safe nutritious food, the behaviour of microbes in food matrices needs to be ascertained. Predicting this behaviour especially with the presence of multiple food pathogens can always be a challenge. In the food industry severe thermal stress is applied to inactivate bacteria including bacterial spores. However, in the pursuit of optimal processing without the risk of over-processing, modelling of microbial processes has become even more important for two reasons. Firstly, to create models that would effectively inactivate bacteria in food with minimal energy requirements and thus reduced production costs and secondly to

prevent undesirable organoleptic properties as a result of over-processing. The latter, also consequently affecting the nutritional quality of foods.

There has been much debate if microbial inactivation follows first order kinetics. However new evidence suggests that a log linear isothermal survival curve of bacterial cells is the exception rather than the rule (Van Boekel, 2002). This is especially the case in bacterial spores, to which *B. sporothermodurans* is expected to mimic (Periago *et al.*, 2004). This suggests that previous methods used to calculate the efficacy of thermal processes need to be revised (Corradini *et al.*, 2005).

2.9 Changing trend in microbial modelling

Traditional methods of determining sterilisation processes have been based on D, *z* and F values. Recent research however suggests a paradigm shift from the traditional based methods. However, there is still the notion among selected food scientists that the excellent safety record experienced especially in the canning industry has been as a result of over processing. Campanella and Peleg (2001) share the above notion and suggest that it's not as a result of an accurate model being used for microbial death kinetics. Nonetheless, if the current trend with respect to a shift towards minimally processed foods is anything to go by, then the traditional first order kinetics methodology needs to be revised. This revision would give rise to safe minimally processed food that would have the full nutritional complement.

This revised model, which is non-linear, can be described by a power law model. The model assumes that thermal heat resistance of most microbial vegetative cells and spores have a Weibull type distribution (Corradini *et al.*, 2005). The traditional thermal inactivation series

based on first order kinetics can be expressed as follows $\log_{10}(N/No) = kt$ – (1). N and No are the momentary and initial number of cells or spores respectively, k is a base 10 logarithmic constant. The D-value as is traditionally calculated is the time needed to reduce the size of the treated population by 10 fold. It is the inverse of k above in equation (1).

The change in trend to the use of the Weibull model can be argued across several lines. Firstly, the linear type model, based on first order kinetics postulates that the probability of a lethal event at a given temperature is constant and does not depend on the exposure duration (Peleg *et al.*, 2008). However, it is expected in most microbial inactivation scenarios that heat will inactivate bacteria, injure others or create resistant ones or a combination of these. Thus, time plays an important role both on the survival and inactivation rates.

2.10 Weibullian Model

The Weibull model is represented as $\log_{10}S(t) = b(T)t^{n(T)}$ – (2), where S(t) is the survival ratio N(t)/No, and N(t) and No are the momentary and initial number of cells or spores respectively and b(T) and t^n temperature dependent coefficients. As the first order kinetics model is a special case of equation (2) where n(T) = 1, models developed by the Weibull distribution is applicable with distributions of first order kinetics, however the reverse doesn't hold same (Corradini *et al.*, 2005). Hence, the familiar D-value can be calculated using a variant of equation (2). In that case the resulting equation becomes $\log_{10}S(t) = -k(t)t^n$ - (3). Thus, the inverse of k(t) becomes the D-value, which may be denoted as $\log_{10}S(t) = - 1/k(t)\ t^n$ – (4).

The Weibullian method for inactivation studies in microbes have been modified into various forms for varying microbes and scenarios. González *et al*. (2009), developed a variant of the

Weibull equation to study *Campylobacter jejuni* in minced chicken meat. The study developed a δ – parameter like that of the D-value in linear models. Albert and Mafart (2005) also developed a similar model out of the Weibull model where the δ – parameter developed also had close characteristics with the D-value. In addition, the latter developed a model that would take care of shoulder and/or tailing phenomena if they occurred. It should be noted that since various microbes behave differently in different media, the Weibull equation should be tailored to respective circumstances.

Since the notion of injured cells were not considered in the above two situations, Corradini & Peleg (2007), addressed the issue still using a modified version of the Weibull method. The method was used to differentiate healthy and injured cells. Though this method goes further with the Weibull method it however fails to consider the effects of recoverable cells. Thus, in such a case as *B. sporothermodurans* spores, the dormant spores cannot be classified as dead but recoverable. In order to consider this situation in microbial modelling, the bacterial spore in question needs to be tested for activity of recovery from injury or dormancy.

2.11 Profiling of cellular proteins

The chemical components of *B. sporothermodurans* is affected by wet heat treatments. Studies prove that the leaking of cellular proteins during heat treatments results in cell death. There is also a suggestion that the amount of protein release seems to be strain specific (Tabit and Buys, 2010). Hence, profiling cellular proteins may give an insight into the various responses to external stimuli including heat of bacterial species.

The SDS-PAGE technique has proved successful in the rapid identification of bacterial species (Berber, 2004). Cellular protein profiling combined with genotypic information may increase the accuracy of bacterial specie identification especially at the strain level. There could also exist important correlations between bacterial cellular proteins and the breakdown of proteins during milk spoilage, a scenario not yet explored in *B. sporothermodurans*.

2.12 Determination of milk spoilage

Milk is highly rich in nutrients and as such very susceptible to spoilage. It provides the perfect medium for spoilage microorganisms to grow. By applying extreme heat to inactivate all bacterial vegetative cells and spores, UHT milk was developed to increase the shelf life of milk to a maximum of 9 months. However, as a result of thermo-resistant bacterial spores as *B. sporothermodurans*, UHT milk may not be devoid of bacterial cells or spores. An industry standard termed commercial sterility used as the benchmark in this respect (Pinto *et al.*, 2018).

The degradation of milk products by various enzyme activities associated with contaminating bacteria can reduce the shelf-life of processed milk (Dogan and Boor, 2003). The control of proteolytic microbes is one of the main challenges of the milk industry because of their breakdown of protein fractions that cause the incidence of milk spoilage. The enzymes involved in the proteolytic activity are responsible for the off-flavours, bitter taste, and gel formation in UHT milk (Montanhini *et al.*, 2013). It is however assumed that the effects of this spoilage can only occur at a certain threshold of CFU ml^{-1} of the contaminating microbe. Studies have shown *B. sporothermodurans* does not normally grow beyond 10^5 CFU ml^{-1}. There is debate whether this levels of growth of *B. sporothermodurans* is enough to cause

significant milk spoilage. What is more certain though is that levels of 10^5 CFU ml^{-1} over several months could affect the keeping quality of UHT milk.

Several methods to check milk spoilage have included laser desorption ionization mass spectrometry (Siciliano *et al.*, 2000), gas chromatography mass spectrometry (GCMS) (Valero *et al.*, 2001) and high performance liquid chromatography (HPLC) (Velso *et al*., 2002).

2.13 Molecular characterisation of *B. sporothermodurans*

2.13.1 Molecular detection of spore formers

The Genus *Bacillus* is of prime importance to man, as the likes of *B. subtilis, B. amyloliquefaciens, B. licheniformis* are involved in the production of important biotechnology products (Daffochio *et al*., 1998). These products include antibiotics, enzymes and solvents (Shafique, 2012). Other species as *B. thuringiensis* and *Lysinibacillus sphaericus* are used as insecticides (Juarez-Perez *et al*., 1997). Though *Bacillus* spore-formers can be a nuisance in the dairy industry, the likes of *B. subtilis, B. clausii, B. cereus, B. coagulans* and *B. licheniformis* are being utilised as probiotics (Hong *et al.*, 2005; Cutting, 2011). *B. sporothermodurans*, with its high heat resistance and non-pathogenic nature could well fit into this probiotic bracket. However, *B. anthracis* and *B. cereus* are pathogenic of plants (Keim *et al.*, 1997) and their presence in food can have devastating consequences, in some cases leading to death.

Generally, the Genus *Bacillus* consists of over 250 species distributed across various terrestrial and aquatic environments (Siefert *et al.*, 2000). Most of the *Bacillus* species produce spores as a means of surviving harsh environmental conditions that would have otherwise inactivated

vegetative cells (Cutting, 2011). In the past this genus had been identified by biochemical analysis often through laborious means. The development of gene sequencing technologies has however improved the ease, reliability and reproducibility of test results (Wu *et al.,* 2006). Until recently the most popular and widely used sequencing technique was the 16S *rRNA* gene sequencing (Wong *et al*., 2003), and to a considerable extent sequencing using protein coding genes. However, in recent years there has been an increase in the use of Next Generation Sequencing Technologies (NGS) as whole genome sequencing as a means of identification and characterisation of bacteria. This change in tide from partial sequencing to WGS is mainly as a result of the level of discrimination the later offers especially at the sub-specie level, as well as the fall in costs for genome sequencing. Presently, identification at the strain level of *B. sporothermodurans* vegetative cells and spores is mainly through molecular techniques utilizing the polymerase chain reaction (PCR), 16S rDNA sequencing and biochemical methods (Heyndrickx *et al*., 2012; Scheldeman *et al*., 2002). In the past, RAPDs (random amplified polymorphic DNA) analysis, ITS-PCR and REP-PCR (repetitive element sequence based PCR) have been utilized. The latter showing more promise at discriminating to the strain level for *B. sporothermodurans* spores. REP-PCR has been effectively used to discriminate Gram positive bacteria such as *Listeria monocytogenes* (Jersek *et al*., 1996) and *B. sporothermodurans* (Herman *et al.,* 1998; De Jonghe *et al.,* 2010). REP-PCR allows the identification of *B. sporothermodurans* by the generation of a typical pattern as well as discriminating clonally related isolates (Herman *et al.*, 1998; Herman and Heyndrickx, 2000).

Mohapatra and Mazumder (2008), compared the efficacy of five different REP-PCR methods in discriminating *Escherichia coli* populations. The $(GTG)_5$ PCR technique yielded the highest average rate of correct classification of 86.8% as compared to BOX-PCR (82.3%), REP-PCR (78.4%), ERIC-PCR (72.6%) and ERIC2-PCR (55.8%) respectively. The $(GTG)_5$ PCR has

since been used successfully in the differentiation of *Lactobacillus* species (De Vuyst *et al*., 2008), *Salmonella enterica* isolates (Rasschaert *et al*., 2005) and for *B. subtilis*, *B. cereus* and *B. licheniformis* (Coorevits *et al.,* 2008; De Jonghe *et al.,* 2008). The $(GTG)_5$ PCR has been effectively used on *B. sporothermodurans* strains isolated from South Africa and Brazil for diversity studies (unpublished data). Pina *et al.* (2005) utilised a combination of the $(GTG)_5$ primer together with $(GAC)_5$ primer to track yeast contamination. The $(GAC)_5$ is yet to be used in *B. sporothermodurans* diversity studies.

2.13.2 PCR identification of *B. sporothermodurans* spores

PCR primers for the specific detection of *B. sporothermodurans* spores were designed by subtractive hybridization by Herman *et al*. (1997) (Figure 2.8). The specificity of these primers has been proven on a large collection of *Bacillus* strains and on strains from relevant taxa. The detection limits of this PCR method are 9, 0.4, and 0.22 CFU ml^{-1} for 1, 10, and 100 ml of milk respectively (Herman *et al.*, 1997). This PCR reaction has been described as heat resistant spore (HRS)-specific PCR, considering that it is specific for *B. sporothermodurans* of UHT milk origin (Scheldeman *et al.*, 2002). The primers were deduced from the 16S rDNA sequence.

Scheldeman *et al*. (2002) subsequently designed PCR specific primers for the identification of *B. sporothermodurans*. Unlike the HRS specific primers earlier described by Herman *et al*. (1997) which detect only *B. sporothermodurans* specifically from UHT milk origin, the primers BSPO-F2 (5′ ACG GCT CAA CCG TGG AG 3′), located at positions 589–605 of the *B. sporothermodurans* sequence (U49078), and BSPO-R2 (5′ GTA ACC TCG CGG TCT A 3′), located at positions 1252–1237, have been used to detect *B. sporothermodurans* from UHT milk, milk products and other non-milk sources and environments, including the dairy farm.

Subsequently these primers have been used in Real Time PCR to detect *B. sporothermodurans* (Tabit and Buys, 2011) isolated from UHT milk (Figure 2.9).

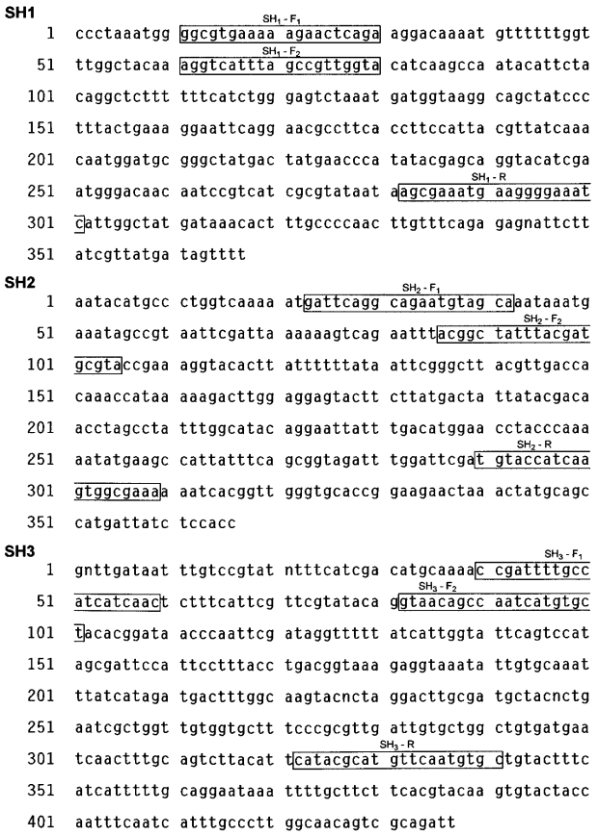


Figure 2.8: PCR primer specific to *B. sporothermodurans* designed using subtractive hybridization (Herman *et al.*, 1997).

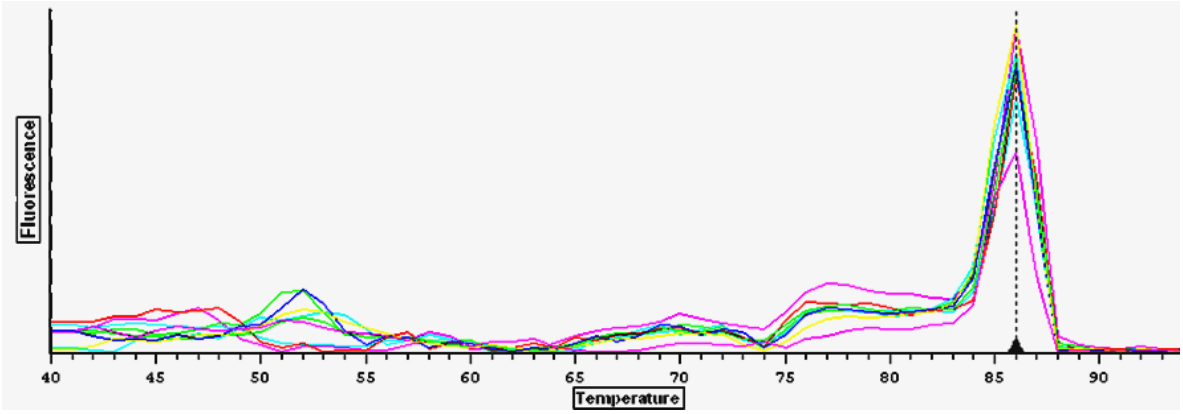


Figure 2.9: Real time PCR melting curve (Tm = 86) specific for *B. sporothermodurans* strains using the BSPO-F2 and BSPO-R2 primers designed by Scheldeman *et al*. (2002). Lines represent the melting curve of different *B. sporothermodurans* isolates (Tabit and Buys, 2011)

2.13.3 RAPD PCR

A Random Amplified Polymorphic DNA (RAPD)-PCR using the primers RAPD I [5'-GTCGTTATGCGGTA-3'] and RAPD II [5'-GAAGCAGCGTGG-3'] have been used to genetically characterize *B. sporothermodurans* in contaminated UHT milk. The generation of DNA fingerprint using RAPD is a quick and reliable method for the identification of *B. sporothermodurans* (Klijn *et al.*, 1997).

2.13.4 REP PCR typing

The repetitive extragenic palindromic polymerase chain reaction (REP-PCR) with the primers REP 1R-I (5' III ICG ICG ICA TCI GGC 3') and REP 2-I (5' ICG ICT TAT CIG GCC TAC 3') or ERIC 1R (5' ATG TAA GCT CCT GGG GAT TCA C 3') and ERIC 2 (5' AAG TAA

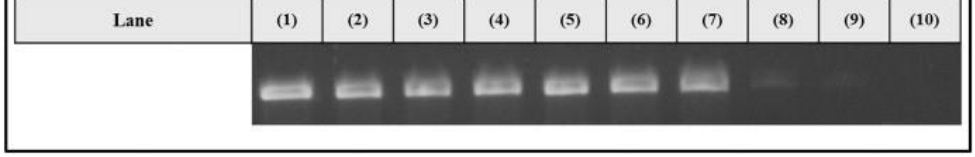
GTG ACT GGG GTG AGC G 3') has been used in the identification and discrimination of clonally related *B. sporothermodurans* isolates from UHT and other milk products. When compared to RAPDs, REP PCR is more reproducible because of the utilisation of a conserved primer and the inflexibility of the annealing temperature (Herman and Heyndrickx, 2000). Alternatively, a (REP)-PCR using the primers REP1R-I (5′ IIIICGICGICATCIGGC 3′) and REP2-I (5′ ICGICTTATCIGGCCTAC 3′), has also been used to characterise genetic heterogeneity in thirty-eight strains of *B. sporothermodurans* isolates (Klijn *et al.*, 1997; Guillaume-gentil *et al.*, 2002).

2.13.5 Propidium monoazide PCR detection

Propidium monoazide (PMA) PCR assay can be used to selectively identify viable cells of *B. sporothermodurans* in UHT milk mixed with dead cells of other bacteria (Cattani *et al.*, 2013). PMA enter dead cells and block their DNA hence preventing them from being amplified. Direct PCR with DNA extracted from UHT milk without culturing cannot discriminate between DNA from living *B. sporothermodurans* and DNA from other dead bacteria. The amplification of DNA from dead bacteria closely related to *B. sporothermodurans* has been found to give a false positive PCR result. On the other hand, PMA-PCR will ensure that only viable *B. sporothermodurans* cells are detected. The concentration of PMA does not interfere with the amplification of the DNA of viable cells while the DNA of dead cells affect the affect the overall PCR results in the absence of PMA; Figure 2.10 (Cattani *et al.*, 2013; Cattani *et al.*, 2016).

A

	Control (without PMA treatment)					PMA treatment (30 µg/mL)				
Ratio viable: dead cells	100:0	75:25	50:50	25:75	0:100	100:0	75:25	50:50	25:75	0:100
Lane	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)



B

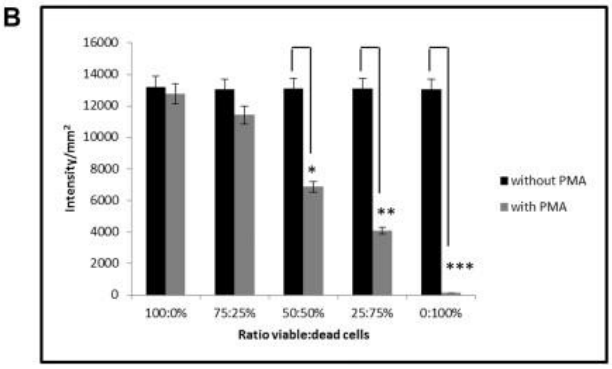


Figure 2.10: The effect of propidium monoazide (PMA) treatment for the detection by PCR of viable and isopropanol-killed *B. sporothermodurans* CBMAI 148 cells mixed at different ratios. (A) PCR products visualized on agarose gel stained with 0.5 µg mL-1 ethidium bromide under UV light. (B) Diagram representing the band intensities from amplicons generated with or without PMA treatment using different ratios of viable: dead cells. *$p < 0.001$; **$p < 0.005$; ***$p < 0.001$ (Cattani et al., 2013).

2.14 Sequencing as a means of *B. sporothermodurans* identification

2.14.1 16S rDNA sequencing

The 16S rDNA sequencing has been used extensively in the characterisation of the genus *Bacillus*. This is especially so for *Bacillus* strains related to the food industry. Yuan *et al.* (2012), used 16S *rRNA* gene to characterize isolated strains from commercial milk powders.

This was done in conjunction with SDS-PAGE of whole cell proteins to effectively characterize the strains involved. This level of combination especially in this case suggests some level of inadequacy of the 16S *rRNA* gene sequencing in effectively characterising isolates. In a similar test, 16S *rRNA* gene sequencing this time together with DNA-DNA hybridisations of representative strains (Scheldeman *et al.,* 2004). These combination of methodologies makes distinguishing closely related isolates more reliable, but again gives an indication of the handicap in the use of 16S *rRNA* gene sequencing as a standalone method. This is especially the case when it relates to sequencing of closely related organisms (Fox *et al.*, 1992). The main shortcoming associated with 16S *rRNA* gene sequencing is its lack of taxonomic resolution to effectively differentiate closely related species.

With respect to the identification of *B. sporothermodurans* strains, (de Silva *et al.,* 1998; Guillaume-gentil *et al.,* 2002; Montanari *et al.,* 2004), used 16S *rRNA* gene sequencing to characterize isolates. The latter test in this case gave low bootstrap values and short branches. The precise evolutionary positions of the isolates could not be decoded using the 16S *rRNA* gene sequences, further confirming the vulnerability of the 16S *rRNA* gene sequencing.

2.14.2 An alternative to 16S rDNA sequencing

As a result of the growing realisation of the handicap of 16S *rRNA* gene sequencing, recent sequencing techniques had utilised selected housekeeping genes (including *rpoB* gene). The *rpoB* gene encodes for the β subunit of RNA polymerase, is common to all bacteria and like the 16S *rRNA* gene is a mosaic of conserved and variable sequence domains (Meintanis *et al.,* 2008). Unlike other housekeeping genes (including *gyrB*, *nifD*, *recA* and *atpD*) the *rpoB* gene has many of its sequences available in public databases including a strain of *B.*

sporothermodurans (USI2006-BC007) isolated in Hong Kong. Compared to the 16S *rRNA* sequencing gene though, the number of *rpoB* genes sequenced is just a minor fraction.

Nonetheless, the *rpoB* gene has some advantages over the 16S *rRNA* gene and has been proposed as a more effective marker over the 16S *rRNA* gene (Walsh *et al.*, 2004). Firstly, the *rpoB* gene is homogenous within cells. It is a single copy gene with long sequences (Ki *et al.*, 2009). Blackwood *et al*. (2004) identified that the *rpoB* gene proved to be the best alternative target with a conserved 4-nucleotide characterisation between *B. cereus* and *B. anthracis*. Consequently, only few members of the *Bacillus* genus have had their *rpoB* genes sequences documented (Ki *et al.*, 2009).

Other studies have confirmed a more accurate discrimination by use of the *rpoB* gene in sequencing the genus *Bacillus* (Meintanis *et al.*, 2008; Ki *et al.*, 2009). However, until recently the 16S *rRNA* gene, as a result of its essential function, ubiquity and evolutionary properties had become the most utilised molecular marker (Hettick *et al.*, 2006) it still has one major flaw. Multiple copies of the 16S *rRNA* gene are usually present in a given bacterium and subsequently differ in sequence (Case *et al*., 2006).

Case *et al.* (2006) found in microbial ecology studies that the *rpoB* gene was comparable to the 16S *rRNA* gene with respect to phylogenetic resolution at all taxonomic levels. However, between closely related organisms at the specie level the former provided better resolution. Other shortfalls to the 16S *rRNA* gene is the lack of recognition of novel taxa, too few sequences deposited in nucleotide databases, species sharing similar and/or identical 16S *rRNA* gene sequences, or nomenclature complications as a result of multiple genomovars allocated to single species or complexes (Janda and Abbott, 2007). These bottlenecks are however not exclusive to the 16 rRNA gene only.

2.15 Other heat resistant *Bacillus* species affecting the dairy industry

To date, *B. sporothermodurans* remains one of the most heat resistant *Bacillus* species of concern especially to the dairy industry. Another spore-forming *Bacillus* specie that affects the dairy industry and has high heat resistance is *B. thermoamylovorans*. Though it was first described as non-spore-forming, a recent study confirmed otherwise (Coorevits *et al.*, 2011). Like *B. sporothermodurans*, *B. thermoamylovorans* is an emerging threat to the dairy industry albeit not as heat resistant as the former. *B. thermoamylovorans* has been implicated in other foods besides dairy (Flint *et al.*, 2016) and looks to follow the footstep of *B. sporothermodurans* with the potential of contaminating milking equipment, animal feed and milk processing equipment (Coorevits *et al.*, 2008; De Jonghe *et al.*, 2008).

B. thermoamylovorans was first isolated in Dakar, Senegal, from palm wine, a local alcoholic beverage popular in West Africa (Combet-Blanc *et al*., 1995). As a result of its niche, and as its name suggests it utilises starch at very high temperatures (Flint *et al.*, 2016), unlike its sibling *B. sporothermodurans* which is unable to hydrolyse starch (Pettersson *et al.*, 1996; Klijn *et al.*, 1997; Montanari *et al.*, 2004). Its colonies are like *B. sporothermodurans*, both having small, white to beige colonies 2 – 3 mm in diameter. Though its optimum temperature is 50 °C as opposed to 37 °C in *B. sporothermodurans*, it has also been isolated at temperatures of about 37 °C (Coorevits *et al.*, 2008; Flint *et al.*, 2016). This shows the high adaptability of these spore-formers and of the possible effects of strain differences. Regarding heat resistance, *B. thermoamylovorans* showed a D_{120} of up to 1.9 s ± 0.2 (Berendsen *et al.*, 2015), compared to D_{140} of up to 7.0 s ± 0.2 (Pettersson *et al.*, 1996). Though *B. sporothermodurans* and *B. thermoamylovorans* affect the quality and value of dairy products, their high heat resistance,

non-pathogenic capacity and the ability to utilise starch at high temperatures (*B. thermoamylovorans*) makes them interesting candidates for use in beneficial biotechnology.

2.16 Bacterial genomics and pan-genomics

2.16.1 Applications of whole genome sequencing

Using the crystallographic data produced by Rosalind Franklin and Maurice Wilkins, Watson and Crick determined the three-dimensional structure of DNA in 1953 (Zallen, 2003). Since then there has been a remarkable growth in the use of DNA based technology to understand biological life. The order of nucleotides in biological matter contains the genetic information that determines and differentiates life. Bacterial genome sequencing is almost 25 years old and since the publication of the first bacterial genome sequence from *Haemophilus influenzae*, genome sequencing has helped tackle various biological questions in the biomedical, food and agriculture based industries.

Bacterial whole genome sequencing involves the determination of the complete DNA sequence of the said organism at a single time (Heather and Chain, 2016), and requires the sequencing of an organisms chromosomal or plasmid DNA. Whole genome sequencing has especially been utilized in the biomedical industry for various clinical and epidemiological studies (Tagini and Greub, 2017; Schürch *et al.,* 2018). Its use in the food industry has especially been extensive over the past few years, with applications in food safety and security (Pritchard *et al.,* 2016; Zheng *et al.,* 2016), and in the selection of starter cultures in the dairy industry (Kelleher *et al.,* 2015). Furthermore, WGS has greatly enhanced insights into the antimicrobial resistance

(AMR) (Gillings *et al.*, 2017; Reding-Roman *et al.*, 2017; Reznick and King, 2017) and studies into various human and food microbiomes (Adu-Oppong *et al.*, 2017; Kim *et al.*, 2017).

WGS and its use during food pathogen outbreaks helps confirm hypothesis quickly, correlate laboratory and epidemiological data and accelerate response due to data sharing (Nadon, 2016). WGS was used successfully in deciphering the Listeriosis outbreak in the United States and Denmark in 2014 and the *Salmonella enteritidis* case in the UK in 2014 (Nadon, 2016). And in 2018 in South Africa, WGS was used to identify the source and strain sequence type (ST6) of the largest Listeriosis outbreak recorded to date. WGS has proved even more useful when incorporated with geographic information, allowing for the effective tracking of food pathogen outbreaks and contaminated food from around the world (FDA, 2016). Furthermore, the influence of WGS has been greatly felt in the area of comparative genome analysis (Poptsova and Gogarten, 2010; Sharma and Satyanarayana, 2013; Magnoerez-Bryan *et al.*, 2015). Insights into bacterial genome diversity (Cordero and Polz, 2014), genome architecture (Bobay and Ochman, 2017) horizontal gene transfer (HGT) and evolution (Gillings, 2017), have sought to explain the intricacies of prokaryotes.

WGS has evolved over the years and various generations of sequencing technologies has sought to increase coverage whiles limiting errors, with the view of providing a swift and cost effective tool for the analysis of bacterial genomes. DNA sequencing should be reproducible and reliable, and the development of the Sanger chain termination method in 1977 and subsequent technologies has since allowed this to be so (Di Resta and Ferrari, 2018). Just about 15 years ago, most genomes were sequenced by the Sanger methods, mostly led by consortia to sequence model organisms such as *E. coli* and *B. subtilis* (Land *et al.*, 2015). It was primarily a factory based model involving robots selecting and growing clones of whole genome shotgun

libraries, isolating the templates for the sequencing and performing sequence reactions and subsequently undertaking electrophoresis (Land *et al.*, 2015). This was the first generation of sequencing technologies and led to WGS information of *Mycobacterium tuberculosis* (Mahadevan *et al.*, 1998), and *Yersinia pestis* (Whitehead *et al.*, 2002). The first generation of sequencing subsequently threw more light on pathogens that where difficult to study *in vitro* including *Treponema pallidum*, *M. leprae* and *Tropheryma whipplei* (Land *et al.*, 2015). The high quality draft genomes obtained from the Sanger method was very labour intensive and could cost as much as $50,000 for fully completed genomes (Land *et al.*, 2015), and as result was mostly undertaken by groups of sequencing centres.

To improve the accessibility of WGS and reduce cost, high throughput Next Generation Sequencing (NGS) was born in the late 2000s (Metzker, 2010) and allowed various technological companies to get involved and create desktop sequencing platforms. The first second-generation sequencing technology was the Roche 454, with Illumina technologies the most commonly used platform now. NGS bases WGS has consequently reduced the cost of sequencing genomes to ≈$1 now. The NGS technologies produce shorter reads than the earlier shotgun sequencing by Sanger and has resulted in the boom of especially bacterial genomes that are available publicly. NGS technologies has extensively been used in the growing areas of microbiome research, with applications in culture independent research into microbial ecology and disease diagnosis (Sheppard *et al.*, 2013; Hanage, 2014)

Recent developments in sequencing technologies has resulted in third and fourth generation sequencing platforms. The single-molecule real-time (SMRT) sequencing platform by Pacific Biosciences (PacBio) offers longer reads than second generation sequencing technologies (Rhoads and Au, 2015). The SMRT sequencing platform normally will produce one contig for

a bacterial genome and as such well suited for unsolved problems in genome, transcriptome and epigenetics research (Rhoads and Au, 2015). As a result of the longer reads it can be used to close gaps in reference sequences and used to detect mutations and discoveries of novel genes and novel isoforms of annotated genes. Though it has a higher error rate than previous generations of sequencing technologies, its combination with previous technologies makes it a powerful tool in bacterial comparative studies and especially in the case of novel bacterial genomes (Roberts *et al.*, 2013). The combination of SMRT and NGS based sequencing technologies thus result in the production of high quality complete genomes experienced with Sanger sequencing but at a fraction of the cost (Roberts *et al.*, 2013; Land *et al.*, 2015).

The fourth generation DNA sequencing technology, like the platform by Pacific Biosciences is based on the single-molecule technique and has the potential to sequence an entire human genome quickly and reliably for less than $1000. This generation of sequencer, based on nanopore technology allow for study of DNA-protein and protein-protein interactions (Feng *et al.*, 2015). Nanopore sequencing offer distinct advantages including label -free, ultra-long reads (10^4-10^6 bases), high throughput and low material requirement (Feng *et al.*, 2015) over previous generations of sequencing. There is also minimal sample preparation and the absence for amplification or modification in the case of nucleotides, polymerases or ligases (Feng *et al.*, 2015). The use of nanopore sequencing technology has resulted in the production of the MinION, the first commercial sequencer using this technology (Feng *et al.*, 2015), was released by Oxford Nanopore Technologies (ONT). MinION identifies DNA bases by measuring the changes in electrical conductivity generated as DNA strands pass through a biological pore (Tyler *et al.*, 2018). A handheld device that fits into the USB port of a laptop, it is portable and affordable. Its size, speed and ease of data production and analysis, makes it suitable for real-

time applications and in remote and resource limited locations (Lu *et al.*, 2016; Tyler *et al.*, 2018)

2.16.2 Impact of genomics on food safety and quality

One of the key concerns of the food processing industry is the contamination and growth of spoilage and pathogenic microbes (Brul *et al.*, 2006). As a result of the ubiquity of these microorganisms, there is always a challenge of minimising the probable contamination routes. In the event of contamination, especially where pathogens are involved, it is critical that the purported organisms are identified and inactivated as soon as possible, or the food discarded. Modern DNA based techniques have the advantage of being quicker and more reliable in detecting microbial contaminants. Hence, the development of genetic markers for various contaminating bacteria, is key to effective and quick identification of a possible food pathogen outbreak (Pecora *et al.*, 2015; Zheng *et al.*, 2016; Allard *et al.*, 2018). Aside its use in food safety and quality, WGS is being increasingly used in the comparison of closely related bacterial species to identify prospective beneficial bacteria species in industry. For example the genomic insights into *B. coagulans* has shed more light into its probiotic potential (Upadrasta *et al.*, 2016) and on the lignin reducing ability of strains of *B. pumilus* isolated from the gut of wood boring *Mesomorphus* sp. (Balsingh *et al.*, 2016). WGS has also been used in purely bacterial comparative studies seeking to identify relatedness amongst species with respect to phenotypic characters as heat resistance and niche adaptation (Krawczyk *et al.*, 2015; Krawczyk *et al.*, 2015; Berendsen *et al.*, 2016; Krawczyk *et al.*, 2016; Wang *et al.*, 2016).

In modern genomic approaches the investigator makes use of the whole set of biomolecules to identify a probable pathogen (Brul *et al.*, 2006). The combination of genomics (total DNA),

transcriptomics (expressed RNA), proteomics (all cellular proteins) and metabolomics (all generated metabolites) in food quality control allow for effective identification and monitoring of unwanted microbes. The genomic era has revolutionised the tracking of contaminants from farm to fork and helps guarantee the identification of contaminants at sections along the entire food supply chain. Recent developments in third generation sequencing technologies as PacBio Sequencing in conjunction with second generation sequencing technologies as Illumina MiSeq, allow for better resolution of bacterial genomes. This subsequently has enlightened scientists to the differences that exist amongst sub-species and of their varied phenotypic characteristics. It is worth noting that, even though genomics has revolutionised food safety and quality, it is important for these techniques to be integrated with classical physiology. This has brought about the term functional genomics where the understanding of cellular processes is used to support genomic data to have a holistic view of bacterial physiology and function (Brul *et al.*, 2006).

2.16.3 Bacterial plasticity and identity

The advent of next generation sequencing technologies has afforded scientists the study of genome architecture and its effects on function (Patel, 2016). The complexity of the bacterial genome is only being understood now with developments in WGS technologies and their ability to identify differences that exists at the strain level. For the bacteria to survive, it is important to for it to have some sort of integrity and stability (Vettone *et al.*, 2014). Ironically however, in order to adapt to its changing environment, a bacterial genome should be able to undergo some extent of transformation. Through various adaptive responses over a period, microbes develop various genotypic characteristics with corresponding phenotypic repair,

offence and defence mechanisms (Laconi *et al.*, 2015). Thus, through the evolutionary past of a microbe's existence it plays that careful balance between stability to preserve its core function and plasticity to adapt to changing environmental conditions. With both these situations ensuring its survival and subsequent ability to colonise various niches. Some of the factors that contribute to bacterial cells adapting to changing environments include mobile DNA elements, horizontal gene transfer, genome rearrangements events (Juhas *et al.*, 2009).

The recent surge of NGS technologies unearthed mobile genetic elements (MGE) as one of the main drivers of genome plasticity. These MGE often made up of viruses, plasmids and together known as the mobilome (Wolf *et al.*, 2001) has been implicated in many bacterial characteristics as high heat resistance and virulence. These MGE's include transposons, insertion sequences (ISs), repetitive extragenic palindromic (REP) sequences, miniature inverted repeat transposable elements (MITEs), transposable bacteriophages and introns (Siguier and File, 2006). Research by Roberts & Mullany (2011) suggests that transposons of the Tn*916*/Tn*1545* family are adept at transferring between species and responsible for antimicrobial resistance in microbes as *Clostridium difficile*, *Staphylococcus aureus* and *Streptococcus pneumoniae*. Recent developments in this area of MGEs have proved that some highly heat resistant spore formers including *B. sporothermodurans* have increased heat tolerance as a result of the Tn*1546*-like transposon that carry the *spoVA*^{2mob} operon identified as conferring high heat resistance to bacterial spores (Berendsen *et al.*, 2015; Berendsen *et al.*, 2016).

Horizontal gene transfer (HGT) is another driver of genome plasticity. Alternatively known as lateral gene transfer (LGT), is the transfer of genetic material from other species which occurs during transformation, conjugation or transduction (Dorman, 2014). HGT can thus infer

virulence from a pathogen to a non-pathogen over several generations. Known to have both beneficial and disruptive effects, extreme rates of HGT can be countered by clustered regularly interspaced short palindromic repeats (CRISPR) (Marrafini and Sontheimer, 2010).

2.16.4 Bacterial pangenome

The developments of high volumes of sequenced data over the past few years has made access to genomes more accessible (Tettelin *et al.,* 2008). Comparative genomics borne out of the influx of sequenced genomes allows for detailed comparisons between strains and amongst species etc. Coined in 2005 (Tettelin *et al.,* 2005) the pangenome is the whole gene repertoire of a group microbes (Medini *et al.,* 2005; Tettelin *et al.,* 2008; Georgiades and Raoult, 2011). The pangenome is made up of the core genome (genes common to all genomes), accessory genome (genes present between two or more but not all groups) and of genes unique to a single species in a study group. Several tools have been developed for pangenomic studies (Xiao *et al.,* 2015) each with their unique selling points. Two of the earliest tools Panseq (Laing *et al.,* 2010) and PanCGHweb (Bayjanov *et al.,* 2010), the latter based on the PanCGH algorithm (Bayjanov *et al*., 2009) were published in 2010. CAMBer (Wozniak *et al.,* 2011), and the Prokaryotic genome Analysis Tool (PGAT) (Brittnacher *et al.,* 2011) were later published in 2011. Subsequently, other tools including Pan-genome analysis pipeline (PGAP) (Zhao *et al.,* 2012), GET_HOMOLOGUES (Contreras-moreira and Vinuesa, 2013) PanCake (Ernst and Rahmann, 2013) and PANNOTATOR (Santos *et al.,* 2013) were outdoored. The Integrated Toolkit for Exploration of microbial Pan-genomes (ITEP) (Benedict *et al.,* 2014), PanGP (Zhao *et al.,* 2014), an updated version of CAMBer called ECAMBer (Wozniak *et al.,* 2014) and the Spine and AGEnt (Ozer *et al.,* 2014) software were developed in 2014, with ClustAGE (Ozer,

2018) a recent tool for the clustering and distribution analysis of bacterial accessory genomic elements. Recently, the PGAP analysis suite was updated with version PGAP-X (Zhao *et al.*, 2018) and a web-based version PGAweb (Zhao *et al.*, 2018). These pangenome analysis tools mentioned are available either as web-based, Windows and Linux based platforms or a combination of these. Their features may include a combination of the clustering of homologous genes and analysing core/accessory genomes, identifying SNPs, plotting pan-genome profiles, determining phylogenetic relationships, annotation and visualization. ITEP, PGAP and GET_HOMOLOGUES process over five or more of the previously mentioned features, with the others offering from one to four of these features (Xiao *et al*., 2015).

One of the relatively more user friendly tools is the 'Spine and AGEnt', which extracts the core and accessory genome of either draft or complete genomes of bacterial strains. Spine determines the core genome based on conserved sequences among the study group. Subsequently, AGEnt, uses an in silico genome subtraction techniques to filter out the core genome sequences thereby detecting the accessory genomic sequences. The Spine and AGEnt process is depicted in Figure 2.11. Written in Perl and utilizing BLAST, ClustAGE can then be used to cluster and identify the distribution of the accessory genomic elements previously identified using Spine and AGEnt (Figure 2.12). The bacterial pan-genome is effective in the analysis of multiple strains or different species to infer common and unique genes in their gene repertoire. This method has been used to profile *Bacillus* species that exist in food microbiomes (Kim *et al.*, 2017) to good effect, identifying the gene repertoire of *B. amyloliquefaciens, B. anthracis, B. cereus, B. subtilis* and *B. thuringiensis*. 'Spine and AGEnt' effectively identified the accessory genome of *Pseudomonas aeruginosa* strains, and inferring accessory genome makeup with genotype (Kung *et al.*, 2010; Ozer *et al.*, 2014).

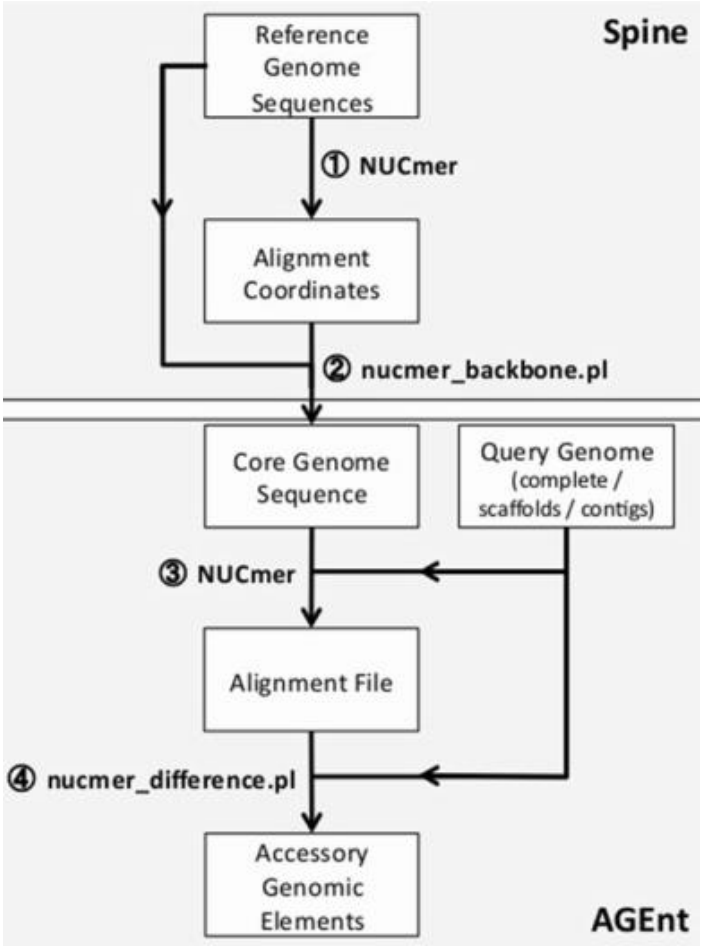


Figure 2.11: Approach to accessory genomic element identification by Spine and AGEnt.Programs used to accomplish the listed steps are indicated by circled numbers: 1, 3, NUCmer (whole-genome aligner); 2, nucmer_backbone.pl (converts coordinates of conserved regions to DNA sequence); 4, nucmer_difference.pl (subtracts regions not aligning to core) (Ozer et al., 2014).

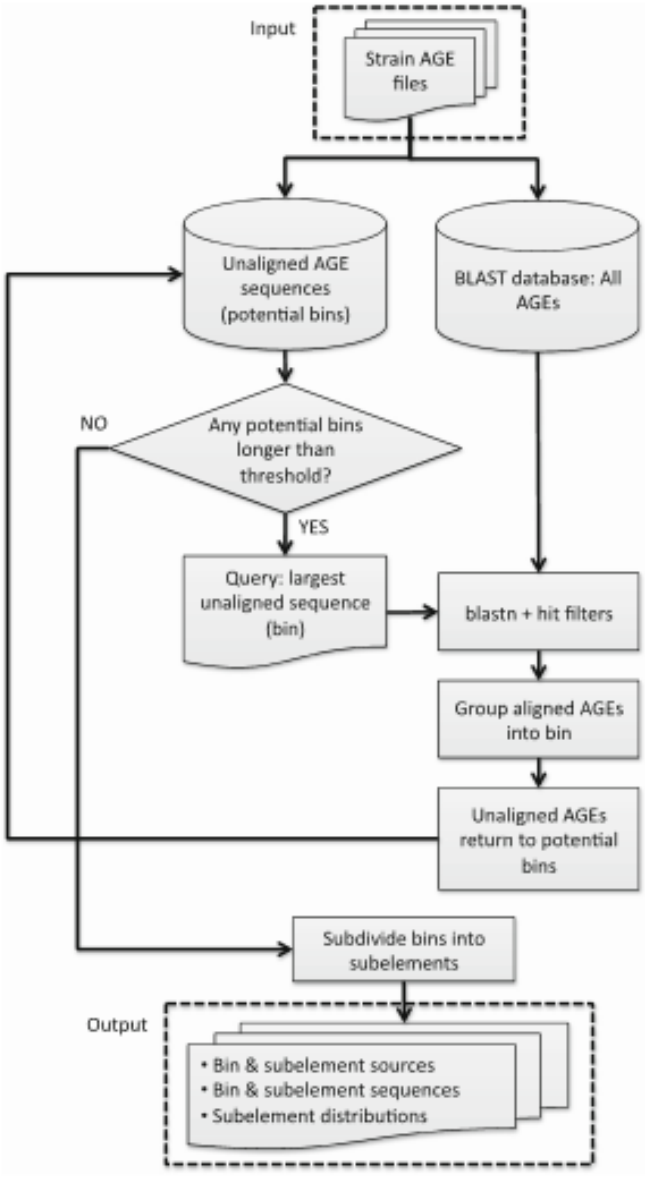


Figure 2.12: Method of clustering accessory genomic elements from nucleotide sequences of bacterial strains (Ozer 2018).

2.17 Hypotheses

The application of the Weibullian equation would provide a more accurate thermal sterilization process and D-value determination for *B. sporothermodurans* than the linear model. The linear type model, based on first order kinetics does not depend on the exposure duration (Peleg *et al.*, 2005), and with no regard to the behaviour and heterogeneity of the microbial population. Survival curves for heat resistant spore forming bacteria are described by a power law model (Corradini *et al.*, 2005) that takes into consideration population type (homogenous or otherwise), and bacterial/spore injury and survival.

B. sporothermodurans HRS and non-HRS strains isolated from UHT milk will show similar heat inactivation profiles. Recovery of cells is done in a medium that favour growth and as a result is insensitive to whether inactivated bacterial spores are either intact or injured (Corradini and Peleg, 2007).

As opposed to 16S rRNA sequencing, WGS will effectively discriminate HRS and non-HRS clones of *B. sporothermodurans* and highlight the possible confirmation of HGT and interaction with other Firmicutes with respect to sporulation, heat resistance and virulence factors (Schroder *et al.*, 1993; Cybulski *et al.*, 2009; McGillivray *et al.*, 2009). 16S rRNA sequencing is limited by the presence of variable copy numbers and sequence variation within closely related taxa (Větrovský and Baldrian, 2013) and consequently closely related taxa of *B. sporothermodurans* previously identified by 16S rRNA sequencing (Kuhnigk *et al.*, 1995) may show differing relatedness with WGS.

Pangenome of *B. sporothermodurans* will exhibit a core to pangenome ratio of less than 89%, depicting an open genome rather than closed one. *B. sporothermodurans* is sympatric, having been shown to exist in non-homogenous niches, and being identified with other *Bacillus* sp. and other Firmicutes in general (Rouli *et al.*, 2015; Mugadza *et al.*, 2018), and as such will possess more CRISPRs (Sheppard *et al.*, 2013), a scenario likely to lead to a high horizontal rate of gene transfer and consequently an open genome (Georgiades and Raoult, 2011).

2.18 Objectives

- To determine the D-value and thermal inactivation dynamics of *B. sporothermodurans* spores using the Weibullian equation to identify differences in heat resistance between an HRS and non-HRS clone.

- To determine the WGS of *B. sporothermodurans* strains and its closest phylogenetic neighbour with the aim of undertaking comparative genomic analysis amongst strains and with other *Bacillus* species.

- To determine the pangenome of *B. sporothermodurans* with the aim of identifying the distribution of COG categories and comparative sizes of the core and accessory genome amongst *B. sporothermodurans* strains and in comparison, with closely related *Bacillus* species.

Chapter 3: DRAFT GENOME SEQUENCE OF *BACILLUS OLERONIUS* DSM 9356^{T} ISOLATED FROM THE TERMITE *RETICULITERMES SANTONENSIS*

Published: Genomics Data 12 (2017) 76-78

3.1 Abstract

Bacillus oleronius strain DSM 9356^T isolated from the Termite *Reticulitermes santonensis* was sequenced to gain insights in relation to its closest phylogenetic neighbor *Bacillus sporothermodurans*. The draft genome of strain DSM 9356^T contains 5,083,966 bp with an estimated G + C content of 35%, 4899 protein-coding genes, 116 tRNAs and 18 rRNAs. The RAST annotation assigned these genes into 462 subsystems, with the maximum number of genes associated with amino acids and derivatives metabolism (14.84%), followed by carbohydrates (13.89%) and protein metabolism subsystems (9.10%). The draft genome sequence and annotation has been deposited at NCBI under the accession number MTLA00000000.

Keywords: Genome sequence, *Bacillus oleronius*, *Reticulitermes santonensis*, Gram-negative, endospore

3.2 Introduction

The genus *Bacillus* is a group of Gram positive, rod-shaped bacteria distributed extensively in the environment. Their ubiquity in nature is because of their ability to produce endospores during adverse conditions. *Bacillus* species includes pathogens of clinical significance, bacterial contaminants in food and as important industrial organisms producing various enzymes. *Bacillus oleronius* is a non-motile endospore-forming bacterium which was originally isolated from the hindgut of the termite *Reticulitermes santonensis* (Feytaud), where it plays a symbiotic role by aiding digestion (Kuhnigk *et al.*, 1995). It is also found in the human skin parasitic mite *Demodex folliculorum*, and is suspected to be related to the development of rosacea, a chronic inflammatory dermatological condition in humans(Jar muda *et al.*, 2012). A school of thought is that *Demodex* mites are vectors for bacteria including *Staphylococcus albus* and *Microsporon canis* (Wolf *et al.*, 1988), and of interest, *B. oleronius* that cause and exacerbate skin lesions (Hsu *et al.*, 2017).

Despite staining Gram-negative, *B. oleronius* has Gram-positive cell wall components shared amongst all *Bacillus* species (Lacey *et al.*, 2007) and thus closely related to other *Bacillus* sp. that contaminate foods. Although initially isolated from the hindgut of the termite and subsequently from mites, *B. oleronius* has been identified as a potential contaminant of milk and dairy products and has been sporadically identified and isolated from fodder, raw milk and milk processing equipment (Vaerewijck *et al.*, 2001; Scheldeman *et al.*, 2002, 2005, 2006). *B. oleronius* is associated with the *Bacillus firmus-lentus* group (Kuhnigk *et al.*, 1995) with its 16S rRNA sequence 95.6% and 95.5% similar to that of *B. lentus* and *B. firmus* respectively. However, its closest phylogenetic neighbor is the highly heat resistant spore forming *B. sporothermodurans* (Heyndrickx *et al.*, 2012), which may survive ultra-high temperature

processing conditions during milk processing. Albeit not as heat resistant as *B. sporothermodurans*, *B. oleronius* spores has been known to survive milk pasteurization, with spores isolated after 30 minutes heating at 100 °C (Vaerewijck *et al.*, 2001). In this study, the type strain RT10 (DSM 9356^T) of *B. oleronius*, procured from the Leibniz Institute in Germany was selected for whole genome sequencing. The principal reason is to enhance understanding of this bacterium in relation to other spore-forming *Bacillus* species of importance to the dairy industry. The dataset has been submitted to NCBI and is reported here.

3.3 Materials and Methods

3.3.1 Culturing, DNA extraction and genome sequencing

Fresh 24h culture of *B. oleronius* obtained from Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH (DSMZ) was inoculated into nutrient agar broth (Oxoid, UK) and incubated at 37 °C for 24 hours. Genomic DNA was extracted using the ZR Bacterial DNA Miniprep kit (Zymo Research, USA). DNA extract was quantified using the Qubit instrument and dsDNA BR Assay kit (Life Technologies, USA). Multiplexed paired-end libraries were prepared using Nextera XT DNA Sample Preparation kit (Illumina, USA). Genome sequencing was carried out on an Illumina MiSeq system (Illumina, San Diego, USA). The paired-end reads were checked for quality, trimmed and *de novo* assembled using the Qiagen CLC Genomics Workbench version 9 (Qiagen, Netherlands). All resultant contigs were then submitted to GenBank, where gene annotation was implemented using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova *et al.*, 2016). The annotation was further uploaded to Rapid Annotation using Subsystem Technology (RAST) for subsystems-based annotation (Aziz *et al.*, 2008; Overbeek *et al.*, 2014; Brettin *et al.*, 2015).

Table 3.1: Specifications of *B. oleronius* DSM 9356^T genome

Organism	*Bacillus oleronius*
Strain	DSM 9356^T
Sequencer	Illumina MiSeq
Data format	Assembled
Experimental factors	Genome sequence of pure microbial culture
Experimental features	Genome sequence followed by assembly and annotation
Consent	N/A
Sample source	île d'Oleron, France

3.4 Results and Discussion

The assembly contains 587 contig sequences of longer than 500 bp, covers 5,083,966 bp with G + C content of 35.00%, an N_{50} of 543,331 bp and a longest contig size of 90,648 bp. The total number of 5,168 genes predicted by PGAP includes 4,899 protein coding genes, 130 pseudo genes, and 139 RNA genes (Table 3.2). The RAST annotation assigned these genes into 462 subsystems, with maximum number of genes associated with amino acids and derivatives metabolism (14.84%), followed by carbohydrates (13.89%) and protein metabolism subsystems (9.10%) (Figure 3.1).

Table 3.2: *B. oleronius* DSM 9356^T genome characteristics and resources.

S. No	Name	Genome characteristics and resources
1	NCBI Bioproject ID	PRJNA362282
2	NCBI Biosample ID	SAMN06237156
3	NCBI genome accession number	MTLA00000000
4	Sequence type	Illumina MiSeq
5	Total number of reads	2,305,932
6	Read length	300
7	Overall coverage	>100x
8	Estimated genome size	5,083,966 bp
9	G+C content (%)	35.00
10	Genes (total)	5168
11	Protein coding genes	4899
12	tRNA coding genes	116
13	rRNA coding genes	18
14	ncRNA coding genes	5
15	Pseudogenes	130

The direct link to the deposited data is found at https://www.ncbi.nlm.nih.gov/nuccore/MTLA00000000. The draft genome sequence of *B. oleronius* DSM 9356^T has been deposited at NCBI under the BioProject number PRJNA362282, BioSample number SAMN06237156 and Accession number MTLA00000000.

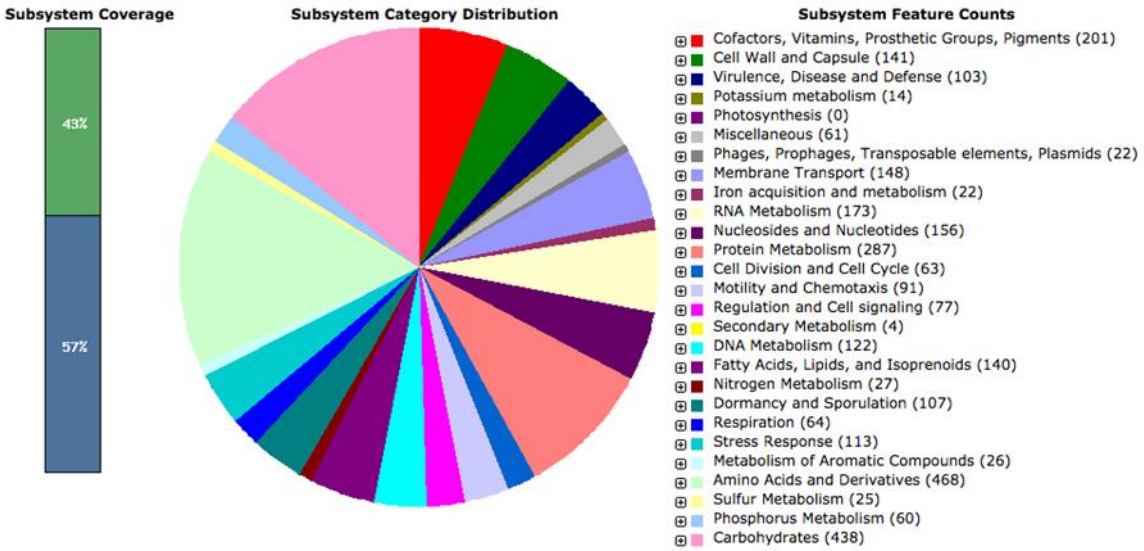


Figure 3.1: Distribution and counts of genes in COG categories for genome of *B. oleronius* strain DSM 9356^{T}

Chapter 4: GENOME SEQUENCES OF BACILLUS SPOROTHERMODURANS STRAINS ISOLATED FROM ULTRA HIGH TEMPERATURE (UHT) MILK

Accepted in Microbiology Resource Announcements

4.1 Abstract

We report the draft genome sequences of 3 *Bacillus sporothermodurans* strains isolated from Ultra High Temperature milk products in South Africa and Brazil and the type strain MB 581 (DSM10599). The genomes will provide valuable information on the molecular dynamics of heat resistance in *B. sporothermodurans*.

Keywords: Genome, food quality, *Bacillus sporothermodurans*, heat resistance

4.2 Introduction

Bacillus sporothermodurans is a thermo-resistant, Gram-positive bacterium that can produce highly heat resistant endospores (HRS), capable of surviving UHT heat treatments (Klijn *et al.*, 1997; Huemer *et al.*, 1998). First detected in UHT milk (Pettersson *et al.*, 1996), it has subsequently been isolated from other dairy products including UHT cream, chocolate milk, evaporated milk and reconstituted milk (Herman and Heyndrickx, 2000). Furthermore, *B. sporothermodurans* has been isolated from non-dairy based foods including Indian curry (Krawczyk *et al.*, 2016), as well as from marine sources (Ki *et al.*, 2009). After heat processing, the surviving spores may germinate and grow in the stored milk (Huemer *et al.*, 1998). The spores of *B. sporothermodurans* grow at low levels ($\approx 10^5$ CFU ml^{-1}) and do not affect the pH of the milk (Klijn *et al.*, 1997), as a result its presence may go unnoticed. If spoilage does occur though, there may be slight changes in colour, off-flavours and the destabilization of casein micelles (Klijn *et al.*, 1997). Consequently, the main concern to the dairy industry is milk quality, concerning non-sterility of milk products and of biofilm formation in milk processing equipment.

B. sporothermodurans closest phylogenetic neighbour, *B. oleronius*, initially confirmed by DNA-DNA hybridization (Scheldeman *et al.*, 2002) and described on the basis of a single strain isolated from the hindgut of the termite *Reticulitermes santonensis* (Kuhnigk *et al.*, 1995), was recently confirmed by whole genome sequencing in a previous study (Owusu-Darko *et al.*, 2017).

4.3 Materials and Methods

4.3.1 Culturing, DNA extraction and genome sequencing

B. sporothermodurans strains SAD and SA01 were isolated from UHT milk produced in South Africa and *B. sporothermodurans* strain BR12 was isolated from UHT milk from Brazil. The type strain *B. sporothermodurans* DSM 10599^{T} was obtained from Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH (DSMZ). Overnight fresh cultures of all four strains of *B. sporothermodurans* derived from single colonies were inoculated into brain heart infusion broth (Oxoid, UK) and incubated at 37 °C for 72 hours. Genomic DNA was extracted using the ZR Bacterial DNA Miniprep kit (Zymo Research, USA) and quantified using the Qubit instrument and dsDNA BR Assay kit (Life Technologies, USA). Multiplexed paired-end libraries were prepared using Nextera XT DNA Sample Preparation kit (Illumina, USA). Genome sequencing was carried out on an Illumina MiSeq system (Illumina, San Diego, USA) with read length of 2 X 300 bp. The paired-end reads were checked for quality (low quality Q < 20 filtered out), trimmed and *de novo* assembled using the Qiagen CLC Genomics Workbench version 9 (Qiagen, Netherlands) and with genome coverage of 100.0x. The resultant contigs were submitted to GenBank, where gene annotation was implemented using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP)(Tatusova *et al.*, 2016). The annotation was further uploaded to Rapid Annotation using Subsystem Technology (RAST) for subsystems-based annotation (Aziz *et al.*, 2008; Disz *et al.*, 2010; Overbeek *et al.*, 2014).

4.4 Results and Discussion

Overall, genomes of *B. sporothermodurans* contain heat shock and hyperosmotic proteins (including DnaJ, GrpE, GroEL), that will have an influence on the heat resistance and consequently the processing dynamics of food products. Additionally, all four strains sequenced contain the biofilm matrix protein component, TasA and its homologs, which have been shown to be the major biofilm matrix component (Branda *et al.*, 2006), especially in *B. subtilis*. Ultimately, the whole genome sequence of *B. sporothermodurans* will help improve our understanding of the heat resistance of this bacterium with the view of improving milk quality. The genomes sequences of all the four strains of *B. sporothermodurans* have been deposited at NCBI under the accession numbers listed in Table 4.1.

Table 4.1: Summary report of the de novo assembly of four *B. sporothermodurans* strains

Organism	**GenBank accession no.**	**SRA accession no.**	**Genome size (bp)**	**G + C content (mol%)**	**No. of coding sequences**	**Coverage (x)**	**No. of contigs**	**N_{50} (kpb)**
***B. sporothermodurans* DSM 10599^{T}**	NAZD01000000	SRR8741694	3,783,858	39.6	4257	226	527	15,402
***B. sporothermodurans* SAD**	NAZB01000000	SRR8732968	3,857,089	36.7	4111	175	110	114,649
***B. sporothermodurans* SA01**	NAZA01000000	SRR8732969	3,414,010	36.7	3768	146	290	22,386
***B. sporothermodurans* BR12**	NAZC01000000	SRR8741693	3,974,872	41.1	4558	193	805	9,377

Chapter 5: GENOME SEQUENCING AND HEAT INACTIVATION KINETICS OF *BACILLUS SPOROTHERMODURANS* STRAINS ISOLATED FROM ULTRA HIGH TEMPERATURE TREATED MILK

5.1 Abstract

B. sporothermodurans, capable of forming spores with high heat resistance causes the problem of non-sterility in UHT milk. In this study, four genomes of *B. sporothermodurans* strains were sequenced, including for the first time, high heat resistant and non-high heat resistant clones. Subsequently thermal inactivation studies were conducted in milk. The size of the genomes ranged from 3.4 - 3.9 Mb. Survival curves for heat inactivation data for both groups of strains showed thermal death times of 16 s for a 6-log reduction of bacterial spores at 135 °C using the linear model. Both the biphasic and Weibull models at 135 °C showed thermal death times of 14 s for a 6-log reduction of bacterial spores. All three mathematical models showed thermal death times of approximately 6 s at 140 °C indicating exceptionally high heat resistance and leading to considerable over-processing in a UHT process at that temperature. In our observations in literature, the only other known bacterial spores with similar high heat resistance, albeit lower, is the recently identified *B. thermoamylovorans*, also isolated from milk products. Information on heat inactivation protocols will enable the validation of UHT processing conditions for spore formers including *B. sporothermodurans*.

Keywords: highly heat resistant endospores, genome, milk quality, heat inactivation, Weibull

5.2 Introduction

Information on the genetic repertoire of genes associated with heat resistance and of heat inactivation properties of *B. sporothermodurans* is essential to the dairy industry, in order to solve problems in relation to the survival of spores in thermal processes. In addition, next generation sequencing technologies is increasingly being used to ascertain the genotypic properties of important bacteria in relation to their phenotypic characteristics in their respective niches (Lücking *et al.*, 2013; Wells-Bennik *et al.*, 2016).

B. sporothermodurans is a thermo-resistant, Gram-positive bacterium capable of highly heat resistance of its endospores (HRS), that may survive UHT heat treatments (Klijn *et al.*, 1997; Huemer *et al.*, 1998). Its closest phylogenetic neighbour, *B. oleronius*, initially confirmed by DNA-DNA hybridization (Scheldeman *et al.*, 2002) and subsequently by whole genome sequencing (chapter 3) was first described on the basis of a single strain isolated from the hindgut of the termite *Reticulitermes santonensis* (Kuhnigk *et al.*, 1995). The heat resistance of *B. sporothermodurans* has been attributed to clonal differences, though laboratory stains are thought to lose their heat resistance over time. The spread of the HRS clone, which is a concern to the dairy industry seems to be caused by reprocessing and circulation of contaminated milk within and between UHT production units. Furthermore, the detection of *B. sporothermodurans* strains has become more widespread throughout different continents, raising considerable concerns and challenges (Scheldeman *et al.*, 2006).

It's currently unclear exactly how the structural and chemical components of *B. sporothermodurans* spores influence their extreme wet heat resistance during heating at UHT temperatures (Klijn *et al*., 1997). However, it is believed that, certain environmental and or sub-lethal stress conditions may additionally affect their heat resistance. Albeit its poor growth,

UHT milk can probably be described as the ecological niche of *B. sporothermodurans* because of the lack of competition due to the high heat processing conditions associated with UHT milk. Consequently, the main concern to dairy industry is milk quality, concerning non-sterility of milk products and of biofilm formation in milk processing equipment.

Information on the heat resistance of *B. sporothermodurans* spores using heat inactivation models is essential for inactivating spores and to prevent over processing resulting in UHT products with poor organoleptic properties. Mathematical models have been proposed to describe the inactivation of microorganisms. Thermal processing is frequently cited as an example where mathematical modelling has been used traditionally for many years to predict the outcome of different processing conditions on the survival of microorganisms (Peleg and Cole, 1998). The inactivation of microorganisms by heat and other processing methods has been traditionally assumed to follow first-order kinetics (Schaffner and Labuza, 1997); the *z* and *D* concepts are used to predict the survival curves. However, significant deviations from non-linearity have been reported in literature (Cerf, 1977; Peleg and Cole, 1998; Van Boekel, 2002). Over time, several models have been proposed to describe these non-linear survival curves. Among them, the Weibull model is gaining popularity due to its flexibility.

This study is the first multiple sequencing and heat inactivation analysis of *B. sporothermodurans* strains to include strains belonging to the HRS clone and vice versa. By sequencing the genome of multiple *B. sporothermodurans* strains, it may be possible to better understand the genetic differences between HRS and non-HRS clones and as well, its evolution with respect to other *Bacillus* species of importance to the food industry.

5.3 Materials and methods

5.3.1 Molecular identification and genome sequencing and analysis

B. sporothermodurans strains SAD and SA01 were isolated from UHT milk produced in South Africa and *B. sporothermodurans* strain BR12 was isolated from UHT milk from Brazil. The type strain *B. sporothermodurans* DSM 10599^{T} was obtained from Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH (DSMZ). Single colonies of overnight pure fresh cultures of all four strains of *B. sporothermodurans* were inoculated into brain heart infusion broth (Oxoid, UK) and incubated at 37 °C for 72 hours. Genomic DNA was extracted using the ZR Bacterial DNA Miniprep kit (Zymo Research, USA) and quantified using the Qubit instrument and dsDNA BR Assay kit (Life Technologies, USA). The PCR reactions were conducted using *B. sporothermodurans* specific primers SH2-F1 (5'-GAT TCA GGC AGA ATG TAG CA-3') and SH2-R1 (5'-TTT GGC CAC TTG ATG GTA CA-3'), (Herman *et al.* 1997) with modifications in PCR cycle conditions, to distinguish HRS and non-HRS strains.

For genome sequencing, multiplexed paired-end libraries were prepared using Nextera XT DNA Sample Preparation kit (Illumina, USA). Genome sequencing was carried out on an Illumina MiSeq system (Illumina, San Diego, USA). The paired end reads (2 X 300 bp) were checked for quality, trimmed and *de novo* assembled using the Qiagen CLC Genomics Workbench version 9 (Qiagen, Netherlands). The resultant contigs were submitted to GenBank, where gene annotation was implemented using the National Centre for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova *et al.*, 2016). The annotation was further uploaded to Rapid Annotation using Subsystem Technology (RAST) for subsystems-based annotation analysis and categorisation

(Aziz, *et al.*, 2008; Disz *et al.*, 2010; Overbeek *et al.*, 2014). Using FigTree v1.4.3, a phylogenetic tree was constructed using conserved protein sequences of the strains sequenced in the present study and related *Bacillus* sp. selected from NCBI Genbank and using a maximum likelihood tree estimation program (Stamatakis, 2014). Coding sequences with homology to DNA sequences labelled as transposons and mobile genetic elements where identified in the COG database.

5.3.2 Spore preparation

A single colony of each isolate of *B. sporothermodurans* strains SAD and SA01 was collected from BHI agar plates and transferred into 1.5 ml Eppendorf tubes containing 1 ml of freshly prepared Brain Heart Infusion broth supplemented with vitamin B_{12} and then incubated for 24 h at 37 °C. From the 24h pure culture, 1 ml of culture was collected from the broth with sterile pipettes and dispensed into the sporulation media, consisting nutrient broth (25 g L^{-1}), bacteriological agar (15 g L^{-1}), vitamin B_{12} (1 mg L^{-1}), $MnSO_4.H_2O$ (8.4 mg L^{-1}) and $CaCl_2.2H_2O$ (1 g L^{-1}), pH 6.8 (Huemer *et al*., 1998). The plates were incubated for at least 10 days at 37 ºC. Spore preparation was undertaken as in a previous study (Scheldeman *et al.,* 2005) and pellets subsequently diluted with physiological saline (8.5 g L^{-1} NaCl) to $\approx 10^8$ spores ml^{-1} prior to use.

5.3.3 Thermal inactivation and enumeration of surviving spores

Sterile skim milk was inoculated with spores of strains SAD and SA01 to a concentration of approximately 2×10^7 spores ml^{-1}. To verify sterility, milk was heated to 121 °C for 2 minutes,

and 1 ml transferred into BHI-B_{12} agar and incubated at 37 °C for 3 days. Thermal resistant glass tubes (Ace glass incorporated, Vineland, NJ, USA; #15) were filled with 5ml of inoculated milk samples and completely submerged in a pre-heated oil bath of set temperature. The process come up time was recorded. Tubes were drawn at predetermined time intervals. Once the desired heating time has elapsed, tubes were cooled in a water/ice bath to about 15 °C to avoid further inactivation. Surviving spores were enumerated by plating on BHI-B_{12} agar at 37 ºC for 3 days prior to counting to allow rejuvenation of injured spores.

5.3.4 Modelling and curve fitting

The data was modelled using the linear, biphasic and Weibull methods. The linear model follows first order kinetics, depicting a linear relationship between the logarithmic decline of bacterial survivors over treatment time (Schaffner and Labuza, 1997). The model is formulated as follows;

$$\log 10 \left(\frac{N}{No}\right) = -\frac{t}{D},$$

Where N_O is the initial number of cells (CFU ml^{-1}), N, the number of survivors after an exposure time t (CFU ml^{-1}), D (decimal reduction time), the time required to destroy 90% of the organisms (min), t, the treatment time (min).

The biphasic model (Cerf, 1977), is based on the principle of a sensitive fast inactivating and resistant slow inactivating populations, leading to a curve with two distinct segments of linearly decreasing populations, the second more resistant population with a less negative slope. This model can be formulated as follows;

$$\log\left(\frac{N}{NO}\right)=\log(f.e-k\max 1t+(1-f)\,e-kmax2t)\left(\binom{n}{k}\right) \ldots\ldots\ldots\ldots 1$$

$$\log\left(\frac{N}{No}\right)=\log(f.\ e\ -\ k\max 1+(1-\ f)\ e\ -\ kmax2t) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

Where *f* is the fraction of the initial population in a major subpopulation, (1 - f) is the fraction of the initial population in a minor (more heat resistant) subpopulation, and k_{max1} and k_{max2} are the inactivation rates of the two populations respectively (Albert and Mafart, 2005; Geeraerd *et al.*, 2005), *N* and N_0 are same as in linear equation above.

The Weibull model (Albert and Mafart, 2005; Coroller *et al.*, 2006) is based on the premise of a heterogenous microbial population with cell death occurring because of the contact time of the cell with external stress conditions with each cell requiring a different contact time to cell inactivation. The dispersion of resistance to conditions to which the cells are exposed is governed by a distribution of the Weibull type. The Weibull distribution is as follows;

$\log 10\ \left(\frac{N}{N0}\right) = -bt^n$ where b and n are the scale and shape factors.

The fitting of experimental data to the corresponding equations through non-linear regression techniques was carried out with Microsoft® Excel GInaFiT (Geeraerd *et al.*, 2005). The goodness of fit of the models was compared by computing the mean square error (MSE) and the regression of coefficient (R^2) (Neter *et al.*, 1996).

5.4 Results

5.4.1 Comparative analysis and distribution of genes of *B. sporothermodurans* strains

The draft genome assemblies of the *B. sporothermodurans* strains comprises two strains belonging to the HRS clone (SA01, isolated from South Africa; BR12, isolated from Brazil) and two belonging to the non-HRS clone (SAD, isolated from South Africa; DSM 10599^{T}, the type strain isolated from Italy). All four strains were isolated from UHT milk. The genome sizes of the four strains ranges from 3.4 - 3.9 Mb, as compared to on average, *B. cereus* (5.2 Mb), *B. anthracis* (5.1 Mb), *B. subtilis* (4.1 Mb), *B. amyloliquefaciens* (3.9 Mb) and *B. pumilus* (3.8 Mb). The number of transposases, mobile genetic elements, and integrases related proteins identified in the COG database ranged from 11- 16, 61 – 141 and 7 – 9 respectively for all four study strains. Further, the HRS strains (SA01 and BR12) exhibited the presence of transposable elements type Tn554 and Tn552. Additional sequencing statistics for the strains used in this study are shown in table 5.1.

A phylogenetic tree of *B. sporothermodurans* strains, its closest phylogenetic neighbor *B. oleronius* and other related *Bacillus* species (Figure 5.1) reveals patterns of ancestry. The tree was deduced using single-copy homology groups by BLAST and the use of FastTree to analyze protein alignments, with a total 1,247 conserved sequences used. The tree was inferred using the maximum likelihood algorithm. The high heat resistant strains in this study (BR12 and SA01) though isolated from different continents cluster together. The distribution and counts of genes as shown in Figure 5.2 depicts majority of the genes involved in essential cellular functions as carbohydrate metabolism (300 avg.), protein metabolism (240 avg.), DNA metabolism (135 avg.), cell wall and capsule activities (115 avg.), dormancy (110 avg.) and stress response (75 avg.). Though non-pathogenic the four *B. sporothermodurans* strains still

on average had 60 COGs dedicated to virulence, disease and defense, regulation and cell signaling having a similar average as the former. Phages, Prophages, TEs, Plasmids and secondary nitrogen metabolism had COG number on average below 20.

Table 5.1: Summary report of the de novo assembly of four B. sporothermodurans strains isolated from UHT milk.

Strain	[n]*B. sporothermodurans* DSM 10599[T]	[n]*B. sporothermodurans* SAD	[h]*B. sporothermodurans* SA01	[h]*B. sporothermodurans* BR12
Genome size (bp)	3,783,858	3,857,089	3,414,010	3,974,892
G + C content (%)	35.88	35.89	35.85	36.31
No. of coding sequences	4,257	4,111	3,768	4,558
No. of tRNAs	88	96	75	96
Contig N50	15,402	114,649	22,386	9,377
GenBank accession no.	NAZD01000000	NAZB01000000	NAZA01000000	NAZC01000000

[n]Non-HRS strains; [h]HRS strains

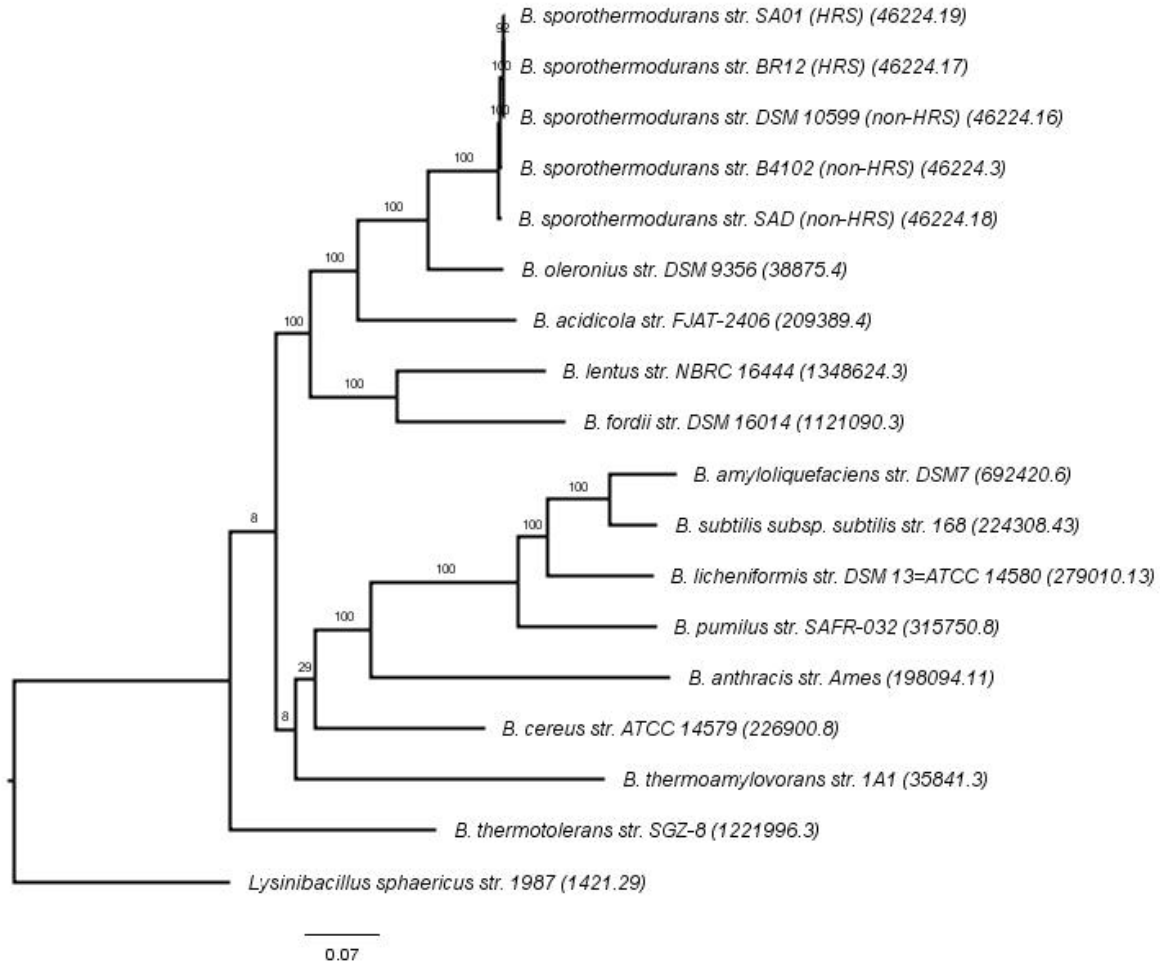


Figure 5.1: Phylogenomic comparison of *B. sporothermodurans* strains reveal patterns of ancestry. The tree was inferred using the conserved protein sequences of genomes of *B. sporothermodurans* strains, *B. oleronius* as well as selected closely related *Bacillus* sp. of importance to the food industry.

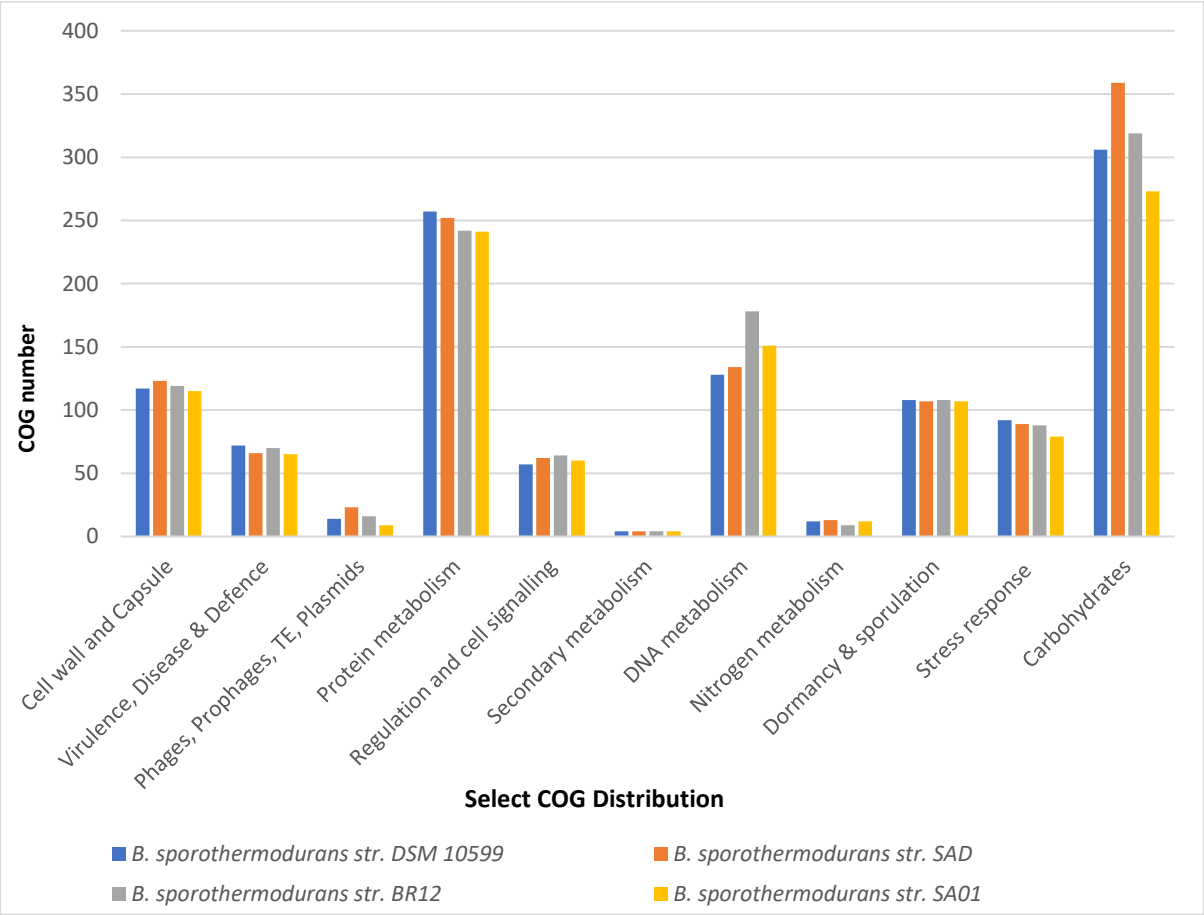


Figure 5.2: Distribution and counts of genes for genomes of *B. sporothermodurans* strain DSM 10599^{T} (non-HRS), *B. sporothermodurans* strain SAD (non-HRS), *B. sporothermodurans* strain BR12 (HRS) and *B. sporothermodurans* strain SA01 (HRS).

5.4.2 Heat inactivation kinetics of two strains of *B. sporothermodurans*

The thermal inactivation dynamics of the two South African strains consisting of SA01 (HRS) and SAD (non-HRS) was determined to compare their heat resistance in milk and to ascertain which inactivation protocol gives the best fit for spore forming bacteria. Consequently, the survival curves were fitted with the linear, Weibull and biphasic models. The survival curves of *B. sporothermodurans* SA01 and *B. sporothermodurans* SAD spores in milk fitted with the

linear, Weibull and biphasic models are presented in figures 5.3 and 5.4. Visual inspection of the survival curves indicates a good fit was obtained by the non-linear models. With respect to the Weibull model, the shapes of the survival curves for all the temperature treatments were similar for the two strains and is characterized by an initial drop in spore population followed by tailing. However, the shapes of the survival curves slightly changed depending on the treatment temperature. At the lowest temperatures studied (Figures 5.3), the survival curves of the two strains showed a slight curvature. At temperatures 130 °C and 135 °C the survival curves were similar and quite pronounced. More specifically for *B. sporothermodurans* SA01 and *B. sporothermodurans* SAD spores at 130 °C (Figure 5.3), tailing started after approximately 30 s and 35 s treatment time respectively, corresponding to 2-$\log_{10}$ reductions. As shown, tailing at 135 °C (Figure 5.4a) started approximately after 10 s for both strains, corresponding to 2-$\log_{10}$ reductions. The survival curves of the two strains at 140 °C (Figure 5.4b) was significantly different from other curves. The curves showed a more pronounced curvature. This was characterized by an initial rapid drop in spore population to 4 $\log_{10}$ reductions at 1 s for both strains. Extending the treatment time to 6 s only increased the reduction by 2-$\log_{10}$ reductions for both strains.

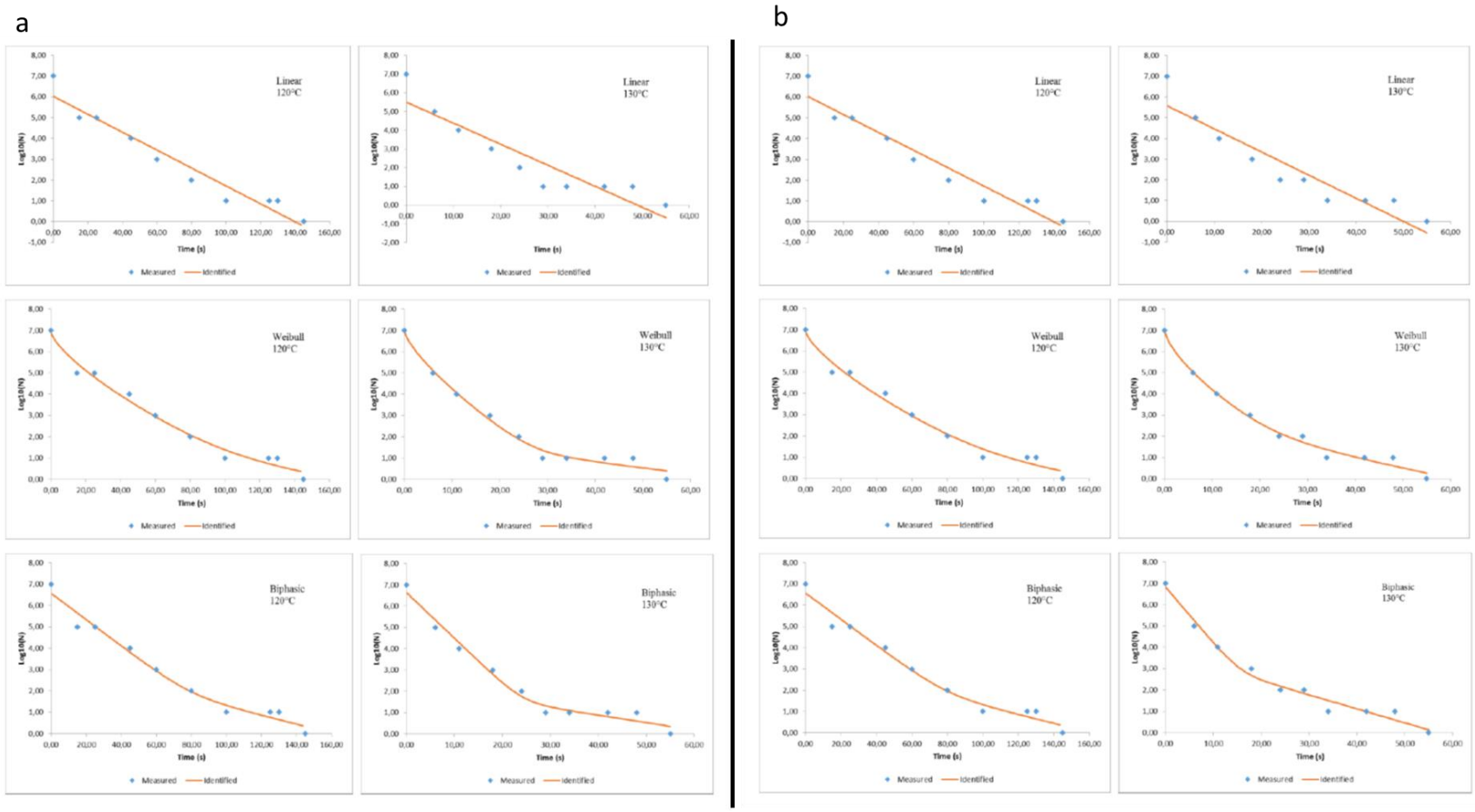


Figure 5.3: Survival curves of (a) *B. sporothermodurans* SA01(HRS) and (b) *B. sporothermodurans* SAD (non-HRS) spores in skimmed milk at 120 °C and 130 °C. Data were fitted with the linear, Weibull and biphasic models. Data points are means of the two replicates.

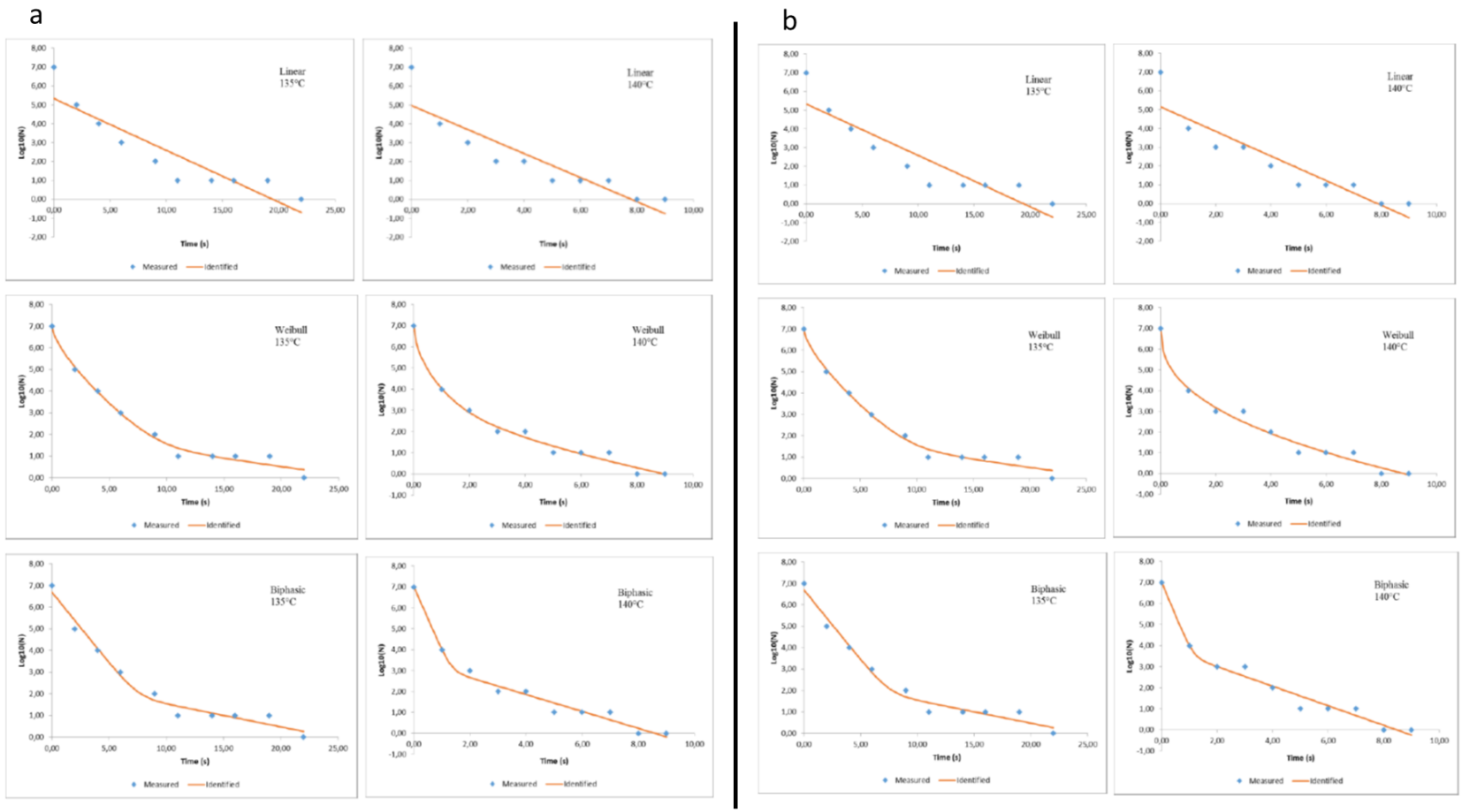


Figure 5.4: Survival curves of (a) *B. sporothermodurans* SA01(HRS) and (b) *B. sporothermodurans* SAD (non-HRS) spores in skimmed milk at 135 °C and 140 °C. Data were fitted with the linear, Weibull and biphasic models. Data points are means of the two replicates.

To compare the goodness of fit for the respective models for *B. sporothermodurans* SA01 and *B. sporothermodurans* SAD spores, mean square error (MSE) and the correlation of coefficient (R^2) were computed, as shown in tables 5.2 and 5.3 respectively. Among the three models, the Weibull model consistently produced a better fit to majority of the survival curves for the two strains. However, at 140 °C, the Weibull and biphasic models had the same R^2 for both strains (Table 5.2 & 5.3). The Weibull model produced the best fit to the survival curves, as indicated by the lowest MSE and the highest R^2 values at the temperatures studied. The linear model produced the poorest fit to the experimental data as inferred by the systematically higher MSE and lowest R^2 values, regardless of temperature applied. For *B. sporothermodurans* SA01 spores, the Weibull and biphasic models had a mean R^2 of 0.99 and 0.98 respectively, and MSE of 0.10 and 0.14 respectively (Table 5.2). The linear model had a mean R^2 value of 0.86 and a mean MSE value of 0.77 (Table 5.2). As shown in Table 8, the mean R^2 for *B. sporothermodurans* SAD spores was 0.99 and 0.98 (Table 5.3) for the Weibull and biphasic models respectively. The MSE was 0.11 for the Weibull and 0.14 for the biphasic model (Table 5.3). The linear model had the lowest mean R^2 of 0.88 and highest mean MSE of 0.67 (Table 5.3).

Table 5.2: Comparison of the goodness- of- fit of the linear, Weibull and biphasic models for the survival curves of *B. sporothermodurans* SA01 (HRS) spores isolated from UHT milk treated at different temperatures in skimmed milk.

Temperature (°C)	*MSE*			R^2		
	Linear	Weibull	Biphasic	Linear	Weibull	Biphasic
120	0.33	0.14	0.17	0.94	0.98	0.97
130	0.82	0.10	0.14	0.85	0.99	0.98
135	0.98	0.09	0.14	0.84	0.99	0.98
140	0.94	0.08	0.10	0.82	0.99	0.99
Mean	0.77	0.10	0.14	0.86	0.99	0.98

Mean square error (MSE) represents the average of the squares of the errors calculated for the goodness of fit of the linear, Weibull and biphasic models for *B. sporothermodurans* SA01. Regression coefficient (R^2) is the proportion of variance for the goodness of fit of the linear, Weibull and biphasic models explained by the temperature differences.

Table 5.3: Comparison of the goodness- of- fit of the linear, Weibull and biphasic models for the survival curves of *B. sporothermodurans* SAD (non-HRS) spore isolated from UHT milk treated at different temperatures in skimmed milk

Temperature (°C)	*MSE*			R^2		
	Linear	Weibull	Biphasic	Linear	Weibull	Biphasic
120	0.33	0.14	0.17	0.94	0.98	0.97
130	0.62	0.10	0.12	0.88	0.99	0.98
135	0.92	0.09	0.14	0.84	0.99	0.98
140	0.78	0.12	0.14	0.84	0.98	0.98
Mean	0.67	0.11	0.14	0.88	0.99	0.98

Mean square error (MSE) represents the average of the squares of the errors calculated for the goodness of fit of the linear, Weibull and biphasic models for *B. sporothermodurans* SAD. Regression coefficient (R^2) is the proportion of variance for the goodness of fit of the linear, Weibull and biphasic models explained by the temperature differences.

Modelling parameters, alpha (α), delta (δ) and shape parameter *p* were computed from the double Weibull model equation (Table 5.4). Generally, higher values and distinct δ_1 and δ_2 at temperature 120 °C, 130 °C and 135 °C were observed for both strains (Table 5.4). At 120 °C both strains had the highest but distinct δ_1 and δ_2 values. Interestingly, the smallest δ_1 was observed at 140 °C for both strains.

Table 5.4: Kinetic parameters, alpha (α), delta (δ) and shape parameter-*p* of the double Weibull model fitted to *B. sporothermodurans* strain SA01 (HRS) spore and *B. sporothermodurans* strain SAD (non-HRS) spore survival curves in milk after thermal treatment

Temperature (°C)	*B. sporothermodurans* SA01				*B. sporothermodurans* SAD			
	α	δ_1	δ_2	*p*	α	δ_1	δ_2	*p*
120	3.39	9.10	29.61	0.73	3.39	9.10	29.61	0.73
130	4.59	2.64	22.15	0.75	2.99	8.25	8.25	0.72
135	4.36	0.83	6.96	0.70	5.36	8.54	8.54	0.71
140	2.27	0.13	0.54	0.55	7.06	0.07	0.07	0.41

The α is the logit transformation of f, the fraction of the original population in the major sub-population. δ_1 and δ_2 are the treatment times for the first decimal reduction of subpopulations 1 and 2 respectively. *p* is a shape factor with a value from negative infinity to positive infinity.

5.5 DISCUSSION

Previous research using 16S *rRNA* sequencing had suggested the closest relatives to *B. sporothermodurans* as *B. oleronius*, *B. lentus*, *B. firmus* and *B. benzoevorans* (Scheldeman *et al.*, 2006). Using WGS sequencing in this study however, the closest relatives are *B. oleronius*,

B. acidicola and *B. coagulans*, with *B. lentus* and *B. firmus* further off. This is expected as operon heterogeneity could result in discrepancies of variable regions when using 16S *rRNA* gene sequencing (Pettersson *et al.*, 1996; Klijn *et al.*, 1997). Also, the HRS strains (SA01 and BR12) cluster together, whilst the non-HRS strains (SAD, DSM 10599^{T}) do likewise. Thus, the two SA strains cluster separately though isolated from similar UHT milk products. This suggests a unique molecular marker responsible for high heat resistance and probably housed in the accessory genome. The primer sequence used for identification of HRS strains and blasted against strains SA01 and BR12, was identified in a mobile element protein sequence relating to a transposase and originating from *Listeria monocytogenes*. This gives indication of probable HGT of mobile genetic elements from other Firmicutes. In general, the HRS strains (SA01 and BR12) exhibited on average twice as much mobile element proteins as the non-HRS strains. Though this does not in any way explain its higher heat resistance, it does suggest higher HGT influence, and thus higher likelihood of having more genetic variance as compared to the non-HRS strains. The Tn552 and Tn554 transposable elements identified in strains SA01 and BR12 code for beta-lactamase and resistance to erythromycin and spectinomycin respectively (Rowland and Dyke, 2013; Bastos and Murphy 2018). Both transposable elements have been identified in *Staphylococcus aureus*, with homologs of Tn552 identified in *B. licheniformis* (Bastos and Murphy, 2018).

In this study the type strain *B. sporothermodurans* DSM 10599^{T} after performing the HRS PCR, does not belong to the HRS clone as was observed in two previous studies (Guillaume-gentil *et al.*, 2002, Scheldeman *et al.*, 2002), albeit from a different culture collection. It should however be noted that HRS PCR used (Herman *et al.*, 1997) is not 100% reliable. Also, *B. sporothermodurans* strains belonging to the HRS clone are known to lose that ability after a period of sub-culturing and cold storage (Scheldeman *et al.*, 2006). The heat inactivation data

for strains SA01 (HRS) and SAD (non-HRS) show similar heat resistance across the various temperatures used in the study. There could be a few reasons for this. Firstly, there is the possibility of lack of gene expression in SA01 with respective to its higher heat resistance capabilities. Secondly, the heat inactivation studies used does not differentiate the state of the heat exposed spores with respect to being intact, injured or inactivated. Consequently, the recovery medium used may recover both intact and depending on the level of injury, some injured cells thus making it difficult to effectively compare the effects of heat application on the different strain spores (Corradini and Peleg, 2007). Future studies may need the use of flow cytometry to further identify the state of heat exposed spores as pertains to HRS and non-HRS strains. A previous study showed differing heat resistance of HRS and non-HRS strains isolated from UHT milk (Scheldeman *et al.*, 2006).

Using a line of best fit through the data points of a non-linear survival curve will be inappropriate and result in considerable errors if the linear equation is used to determine inactivation time of spores, since the linear model did not produce a good fit to the survival curves in terms of MSE and R^2 values. Case in point, if the target reduction at 140 °C of *B. sporothermodurans* spores in milk is 6 $\log_{10}$, the processing time would be 6 s for both strains when using the double Weibull model. However, based on the linear model, the corresponding time would be approximately 8 s for both strains. The relevant figures for the biphasic model would be approximately 6.1 s for both strains. Therefore, the processing conditions determined based on the linear model especially would result in over-processing of UHT milk. The degree of inactivation of *B. sporothermodurans* spores depended on the intensity of the heat and holding time. As the temperature increased and treatment time increased, the degree of destruction also increased.

The non-linear trend of the survival curves suggests that the linear approach can infer inaccuracies. Many explanations have been proposed, such as biological heterogeneity among the spores in subpopulations with individual inactivation kinetics, even in the case of a pure culture. As a result, inactivation curves could be regarded as the cumulative form of underlying distribution of individual bacterial inactivation times (Peleg and Cole, 1998). Additionally, spores may change thermo-resistance, a consequence of adaptations to the stress that makes the remaining spore subpopulation more resistant (Peleg and Cole, 1998; Van Boekel, 2002). This phenomenon indicates that it is not desirable to extend the treatment time to increase heat inactivation of these spore populations. A more effective way is to increase the temperature so that treatment time can be substantially decreased.

Since the double Weibull model produced a reasonably good fit, its modelling parameters α, δ and *p* were used to explain the inactivation kinetics. The shape factors of the double Weibull model indicated that all the survival curves of the two strains fitted with this model were concave upward ($p<1$). These observations suggest that, the population was composed of two groups with different resistances to stress. Distinct δ_1 and δ_2 for all the survival curves were observed indicating that the two subpopulations behaved differently in both strains (Ibekwe *et al.,* 2011). All the survival curves disclosed that $\delta_1 < \delta_2$ suggesting that subpopulation 1 (δ_1) was the more sensitive subpopulation and did not activate or activated the mechanisms of resistance to an insignificant extent. Subpopulation 2 (δ_2) corresponded to the most resistant spores population which had restricted metabolic activity thus develops the mechanisms of resistance (Coroller *et al.,* 2006). The initial sharp decrease in spore concentration at 140 °C for both strains might be largely attributed to the faster decline of subpopulation (δ_1) as shown with reduced δ_1 (Ibekwe *et al.,* 2011). However, at 120 °C, the high δ_1 for both strains were characterized by a steady decrease in the spore population of subpopulation 1. The higher δ_2

dominated the spore population, leading to a slower decline phase in the more resistant subpopulation. This slower decline was also explained by the slight curvature at 120 °C for both strains.

5.6 Conclusion

This study provides evidence that *B. sporothermodurans* spores fit better to the non-linear inactivation models, and this may well be the case for especially other spore formers. Thus, past inactivation protocols based on first order kinetics may need to be revised especially in the dairy industry. The similar heat inactivation characteristics of HRS and non-HRS strains may be as a result of the inability to differentiate intact, injured and inactivated spores in this study or pertaining to gene expression. This study thus shows the importance of the doubling up of heat inactivation studies with molecular based approaches to elucidate the strain effects of bacterial heat resistance.

Chapter 6: COMPARATIVE GENOME ANALYSIS OF *BACILLUS SPOROTHERMODURANS* WITH ITS CLOSEST PHYLOGENETIC NEIGHBOUR, *BACILLUS OLERONIUS*, AND SELECTED *BACILLUS CEREUS* AND *BACILLUS SUBTILIS* GROUPS

6.1 Abstract

Bacillus sporothermodurans presently possess one the most highly heat resistant spores (HRS) in prokaryotes. The spores can withstand ultra-high temperature (UHT) processing, with over processing of the milk to inactivate spores resulting in undesirable organoleptic properties. In this study, multiple whole genome sequences of *B. sporothermodurans* strains including those belonging to the HRS clone was compared with other *Bacillus* species. *B. sporothermodurans*, together with its closest phylogenetic neighbour *B. oleronius*, was clustered with *B. thermoamylovorans* and *Quasibacillus thermotolerans*. The *cod*Y, *clp*X and *sod*A1 virulence genes found in *B. anthracis* and other *B. cereus* group members were identified in *B. sporothermodurans* with an average homology of 84% to those in *B. anthracis*. Homologs of *Listeria monocytogenes* virulence genes *clp*P, *pur*A, *pur*B, and *fur* were similarly identified in the genome of *B. sporothermodurans*. Single nucleotide polymorphism amongst strains of *B. sporothermodurans*, a minimum and maximum of 97 and 11,255, respectively, indicate both minimal and extensive genome differences amongst strains. Protein cluster analysis of all twelve strains resulted in 21,422 clusters overall, with 482 clusters common to all. Protein clusters responsible for heat resistance showed differences in amino acid sequence of proteins such as GrpE, responsible together with DnaK and DnaJ clusters for protein repair during unfavourable conditions. The genome sequence of *B. sporothermodurans* revealed similarities to other *Bacillus* species, which invariably can have both positive and negative connotations to industry.

Keywords: heat shock, spore coat, comparative genomics, antibacterial, protein cluster

6.2 Introduction

Comparative genome analysis typically consists of identification of homologs (orthologues and paralogues), thus elucidating bacterial rearrangements, plasticity, epigenetics and polymorphisms. The afore mentioned bacterial analysis has especially taken off since the gains in Next Generation Sequencing (NGS) technologies. In recent times, NGS technologies evaluating whole genomes has made it possible to study the architecture of bacterial species with the aim of identifying phylogeny and evolutionary trends.

Whole genome sequencing (WGS) of particularly industrially important organisms and bacterial pathogens provide critical answers for the biotechnology industry, for human health and wellbeing, as well as the environment. With the advent of bacterial bioprospecting, comparative genomics is key to identify new sources of important bacterial metabolites of benefit to the food, agriculture, health and chemical industries. The study of evolutionary trends and pathogenesis of important disease-causing bacterial species would allow the development of new antibiotics and detailed knowledge of genetic mechanisms and possible use of recombinant technology for pathogen control. Through NGS technologies, there are significant data being generated, enabling the understanding of moderate to high conservation of bacterial species via numerous proteins and unravelling the mechanisms of their adaptation to new niches and competition with other bacteria or forms of life (Sharma and Satyanarayana, 2013).

The genus *Bacillus* comprises low G + C Gram positive bacteria that are readily found in nature. Their ability to form spores during unfavourable conditions enables their survival in harsh environmental conditions that will, otherwise, kill vegetative bacteria (Nicholson *et al.*, 2000). They are probably the most studied group as a result of their importance to industry in the production of various biomolecules as riboflavin, β-lactamase and various other enzymes.

Bacillus is able to present high growth rates and short fermentation cycles, and hence are good candidates for extracellular enzymes production (Sharma and Satyanarayana, 2013). Some industrially important *Bacilli* include *B. subtilis, B. licheniformis, B. amyloliquefaciens* and *B. megaterium,* and although *Bacilli* are generally regarded as safe (GRAS) in the Food and Drug Administration's (FDA) list, a few species, such as *B. cereus* and *B. anthracis,* infect humans causing food borne illness and anthrax, respectively. Therefore, illustrating both the usefulness and danger of *Bacillus* species (Maughan and Van der Auwera, 2011).

B. sporothermodurans is an emerging highly heat resistant spore (HRS)-forming bacterium of interest to the dairy industry. Spores of this organism are known to survive in foods treated by ultra-high temperature (UHT), affecting its quality and causing significant economic losses. Therefore, there is a growing interest to identify evolutionary associations and understand its propensity for high heat resistance as well. However, not all strains have the high heat resistant attribute, the highly heat resistant spore (HRS) clone. In this study, we sequenced *B. sporothermodurans* strains isolated from UHT milk from South Africa and Brazil, a type strain and *B. oleronius,* its closest phylogenetic neighbour (Heyndrickx *et al.*, 2012). By using molecular comparisons, we report on various antibacterial factors and protein family clusters shared with other *Bacillus* species. To the best of our knowledge, this is the first study comparing *B. sporothermodurans* genomes with other *Bacillus* species.

6.3 Materials and methods

6.3.1 Bacterial isolation and identification

B. sporothermodurans strains SAD and SA01 were isolated from UHT milk produced in South Africa and *B. sporothermodurans* strain BR12 was isolated from UHT milk from Brazil. The type strain *B. sporothermodurans* DSM 10599^T and *B. oleronius* DSM 9356^T were obtained from Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH (DSMZ). Growth of *B. sporothermodurans* isolates was performed on Brain Heart Infusion agar (Oxoid, UK) and incubated at 37 °C for 48 hours. *B. oleronius* was grown in Nutrient Agar (Oxoid, UK) at 37 °C for 24 hours. Genomic DNA was extracted using the ZR Bacterial DNA Miniprep kit (Zymo Research, USA). DNA was quantified using the Qubit instrument and dsDNA BR Assay kit (Life Technologies, USA). Molecular confirmation of *B. sporothermodurans* strains and those belonging to the HRS clone was undertaken using *B. sporothermodurans* specific primers of BSPO-F2 (5'-ACG GCT CAA CCG AG-3') and BSPO-R2 (5'-GTA ACC TCG CGGTCT A-3') (Scheldeman *et al.*, 2002) and SH2-F1 (5'-GAT TCA GGC AGA ATG TAG CA-3') and SH2-R1 (5'-TTT GGC CAC TTG ATG GTA CA-3') (Herman *et al.*, 1997), respectively.

6.3.2 Genome sequencing and analysis

Bioinformatics tools were used to identify genes and protein families (Aziz *et al.*, 2008; Wattam *et al.*, 2014) of *B. sporothermodurans* strains, *B. oleronius* DSM 9356^T, and selected *B. subtilis* and *B. cereus* group members. Genome sequences for *B. sporothermodurans* strains SAD, SA01, BR12, DSM 10599^T and B4102 with genome identification (46224.18, 46224.19,

46224.17, 46224.16 and 46224.3 respectively), and *B. oleronius* DSM 9356^{T} (38875.5) were used in this study. *B. subtilis* group members included were *B. subtilis* subsp. *subtilis* str. 168 (224308.262), *B. licheniformis* DSM 13 = ATCC 14580^{T} (279010.13), and *B. amyloliquefaciens* DSM 7^{T} (692420.6). *B. cereus* group members included *B. cereus* ATCC 10987^{T} (222523.39), *B. anthracis* str. 'Ames Ancestor'^{T} (261594.21) and *B. thuringiensis* strain ATCC 10792^{T} (1428.478).

For *B. sporothermodurans* strains SAD, SA01, BR12, DSM 10599^{T} and *B. oleronius* DSM 9356^{T}, multiplexed paired-end libraries were prepared using Nextera XT DNA Sample Preparation kit (Illumina, USA). An Illumina MiSeq system (Illumina, San Diego, USA) was used to carry out sequencing. The paired-end reads were checked for quality, trimmed and de novo assembled using the Qiagen CLC Genomics Workbench version 9 (Qiagen, Netherlands). Resultant contigs were submitted to GenBank, and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) was used for gene annotation (Tatusova *et al.*, 2016). *B. sporothermodurans* B4102 (isolated from Indian curry in the Netherlands) (Krawczyk *et al.*, 2016), and the select *B. subtilis* and *B. cereus* group members were sourced from NCBI. Subsystem based annotation was undertaken for all 12 *Bacillus* species employing the Rapid Annotation using Subsystem Technology (RAST) (Aziz *et al.*, 2008; Overbeek *et al.*, 2014; Brettin *et al.*, 2015).

A phylogenetic tree was constructed using conserved protein sequences of the strains sequenced in the present study and *B. cereus* and *B. subtilis* groups and alkaliphiles (*B. akibai*, *B. hemicellulosilyticus*, *B. halodurans*) were selected from NCBI and using a maximum likelihood tree estimation program (Stamatakis, 2014) and FigTree v1.4.3. (Rambaut, 2016). The Phyre2 suite of tools was used to predict the tertiary structure of protein GrpE (Kelley *et al.*, 2015) at 100% model confidence for all four protein structures. Protein sequence alignment

was undertaken with BioEdit v7.2.1.(Hall, 1999). Conservation analysis was undertaken using the Jensen-Shannon divergence approach (Capra and Singh, 2007) and amino acid conservation was based on a relative conservation sensitivity of 75%. Single nucleotide polymorphisms (SNPs) were identified using CSI Phylogeny (Kaas *et al.*, 2014). CSI an online based server was used to call and filter SNPs using WGS reads of the study strains. CSI Phylogeny then validates WGS reads and infers relationships based on concatenated alignment of SNPs.

6.3.3 Accession numbers

All *B. sporothermodurans* strains and *B. oleronius* genome sequences used in this study have been deposited in GenBank of the National Centre for Biotechnology Information (NCBI) under the following accession numbers: NAZD01000000, NAZB01000000, NAZA01000000, NAZC01000000, MTLA01000000.

6.4 Results

6.4.1 Genome sequences and circular maps shows wide variations of selected *Bacillus* species

To compare *B. sporothermodurans* strains with other closely related *Bacillus* species, the genomic features of *B. sporothermodurans* strains as determined and deposited to GenBank was compared to type strains of others *Bacillus* species. The comparative genome features of *B. sporothermodurans* strains (SAD, SA01, BR12, DSM 10599^{T}, B4102), and members from *B. subtilis* group (*B. subtilis*, *B. licheniformis*, *B. amyloliquefaciens*) and *B. cereus* group (*B.*

cereus, *B. anthracis*, *B. thuringiensis*) are shown in Table 10. The average genome size of the *B. sporothermodurans* strains is 3.7 Mb, on average 25% smaller than that of *B. oleronius*, 10% smaller than the *B. subtilis* group strains and 30% smaller than the *B. cereus* group strains. The G + C content of the *B. sporothermodurans* strains, 36.08% on average, is approximately 1% larger than *B. oleronius* and the *B. cereus* group strains, and 9% lower than that of the *B. subtilis* group strains (45.46%). The presence of plasmids was detected only for the *B. cereus* group strains (Table 10).

The genome alignment analysis indicates the presence of three genes encoding virulence factors in *B. sporothermodurans* SA01 and *B. subtilis* subsp. *subtilis* str. 168, in comparison with 11 detected in *B. cereus* ATCC 10987. Concerning antimicrobial resistance genes, 33 were detected in *B. sporothermodurans*, 51 in *B. subtilis* and 46 in *B. cereus* (Figure 6.1).

Table 6.1: A comparison of the genomic features of *B. sporothermodurans* strains with *B. oleronius* and strains from *B. subtilis* and *B. cereus* groups.

Bacillus species	Features				
	Contigs (plasmids)	DNA sequence (bp)	G + C content (%)	Total CDSs	tRNA
B. sporothermodurans SAD	108(0)	3,857,089	35.9	4,111	96
B. sporothermodurans SA01	286(0)	3,414,010	36.6	3,768	75
B. sporothermodurans BR12	800(0)	3,974,872	36.3	4,558	96
B. sporothermodurans DSM 10599^T	511(0)	3,783,858	35.9	4,257	88
B. sporothermodurans B4102	176(0)	4,047,827	35.7	4,359	87
B. oleronius DSM 9356^T	587(0)	5,083,966	35.0	5,530	116
B. subtilis subsp. *subtilis* str. 168	1(0)	4,215,385	43.5	4410	86
B. licheniformis DSM 13^T	1(0)	4,222,597	46.2	4461	72
B. amyloliquefaciens DSM 7^T	1(0)	3,980,199	46.1	4,206	94
B. cereus ATCC 10987^T	1(1)	4,871,269	35.4	5,181	39
B. anthracis str. 'Ames Ancestor'^{T}	3(2)	5,503,926	35.2	6,010	95
B. thuringiensis strain ATCC 10792^T	7(6)	6,445,173	35.0	8,823	107

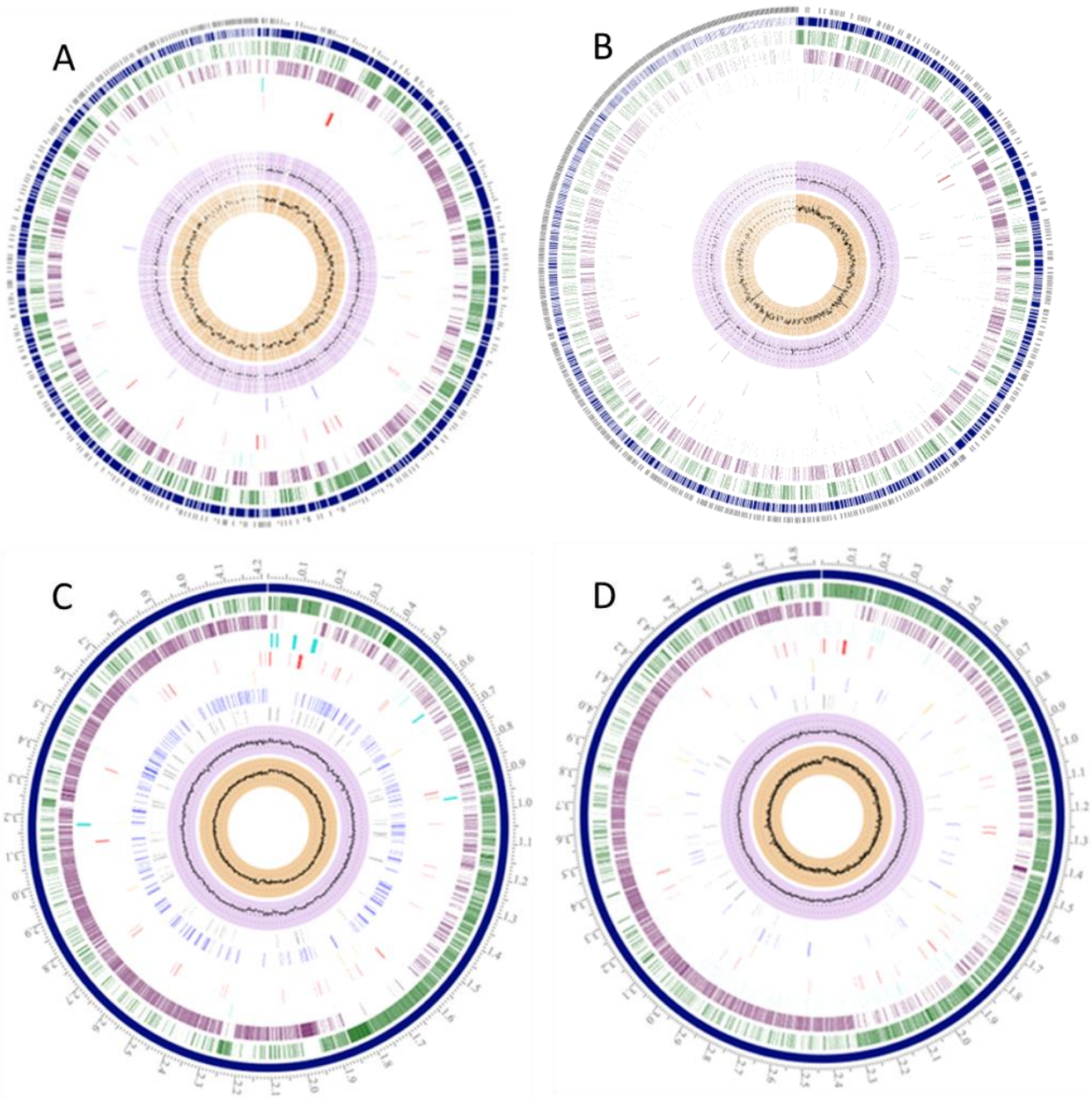


Figure 6.1: *B. sporothermodurans* SA01 (A), *B. oleronius* DSM 9356^{T} (B), *B. subtilis* subsp. *subtilis* str. 168 (C) and *B. cereus* ATCC 10987 (D) possess unique genomic regions in comparison with each other. The outer circle designates the genome's coordinates in mega base pairs (Mbp). The blue circle denotes the number of contigs and the green and purple circles representing the coding sequences (CDS), forward and reverse respectively, with white spaces between the CDS accounting for hypothetical proteins. Non CDS features are shown by the

light blue circle. The proceeding red and orange circles represent regions encoding antimicrobial determinants and virulence factors, respectively. The two inner black circles depict the G + C content and skew.

6.4.2 Clusters of Orthologous Groups (COG) categories of *Bacillus* species

The functional annotations of the *Bacillus* strains sequenced are indicated in Figure 6.2. On average *B. sporothermodurans* strains had 61% of their CDSs assigned to functions, as opposed to 57% in the *B. oleronius* and on average 71% for the *B. cereus* and 75% for the *B. subtilis*. In general, 40% of the genes were assigned as involved in metabolism, and a further 40% of genes could be associated with protein processing, energy and cellular processes in similar proportions. The remaining 20% sequences encoding proteins related to DNA and RNA processing, stress response-defence-virulence, membrane transport, cell envelope and regulation and cell signalling subsystems. Several variations in the subsystem portfolio amongst *Bacillus* groups exist, including the presence of genes related to invasion, intracellular resistance, host pathogen interaction, toxins and superantigens that were only present in the *B. cereus* group strains.

6.4.3 Phylogenetic analysis depicts distinct clusters of *B. sporothermodurans* separate from the *B. subtilis* group

Figure 6.3 shows the phylogenetic relationship of *Bacillus* species, including the most studied and newly emerging ones. *B. oleronius* showed the higher relatedness to the *B.*

sporothermodurans strains and both *Q. thermotolerans* and *B. thermoamylovorans*, two highly heat resistant bacteria, are also present in the *B. sporothermodurans* cluster (cluster A). Other clusters include the *B. subtilis* group (cluster B) and *B. cereus* group (cluster C). Cluster D mostly consists of alkaliphiles, including *B. akibai* and *B. hemicellulosilyticus* used in the degradation of cellulose and starch in the biotechnology industry.

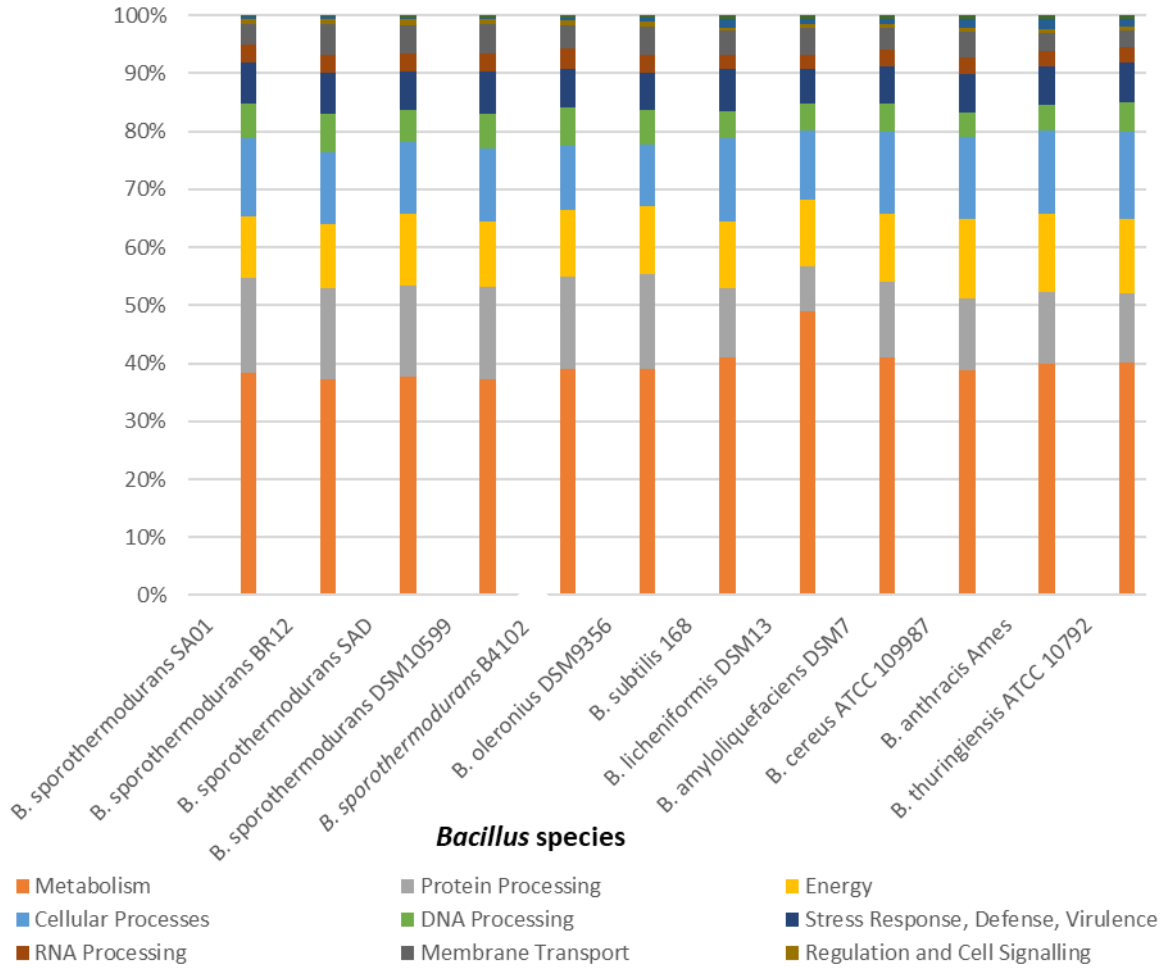


Figure 6.2: Functional annotation of *Bacillus* genomes. Graph shows the percentage of genes in each subsystem found in each species. *B. sporothermodurans* strains SA01, SAD and BR12 are from the present study and isolated from UHT milk from South Africa and Brazil respectively. The type strain DSM 10599^{T} was procured from DSMZ and isolated from UHT milk in Italy. Remainder of *Bacillus* strains sourced from NCBI GenBank.

6.4.4 Single nucleotide polymorphism (SNP) analysis

The phylogenetic tree (Figure 6.4a) arising from SNP calling using the type strain *B. sporothermodurans* as a reference clustered the two *B. sporothermodurans* strains belonging to the high heat resistant clones together. The number of SNPs between those strains was 183 as opposed to 11,025 SNPs for the two most distant strains.

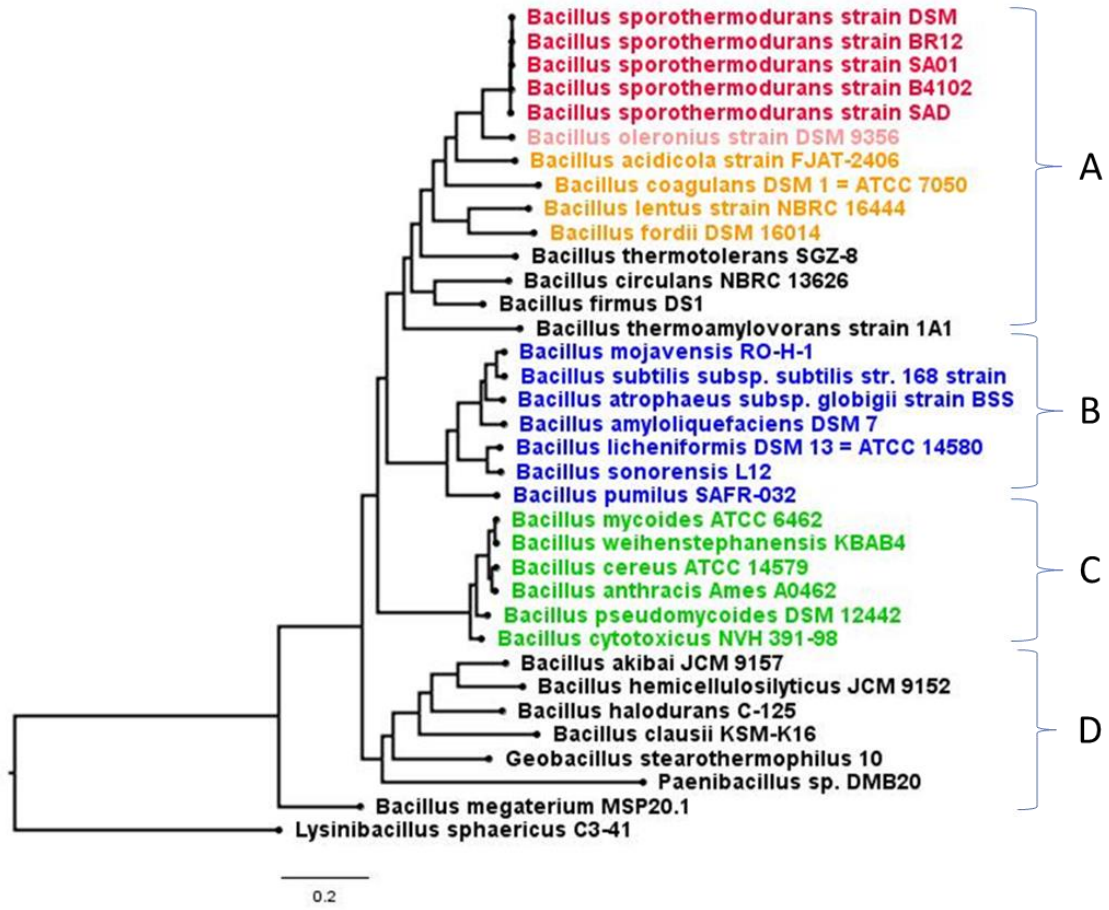


Figure 6.3: Phylogenetic analysis of *Bacillus* species showed distinct clusters for *B. sporothermodurans* and its group (A), *B. subtilis* group (B) and *B. cereus* group (C). Tree was constructed using the RAxML method and processed using FigTree. Tree was rooted with *Lysinibacillus sphaericus* C3-41

6.4.5 Protein clusters involved in heat resistance, spore coats and cell walls

Protein families were identified using the Pathosystems Resource Integration Centre (PATRIC) facility and protein families responsible for heat resistance are shown in figure 6.5. A combined core of 482 protein clusters common to all 12 *Bacillus* species were identified. *B. sporothermodurans* and its closest relative, *B. oleronius*, showed 679 and 2,559 clusters, respectively. The *B. subtilis* and *B. cereus* groups had 659 and 2476 clusters, respectively. *B. sporothermodurans* had a total of 38, 68 and 431 clusters common to the *B. subtilis* group, *B. cereus* group and *B. oleronius*, respectively.

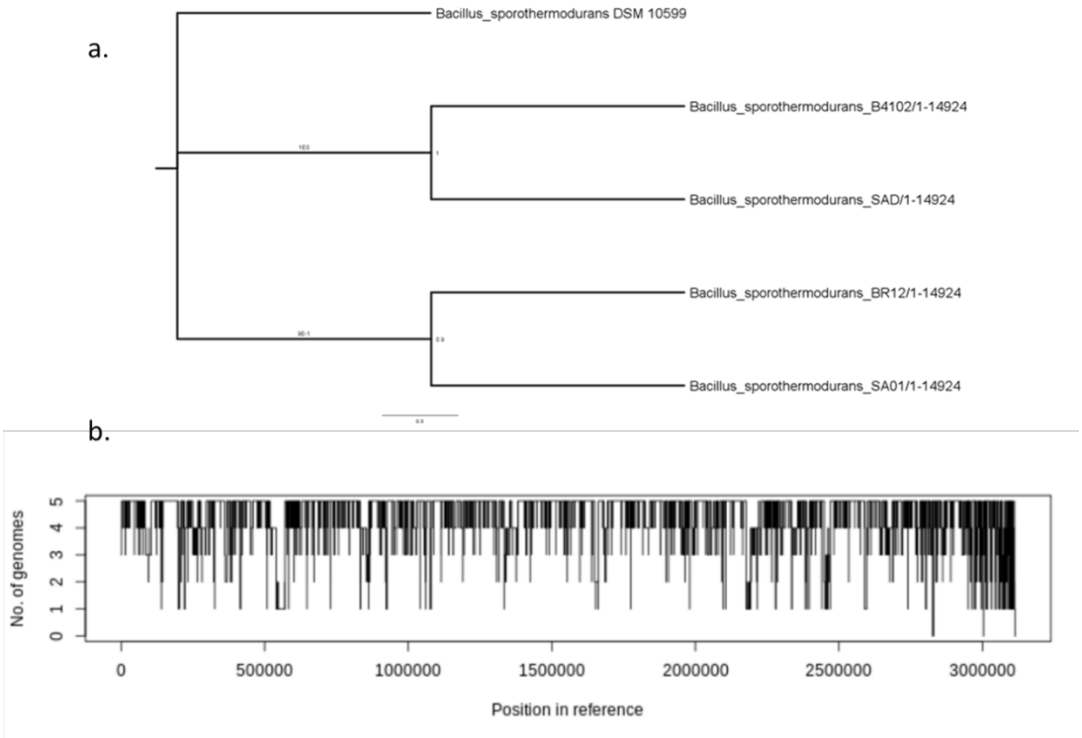


Figure 6.4: Single nucleotide polymorphism (SNP) analysis showing (a) phylogeny of five *B. sporothermodurans* strains SAD, SA01, BR12, DSM 10599^T and B4102, type strain DSM 10599^T used as a reference. The phylogenetic tree was derived from the Newick files of CSI phylogeny and processed with FigTree, (b) genome position plot of SNPs by novel SNP analysis using CSI phylogeny

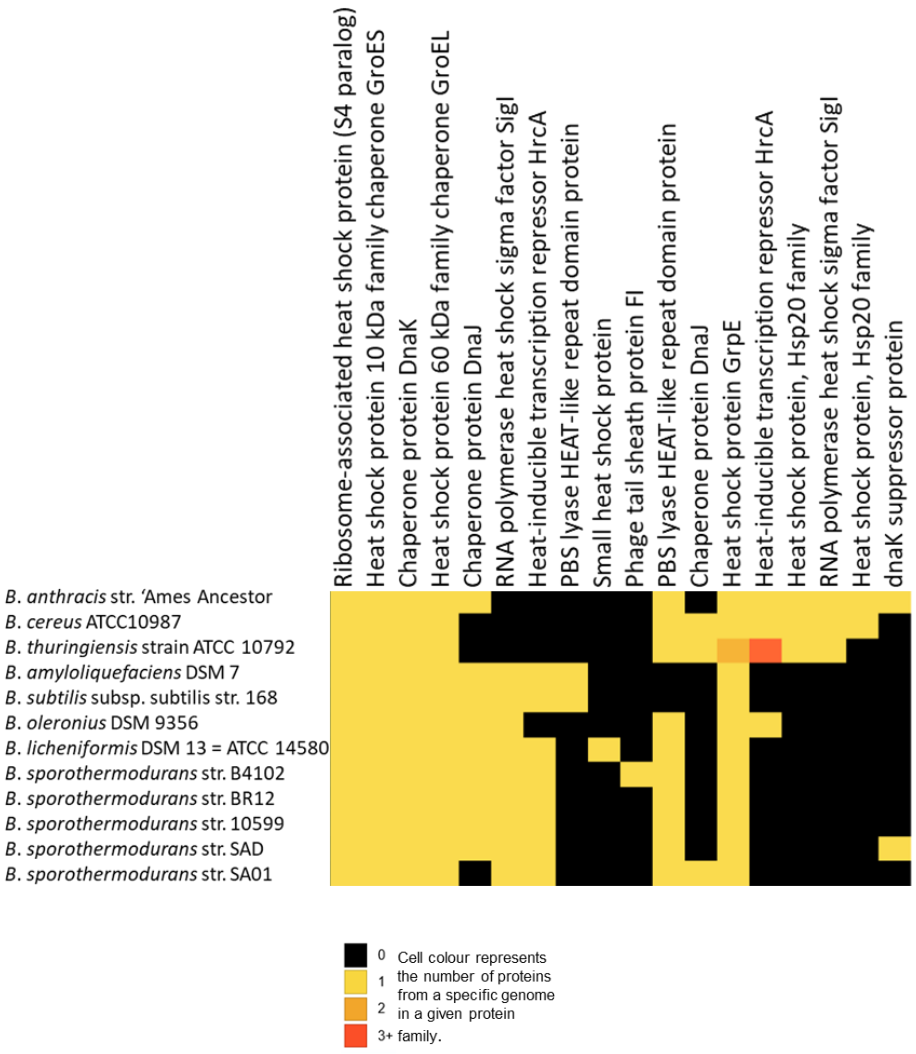
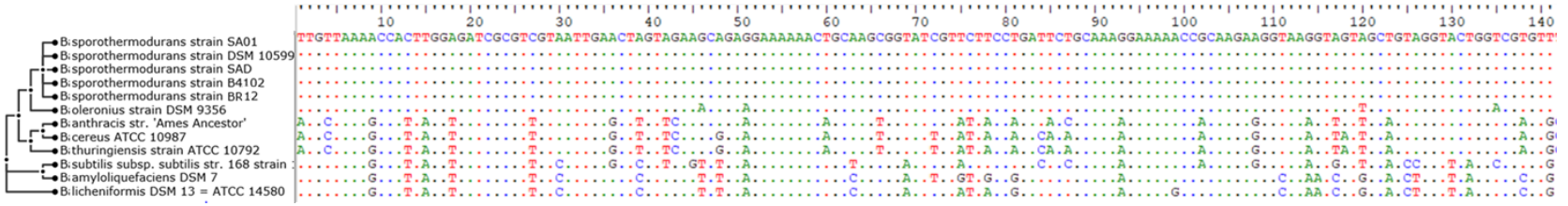


Figure 6.5: Heat map of genomes of *Bacillus* species and associated protein clusters responsible for heat resistance. The heat map was inferred using the PATRIC protein families' utility and clustered by genomes using the Pearson correlation and by pairwise average-linkage. Cell colour represents the number of proteins from a specific genome in each protein family.

Figure 6.6 shows the amino acid and nucleotide sequence alignments of the heat shock protein GrpE (A) shown in the heat map in Figure 6.5. The protein sequence alignment depicts areas of conservation and mutations in the *Bacillus* species under study, and no differences were detected among the *B. sporothermodurans* strains. Although there is a large degree of

conservation in the heat shock protein GrpE amongst the study strains as is evidenced in the amino acid sequences in Figure 6.6 B, differences in nucleotide and consequently amino acids was apparent between *B. sporothermodurans* strains, *B. cereus* and *B. subtilis* groups. GrpE, a heat shock protein in the *dna*K gene cluster, is shown in Figure 6.7, together with associated genes responsible for heat shock. There were no differences observed in the GrpE protein region of the *dna*K gene cluster with respect to the *B. sporothermodurans* strains. Conserved genes are given the same number and shown with a grey background colour.

A.



B.

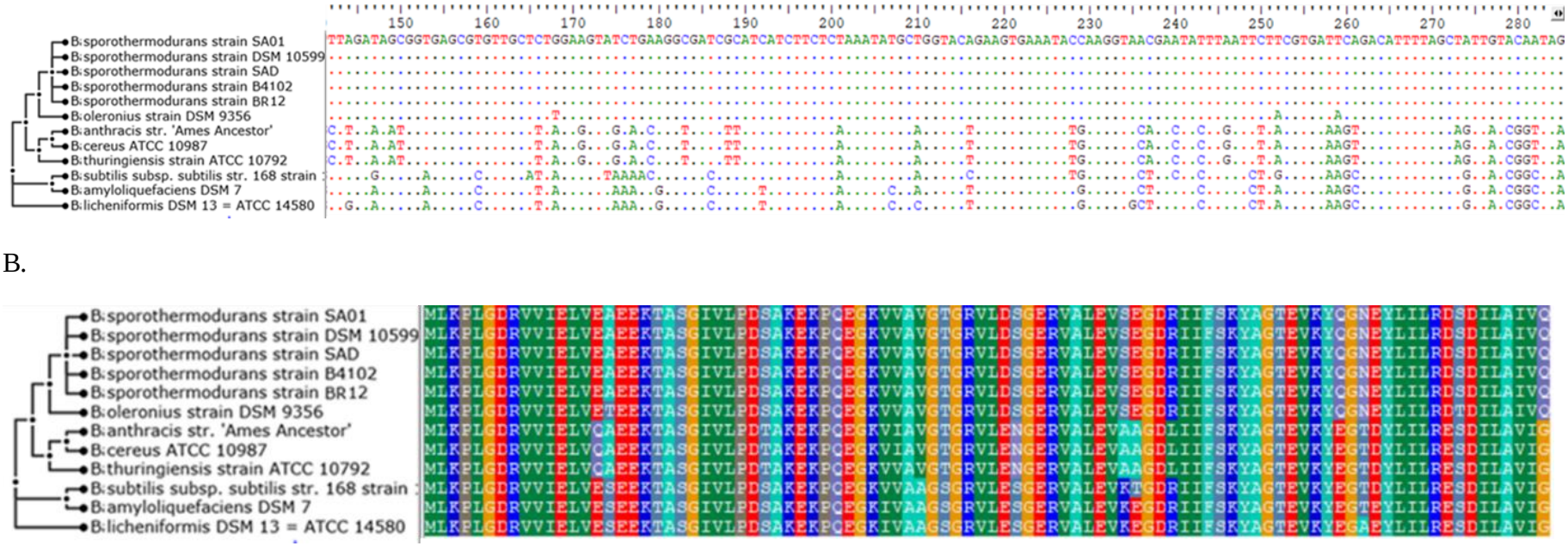


Figure 6.6: Nucleotide sequence alignment of the heat shock protein, GrpE (A), showing conserved regions and the sequence alignment of its corresponding amino acid sequence (B), indicating the effects of the various mutations of the nucleotides on the protein sequence. Sequences were aligned with BioEdit v7.

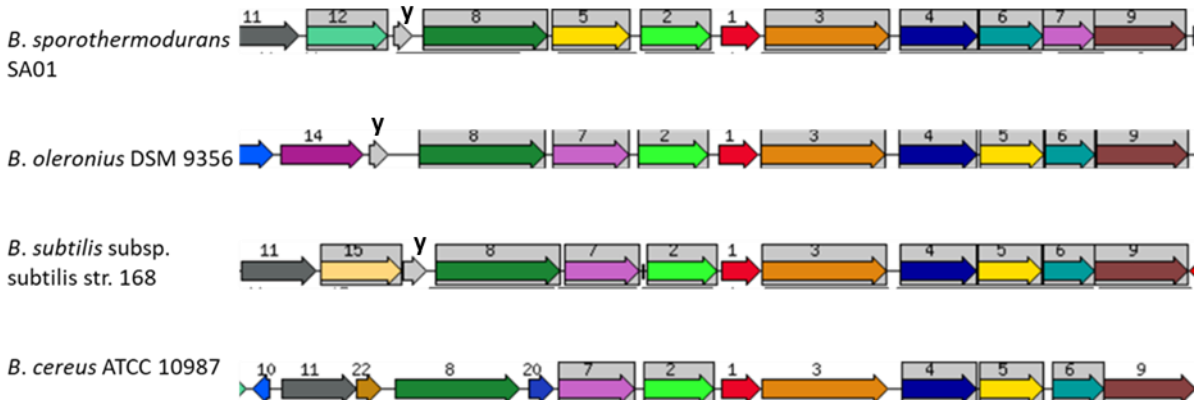


Figure 6.7: Comparison of chromosomal regions of the heat shock protein GrpE (1; red) of *B. sporothermodurans* SA01, *B. oleronius* DSM 9356^{T}, *B. subtilis* subsp. *subtilis* str. 168 and *B. cereus* ATCC 10987^{T}. Sets of genes with similar sequence are grouped with the same number and colour. Genes whose relative position are conserved in at least four other species are functionally coupled and share grey background boxes. Proteins of interest in the region are as follows: **2**, heat inducible transcription repressor HrcA; **3**, chaperon protein DnaK; **4**, chaperon protein DnaJ; **5**, hypothetical radical SAM family enzyme in heat shock genes cluster, **6**, ribosomal protein L11 methyltransferase; **7**, ribosomal RNA small subunit methyltransferase; **8**, translation elongation factor LepA; **9**, tRNA-t(6)A37 methylthiotransferase; **20**, transcriptional regulator, HxIR family; **22**, hypothetical protein; **y**, *yqe*P / *yqx*A gene.

Figure 6.8 depicts for the first time the tertiary protein structures of GrpE from *B. sporothermodurans* and from its closest phylogentic neighbour, *B. oleronius*, in comparison with *B. cereus* and *B. subtilis*. The GrpE protein structure of the four *Bacillus* species on average has a 64% identity with the GroES family of proteins and 37.5% average identity to the human mitochondrial chaperonin symmetrical football complex. Conservation analysis was undertaken using the Jensen-Shannon divergence approach (Nallapareddy *et al.*, 2002) and amino acid conservation was based on a relative conservation sensitivity of 75%.

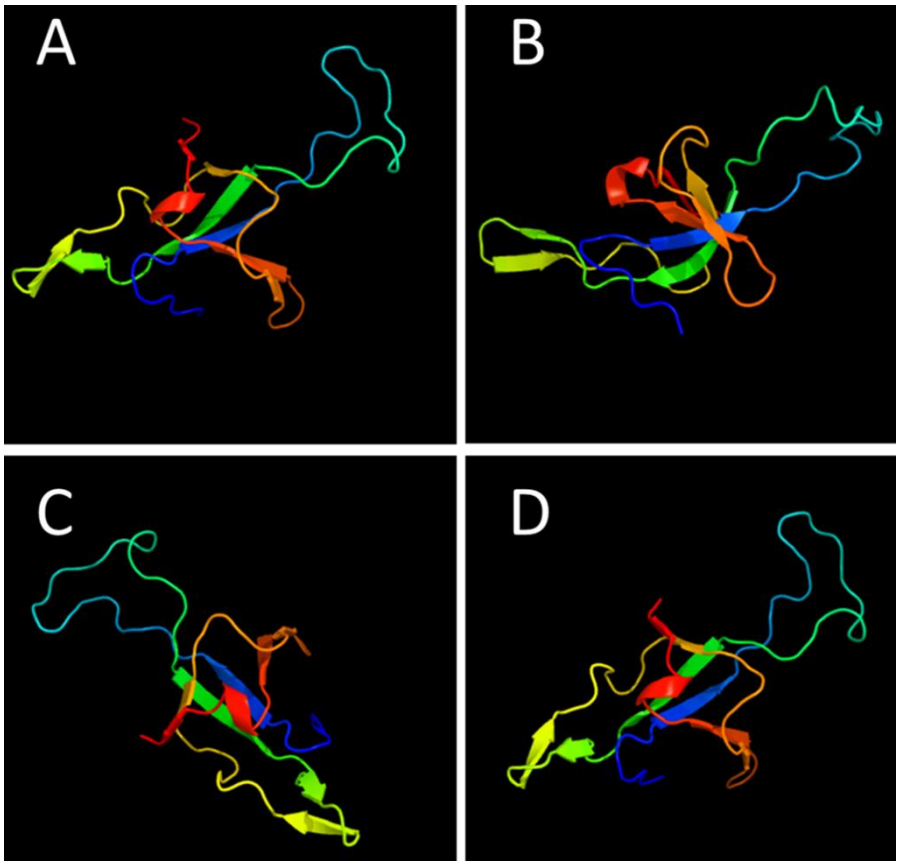


Figure 6.8: Comparison of predicted tertiary structure of the heat shock protein GrpE as identified in *B. sporothermodurans* strains SA01, SAD, BR12, DSM 10599^{T}, and B4102 (A), *B. oleronius* DSM 9356^{T} (B), *B. cereus* ATCC 10987^{T} (C) and *B. subtilis* subsp. *subtilis* str. 168 (D). Models were predicted using the Phyre2 suite of tools and at 100.0% model confidence for all four protein structures. Protein sequence alignment undertaken with BioEdit v7.2.1.

6.5 DISCUSSION

The various genomes compared in this study differ in size and only strains from the *B. cereus* group showed plasmids. The presence of plasmids in the *B. cereus* group is especially expected,

since they often carry genes for antibiotic resistance and virulence (Anderson *et al.*, 2005). Mostly through conjugation (Smillie *et al.*, 2010), these plasmids form part of the mobilome. The number of plasmids detected for *B. thuringiensis* is 6, but 9 were shown in other studies (Anderson *et al.*, 2005). The *B. thuringiensis* plasmids are known to include the toxin-carrying plasmid pBtoxis (FEMS 12) used in pest control, while the *B. cereus* plasmid is the pBClin15 (FEM 5), *B. anthracis* carries two plasmids, pXO1 and pXO2, responsible for anthrax. Though *B. sporothermodurans* strains in this study did not present plasmids, strains BR12, SAD, DSM 10599^{T} show the *Listeria* Pathogenicity Island LIPI-1. Since plasmids and genomic islands are part of the mobilome, the evolutionary future of *B. sporothermodurans* may include the presence of plasmids through horizontal gene transfer carrying either antimicrobial resistance or virulence elements.

B. cereus group showed more unique genes (*B. thuringiensis* having as much as ≈7%, for instance) when compared with an average 1% found for the *B. sporothermodurans*, *B. oleronius* and *B. subtilis* groups. The increased unique sequences in the *B. cereus* group can be attributed to the presence of plasmids and prophages, which usually have DNA content without similarity with known proteins (Anderson *et al.*, 2005).

B. sporothermodurans shares all its virulence genes repertoire with its closest phylogenetic neighbour *B. oleronius*, which is considered a pathogen, and together with the microscopic mite *Demodex folliculorum* is involved in the skin condition rosacea (Kuhnigk *et al.*, 1995). The virulence genes in *B. sporothermodurans* strains include the homologs *cod*Y, *clp*X, and *sod*A1 identified in *B. anthracis*, *clp*P, *pur*A, *pur*B, and *fur* from *L. monocytogenes* and *rec*A from *S. aureus*. All genes identified have an identity of between 80 and 87% with homologs in the source organisms.

The CodY protein regulates toxin gene expression in *B. anthracis* by controlling the AtxA protein, with studies indicating that the disruption of the *codY* gene may halt virulence in *B. anthracis* (Van Schaik *et al.,* 2009). Thus, the activity of CodY is required for full virulence of toxicogenic *B. anthracis*. Similarly, the absence of the *cIp*X gene reduces lethality of *B. anthracis* by making it susceptible to antimicrobial peptides *in vitro* (McGillivray *et al.,* 2009). The ClpX protein is a member of the Clp/Hsp100 family of ATP-dependent chaperones and, in conjunction with ClpP, works to degrade protein substrates bearing specific recognition signals (LaBreck *et al.,* 2017). ClpX targets specific proteins for degradation directly or with substrate-specific adaptor proteins. Thus, the ClpB, ClpA, ClpX protein complex help in the tolerance to extreme temperatures in various prokaryotes, by remodelling and degrading aggregated proteins (LaBreck *et al.,* 2017). The *sod*A1 gene found in the exosporium of matured spores protects against oxidative stress (Cybulski *et al.,* 2009), whilst also contributing to pathogenicity in *B. anthracis*. At similarity levels around 84%, these natively *B. anthracis* genes are missing essential components to offer virulence in *B. sporothermodurans*. The *pur*A and *pur*B genes help in the growth, multiplication and colonization of *L. monocytogenes* in the gastrointestinal tract and subsequent systemic infection of the spleen and liver (Faith *et al.* 2012), with the *fur* gene affecting virulence in the aforementioned organism (Rea *et al*., 2004; Schauer *et al.,* 2010). The Fur protein complex in *L. monocytogenes* is involved in the acquisition of iron from the human host and hence essential for its survival (Ledala *et al.,* 2010). *B. sporothermodurans* strains share homologs of the *pur*B, *pur*A, *clp*X, *cod*Y and *rec*A genes with the *B. subtilis* groups with identities of 81 – 89%. The *rec*A gene is responsible for genetic recombination and DNA repair and forms part of the SOS response of coordinated activation of diverse unlinked genes involved in DNA repair, error-prone DNA replication, in response to severe DNA damage (Bayles *et al.,* 1994).

B. sporothermodurans strains clusters with *B. oleronius* and other closely related *Bacillus* species, such as *B. lentus* and *B. firmus* (Figure 3, cluster A), which were initially considered as their closest relatives (Kuhnigk *et al.*, 1995), through 16S rDNA partial sequencing. Whole genome phylogenetic analysis suggests *B.* firmus is more distant to *B. sporothermodurans*, and *B. fordii* a closer relative. *B. thermoamylovorans,* also an emerging highly heat resistant *Bacillus* species isolated from milk and found in the *B. sporothermodurans* cluster, is known to be almost as heat resistant as *B. sporothermodurans* (Flint *et al.*, 2016). Similarly, the moderately heat resistant *B. thermotolerans*, initially thought to be closely related to *B. firmus* (Yang *et al.*, 2013), was included in the cluster A, albeit recent calls for it to be given a new genus, *Quasibacillus thermotolerans* (Verma *et al.*, 2017). Due to the relatedness of the whole genome sequence of *Q. thermotolerans* to *B. coagulans*, *B. firmus* and *B. lentus* as evidenced in this present study, it is of opinion that it remains in the genus *Bacillus*. There is however a contradiction between Figure 5.1 and Figure 6.3 as to the position of *Q. thermotolerance*. This is a result of the differences in the number and organisms used as outgroups for the rooting of the two trees. The categorisation of bacterial populations are easier applied by using sequence similarity thresholds without being necessarily biologically relevant (Bobay and Ochman, 2017).

High heat resistance in *B. sporothermodurans* has been attributed to strain differences. Consequently, single nucleotide polymorphisms (SNPs) in addition to the pangenome has been advocated for use in tracking clonal strains. Thus, this combination of resources may help remove the stumbling block of the high number of genes with unknown function and give a super-resolution view into bacterial subpopulations (McNally *et al.*, 2016). Given the proposed function of the pangenome and SNPs in adoptive evolution, the latter may be employed as a predictive genotypic marker to identify characteristics, such as high heat resistance attributed

to strain differences, offering much needed insight into niche specification and adaptation (Mcinerney *et al*., 2017). The minimum and maximum number of SNPs in this study amongst the *B. sporothermodurans* strains are 97 and 11255, respectively, indicating an extensive range of genome differences amongst strains. SNPs between the reference strain (DSM 10599^{T}) and the other two strains belonging to the HRS clone, SA01 and BR12, were of 97 and 112, respectively, and 183 between BR12 and SA01. Though strain DSM 10599^{T} did not amplify the HRS clone feature through PCR, it is closest to the two strains belonging to the HRS clone. This is because bacterial species rely on a relatively small number of unique regions to generate variation (Raskin *et al.*, 2006). Higher numbers of SNPs were found comparing strain B4102 (isolated from Indian curry) with the remaining strains (isolated from UHT milk), indicating a possible influence of niche adaptation on DNA sequence variation.

To maintain cellular proteostasis, cells need various chaperone pathways that enhance inherent protein folding (LaBreck *et al.*, 2017). These chaperones target misfolded, unfolded and aggregated polypeptides for reactivation or degradation (Bukau and Horwich, 1998; Wickner *et al*., 1999; Stoecklin and Bukau, 2013). The misfolding of proteins, a consequence of the effects of environmental stresses, such as increased heat exposure, causes polypeptide elongation (LaBreck *et al.*, 2017). The importance of these chaperone pathways is especially important in organisms, such as *B. sporothermodurans* and other *Bacillus* species, exposed to increased heat conditions during food processing. Again, the presence of disulphide bonds is known to play a critical role in the structural stabilization of intracellular proteins, with thermophiles normally rich in these bonds (Beeby *et al.*, 2005; Ladenstein and Ren, 2006). The response to heat shock, including such extreme conditions as thermal processing of food, is mediated by a family of proteins called heat shock proteins (HSPs). Apart from heat shock, HSPs are known to be also expressed by other physiological stresses, such as exposure to cold,

UV light or during cell healing (Hecker *et al*., 1996). The main heat shock proteins involved in prokaryotes are the GroES, GrpE, DnaJ, GroEL, DnaK, HtpG and ClpB, ClpA, ClpX group (Liberek and Georgopoulos, 1993; Schroder *et al.*, 1993; Gamer *et al.*, 1996; Hecker *et al.*, 1996; Bukau and Horwich, 1998).

Apart from heat shock protein HSP20 (present only in the *B. cereus* group), heat inducible transcription repressor HrcA (present only in *B. subtilis* and *B. sporothermodurans*), DnaK suppressor protein (detect only in *B. amyloliquefaciens* and *B. subtilis*), the remainder of the clusters had homologs in all *Bacillus* groups evaluated in this study. HSP20, also known as small heat shock proteins (sHSPs), comprises proteins between 10 – 20 kDa. They act as chaperones that protect other proteins against heat induced denaturation and aggregation. *hrc*A is the first cistron of the *B. subtilis dna*K operon and encodes a negative regulator of class I heat shock genes (*dna*K and *gro*E operons) (Hecker *et al*., 1996; Schulz and Schumann, 1996). DnaK, DnaJ and GrpE form a cellular chaperone machinery involved in the repair of heat-induced protein damage (Schroder *et al.*, 1993). DksA, a 163-chain amino acid protein, homolog of YocK and TraR, was initially thought of acting as a multicopy suppressor of *dna*K, *dna*J, *grp*E and *prc* (Eichenberger *et al.*, 2003). The DksA protein binds directly to RNA polymerase, affecting transcript elongation and augmenting the effect of the alarmone ppGpp on transcription initiation (Cashel and Potrykus, 2013). However, recent research suggests that the suppression of *dna*K and *dna*J is rather by an unidentified product whose promoter is regulated by DksA (Chandrangsu *et al.*, 2012). DksA, an RNA polymerase (RNAP) binding transcription factor that controls expression of a large number of genes in concert with the small-molecule “alarmone” ppGpp, also aids in the resolution of conflicts between RNAP and DNA polymerase (DNAP) during genome replication (Chandrangsu *et al.*, 2012).

Many of these protein clusters shared amongst the *Bacillus* species are homologs of each other with differences along the amino acid chain. In this sense, Figure 6 depicts nucleotide and amino acid alignment of GrpE, a chaperone part of the *dna*K operon that in conjunction with GroEL (a 60 kDA family chaperone) is responsible for renaturation of heat denatured proteins (Schroder *et al.*, 1993).

Interestingly, there are differences in the arrangement of genes in the chromosomal region of GrpE among the *Bacillus* species studied (represented in Figure 6.7). Although GrpE DnaK and DnaJ are present in similar positions in all species, the hypothetical radical SAM family enzyme (gene 5) swaps locations with the ribosomal RNA small subunit methyltransferase (gene 7) in *B. sporothermodurans* only. The effects of this change to the heat resistance or other biological functions of *B. sporothermodurans* is not known. Additionally, the ribosomal RNA small subunit methyltransferase in *B. sporothermodurans* consists of a 251 amino acid chain as compared to 379 in *B. oleronius* and *B. cereus* and 367 in *B. subtilis*. The amino acid chain length for the hypothetical radical SAM family enzyme and ribosomal protein L11 methyltransferase in *B. sporothermodurans* is 379 and 313 respectively as compared to 313 and 251 respectively in *B. oleronius*, 312 and 257 respectively in *B. subtilis* and 313 and 250 respectively in *B. cereus*. The genes in the dnaK protein cluster is highly conserved in the *B. sporothermodurans* strains and exhibited identical amino acid length and configuration. The hypothetical radical SAM family enzyme may improve the action of *hrc*A in *B. sporothermodurans*. HrcA aids in the transcription of proteins produced to protect cellular proteins from denaturation (Liu *et al.*, 2005). It has been shown to act as a thermo-sensor (Hitomi *et al.*, 2003) in *B. subtilis* and *B. thermoglucosidasius*, the latter possessing significantly higher heat resistance than the former. The translation elongation factor LepA (gene 8) thus acting on the hypothetical radical SAM family enzyme in *B. sporothermodurans*

is required for accurate and efficient protein synthesis, inducing a back-translocation after a defective translocation reaction (Gruber and Gross, 2003; Qin *et al.,* 2006). Consequently, this maintains the integrity and function of cellular proteins. The LepA protein is, however, present in all bacteria.

Other differences include the *yqe*P gene, synonymous to the *yqx*A gene, is uncharacterised and hence its function unknown and acting on the translation elongation factor LepA in *B. sporothermodurans*, *B. oleronius* and *B. subtilis*, but not *B. cereus*. *B. cereus*, however, possesses the transcriptional regulator HxIR, a DNA binding protein, which directs gene activity by the conversion of DNA to RNA (Gruber and Gross, 2003). It is responsible for detoxification of formaldehyde, with its interruption resulting in the intracellular accumulation of formaldehyde, thus, decreasing bacterial survival (Willenbacher *et al.,* 2016). These differences may well be involved in the efficiency of transcription regulation and the protection of cellular proteins, and, consequently, result in different resistance to stresses, including heat shock.

Besides the sequence variation on the chromosomal region of *grp*E, it could be verified that the predicted tertiary protein structure of GrpE from the *Bacillus* species studied showed differences as well (Figure 8). GrpE from *B. sporothermodurans* and *B. oleronius* presents almost identical amino acid chains, only differing on the 16th amino acid, where the former's alanine is replaced by threonine in the latter. Conservation analysis shows 10 out of the 94 amino acids conserved in *B. sporothermodurans*, same as in *B. oleronius*, and as compared to 6 and 9 conserved amino acids in *B. cereus* and *B. subtilis*, respectively. These conserved amino acid residues are essential for the appropriate structure and function of the GrpE (Capra and Singh, 2007) with mutations likely to affect it. The detection of 'pockets', often found to

harbour active sites along the protein structure, identified 13 likely amino acid active sites in the GrpE protein for both *B. sporothermodurans* and *B. oleronius*. The exact locations of 'pockets' were identical for both *B. sporothermodurans* and *B. oleronius* along the amino acid chain. On the other hand, *B. cereus* and *B. subtilis* exhibited 38 and 12 pockets, respectively. The high numbers of likely amino acid active sites in *B. cereus* may be correlated with roles in enzymatic reactions involving its numerous virulence genes. GrpE is known to take part in virulence of the pathogens, including *Escherichia coli* and *Clostridium* sp. (Liberek and Georgopoulos, 1993).

Many bacteria of the orders Bacillales and Clostridiales survive unfavourable conditions by the formation of spores. Bacterial spore layers consist of the exosporium, spore coats, outer membrane, cortex, germ cell wall, inner membrane and central core (Setlow, 2006). The spore protects against both dry and wet heat treatments. In the latter, inactivation of the spore can occur due to the penetration of hot moisture with subsequent enzyme inactivation and the leaking out of vital cellular constituents, such as dipicolinic acid (DPA) (Tabit and Buys, 2010). The spore coat in the *B. cereus* group is well characterised, with the coat and exosporium very much conserved (Anderson *et al.*, 2005). The coat protein genes *cot*A, *cot*C, *cot*G, *cot*I, *cot*O, *cot*R, *cot*S, *cot*T, *cot*U, *cot*V, *yuz*C, *ytr*L, *yhe*D and *ykn*T found in *B. sporothermodurans*, *B. oleronius* and *B. subtilis* group had no such homologs in the *B. cereus* group. CotG is responsible for the assembly of CotB (Driks, 1999). However, *saf*A, *yaa*H, *yab*G, *spo*VID, *cot*B, *cot*D, *cot*E, *cot*H, *cot*JA, *cot*JB, *cot*JC, *cot*M, *cot*Y, *spo*IVA, *spo*VM, *spo*VID, *saf*A, *yab*G and *yaa*H genes are found in all species (Anderson *et al.*, 2005), with *cot*E and *cot*M required for outer coat assembly and control, respectively (Driks, 2002). The spore coat proteins YuzC, Ytrl, YheD, YknT, CotR and CotO are shared amongst the *B. subtilis* group, *B. sporothermodurans* and *B. oleronius*. Similarly, the spore proteins SpoIVA, SpoVM, SpoVID,

SafA, CotJC, YabG and YaaH have homologs in the *B. subtilis* group, *B. cereus* group *B. sporothermodurans* and *B. oleronius* (Driks, 1999, 2002). Important and specific functions of many of these proteins are still not thoroughly understood as well as the known effects of the deletions of any one coat protein on the overall sporulation process. However, the morphogenetic proteins SpoIVA and CotE have specific known roles in coat assembly, with the absence of the former resulting in a significant deformation of the bacterial coat (Driks, 2002). Conserved in all *B. cereus* group organisms are the *exs*B, *exs*C, *exs*D, *exs*E, *exs*F and *exs*J genes responsible for exosporium proteins (Bailey-smith *et al.*, 2005).

6.6 Conclusion

Comparative analysis of *B. sporothermodurans* with other *Bacillus* species show the close relatedness of these *Bacillus* species and of the propensity for HGT of genes. *B. sporothermodurans* and other *Bacillus* species share sets of core genes with differing percentage identities that may invoke varying phenotypic responses in different bacteria. Thus, though *B. sporothermodurans* share some virulence genes with some pathogenic bacteria, it does not presently show pathogenic potential. However, this may not always be the case in its long term evolutionary future as a result of the effects of mutations, HGT and other genetic events. HRS and non-HRS strains of *B. sporothermodurans* do not show differences in their heat shock protein cluster, that fuels suggestions of the influence of MGEs in their differing phenotypic responses.

Chapter 7: PANGENOME ANALYSIS OF *BACILLUS SPOROTHERMODURANS*, AN EMERGING HIGHLY HEAT RESISTANT SPORE-FORMER

Submitted to Scientific Reports

7.1 Abstract

Bacillus sporothermodurans is an emerging highly heat resistant endospore-forming bacterium that affects the quality of ultra-high temperature (UHT) milk products. Diverse strains of *B. sporothermodurans* are known to have different heat resistance and spoilage potential in UHT milk. In this study, we have described for the first time the pangenome of *B. sporothermodurans*, with the inclusion of strains belonging to both the highly heat resistant spore (HRS) and non-HRS clones. The core genome covers 81.9% (69.9 – 92.2%) of the *B. sporothermodurans* pangenome, with the remaining 18% (7.8 – 30.1%) representing the accessory genome. Approximately 44% of the subsystem categories for the core genome was attributed to amino acid, carbohydrates and protein metabolism with 50% of the accessory genome accounting for DNA metabolism, membrane transport and mobile genetic elements. The core/pangenome ratio of 55% suggests that *B. sporothermodurans* has an open pangenome with sympatric characteristics. The presence of a complete prophage in strain B4102, which incidentally exhibited three times as much clustered regularly interspaced short palindromic repeats (CRISPRs) was identified. Our findings provide a stepping-stone for the extensive characterization of *B. sporothermodurans*, especially as related to its high heat resistance to ultra-high temperature processing conditions in the dairy industry.

Keywords: pangenome, accessory genes, core genome, mobile genetic elements, prophage

7.2 Introduction

Bacillus sporothermodurans is an endospore forming bacteria, with high heat resistance that may survive ultra-high temperature (UHT) processing of dairy products (Pettersson *et al.*, 1996). The spore of *B. sporothermodurans* can typically withstand temperatures of 140 °C for a few seconds (Huemer *et al.*, 1998), which coincides with the processing times of UHT milk. *B. sporothermodurans* has been isolated in UHT processed milk products, evaporated milk (Herman *et al.*, 1997), and other food products including soy, canned soups and Indian curry (Krawczyk *et al.*, 2016). Its spores have been found to be more resistant than other heat resistant spores at temperatures above 130 °C, with D140 ranging from 3.4–7.9 s and Z-values ranging from 13.1 – 14.2 °C (Huemer *et al.*, 1998). Increasing the processing temperature or the holding time to inactivate *B. sporothermodurans* negatively impacts the organoleptic and sensory quality of UHT milk (Claeys *et al.*, 2001). Although *B. sporothermodurans* is not known to be pathogenic, their presence in UHT milk constitute a quality problem, resulting in reprocessing or discarding of milk products, leading to considerable economic losses (Scheldeman *et al.*, 2006; Burgess, Lindsay and Flint, 2010).

Over the years *B. sporothermodurans* has been increasingly identified in the UHT processing supply chain, with the contaminating spores associated with the problem of non-sterility (Aouadhi *et al.*, 2014; Pinto *et al.*, 2018). Much of the contamination of UHT products worldwide has been attributed to the highly heat-resistant spore (HRS) clone. Various molecular studies have shown heterogeneity of *B. sporothermodurans* strains isolated from diverse sources, and sometimes confusingly exhibiting phenotypic characteristics like *B. oleronius*, its closest phylogenetic neighbour, and *B. fordii* (Guillaume-gentil *et al.*, 2002, Scheldeman *et al.*, 2002, 2006, Heyndrickx *et al.*, 2012). The availability and characterization

of additional genomes of *B. sporothermodurans*, specifically of strains presenting different phenotypic traits, will help decipher the physiologic complexity of this increasingly important organism.

With the improvements and reduction in costs of next generation sequencing (NGS) technologies, whole genome comparative studies are progressively being used to identify the plasticity that exists within genomes. Because of the complexity of bacterial genomes, deciphering their architecture and dynamics is of prime importance in identifying elements that may infer relevant phenotypic characteristics, including heat resistance. The *B. sporothermodurans* genome contains between 3,768 – 4,558 protein coding genes and because of evolution drivers, some strains may contain genes not found in closely related taxa (Tettelin *et al.,* 2005; Wiedenbeck and Cohan, 2011). This modification of gene content in bacteria can generally arise from gene loss and gain through duplication or horizontal gene transfer (HGT) (Rocha, 2008; Kuo and Ochman, 2009; Ozer, 2018) and can be studied using pangenomics. The bacterial pangenome (or supra-genome) describes the complete repertoire of genes in a clade or study group (Medini *et al.,* 2005; Tettelin *et al.,* 2008; Georgiades and Raoult, 2011; Rouli *et al.,* 2015), which is usually made up of species or strains of a species. The pangenome consists of the core genome, genes shared by all or nearly all strains, the accessory genome, genes present between two and n-1 strains (Medini *et al.,* 2005; Tettelin *et al.,* 2008).

The core genome mainly comprises genes for basic survival and its major phenotypic traits. In contrast, the accessory genome is significant in the ability of bacteria to persevere especially in adverse environmental conditions. The bacterial accessory genome can infer important phenotypic characteristics, including virulence, antibiotic resistance and heat resistance(Ozer *et al.,* 2014; Rouli *et al.,* 2015). Although these phenotypic characteristics may be engrained in

the core genome of specific bacterial strains, some may be acquired through HGT and thus cause a deviation to known phenotypic characteristics. Horizontally transferred genetic elements, collectively known as the mobilome(Burrus and Waldor, 2004; Kung *et al.*, 2010; Ozer, 2018), and together with other elements not part of the bacterial core genome and constituting the accessory genome can drive diversity, evolution and environmental niche adaptation(Stokes and Gillings, 2011; Gillings, 2017; Ndejjo *et al.*, 2017). The pangenomes of other industrially important *Bacillus* species that exist in food microbiomes such as *B. amyloliquefaciens, B. anthracis, B. cereus, B. subtilis* and *B. thuringiensis* have been studied (Bazinet, 2017; Kim *et al.*, 2017).

This study, for the first time identifies the pangenome of *B. sporothermodurans* strains, including both HRS and non-HRS clones. Identification of the core and particularly the accessory genome of *B. sporothermodurans* is key to eventually decoding its evolution and adaptation to especially harsh environments including high heat processing conditions. The pangenome of *B. sporothermodurans* will also enable comparative studies with other available pangenomes of *Bacillus* species that exist in the food microbiome.

7.3 Materials and methods

7.3.1 Growth conditions and molecular confirmation of strains

The strains involved in the study were isolated from UHT milk produced in South Africa and Brazil, with the type strain from Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH (DSMZ). Cultivation of all isolates was performed on Brain Heart Infusion agar (Oxoid, UK) and incubated at 37 °C for 48 hours. Genomic DNA

was extracted using the ZR Bacterial DNA Miniprep kit (Zymo Research, USA). DNA extract was quantified using the Qubit instrument and dsDNA BR Assay kit (Life Technologies, USA). Molecular confirmation of *B. sporothermodurans* strains and those carrying the HRS clone was undertaken using *B. sporothermodurans* specific primers of BSPO-F2 (5'-ACG GCT CAA CCG AG-3') and BSPO-R2 (5'-GTA ACC TCG CGGTCT A-3')(Scheldeman *et al.*, 2002) and SH2-F1 (5'-GAT TCA GGC AGA ATG TAG CA-3') and SH2-R1 (5'-TTT GGC CAC TTG ATG GTA CA-3') (Herman *et al.*, 1997), respectively.

7.3.2 Genome Sequencing

Multiplexed paired-end libraries were prepared using Nextera XT DNA Sample Preparation kit (Illumina, USA). An Illumina MiSeq system (Illumina, San Diego, USA) was used to carry out sequencing. The paired-end reads were checked for quality, trimmed and *de novo* assembled using the Qiagen CLC Genomics Workbench version 9 (Qiagen, Netherlands). Resultant contigs were submitted to GenBank, and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova *et al.*, 2016) was used for gene annotation. Subsystem based annotation was undertaken employing the Rapid Annotation using Subsystem Technology (RAST) (Aziz *et al.*, 2008; Overbeek *et al.*, 2014; Brettin *et al.*, 2015).

The annotated draft genomes of *B. sporothermodurans* strains SAD, SA01, BR12 and DSM 10599^{T} sequenced in this study, in addition to the genome from strain B4102, which was obtained from the NCBI GenBank and annotated with RAST, were used for the characterisation of core and accessory genomes. Prophage sequences were predicted using PHASTER (Arndt *et al.*, 2016).

7.3.3 Identification of core and accessory genomes

In this study the core genome was determined based on the conserved sequences among five draft genomes of *B. sporothermodurans* using the Spine software(Ozer *et al.*, 2014). The accessory genome was deduced by a genome subtraction program, AGEnt, which filtered out the core genomic sequences and subsequently clustered with ClustAGE (Ozer *et al.*, 2014; Ozer, 2018). The Spine software was used to identify the core genome by first performing genome alignments of the strains using the NUCmer function of the MUMmer software package (ver.0.3.1). The core genome was defined as corresponding to sequences present in at least 80% of the strains analysed. Sequences presenting alignments of at least 85% sequence identity with strain SAD sequences, chosen as reference genome for this study, were considered as homologs. An individual annotated gene of *B. sporothermodurans* was determined as belonging to the core genome if ≥ 50% of the gene nucleotide sequence was present within the core coordinate set. The nucleotide core genome and pangenome sizes were derived from the alignments generated by Spine and the pangenome plotted using the GView program (Stothard *et al.*, 2018).

The accessory genome was defined as all sequences found in the *B. sporothermodurans* strains that were not considered as part of the core genome. This analysis was carried out with the AGEnt algorithm, which works with the NUCmer function of the MUMmer software package (Ozer *et al.*, 2014). An individual gene was categorised to belonging to the accessory genome if ≥ 50% of its sequence was contained in the accessory region.

7.3.4 Clustering and distribution of accessory genomic elements

The nucleotide sequence inputs of accessory genomic elements (AGEs), previously from the Spine and AGEnt softwares, was incorporated into ClustAGE algorithm (Ozer, 2018). The ClustAGE algorithm involves the following steps of defining 'bins' and 'subelements' and identifies AGEs within the data set and defines the distribution of discrete AGEs among the genomes concerned. The distribution of AGEs was clustered using InteractiVenn (Heberle *et al.*, 2015) to identify relatedness amongst the five strains in terms of common AGEs.

7.3.5 Data Availability

All genome sequences used in this study have been deposited in GenBank of the National Center for Biotechnology Information (NCBI) under the following accession numbers: NAZD01000000, NAZB01000000, NAZA01000000, NAZC01000000.

7.4 Results

7.4.1 Determination and characterisation of the *B. sporothermodurans* core genome

Five strains were used to infer the core genome of *B. sporothermodurans*, including two isolated from UHT milk in South Africa (SAD and SA01), one from UHT milk in Brazil (BR12), the type strain isolated from UHT milk from Italy (DMS10599^T) and the last strain (B4102) isolated from Indian curry and sourced from GenBank. The genome sizes varied from 3.4 Mbp to 4.0 Mbp (Table 7.1). The core genome of *B. sporothermodurans* is defined as sequences representative in ≥ 80% of the reference genome (4 out of the 5 strains in the present

study). The core genome ranges from 69.9 to 92.2% of the total genome with the average G + C content of 36.7%, which was higher than that of the accessory genome (Table 7.2).

7.4.2 Determination and characterisation of *B. sporothermodurans* accessory genome

After determination of the core genome, the remaining sequences of the *B. sporothermodurans* strain constitute the accessory genome. Average size of the *B. sporothermodurans* accessory genome was 661 kbp, which represents on average 18% of the pangenomes analysed. Except for *B. sporothermodurans* BR12, the average G + C content of the accessory genome of the study strains was lower than that of the core genome. Detailed information of the accessory genome of each strain are shown in Table 12.

Table 7.1: Genome characteristics of *B. sporothermodurans* strains used in pangenome study

Genome characteristics	*B. sporothermodurans* strains				
	SAD	SA01	BR12	DSM 10599^T	B4102
Total contig size (kbp)	3,898	3,484	4,047	3,783	3,358
Contig N50 (bp)	123,945	22,616	14,323	15,396	21,499
G + C Content (%)	36.0	35.9	36.4	35.9	36.0
Coding sequences (#)	3,979	3,630	4,093	3,897	3,399
Accessory genome size (bp)	1,088,822	225,100	470,557	304,132	1,217,307
% of total contig size	27.9	6.5	11.6	8.0	36.2
Average accessory G + C (%)	33.9	34.7	38.3	35.1	33.3
# of subsystems[a]	437	414	425	431	421
#RNAs	135	115	152	100	97

[a]Subsystem annotation using RAST

The sizes of the accessory genomic elements range from 1 bp to 238,781 bp, with majority of them between 1 – 5,000 bp (Figure 7.1). Determination of prophage sequences using PHASTER yielded no known sequences in the core genome. In the accessory genome of two strains, there was evidence of up to 4 prophage sequences. Strain B4102 exhibited one intact, one incomplete and two questionable prophage sequences while strain SAD presented two incomplete and two questionable prophage sequences. The one intact and two incomplete prophage sequences of the two respective strains are represented in Figure 7.8.

Table 7.2: Core and accessory genome characteristics of *B. sporothermodurans* strains SAD, SA01, BR12, DSM10599 and B4102.

	Core[a] Avg. (range)	Accessory Avg. (range)
Size (kbp)	2,777 (2,655 – 2,829)	661 (225 – 1,217)
% of total genome	81.9 (69.9 - 92.2)	18.0 (7.8 - 30.1)
Average G + C %	36.7 (36.3 - 36.8)	35.1 (33.3 – 38.3)
Subsystem counts[b]	2,356	70
Segment length in bases	847 (11 – 9502)	450 (10 – 80295)
Predicted phage genes[c]	0	>1< 2 (0 – 4)

[a]Core genome defined in this study as the sequence present in ≥ 80% of the five genomes.

[b]Number of subsystem feature counts using RAST.

[c]Prophage sequences (intact, questionable, and incomplete) predicted by PHASTER

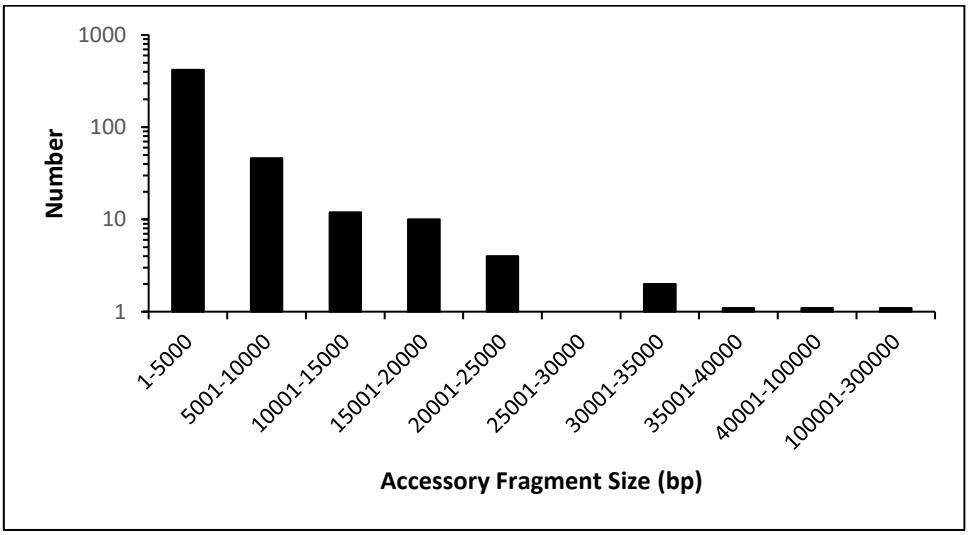


Figure 7.1: Size of accessory genomic regions of *B. sporothermodurans* strains. The size distribution of all the accessory genomic elements extracted from the pangenome of the strains used in this study is plotted. The range 25,001-30,000 had no genetic element. Ranges 35,001-40,000 bp, 40,001-100,000 bp and 100,001-300,000 bp had one genetic element each, which is shown in the plot by a bar of artificially non-zero height on the log scale.

7.4.3 Subsystem comparison of *B. sporothermodurans* core and accessory genomes

Subsystem categories of the core genome using RAST were mainly prominent with amino acids and derivatives (16.0%), carbohydrates (12.0%), protein metabolism (10.3%), and cofactors, vitamins, prosthetic groups and pigments (6.5%). The accessory genome subcategories centred mainly on DNA metabolism (28.5%), membrane transport (12.4), mobile genetic elements (phages, prophages, transposable elements and plasmids) (10.7%), and virulence, disease and defence (9.3%) and cell wall and capsule and amino acids and derivatives; 7.6% and 7.1% respectively (Figure 7.2). In absolute numbers, the core genome on average had 5 – 30 times more subsystem feature counts as compared to the accessory

genome with respective to the all but one of the subsystem categories. The exception, in the case of the phages, prophages, transposable elements and plasmids subsystem category which contained 6 subsystem feature counts in the core genome as opposed to 8 in the accessory genome. The combined total number of feature counts for all subsystem categories and for all strains in this study amounted to 2,357 and 71 for the core and accessory genome respectively.

The pangenome of *B. sporothermodurans* was examined and the locations of core and accessory genomic elements of each strain were identified (Figure 7.3). Albeit the distribution of the accessory elements was quite broad, there was some concentration of accessory genetic elements in certain parts of the genome. Also, there was a difference in the size distribution amongst the accessory elements of *B. sporothermodurans* strains. The choice of the Spine and AGEnt programs is as a result of their effectiveness in identifying the accessory genetic elements, even for draft genomes, as well as their accuracy and flexibility, as compared to others including Island viewer, Panseq and the Regions of Genomic Plasticity (RGPs)(Ozer *et al.* 2014).

7.4.4 AGE bin representation and subelement sequence distribution

The total size of bin representatives (longest contiguous AGE in the data set) of the five strains was 2,143,422 bp with an average bin length range of 703 – 1,999 bp. The average number of bin representatives for the five strains was 345 and their average bin representative size amounting to 428,684 bp (Table 7.3).

Among the studied strains, 65.5% of the accessory sequences were unique to subelements found in only one genome (Figure 7.4a). Strain B4102 had the highest representation of unique AGE sequence (40.6%), followed by strain SAD with 32.5% (Figure 7.4b).

Table 7.3: Accessory genomic element bin representative features of *B. sporothermodurans* strains

Strain	# bin representatives	Total size of bin representatives in bp	Average bin representative size in bp (min-max)
B. sporothermodurans SAD	792	960,924	1213 (200 - 43448)
B. sporothermodurans B4102	347	693,921	1999 (202 - 80295)
B. sporothermodurans SA01	158	136,405	863 (216 - 2991)
B. sporothermodurans DSM10599	139	97,779	703 (204 - 3810)
B. sporothermodurans BR12	292	254,393	871 (201 - 8472)
Total	1728	2,143,422	
Average	345	428,684	

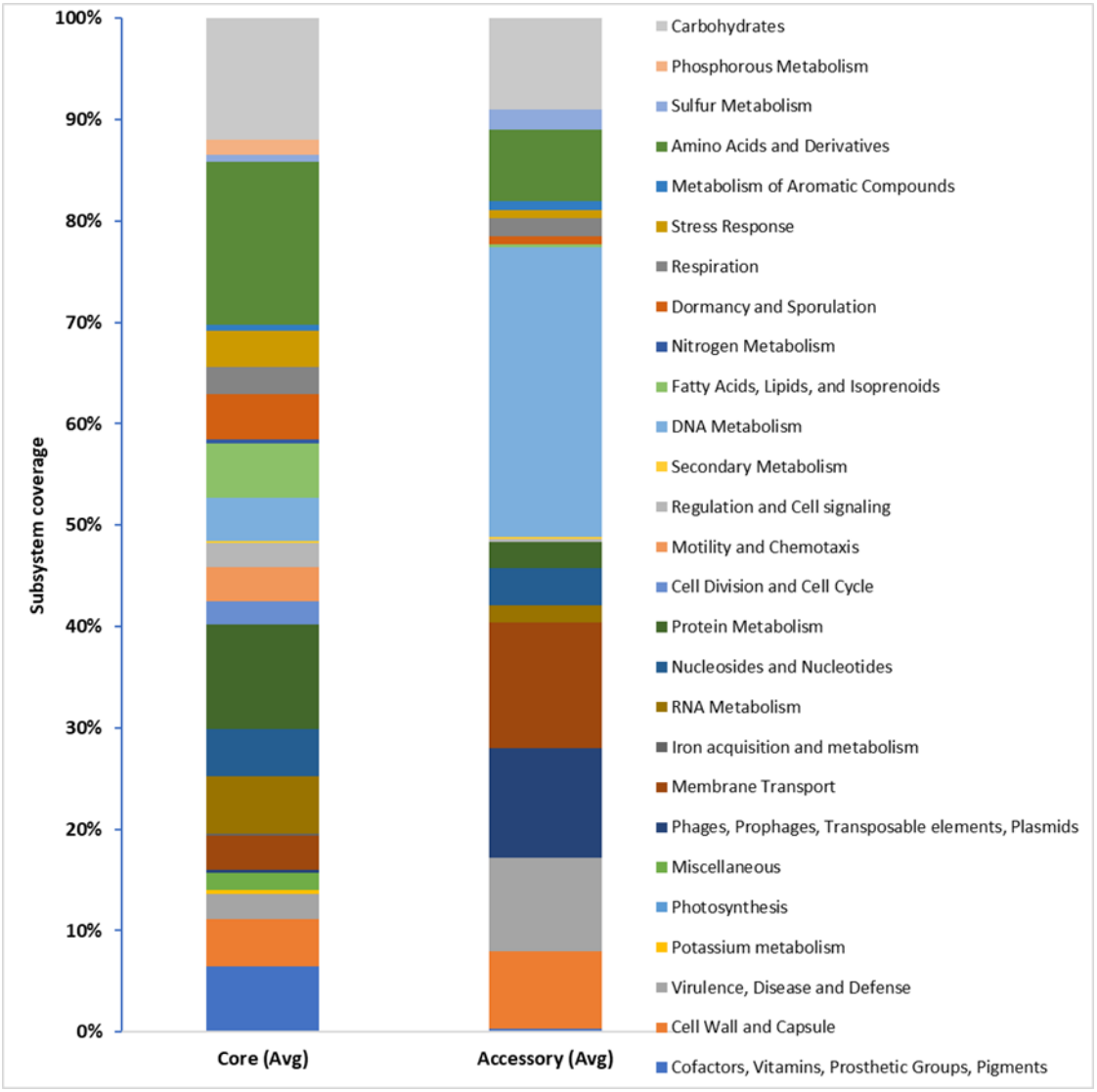


Figure 7.2: RAST subsystem annotation of core and accessory genes of *B. sporothermodurans* strains. Subsystem feature distribution of the core and accessory genomes of *B. sporothermodurans* strains used in this study. Each category is presented as a percentage of the total feature distribution in the core and accessory genomes, respectively.

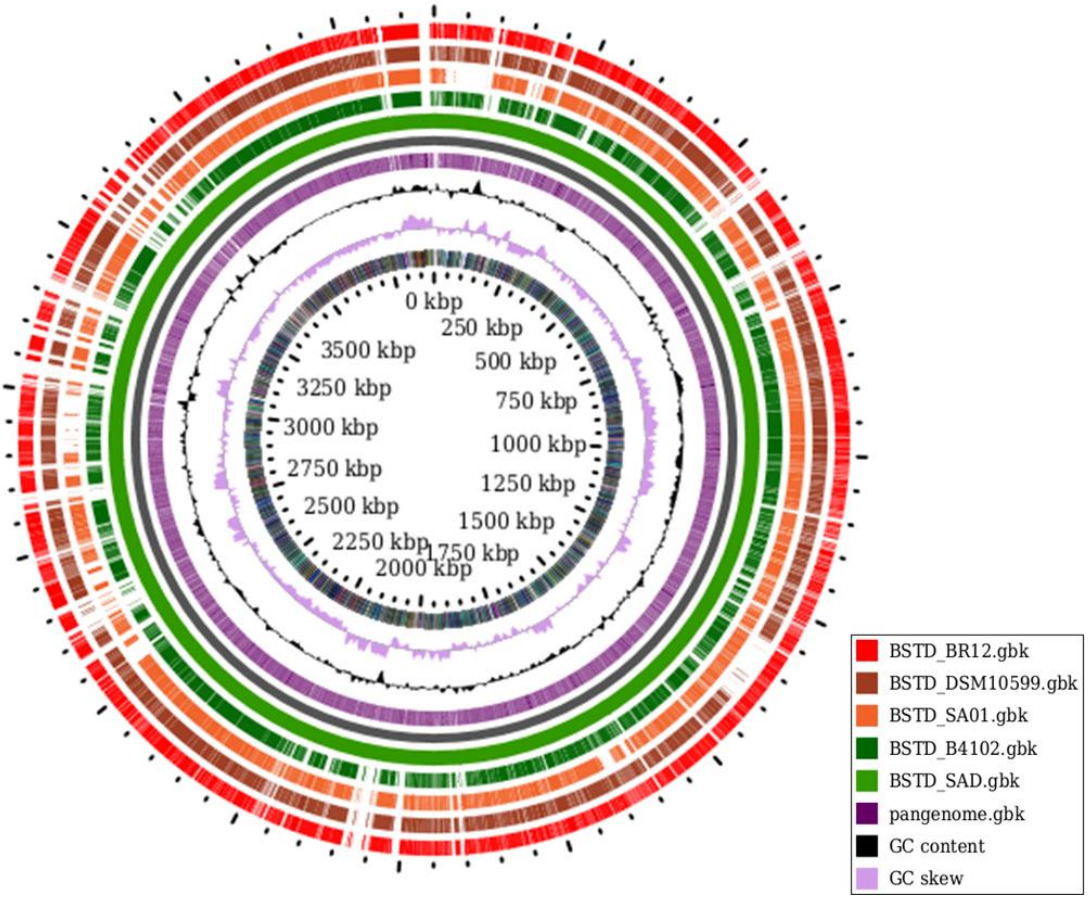


Figure 7.3: The pangenome of five *B. sporothermodurans* strains generated with GView. The three inner plots show from inside to outside a ruler depicting size of genomes in kbp, GC skew and GC content of each region or below the mean GC content of the pangenome.

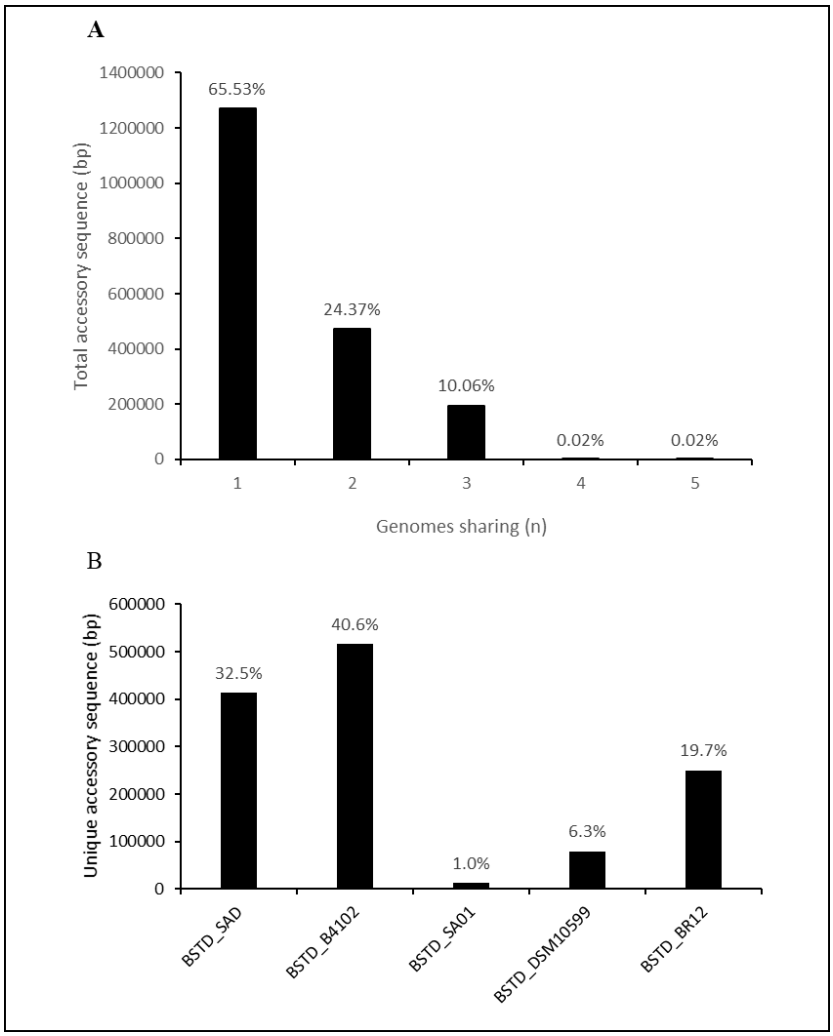


Figure 7.4: AGE subelement sequence distribution. (a) Total amount of subelement sequence (bp) shared among the *B. sporothermodurans* strains. The percentage of the total subelement sequences of the various *B. sporothermodurans* strains shared by the given number of strains is depicted on the bars. (b) Total amount of unique subelement sequences (bp) found in only one out of the five genomes. The percentage of the total unique subelement sequences among the five strains found within an indicated strain is depicted on the bars. *B. sporothermodurans* strains are abbreviated as BSTD for convenience.

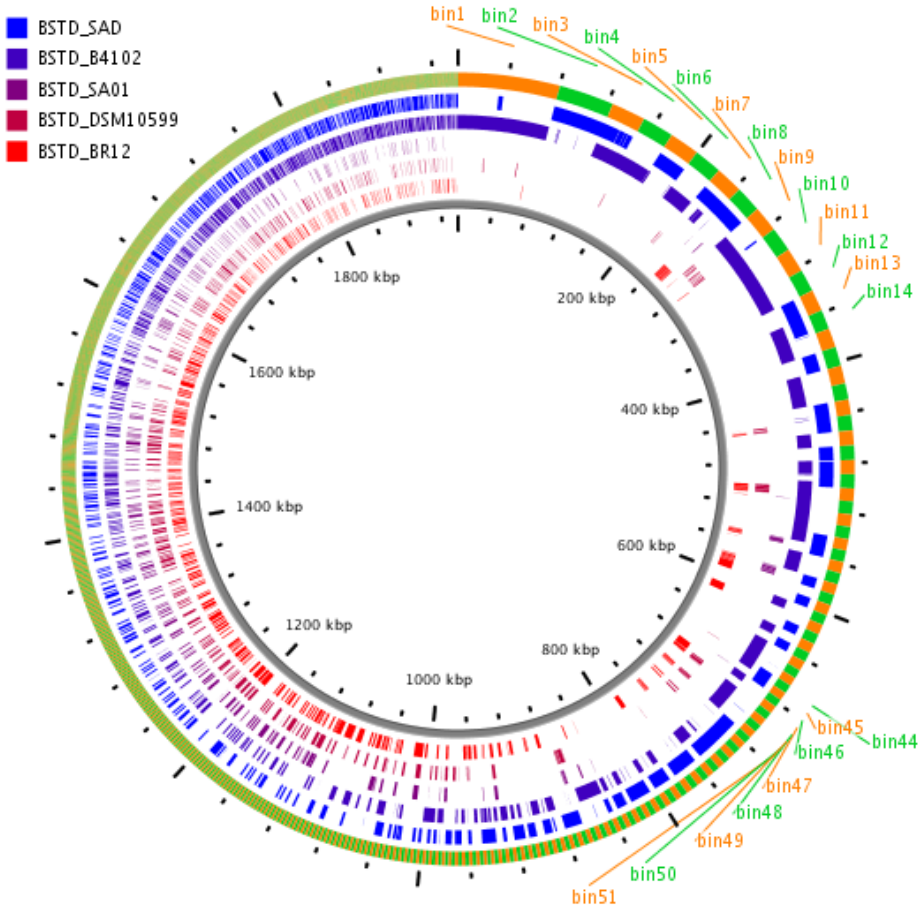


Figure 7.5: Pan accessory genome distribution among five *B. sporothermodurans* strains. The outer ring depicts ClustAGE bins ordered clockwise from largest to smallest, showing bin borders by alternating orange and green colours. The inner rings show accessory genomic elements distribution of each strain. The cumulative size (bp) of the accessory genome bin representatives are shown by the ruler in the centre of the figure. Figure was generated using the ClustAGE Plot utility.

7.4.5 Pan accessory genome distribution

The pan accessory genome distribution by ClustAGE depicts smaller AGE of the strains clustered from the 1,000 kbp to the 1,800 kbp side of the cumulative size of the accessory genome bin representatives (Figure 7.5). ClustAGE outputs AGE figures that can be used to visualize and compare relative similarity of total accessory genome content among strains in the population studied (figure 7.6).

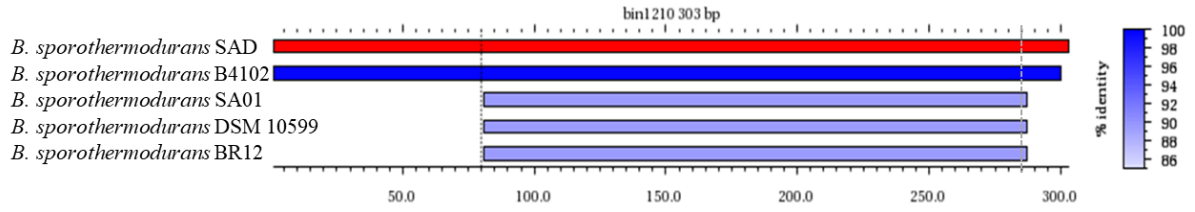


Figure 7.6: An example of AGE figure (bin1210) generated by ClustAGE. The red strip indicates the strain which was the source of the bin representative sequence. The blue strips depict the distribution of accessory element alignment and the intensity of the colour corresponds to its percent nucleotide identity to the blue gradient shown on the right. The vertical dashed line refers to subelement border.

The InteractiVenn program was utilised to identify the interactions amongst the strains with respect to their AGE profile (figure 7.7A). As calculated with InteractiVenn, the percentage distribution of the total AGEs unique to one out of the five strains was SA01, 6%; SAD, 26%; BR12, 48%; DSM10599, 29% and B4102, 24%. The binary tree (Figure 7.7A) further gives an insight to the relatedness between the five strains in connection to their AGE. The binary tree (Figure 7.7B) clusters the two highly heat resistant strains (SA01, BR12) together as well as DSM10599, the type strain.

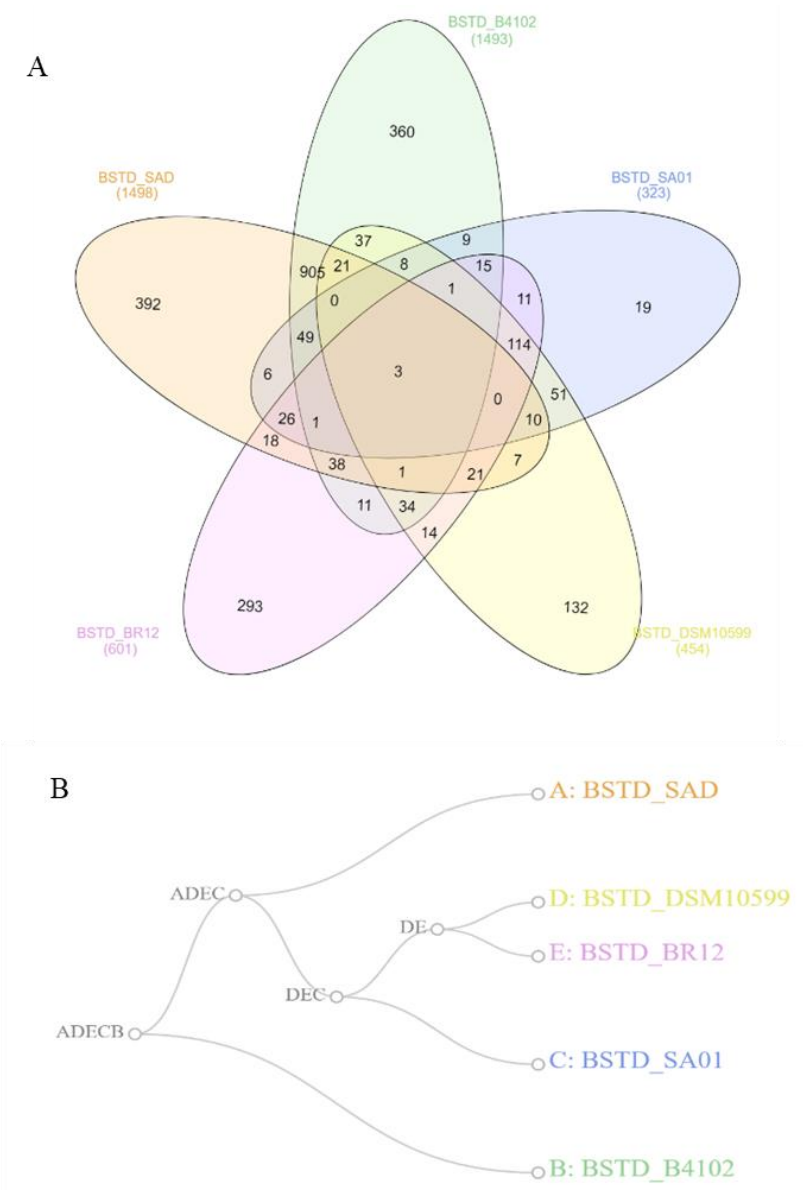


Figure 7.7: InteractiVenn output. **(A)**Venn diagram and **(B)** binary tree showing the interaction of accessory genomic elements derived from ClustAGE.

Figure 7.8: Linear genome view of coloured chromosomal maps of prophages generated by PHASTER. (a) Intact prophage region of *B. sporothermodurans* strain B4102 accessory genome (b,c) Incomplete prophage regions of *B. sporothermodurans* strain SAD accessory genome. Colour coded key shows the class of associated proteins identified in the prophage sequence.

7.5 Discussion

The core genome mostly comprises genes responsible for the basic characteristics of a species required for survival as well as its major phenotypic traits. Conversely, the accessory or dispensable genes and in some cases unique genes, contribute to evolution (Kung *et al.*, 2010). Though the accessory genes may not be essential for growth and major metabolic processes, they may confer discriminatory advantages such as antibiotic resistance, heat resistance, pathogenicity or adaptation to new niches (Raskin *et al.*, 2006; Mosquera-Rendón *et al.*, 2016). The recent surge in the use of genome sequencing amongst closely related clades have revealed widespread genomic intra-species diversity. As such the bacterial pangenome is being used by several groups to better describe the gene repertoire of bacterial strains, a test to the boundaries of what indeed is a species or strain.

The present study, as far as we are aware, is the first attempt to study the pangenome of *B. sporothermodurans*, an emerging highly heat resistant spore former, mainly associated with the dairy industry. The accessory genome of *B. sporothermodurans* comprised on average 18% (8 – 30%) of the total genome size as compared to *B. amyloliquefaciens*, 25%; *B. anthracis*,26%; *B. cereus*, 69%; *B. subtilis*, 74% and *B. thuringiensis*, 61%, (Kim *et al.*, 2017). This is in comparison to accessory genomes in other bacterial genera including *E. coli* (20%), *L. monocytogenes* (12 – 23%) and *S. aureus* (25%) (Ozer *et al.*, 2014). Compared to the microbes mentioned above, the accessory genome of *B. sporothermodurans* is relatively smaller, which, in part, can be due to the limited number of genomes used in this study. As more genomes are available and added to analyse the pangenome of *B. sporothermodurans*, the accessory genome will likely increase, with the reverse the case for the core genome. From the binary tree showing the interaction of accessory genomic elements derived from ClustAGE,

strain B4102 was the most distant, possibly representing a genetic range significantly different, maybe partly due to its isolation source (Indian curry) as compared to UHT milk, the common source of all other strains analysed in this study.

The study included strains belonging or not to the HRS clone, which could assist in categorising the accessory genes responsible for heat resistance in *B. sporothermodurans*. The repertoire of all 17 genes associated with the heat shock *dnaK* gene cluster was contained in the core genome for all the study strains, suggesting a high level of conservation. There was no difference in the number of heat shock proteins and of genes relating to the *dnaK* cluster amongst the study strains. This suggests that the highly heat resistant strains (BR12 and SA01) belonging to the HRS clone could have their higher heat resistance attributed to mobile genetic elements (MGEs) specifically transposons (Berendsen *et al.*, 2016). The DnaK chaperone with co-chaperones DnaJ and GrpE, are responsible for hyperosmotic and heat shock by preventing the aggregation of stress-denatured proteins and by disaggregating proteins (Schroder *et al.*, 1993; Gamer *et al.*, 1996). Class 1 heat shock genes (*hrcA*, *groE*, *grpE*, *dnaK* and *dnaJ*) involved in the heat resistance in a *B. subtilis* study (Schulz and Schumann, 1996) were identified specifically in the core genomes in this study. The overproduction of the HrcA protein during heat shock is known to lead to the complete halting of vegetative growth (Schulz and Schumann, 1996) and thus the onset of dormancy. Proteins associated with the dormancy and sporulation subsystem from RAST analysis were identified in the core regions of all study strains, except for an additional homolog of the SpoVs family that was identified in the accessory genome of strains SA01, SAD and B4102. The SpoVs protein family is highly conserved in spore-forming *Bacilli* and *Clostridia* (Galperin *et al.*, 2012). Specific proteins responsible to produce dipicolinic acid and for small acid-soluble spore proteins where also housed in the core genomes of the study strains and contained a similar repertoire for all five

study strains. Dipicolinic acid albeit conferring resistance of spore DNA to wet and dry heat, desiccation and hydrogen peroxide makes the spore susceptible to UV light (Magge *et al.*, 2008; Tabit and Buys, 2010; Aouadhi *et al.*, 2016). On the contrary small acid-soluble spore proteins confers resistance to UV light to the spore DNA, in addition to the aforementioned benefits of dipicolinic acid (Hayes and Setlow, 1997; Magge *et al.*, 2008).

The accessory genome's DNA metabolism subsystem coverage of 28.5% for the combined study strains is high compared to 4.3% allocated to the core genome of the study strains. In absolute number terms though, there where averagely five times more genes responsible for DNA metabolism in the core genome as opposed to the accessory genome. The high percentage of 28.5% is mainly attributed to CRISPRs and other genes that are yet to be categorised and likely of MGE origin. Of the 37 categorised genes associated to the accessory genome for DNA metabolism, 31 of them are CRISPRs. Strain SA01 did not have any CRISPRs, with DSM10599 and BR12 both having five each, SAD having six with B4102 having fifteen. The high CRISPRs content of B4102 could be as a result of exposure to frequent foreign DNA invasion events (César *et al.*, 2011). Incidentally B4102 revealed the only complete prophage (Fig. 7.8a) amongst the strains in this study. The repertoire of CRISPRs mainly belonged to the Cas1, Cas2, Cas3, Csd1, Cas5d and CasReg families. In line with the guns-for-hire paradigm, these CRISPR-Cas systems are known to capture MGEs, especially transposons and vice versa for bacterial self-defence and other evolutionary bacterial events (Koonin 2016; Peters *et al.*, 2017). Notable dormancy and stress response subsystem sub-categories namely biofilm matrix protein component and homologs and the cold shock, CspA family of proteins respectively were almost entirely evident in the core genome except in the case of the SpoVs protein family previously mentioned.

The gain and loss of genes in bacteria invariably has a strong bearing on the size of the pangenome. When there is a change in an ecosystem, some gene functions may become redundant and may eventually be lost leading to a reduction in the pangenome (Mcinerney *et al.*, 2017). The bacteria in question could become very adapted to the change in niche and eventually exhibit a closed genome as is the case in *B. anthracis* (Medini *et al.*, 2005). On the other hand, a bacterium existing in very diverse environment where it interacts with many other microbes may gain genes, which will increase its pangenome and will consequently exhibit an open genome. Bacterial species with a core/pangenome ratio ≥ 89% were found to have closed pangenomes and usually living alone in their ecological niche (allopatric) (Rouli *et al.*, 2015). A core/pangenome ratio of 55% was computed for *B. sporothermodurans*, suggesting that this microorganism possesses an open pangenome, in agreement with its ecology characterized by the life in a large community with other Firmicutes in its niche (sympatric) (Rouli *et al.*, 2015). Matrix assisted laser desorption ionisation (MALDI) analysis of various dairy samples (exclusive of UHT milk) in parallel studies by this group has found *B. sporothermodurans* existing in communities inclusive of but not limited to *B. subtilis, B. cereus, B. licheniformis, B. pumilus* and *Paenibacillus* sp. (Mugadza, 2018).

Antibiotic resistant protein systems for fosfomycin, vancomycin, fluoroquinolones and beta-lactamase were identified in both the core and accessory genomes of study strains. Notably, only strains SAD and BR12 exhibited functioning subsystems for beta-lactamase (BL). The other strains, though showing the presence of BL protein systems did not have the precursors for its activity but rather the 789 bp BL gene is sandwiched by the lipolytic enzyme YcsK and the oligopeptide transport ATP binding protein OppF.

Though *B. sporothermodurans* is not known to be pathogenic, however, its accessory genome exposed the presence of prophages. The prophages identified included intact, questionable, or incomplete phage regions. Incidentally RAST subsystem annotation of the accessory genome exhibited a considerably higher percentage of its genome dedicated to virulence, disease and defence as opposed to its core genome.

The prophages identified in two out of the five strains in the accessory genome could have been acquired through HGT of MGEs in a sympatric environment. Strain B4102 exhibited an intact prophage, 23.8 kb in length and consisting of phage proteins notably from *L. monocytogenes*, *B. cereus*, *B. anthracis*, *B. thuringiensis*, *Clostridium botulinum*, *B. subtilis*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, *S. aureus*, *Enterococcus faecalis.*

Mobile genetic elements may allow bacteria to acquire antibiotic resistance, adapt to new niches or become pathogenic (Roberts and Mullany, 2011). This could invariably affect the non-pathogenic dynamic of *B. sporothermodurans* in the future, and can be explained by the Court Jester theory, which postulates that species diversity with respect to evolution of pathogenic bacteria depends on the fluctuations in climate, niche and food supply (Benton, 2009). Thus, the association of *B. sporothermodurans* with other pathogenic Firmicutes as *B. cereus*, *L. monocytogenes*, *C. botulinum*, *Enterococcus faecium*, could possibly infer pathogenicity-related genes through HGT in its evolutionary future. Additionally, the presence of phages is known to cause fermentation failures of cheese starter cultures (Atamer *et al.*, 2013). Based on the behaviour of genes in the accessory genome and of inferences to the Court Jester theory, we would postulate a theory of our own, the Sleeper Cell theory, which suggests that genes and other mobile genetic elements acquired through HGT may remain dormant until activated through acquisition of the complete gene compliment and in conjunction with

variation in environmental condition or niche to be expressed. This is especially applicable in the case of the transition of prophages to phages. This theory could be related to the apparent dormancy of genes in a spore and prior to germination during favourable conditions.

Pangenome analysis of bacteria, such as *B. sporothermodurans*, provides profound genomic comparative abilities that pushes the boundaries of the classical definition of a species. With respect to phylogenetic comparison, the pangenome provides 100% coverage as opposed to multilocus sequence typing – MLST (~ 0.2%) and 16s rRNA (~ 0.07%). Thus, deciphering the pangenome of *B. sporothermodurans* enables reliable reconstructing of the phylogenies and evolutionary history of bacterial populations to unearth genetically distinct features that infer phenotypic responses of interest.

7.6 Conclusion

This study successfully shows for the first time the pangenome of *B. sporothermodurans* and includes crucially both HRS and non-HRS strains from different countries. Pangenomics is an important exercise to help identify differences amongst closely related *B. sporothermodurans* strains. Information from this study shows the effects of MGE in affecting gene selection and of subsequent adaptation to differing ecological niches. The repertoire of subsystem categories and of genes in the core and accessory genomes show a pattern of MGE related systems based predominantly in the accessory genome of *B. sporothermodurans* strains in this study. In-depth analysis of the various genes within the pangenome will enable the identification of strain effects with respect to heat resistance, spore formation and other phenotypic characteristics of importance to the dairy industry.

Chapter 8: GENERAL DISCUSSION

Aerobic spore-forming *Bacillus* species with high heat resistance are of critical importance to the food processing industry. Spore-forming *Bacillus* sp. offer several challenges, as their ubiquity means that they can contaminate food at several points of the food processing chain. Critically, spores are able to withstand processing temperatures that are normally able to inactivate vegetative cells (Scheldeman *et al.*, 2006). Also, some spore-formers after contaminating processing equipment, form biofilms, making cleaning regimes challenging and subsequently allowing for future contamination of food products (Jindal *et al.*, 2016, 2018). Lastly, there is consensus that there is an emergence of highly heat resistant bacterial spores, probably adapted to, or selected for by new food processing technologies or ingredients (Scheldeman *et al.*, 2006), and possibly a consequence of the lateral movement of heat resistant genes or mobile genetic elements.

The contamination of dairy products by *B. sporothermodurans* has been encountered by various companies worldwide with the highly resistant spore clone (HRS) responsible for majority of cases (Guillaume-gentil *et al.*, 2002). Other *B. sporothermodurans* strains isolated from raw milk, feed concentrate, and silage were genetically heterogenous to the HRS clone, with *B. oleronius* its closest phylogenetic neighbour (Kuhnigk *et al.*, 1995; Heyndrickx *et al.*, 2012) also isolated from raw milk and feed concentrate (Vaerewijck *et al.*, 2001; Guillaume-gentil *et al.*, 2002). The latter, having only previously been isolated from the hindgut of the termite. The overall objective of this study was to use WGS to conduct a comparative genetic study of *B. sporothermodurans* consisting of both HRS and non-HRS strains and together with its closest relative *B. oleronius* and related Firmicutes, with respect to heat resistance, spore formation and virulence genes. Consequently, this chapter will review the methodologies used

in this study and the general implications of the results especially with respect to the food industry.

8.1 Review of methodology

8.1.1 Thermal death kinetics

Microbial inactivation initially thought to be based on first order kinetics is now known to favour the Weibullian kinetic theory especially for most bacterial spore-formers. In this study the linear, biphasic and Weibull models were used to fit inactivation data for *B. sporothermodurans* spores. *B. sporothermodurans* from UHT milk were initially grown on BHI agar supplemented with vitamin B_{12}. Since *B. sporothermodurans* is challenging to grow, most studies involving this organism (Huemer *et al.*, 1998; Tabit and Buys, 2011; Aouadhi *et al.*, 2014), have used BHI, as it is nutrient rich. BHI can be supplemented with vitamin B_{12}, as it is known to regulate the production of DNA and some vital proteins and hence plays a vital role in microbial cellular metabolism and growth. In the dairy industry, modified plate count agar (PCA) is normally used to test milk after processing instead of the use of BHI supplemented with B_{12}, resulting in the possibility of false negatives for *B. sporothermodurans* as a result of poor growth on other bacterial media (Montanari *et al.*, 2004). Modified BHI agar allows the detection of very small number of CFU and makes colonies readily visible.

Spore preparation was based on a previous study (Huemer *et al.*, 1998), with some modifications. *B. sporothermodurans* plates were incubated at 37 °C for 10 days, instead of 7 days in the previous study. The reason being, the incubator used in the present study had a powerful fan which meant that the agar plates dried quicker than expected and hence the BHI

agar plates were made thicker than usual to allow enough time for spore formation. Also, the harvested spores were heated at 95 °C for 30 minutes as compared to 80 °C for 13 minutes in the previous study. It was implied that the increased temperature and time regime in the present study apart from killing vegetative cells, would ensure that any contaminating spores other than *B. sporothermodurans* spores were inactivated (Vaerewijck *et al.*, 2001). The bacterial spores where further stored in sterile distilled water at about 5 °C to allow for spore maturation (van Zuijlen *et al.*, 2010).

Thermal inactivation of *B. sporothermodurans* spores was undertaken using the method by Huemer *et al.* (1998). Other methods that has been used in similar studies include the capillary method (Stern and Proctor, 1954) and the use of miniature (lab scale) UHT processing units. In the context of the present study, the capillary method which needs special capillary tubes that needs to be sealed by flaming was not going to be appropriate and the miniature UHT processing facility was not available during this study. Spore inactivation studies was undertaken in skim milk rather than in distilled water as in some studies (van Zuijlen *et al.*, 2010), as the present study intended to mimic the processing medium as is, in industry. Besides, thermal inactivation studies in distilled water normally produces higher inactivation spore rates as compared to milk, with the latter offering a protective function to vegetative cells and spores (Aouadhi *et al.*, 2013). In a strict sense though, distilled water has been used mostly in studies where the general heat inactivation properties of spores and independent of a food medium is sought.

The linear, biphasic and basic Weibull, the most commonly used inactivation models were used in the present study. Various other models used in similar studies include variants of the Weibull model including the Weibull-Log method (Corradini *et al.*, 2008), modified Weibull

with secondary model (González *et al.,* 2009), Weibull method based on continuous and incremental models (Chen *et al.,* 2007) and the implied form of the Geeraerd model (van Zuijlen *et al.,* 2010). The present study was a proof of concept that *B. sporothermodurans* spores follow better the basic Weibull method and as well to identify if any difference in heat inactivation existed between an HRS and non-HRS strain. The other models prior mentioned are variants which fit ideally into modelling of inactivation data input into control systems and programs and as such not necessarily useful in the present study.

8.1.2 Molecular detection and whole genome sequencing

DNA extraction for the present study was undertaken using the ZR Bacterial DNA Miniprep kit (Zymo Research, USA). In the case for DNA for WGS, single colonies were inoculated into an eppendorf tube with BHI broth supplemented with vitamin B_{12} and incubated for a longer time (up to 4 days) to get better growth especially for WGS. Similar *B. sporothermodurans* studies (Scheldeman *et al.,* 2002) have used the bacterial boiling method for DNA extraction. Though this method is adequate for end point PCR reactions, it will not suffice for WGS protocols as result of the likelihood of increased impurities. Using the bacterial boiling method would require a further step of DNA cleaning and concentrating using a commercial kit. DNA extraction with the ZR Bacterial DNA miniprep kit produces good quality DNA with A_{260}/A_{280} >1.8. DNA quality of between 1.65 and 1.91 was obtained in the present study which was of good enough quality for WGS. PCR detection for *B. sporothermodurans* and the HRS strains was identified with primers from previous studies (Scheldeman *et al.,* 2002) and (Herman *et al.,* 1997) respectively. These primers derived from the 16S *rRNA* gene and subtractive hybridization of *B. sporothermodurans* strains do have some inconsistencies. Firstly, the 16S

rRNA gene although highly conserved in bacteria can give low bootstrap values and short branches and consequently affect the precise evolutionally positions of closely related bacteria. Secondly, PCR based on subtractive hybridization primers will miss out on some *B. sporothermodurans* strains as its not based on a conserved gene within a particular species (Scheldeman *et al.*, 2002). The availability of WGS data for *B. sporothermodurans* in this study should provide the avenue for better discriminatory primers in the future.

WGS was undertaken by the sequencing workhorse for the past few years, Illumina MiSeq, with this study providing the first WGS of *B. sporothermodurans* and *B. oleronius* strains used in any comparative genomic study. This study would have ideally wished to include WGS from the PacBio sequencing platform as the later offer longer reads of 1500 bases as opposed to 150 of the former (Swerdlow *et al.*, 2012). However, a myriad of factors including issues with quality and quantity of DNA from *B. sporothermodurans* strains as a result of the generally poor growth (in the context for PacBio sequencing) hampered progress with the PacBio sequencing. Typical DNA requirements for Illumina MiSeq ranges from 50-1000 ng as opposed to approximately 1 µg with PacBio (Besser *et al.*, 2018). Thus, the quality of sequencing of some of the strains could have been better with two of them having a few hundred contigs, which may have an effect on the phylogenetic studies. Though PacBio sequencing is good for *de novo* sequencing and identifying unique sequences it produces more incorrect SNP calls (Swerdlow *et al.*, 2012) and more error rates in general (Besser *et al.*, 2018). The best case scenario in this study would have been a combination of the two platforms.

8.1.3 Pangenome analysis

Several software tools for pangenomics are either available online, or as standalone Windows or Linux based platforms. In the present study, ClustAGE, together with the Spine and AGEnt suite of tools for core and accessory genome identification were used (Ozer *et al.*, 2014; Ozer, 2018). This choice was informed by a few factors, firstly the chosen program works particularly well for draft genomes and secondly has an effective Windows based version which is particularly useful for deficient Linux users. Also, the software of choice is able to accurately predict 90% of the accessory genome as opposed to on average 70% by Panseq, IslandViewer and regions of genomic plasticity (RGP) (Ozer *et al.*, 2014). ClustAGE, Spine and AGEnt software suite is one of the most recent pangenome platforms that have built on some short comings of earlier platforms. In the present study, the sequence identity for study strains was set as 85%, which allows for more AGE fragments to serve as bin representatives. This leads to a decrease in the average length of bin representatives compared to longer average length bin representative for a more restrictive cut-off of 90% (Ozer, 2018). As there were 5 *B. sporothermodurans* strains for this study, the core genome was defined as corresponding to sequences present in at least 80% of the strains analysed, equivalent to 4 out of 5 strains.

8.2 Genetic determinants of *B. sporothermodurans* phenotypic plasticity

Since the recent advances in high-throughput sequencing technologies, there has been growing efforts to understand the various genetic attributes that direct the phenotypic responses of microbial populations. There's also much interest in the genetic and subsequent phenotypic variations displayed by strains of the same species. Despite these developments, it is challenging to identify specific genetic elements that cause varying phenotypic responses as

opposed to those that are just hitchhiking. In the current study, WGS and comparative genetic analysis was used to identify relatedness of *B. sporothermodurans* strains as well as to related *Bacillus* sp. *B. oleronius* was confirmed as the closest phylogenetic neighbour to *B. sporothermodurans* in the present study. The presence of the membrane-associated zinc metalloprotease identified in both *B. oleronius* and *B. sporothermodurans* and found in the hindgut of termites and mites suggest that the insect intestine or that of Arachnids may have been the natural habitat of the common ancestor of both species. Thus, it could be postulated that the termite-fodder relationship may have provided the initial route of contamination of milk products by *B. sporothermodurans* spores. Homologs of the zinc metalloprotease are thought to provide the ability to cleave the intestinal mucin (ITM) of invertebrates, ITM being a major protein component of the insect peritrophic membrane (Ivanova *et al.*, 2003). On the other hand, since these metalloproteases are extensively found in both Gram positive and negative bacteria they could just as well be a necessary enzyme required for biological processes of the bacteria who produce them (Häse and Finkelstein, 1993).

Genome analysis in the present study show *B. sporothermodurans* and *B. oleronius* have more than twice as much of their CDS assigned to functions as compared to the *B. cereus* and *B. subtilis* groups despite the latter two being two of the most studied bacterial groups. Assigning genes to functions is one of the major bottlenecks in the NGS era. Generally, functions are either assigned based on the similarity of the CDS to known gene sequences or by identification of transcriptional *cis*-regulatory elements associated to specific pathways (Zick *et al.*, 2005). Since *B. sporothermodurans* and *B. oleronius* are among the least studied bacteria presently, there is the possibility that these two bacteria have evolved having genes common to several other Firmicutes hence sharing gene function with Firmicutes and other bacteria. During evolution, functional proteins are preserved, eliminated (Pellegrini *et al.*, 1999) or gained in a

new specie and *B. sporothermodurans* being a sympatric organism seems to share through evolution, an array of genes common to other characterized bacteria.

Most Firmicutes and for that matter bacteria in general have mechanisms that help preserve their existence when an external deleterious stimulus is encountered. The DnaK chaperone with co-chaperones DnaJ and GrpE, which are responsible for hyperosmotic and heat shock by preventing the aggregation of stress-denatured proteins and by disaggregating proteins (Schroder *et al.,* 1993, Gamer *et al.,* 1996) was present in all *B. sporothermodurans* strains sequenced in the study. The action of the DnaK chaperone is to enhance the action of the heat shock proteins GroEL and GroES, however these heat shock proteins are present in most Firmicutes. Genes coding for the heat shock proteins are highly conserved in the *B. sporothermodurans* as they were identified in their core genomes. As such the increased high heat resistance of *B. sporothermodurans* HRS strains may be attributed to the mobilome found in the accessory genome.

The mobilome, an amalgam of all MGEs is particularly prone to HGT and normally reside in the fluid part (accessory) of the genome. In the present study 50% of the accessory genome accounted for DNA metabolism, membrane transport and most importantly mobile genetic elements. Though some of these MGEs may just be hitchhikers and not contribute to any variations in phenotypic response, others could offer transient phenotypic characteristics as heat resistance. There is a notion that chance favors the prepared genome and as such a bacterium with the required accessory genetic elements needed for heat resistance for example, is likely to survive harsh heat stimuli when encountered. These accessory genetic elements may be in the form of transposons, integrases, mobile element proteins, CRISPRs to mention a few. Having an open genome, *B. sporothermodurans* is likely to gain genes, as such the pangenome

will keep on growing. However since gene gain is balanced with a continuous mutational bias towards gene deletions, it means there exists some control especially among bacteria with open genomes (Kuo and Ochman, 2009). Although there is consensus that these accessory genes have adaptive, neutral or harmful effects on survival, there are questions to the notion that these genes are mostly beneficial to bacterial survival (Mcinerney *et al.*, 2017; 2017; Vos and Eyre-walker, 2017).

In the present study the *B. sporothermodurans* HRS feature was linked to two transposases, Tn554 and Tn552 which both have antibiotic resistance effects (Rowland and Dyke, 2013; Bastos and Murphy, 2018). The Tn5 family of transposases belongs to the RNase superfamily of proteins and are known to be inertly inactive and are charged with preventing deleterious mutations as a result of transposable elements (Lovell *et al.*, 2002; Aziz *et al.*, 2010). This may suggest that the Tn5 linked with a mobile element protein sequence in *B. sporothermodurans* HRS strains may assist in the regulation of certain transposons linked with increased high heat resistance. It has previously been reported that one such transposon, Tn*1546*-like carrying the *spoVA*^{2mob} operon is responsible for the increased high heat resistance in *B. subtilis* and *B. sporothermodurans* (Berendsen *et al.*, 2016). Consequently, the activity of these transposable elements may be induced by environmental stresses like high heat application (Capy *et al.*, 2000; Miller and Capy, 2006). In general terms the induction of defense genes normally lead to short term responses whiles that of transposable elements mobility leads to genetic variability and subsequently to a longer term response to stress (Capy *et al.*, 2000).

Though it is not clear how many Firmicutes have these transposons responsible for increased heat resistance, several of them possess spore forming attributes that also helps in their survival. Spore formation amongst *Bacillus* sp. are very much conserved with respect to most of the

genes responsible (Galperin *et al.*, 2012). In the present study, the spore proteins SpoIVA, SpoVM, SpoVID, SafA, CotJC, YabG and YaaH are shared amongst all the *B. subtilis*, *B. cereus*, *B. sporothermodurans* and *B. oleronius* and as such where mostly restricted to the core genome of *B. sporothermodurans*.

Aside the problems of heat resistance attributed to genes conferring high heat resistance and spore formation, occurrence of biofilms in the milk industry is of high importance, as this leads to milk contamination along the UHT processing line with especially bacterial endospores. *B. sporothermodurans* is known to form biofilms on stainless steel and modified stainless steel fabricated surfaces designed to reduce biofilm production in the dairy industry (Jindal *et al.*, 2016). All four *B. sporothermodurans* strains in the present study possess the TasA/CalY family of proteins, found in the family Bacillaceae and responsible for biofilm production. TasA has been reported to be responsible for biofilm production in *B. subtilis* (Branda *et al.*, 2006) and *B. licheniformis*, whiles CaIY is known to code for biofilm formation in *B. cereus* (Grass *et al.*, 2004) and *B. thuringiensis*. Although *B. sporothermodurans* does possess the protein SipW, responsible for the secretion of TasA, it is devoid of TasA and CalY. It however possesses the Prot3 protein, a homolog of the TasA and probably exported out of the cell by SipW as in the case of TasA. *B. cereus*, *B. thuringiensis*, *B. subtilis* and *B. licheniformis* do not possess the Prot3 protein found in *B. sporothermodurans*, they do however possess other homologs of TasA in the form of Prot1 and Prot2.

Studies conducted on stainless steel and modified stainless steel surfaces during the production of UHT milk point to the formation of biofilms by *B. sporothermodurans* albeit not as extensive as *B. licheniformis* (Jindal *et al.*, 2016), and as compared to *B. subtilis*, *B. cereus*, and *B. thuringiensis*. The reason for the lower biofilm formation in *B. sporothermodurans* is more

likely the result of its generally poor growth (Klijn *et al.*, 1997) rather than the type of TasA/CalY family of proteins present.

8.3 Heat resistance of *B. sporothermodurans* in UHT milk

Molecular advancements including the use of WGS in combination with phenotypic assays helps the food industry better understand variations in bacterial populations and hence identify innovative ways for bacterial inactivation. In the present study, the use of heat inactivation studies in combination with molecular characterisation helps explain the behaviour of HRS and non-HRS strains of *B. sporothermodurans*. Though the number of strains is limited, the similar heat inactivation data of both the HRS and non-HRS strains suggest the influence of gene expression or the need for further differentiation of cell injury with flow cytometry (Mathys *et al.,* 2007). From the present study, it is evident that UHT temperatures presently used in industry may not be adequate to inactivate *B. sporothermodurans* spores, as to obtain a 6 $\log_{10}$ reduction at 140 °C took 6 s for the more effective Weibull method and 8 s for the linear model. UHT processing conditions presently in industry ranges from 2-5 s (Kessler, 1981) or 2-10 s (Pinto *et al.,* 2018), between 135 and 140 °C, thus may not be adequate in some instances. The use of a combination of temperature and nisin (Aouadhi *et al.,* 2014) and ultra-high pressure homogenisation (Amador Espejo *et al.,* 2014) has produced much effective inactivation especially in *B. sporothermodurans* spores. Studies by Aouadhi *et al.* (2014), achieved a 5 $\log_{10}$ reduction for *B. sporothermodurans* spores in milk using a combination of nisin (15-184 UI mL^{-1}) and temperature (73–106 °C). Using ultra high pressure homogenisation (UHPH) at 300 MPa with a maximum inlet temperature of 85 °C, Amador Espejo *et al.* (2014) also achieved a 6 $\log_{10}$ reduction of *B. sporothermodurans* spores in milk. Interestingly, *B.*

sporothermodurans was more susceptible to UHPH than *G. stearothermophilus* and *B. subtilis*. This shows the promise of reduced temperature in combination with other hurdles in bacterial spore inactivation as well as eliminating UHT milk products of the unpleasant burnt flavour perceived by consumers. It is however imperative that in the first instance, the tracking, identification and characterisation of *B. sporothermodurans* coupled with knowledge of spore formation and of their impact on biofilm formation and contamination along the milk supply chain is effective to pick up contamination when it does occur (Postollec *et al.*, 2012; Gauvry *et al.*, 2017). In the long run, the dynamics of high heat resistant strains, their control and evolutionary intricacies is of paramount importance to the food industry to avoid a case where super high heat resistant bacteria emerge. An unfortunate case which will in no doubt have terrible ramifications on food quality and safety and food security on a whole.

Chapter 9: CONCLUSIONS AND RECOMMENDATIONS

All three objectives of this study as stated in 2.18 were achieved. This study confirms *B. oleronius* as the closest phylogenetic neighbour of *B. sporothermodurans*. That using WGS, confirms *B. coagulans* and *B. acidicola* are phylogenetically closer to *B. sporothermodurans* than *B. lentus* and *B. firmus*, as was previously known with the use of 16S *rDNA* sequencing. However, comparing phylogeny based on only one conserved gene like 16S *rRNA* is not as exhaustive as using WGS. The TasA/CaIY family of proteins that has been confirmed in *B. subtilis* to be responsible for biofilm formation is present in *B. sporothermodurans* strains. Identification of Tn554 and Tn552 transposases in the mobile element protein linked to the HRS primer sequence shows its probable influence in heat resistance in *B. sporothermodurans*.

UHT milk is processed between 135 – 140 °C for 2 – 5 s, however heat inactivation at 140 °C for a 6 $\log_{10}$ reduction of *B. sporothermodurans* spores is 6 s for the Weibull, 6.1 s for biphasic and 8 s for the linear models in this study. Thus, industrial heat inactivation protocols used in UHT milk processing may not be capable of inactivating milk contaminated with *B. sporothermodurans* spores in the production of commercially sterile milk products.

Genes in the DnaK protein cluster is highly conserved in *B. sporothermodurans* as such the nucleotide and amino acid alignment for the GrpE chaperone has the same sequence in all *B. sporothermodurans* strains in this study.

B. sporothermodurans strains in this study have a core/pangenome ratio of 55%. This suggests that it possesses an open genome and makes it more prone to HGT events. *B. sporothermodurans* has an accessory genome of 18% on average, considerably smaller than other *Bacillus* species and in part as a result of the limited number of genomes used.

This study scratches the surface of the genetic characterisation of *B. sporothermodurans* strains and further studies could address the following. The use of more strains and isolated from different sources along the milk supply chain will give further insight on genetic heterogeneity and subsequently give a better picture of the *B. sporothermodurans* pangenome. Also, additional strains with different heat characteristics will help unravel the genetic basis of the phenotypic characteristics of the highly heat resistance shown by spores of some strains. Additionally, other UHT resistant species like *Paenibacillus lactis* and *B. thermoamylovorans* could be studied alongside *B. sporothermodurans* to identify similar genetic characteristics.

The use of PacBio sequencing instead of the MiSeq platform will allow the identification of especially MGEs present in the sequenced genomes as a result of the increased sequence lengths. The use of flow cytometry to sort out injured and intact cells would allow for the accurate determination of spore inactivation state after UHT treatment.

Chapter 10: REFERENCES

Adu-Oppong, B., Gasparrini, A.J., and Dantas, G., 2017. Genomic and functional techniques to mine the microbiome for novel antimicrobials and antimicrobial resistance genes. *Annals of the New York Academy of Sciences*, 1388 (1), 42–58.

Albert, I. and Mafart, P., 2005. A modified Weibull model for bacterial inactivation. *International Journal of Food Microbiology*, 100 (1–3), 197–211.

Allard, M.W., Bell, R., Ferreira, C.M., Gonzalez-Escalona, N., Hoffmann, M., Muruvanda, T., Ottesen, A., Ramachandran, P., Reed, E., Sharma, S., Stevens, E., Timme, R., Zheng, J., and Brown, E.W., 2018. Genomics of foodborne pathogens for microbial food safety. *Current Opinion in Biotechnology*, 49, 224–229.

Amador Espejo, G.G., Hernández-Herrero, M.M., Juan, B., and Trujillo, A.J., 2014. Inactivation of Bacillus spores inoculated in milk by Ultra High Pressure Homogenization. *Food Microbiology*, 44, 204–210.

Anderson, I., Sorokin, A., Kapatral, V., Reznik, G., Bhattacharya, A., Mikhailova, N., Burd, H., Joukov, V., Kaznadzey, D., Walunas, T., Markd'Souza, Larsen, N., Pusch, G., Liolios, K., Grechkin, Y., Lapidus, A., Goltsman, E., Chu, L., Fonstein, M., Ehrlich, S.D., Overbeek, R., Kyrpides, N., Ivanova, N., D'Souza, M., Larsen, N., Pusch, G., Liolios, K., Grechkin, Y., Lapidus, A., Goltsman, E., Chu, L., Fonstein, M., Ehrlich, S.D., Overbeek, R., Kyrpides, N., and Ivanova, N., 2005. Comparative genome analysis of Bacillus cereus group genomes with Bacillus subtilis. *FEMS Microbiology Letters*, 250 (2), 175–184.

Aouadhi, C., Maaroufi, A., and Mejri, S., 2014. Incidence and characterisation of aerobic spore-forming bacteria originating from dairy milk in Tunisia. *International Journal of Dairy Technology*, 67 (1), 95–102.

Aouadhi, C., Rouissi, Z., Kmiha, S., Mejri, S., and Maaroufi, A., 2016. Effect of sporulation conditions on the resistance of Bacillus sporothermodurans spores to nisin and heat. *Food Microbiology*, 54, 6–10.

Aouadhi, C., Rouissi, Z., Mejri, S., and Maaroufi, A., 2014. Inactivation of Bacillus sporothermodurans spores by nisin and temperature studied by design of experiments in water and milk. *Food Microbiology*, 38, 270–275.

Aouadhi, C., Simonin, H., Maaroufi, A., and Mejri, S., 2013. Optimization of nutrient-induced germination of Bacillus sporothermodurans spores using response surface methodology. *Food Microbiology*, 36 (2), 320–326.

Aouadhi, C., Simonin, H., Prevost, H., de Lamballerie, M., Maaroufi, A., and Mejri, S., 2013. Inactivation of Bacillus sporothermodurans LTIS27 spores by high hydrostatic pressure and moderate heat studied by response surface methodology. *LWT - Food Science and Technology*, 50 (1), 50–56.

Aouadhi, C., Simonin, H., Prévost, H., de Lamballerie, M., Maaroufi, A., Mejri, S., Pr??vost, H., de Lamballerie, M., Maaroufi, A., and Mejri, S., 2012. Optimization of pressure-induced germination of Bacillus sporothermodurans spores in water and milk. *Food Microbiology*, 30 (1), 1–7.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids*

research, 44 (W1), W16–W21.

Atamer, Z., Meike, S., Neve, H., Heller, K.J., and Hinrichs, J., 2013. Review: Elimination of bacteriophages in whey and whey products. *Frontiers in Microbiology*, 4 (JUL), 1–9.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O., 2008. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, 9 (1), 75.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., Meyer, F., Goesmann, A., McHardy, A., Bartels, D., Bekel, T., Clausen, J., Kalinowski, J., Linke, B., Rupp, O., Giegerich, R., Bryson, K., Loux, V., Bossy, R., Nicolas, P., Chaillou, S., Guchte, M. van de, Penaud, S., Maguin, E., Hoebeke, M., Bessieres, P., Vallenet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., Cruveiller, S., Lajus, A., Pascal, G., Scarpelli, C., Medigue, C., Overbeek, R., Begley, T., Butler, R., Choudhuri, J., Chuang, H., Cohoon, M., Crecy-Lagard, V. de, Diaz, N., Disz, T., Edwards, R., Tatusov, R., Natale, D., Garkavtsev, I., Tatusova, T., Shankavaram, U., Rao, B., Kiryutin, B., Galperin, M., Fedorova, N., Koonin, E., Schneider, M., Tognolli, M., Bairoch, A., Kanehisa, M., Goto, S., Haft, D., Loftus, B., Richardson, D., Yang, F., Eisen, J., Paulsen, I., White, O., Overbeek, R., Bartels, D., Vonstein, V., Meyer, F., Wu, C., Shivakumar, S., Lowe, T.,

Eddy, S., Delcher, A., Harmon, D., Kasif, S., White, O., Salzberg, S., DeJongh, M., Formsma, K., Boillot, P., Gould, J., Rycenga, M., Best, A., Becker, S., and Palsson, B., 2008. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, 9 (1), 75.

Aziz, R.K., Breitbart, M., and Edwards, R.A., 2010. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Research*, 38 (13), 4207–4217.

Bailey-smith, K., Southworth, T.W., Proctor, J., and Moir, A., 2005. The ExsA Protein of. *Society*, 187 (11), 3800–3806.

Balsingh, J., Radhakrishna, S., and Ulaganathan, K., 2016. Draft genome sequence of Bacillus pumilus ku-bf1 isolated from the gut contents of wood boring Mesomorphus sp. *Frontiers in Microbiology*, 7 (JUN), 1–3.

Baril, E., Coroller, L., Postollec, F., Leguerinel, I., Boulais, C., Carlin, F., and Mafart, P., 2011. The wet-heat resistance of Bacillus weihenstephanensis KBAB4 spores produced in a two-step sporulation process depends on sporulation temperature but not on previous cell history. *International Journal of Food Microbiology*, 146 (1), 57–62.

Bastos, M.C. and Murphy, E., 2018. Transposon Tn554 encodes three products required for transposition. *The EMBO Journal*, 7 (9), 2935–2941.

Bayjanov, J.R., Siezen, R.J., Hijum, S.A.F.T. Van, and van Hijum, S.A.F.T., 2010. PanCGHweb: A web tool for genotype calling in pangenome CGH data. *Bioinformatics*, 26 (9), 1256–1257.

Bayjanov JR, Wels M, Starrenburg M, van Hylckama Vlieg JE, Siezen RJ, M.D., 2009. PanCGH: a genotype-calling algorithm for pangenome CGH data. *Bioinformatics*, (25),

309–14.

Bayles, K.W., Brunskill, E.W., landolo, J.J., Hruska, L.L., Huang, S., Pattee, P.A., Smiley, B.K., and Yasbin, R.E., 1994. A genetic and molecular characterization of the recA gene from Staphylococcus aureus. *Gene*, 147 (1), 13–20.

Bazinet, A.L., 2017. Pan-genome and phylogeny of Bacillus cereus sensu lato. *BMC Evolutionary Biology*, 17 (1), 176.

Beeby, M., Connor, B.D.O., Ryttersgaard, C., Boutz, D.R., Perry, L.J., and Yeates, T.O., 2005. The Genomics of Disulfide Bonding and Protein Stabilization in Thermophiles. *PLoS Biology*, 3 (9), 1549–1558.

Benedict, M.N., Henriksen, J.R., Metcalf, W.W., Whitaker, R.J., and Price, N.D., 2014. ITEP: An integrated toolkit for exploration of microbial pan-genomes. *BMC Genomics*, 15 (1), 8–19.

Benton, M.J., 2009. The Red Queen and the Court Jester: Species Diversity and the Role of Biotic and Abiotic Factors Through Time. *Science*, 323 (5915), 728 LP – 732.

Berendsen, E.M., Boekhorst, J., Kuipers, O.P., and Wells-Bennik, M.H.J., 2016. A mobile genetic element profoundly increases heat resistance of bacterial spores. *The ISME Journal*, 10 (11), 2633–2642.

Berendsen, E.M., Wells-Bennik, M.H.J., Krawczyk, A.O., de Jong, A., van Heel, A., Eijlander, R.T., and Kuipers, O.P., 2016. Draft Genome Sequences of 10 Bacillus subtilis Strains That Form Spores with High or Low Heat Resistance. *Genome announcements*, 4 (2), 4–5.

Berendsen, E.M., Wells-bennik, M.H.J., Krawczyk, A.O., Jong, A. De, Heel, A. Van, Holsappel, S., Eijlander, R.T., and Kuipers, O.P., 2016. Draft Genome Sequences of Seven Thermophilic Spore-Forming Bacteria Isolated from Foods That Produce Highly Heat-Resistant Spores, Comprising Geobacillus spp., Caldibacillus debilis, and Anoxybacillus flavithermus. *Genome announcements*, 4 (3), 4–5.

Berendsen, E.M., Zwietering, M.H., Kuipers, O.P., and Wells-Bennik, M.H.J., 2015. Two distinct groups within the bacillus subtilis group display significantly different spore heat resistance properties. *Food Microbiology*, 45 (PA), 18–25.

Besser, J., Carleton, H.A., Gerner-Smidt, P., Lindsey, R.L., and Trees, E., 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical Microbiology and Infection*, 24 (4), 335–341.

Blackwood, K.S., Turenne, C.Y., Harmsen, D., and Kabani, A.M., 2004. Reassessment of Sequence-Based Targets for Identification of Bacillus Species Reassessment of Sequence-Based Targets for Identification of Bacillu s Species. *Journal of clinical microbiology*, 42 (4), 1626–1630.

Bobay, L.M. and Ochman, H., 2017. The evolution of bacterial genome architecture. *Frontiers in Genetics*.

Van Boekel, M. A.J.S., 2002. On the use of the Weibull model to describe thermal inactivation of microbial vegetative cells. *International Journal of Food Microbiology*, 74 (1–2), 139–159.

Branda, S.S., Chu, F., Kearns, D.B., Losick, R., and Kolter, R., 2006. A major protein component of the Bacillus subtilis biofilm matrix. *Molecular Microbiology*, 59 (4),

1229–1238.

Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomason, J.A., Stevens, R., Vonstein, V., Wattam, A.R., and Xia, F., 2015. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, 5 (1), 1–6.

Brittnacher, M.J., Fong, C., Hayden, H.S., Jacobs, M.A., Radey, M., and Rohmer, L., 2011. PGAT: A multistrain analysis resource for microbial genomes. *Bioinformatics*, 27 (17), 2429–2430.

Brul, S., Schuren, F., Montijn, R., Keijser, B.J.F., van der Spek, H., and Oomes, S.J.C.M., 2006. The impact of functional genomics on microbiological food quality and safety. *International journal of food microbiology*, 112 (3), 195–9.

Bukau, B. and Horwich, A.L., 1998. The Hsp70 and Hsp60 chaperone machines. *Cell*, 92 (3), 351–366.

Burgess, S.A., Flint, S.H., and Lindsay, D., 2014. Characterization of thermophilic bacilli from a milk powder processing plant. *Journal of Applied Microbiology*, 116 (2), 350–359.

Burgess, S.A., Lindsay, D., and Flint, S.H., 2010. Thermophilic bacilli and their importance in dairy processing. *International Journal of Food Microbiology*, 144 (2), 215–225.

Burrus, V. and Waldor, M.K., 2004. Shaping bacterial genomes with integrative and conjugative elements. *Research in Microbiology*, 155 (5), 376–386.

Caamano-Antelo, S., I.C. Fern?andez-No, K. B€ohme, M. Ezzat-Alnakip, M. Quintela-Baluja, J. Barros-Vel?azquez, P.C.-M., 2015. Genetic discrimination of foodborne pathogenic and spoilage Bacillus spp. based on three housekeeping genes. *Food Microbiology*, 46, 288–298.

Campanella, O.H. and Peleg, M., 2001. Theoretical comparison of a new and the traditional method to calculate Clostridium botulinum survival during thermal inactivation. *Journal of the Science of Food and Agriculture*, 81 (11), 1069–1076.

Capra, J.A. and Singh, M., 2007. Predicting functionally important residues from sequence conservation, 23 (15), 1875–1882.

Capy, P., Gasperi, G., Biémont, C., and Bazin, C., 2000. Stress and transposable elements: Co-evolution or useful parasites? *Heredity*, 85 (2), 101–106.

Cashel, M. and Potrykus, K., 2013. Stringent Response. *Brenner's Encyclopedia of Genetics: Second Edition*, 62 (6), 573–575.

Cattani, F., Barth, V.C., Nasário, J.S.R., Ferreira, C.A.S., and Oliveira, S.D., 2016. Detection and quantification of viable Bacillus cereus group species in milk by propidium monoazide quantitative real-time PCR. *Journal of Dairy Science*, 99 (4), 2617–2624.

Cattani, F., Ferreira, C.A.S., and Oliveira, S.D., 2013. The detection of viable vegetative cells of Bacillus sporothermodurans using propidium monoazide with semi-nested PCR. *Food Microbiology*, 34 (1), 196–201.

Cerf, 0, 1977. Tailing of Survival Curves of Bacterial Spores. *Journal ofApplied Bacteriology*, 42, 1–19.

César, C.E., Álvarez, L., Bricio, C., van Heerden, E., Littauer, D., and Berenguer, J., 2011. Unconventional lateral gene transfer in extreme thermophilic bacteria. *International Microbiology*, 14 (4), 187–199.

Chandrangsu, P., Wang, L., Choi, H., and Gourse, R.L., 2012. Suppression of a dnaKJ Deletion by Multicopy dksA Results from Non-Feedback-Regulated Transcripts That Originate Upstream of the Major dksA Promoter. *Journal of Bacteriology*, 194 (6), 1437–1446.

Chen, G., Campanella, O.H., and Corvalan, C.M., 2007. A numerical algorithm for calculating microbial survival curves during thermal processing. *Food Research International*, 40 (1), 203–208.

Claeys, W.L., Ludikhuyze, L.R., and Hendrickx, M.E., 2001. Formation kinetics of hydroxymethylfurfural, lactulose and furosine in milk heated under isothermal and non-isothermal conditions. *The Journal of dairy research*, 68 (2), 287–301.

De Clerck, E. and De Vos, P., 2004. Genotypic diversity among Bacillus licheniformis strains from various sources. *FEMS Microbiology Letters*, 231 (1), 91–98.

Contreras-moreira, B. and Vinuesa, P., 2013. GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis, 79 (24), 7696–7701.

Coorevits, A., De Jonghe, V., Vandroemme, J., Reekmans, R., Heyrman, J., Messens, W., De Vos, P., and Heyndrickx, M., 2008. Comparative analysis of the diversity of aerobic spore-forming bacteria in raw milk from organic and conventional dairy farms. *Systematic and applied microbiology*, 31 (2), 126–40.

Coorevits, A., Logan, N.A., Dinsdale, A.E., Halket, G., Scheldeman, P., Heyndrickx, M.,

Schumann, P., van Landschoot, A., and de Vos, P., 2011. Bacillus thermolactis sp. nov., isolated from dairy farms, and emended description of bacillus thermoamylovorans. *International Journal of Systematic and Evolutionary Microbiology*, 61 (8), 1954–1961.

Cordero, O.X. and Polz, M.F., 2014. Explaining microbial genomic diversity in light of evolutionary ecology. *Nature Reviews Microbiology*, 12 (4), 263–273.

Coroller, L., Leguerinel, I., Mettler, E., Savy, N., and Mafart, P., 2006. General model, based on two mixed weibull distributions of bacterial resistance, for describing various shapes of inactivation curves. *Applied and Environmental Microbiology*, 72 (10), 6493–6502.

Corradini, M.G., Normand, M.D., and Peleg, M., 2005. Calculating the efficacy of heat sterilization processes. *Journal of Food Engineering*, 67 (1–2), 59–69.

Corradini, M.G., Normand, M.D., and Peleg, M., 2008. Prediction of an organism's inactivation patterns from three single survival ratios determined at the end of three non-isothermal heat treatments. *International Journal of Food Microbiology*, 126 (1–2), 98–111.

Corradini, M.G. and Peleg, M., 2007. A Weibullian model for microbial injury and mortality. *International Journal of Food Microbiology*, 119 (3), 319–328.

Cutting, S.M., 2011. Bacillus probiotics. *Food Microbiology*, 28 (2), 214–220.

Cybulski, R.J., Sanz, P., Alem, F., Stibitz, S., Bull, R.L., and O'Brien, A.D., 2009. Four superoxide dismutases contribute to Bacillus anthracis virulence and provide spores with redundant protection from oxidative stress. *Infection and Immunity*, 77 (1), 274–285.

Disz, T., Akhter, S., Cuevas, D., Olson, R., Overbeek, R., Vonstein, V., Stevens, R., and

Edwards, R. a, 2010. Accessing the SEED genome databases via Web services API: tools for programmers. *BMC bioinformatics*, 11, 319.

Dogan, B. and Boor, K.J., 2003. Genetic Diversity and Spoilage Potentials among Pseudomonas spp . Isolated from Fluid Milk Products and Dairy Processing Plants Genetic Diversity and Spoilage Potentials among Pseudomonas spp . Isolated from Fluid Milk Products and Dairy Processing Plants. *Applied and Environmental Microbiology*, 69 (1), 130–138.

Driks, A., 1999. Bacillus subtilis spore coat introduction of molecular genetics to the problem of spore. *Society*, 63 (1), 1–20.

Driks, A., 2002. Maximum shields: The assembly and function of the bacterial spore coat. *Trends in Microbiology*, 10 (6), 251–254.

Eichenberger, P., Jensen, S.T., Conlon, E.M., Van Ooij, C., Silvaggi, J., González-Pastor, J.E., Fujita, M., Ben-Yehuda, S., Stragier, P., Liu, J.S., and Losick, R., 2003. The σE regulon and the identification of additional sporulation genes in Bacillus subtilis. *Journal of Molecular Biology*, 327 (5), 945–972.

Elliott, A.J., Datta, N., Amenu, B., and Deeth, H.C., 2005. Heat-induced and other chemical changes in commercial UHT milks. *Journal of Dairy Research*, 72 (4), 442–446.

Ernst, C. and Rahmann, S., 2013. PanCake : A Data Structure for Pangenomes. *Genome Informatics*, 2013, 35–45.

Esteban, M.D., Huertas, J.P., Fernández, P.S., and Palop, A., 2013. Effect of the medium characteristics and the heating and cooling rates on the nonisothermal heat resistance of Bacillus sporothermodurans IC4 spores. *Food Microbiology*, 34 (1), 158–163.

Faith, N.G., Kim, J.-W., Azizoglu, R., Kathariou, S., and Czuprynski, C., 2012. Purine Biosynthesis Mutants (purA and purB) of Serotype 4b Listeria monocytogenes Are Severely Attenuated for Systemic Infection in Intragastrically Inoculated A/J Mice. *Foodborne Pathogens and Disease*, 9 (5), 480–486.

FDA, 2016. Whole Genome Sequencing Cracking the code for foodborne illness. *FDA Consumer Hea l th In fo rma t i on / U. S. Food and Dr ug Admi n i s t r a t i on*, (June), 1–3.

Feng, Y., Zhang, Y., Ying, C., Wang, D., and Du, C., 2015. Nanopore-based fourth-generation DNA sequencing technology. *Genomics, Proteomics and Bioinformatics*, 13 (1), 4–16.

Flint, S., Gonzaga, Z.J., Good, J., and Palmer, J., 2016. Bacillus thermoamylovorans – a new threat to the dairy industry – A review. *International Dairy Journal*, 65, 38–43.

Fox, G.E., Wisotzkey, J.D., and Jurtshuk, P., 1992. How Close Is Close : 1 s rRNA Sequence Identity May Not Be Sufficient To Guarantee Species Identity. *International Journal of Systematic Bacteriology*, 42, 166–170.

Galperin, M.Y., Mekhedov, S.L., Puigbo, P., Smirnov, S., Wolf, Y.I., and Rigden, D.J., 2012. Genomic determinants of sporulation in Bacilli and Clostridia: Towards the minimal set of sporulation-specific genes. *Environmental Microbiology*, 14 (11), 2870–2890.

Gamer, J., Multhaup, G., Tomoyasu, T., Mccarty, J.S., Rudiger, S., Schonfeld ', H.-J., Schirra, C., Bujard, H., and Bukau, B., 1996. A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the Escherichia coli heat shock transcription factor &2. *The EMBO Journal*, 15 (3), 607–617.

Garcia, D., der Voort, M. Van, and Abee, T., 2010. Comparative analysis of Bacillus weihenstephanensis KBAB4 spores obtained at different temperatures. *International Journal of Food Microbiology*, 140 (2–3), 146–153.

Gauvry, E., Mathot, A.G., Leguérinel, I., Couvert, O., Postollec, F., Broussolle, V., and Coroller, L., 2017. Knowledge of the physiology of spore-forming bacteria can explain the origin of spores in the food environment. *Research in Microbiology*, 168 (4), 369–378.

Geeraerd, a. H., Valdramidis, V.P., and Van Impe, J.F., 2005. GInaFiT, a freeware tool to assess non-log-linear microbial survivor curves. *International Journal of Food Microbiology*, 102 (1), 95–105.

Georgiades, K. and Raoult, D., 2011. Defining pathogenic bacterial species in the genomic era. *Frontiers in Microbiology*, 1 (JAN), 1–13.

Gillings, M.R., 2017. Lateral gene transfer, bacterial genome evolution, and the Anthropocene. *Annals of the New York Academy of Sciences*, 1389 (1), 20–36.

Gillings, M.R., Paulsen, I.T., and Tetu, S.G., 2017. Genomics and the evolution of antibiotic resistance. *Annals of the New York Academy of Sciences*, 1388 (1), 92–107.

González, M., Skandamis, P.N., and Hänninen, M.L., 2009. A modified Weibull model for describing the survival of Campylobacter jejuni in minced chicken meat. *International Journal of Food Microbiology*, 136 (1), 52–58.

Grass, G., Schierhorn, A., Sorkau, E., Rücknagel, P., Nies, D.H., Fricke, B., Mu, H., and Ru, P., 2004. Camelysin Is a Novel Surface Metalloproteinase from Bacillus cereus Camelysin Is a Novel Surface Metalloproteinase from Bacillus cereus. *Infection and*

Immunity, 72 (1), 219–228.

Grijspeerdt, K., Mortier, L., De Block, J., and Van Renterghem, R., 2004. Applications of modelling to optimise ultra high temperature milk heat exchangers with respect to fouling. *Food Control*, 15 (2), 117–130.

Gruber, T.M. and Gross, C.A., 2003. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annual Review of Microbiology*, 57 (1), 441–466.

Guillaume-gentil, O., Scheldeman, P., Marugg, J., Herman, L., Joosten, H., and Heyndrickx, M., 2002. Genetic heterogeneity in Bacillus sporothermodurans as demonstrated by ribotyping and repetitive extragenic palindromic-PCR fingerprinting. *Applied and Environmental Microbiology*, 68 (9), 4216–4224.

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.*, 41 (41), 95–8.

Hanage, W., 2014. Microbiology: microbiome science needs a healthy dose of scepticism. *Nature*, 512, 247–248.

Häse, C.C. and Finkelstein, R.A., 1993. Bacterial extracellular zinc-containing metalloproteases. *Microbiological reviews*, 57 (4), 823–37.

Hayes, C.S. and Setlow, P., 1997. Analysis of deamidation of small , acid-soluble spore proteins from Bacillus subtilis in vitro and in vivo . Analysis of Deamidation of Small , Acid-Soluble Spore Proteins from Bacillus subtilis In Vitro and In Vivo Downloaded from http://jb.asm.org/ on A. *J. Bacteriol*, 179 (19), 6020–6027.

Heather, J.M. and Chain, B., 2016. The sequence of sequencers: The history of sequencing

DNA. *Genomics*, 107 (1), 1–8.

Heberle, H., Meirelles, G.V., Silva, F.R., Telles, G.P., and Minghim, R., 2015. InteractiVenn : a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, 16 (169), 1–7.

Hecker, M., Schumann, W., and Völker, U., 1996. Heat-shock and general stress response in *Bacillus subtilis*. *Molecular microbiology*, 19 (3), 417–28.

Herman, L. and Heyndrickx, M., 2000. The presence of intragenically located REP-like elements in Bacillus sporothermodurans is sufficient for REP-PCR typing. *Res. Microbiol.*, 151, 255–261.

Herman, L., Heyndrickx, M., and Waes, G., 1998. Typing of Bacillus sporothermodurans and other Bacillus species isolated from milk by repetitive element sequence based PCR. *Letters in Applied Microbiology*, 26, 183–188.

Herman, L.M.F., Vaerewijck, M.J.M., Moermans, R.J.B., Waes, G.M.A.V.J., and Waes, G.M.A.V.J., 1997. Identification and detection of Bacillus sporothermodurans spores in 1, 10, and 100 milliliters of raw milk by PCR. *Applied and Environmental Microbiology*, 63 (8), 3139–3143.

Hettick, J.M., Kashon, M.L., Slaven, J.E., Ma, Y., Simpson, J.P., Siegel, P.D., Mazurek, G.N., and Weissman, D.N., 2006. Discrimination of intact mycobacteria at the strain level: a combined MALDI-TOF MS and biostatistical analysis. *Proteomics*, 6 (24), 6416–25.

Heyndrickx, M., 2011a. The Importance of Endospore-Forming Bacteria Originating. *Applied and Environmental Soil Science*, 2011, 1–11.

Heyndrickx, M., 2011b. Dispersal of Aerobic Endospore-forming Bacteria from Soil and Agricultural Activities to Food and Feed.

Heyndrickx, M., Coorevits, A., Scheldeman, P., Lebbe, L., Schumann, P., Rodŕiguez-Diaz, M., Forsyth, G., Dinsdale, A., Heyrman, J., Logan, N.A., and de Vos, P., 2012. Emended descriptions of Bacillus sporothermodurans and Bacillus oleronius with the inclusion of dairy farm isolates of both species. *International Journal of Systematic and Evolutionary Microbiology*, 62 (2), 307–314.

Hitomi, M., Nishimura, H., Tsujimoto, Y., Matsui, H., and Watanabe, K., 2003. Identification of a helix-turn-helix motif of Bacillus thermoglucosidasius HrcA essential for binding to the CIRCE element and thermostability of the HrcA-CIRCE complex, indicating a role as a thermosensor. *Journal of Bacteriology*, 185 (1), 381–385.

Hong, H.A., Le, H.D., and Cutting, S.M., 2005. The use of bacterial spore formers as probiotics. *FEMS Microbiology Reviews*, 29 (4), 813–835.

Hsu, C.-K., Hsu, M.M.-L., and Lee, J.Y.-Y., 2017. Demodicosis: A clinicopathological study. *Journal of the American Academy of Dermatology*, 60 (3), 453–462.

Huemer, I.A., Klijn, N., Vogelsang, H.W.J., and Langeveld, L.P.M., 1998. Thermal death kinetics of spores of Bacillus sporothermodurans isolated from UHT milk. *International Dairy Journal*, 8 (10–11), 851–855.

Ibekwe, A.M., Papiernik, S.K., Grieve, C.M., and Yang, C.H., 2011. Quantification of persistence of escherichia coli O157:H7 in contrasting soils. *International Journal of Microbiology*, 2011.

Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V.,

Bhattacharyya, A., Reznik, G., Mikhailova, N., Lapidus, A., Chu, L., Mazur, M., Goltsman, E., Larsen, N., Souza, M.D., Walunas, T., Grechkin, Y., Pusch, G., Haselkorn, R., Fonstein, M., Ehrlich, S.D., Overbeek, R., and Kyrpides, N., 2003. Genome sequence of Bacillus cereus and comparative analysis with Bacillus anthracis, 423 (May), 87–91.

Janda, J.M. and Abbott, S.L., 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of clinical microbiology*, 45 (9), 2761–2764.

Jarmuda, S., O'Reilly, N., Zaba, R., Jakubowicz, O., Szkaradkiewicz, A., and Kavanagh, K., 2012. Potential role of Demodex mites and bacteria in the induction of rosacea. *Journal of Medical Microbiology*, 61 (PART 11), 1504–1510.

Jindal, S., Anand, S., Huang, K., Goddard, J., Metzger, L., and Amamcharla, J., 2016. Evaluation of modified stainless steel surfaces targeted to reduce biofilm formation by common milk sporeformers. *Journal of Dairy Science*, 99 (12), 9502–9513.

Jindal, S., Anand, S., Metzger, L., and Amamcharla, J., 2018. Short communication: A comparison of biofilm development on stainless steel and modified-surface plate heat exchangers during a 17-h milk pasteurization run. *Journal of Dairy Science*, 101 (4), 2921–2926.

De Jonghe, V., Coorevits, A., De Block, J., Van Coillie, E., Grijspeerdt, K., Herman, L., De Vos, P., and Heyndrickx, M., 2010. Toxinogenic and spoilage potential of aerobic spore-formers isolated from raw milk. *International journal of food microbiology*, 136 (3), 318–25.

De Jonghe, V., Coorevits, A., Vandroemme, J., Heyrman, J., Herman, L., De Vos, P., and Heyndrickx, M., 2008. Intraspecific genotypic diversity of Bacillus species from raw milk. *International Dairy Journal*, 18 (5), 496–505.

Juhas, M., Van Der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., and Crook, D.W., 2009. Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews*, 33 (2), 376–393.

Kaas, R.S., Leekitcharoenphon, P., Aarestrup, F.M., and Lund, O., 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS ONE*, 9 (8), 1–8.

Keim, P., Kalif, a, Schupp, J., Hill, K., Travis, S.E., Richmond, K., Adair, D.M., Hugh-Jones, M., Kuske, C.R., and Jackson, P., 1997. Molecular evolution and diversity in Bacillus anthracis as detected by amplified fragment length polymorphism markers. *Journal of bacteriology*, 179 (3), 818–824.

Kelleher, P., Murphy, J., Mahony, J., and van Sinderen, D., 2015. Next-generation sequencing as an approach to dairy starter selection. *Dairy Science and Technology*, 95 (5), 545–568.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling , prediction and analysis. *Nature Protocols*, 10 (6), 845–858.

Ki, J.-S.S., Zhang, W., and Qian, P.-Y.Y., 2009. Discovery of marine Bacillus species by 16S rRNA and rpoB comparisons and their usefulness for species identification. *Journal of Microbiological Methods*, 77 (1), 48–57.

Kim, Y., Koh, I., Young Lim, M., Chung, W.H., and Rho, M., 2017. Pan-genome analysis of Bacillus for microbiome profiling. *Scientific Reports*, 7 (1), 1–9.

Klijn, N., Herman, L., Langeveld, L., Vaerewijck, M., Wagendorp, A.A., Huemer, I., and Weerkamp, A.H., 1997. Genotypical and phenotypical characterization of Bacillus sporothermodurans strains, surviving UHT sterilisation. *International Dairy Journal*, 7 (6–7), 421–428.

Klijn, N., Hermanb, L., Langeveld, L., Vaerewijck, M., Weerkamp, A.H., and Wagendorp, A.A., 1997. Genotypical and Phenotypical Characterization of Bacillus sporothermodurans Strains , Surviving UHT Sterilisation. *Int. Dairy Journal*, 6946 (97), 421–428.

Kmiha, S., Aouadhi, C., Klibi, A., Jouini, A., Béjaoui, A., Mejri, S., and Maaroufi, A., 2017. Seasonal and regional occurrence of heat-resistant spore-forming bacteria in the course of ultra-high temperature milk production in Tunisia. *Journal of Dairy Science*.

Koonin, E. V, 2016. Viruses and mobile elements as drivers of evolutionary transitions. *Phil. Trans. R. Soc*, (371), 1–13.

Krawczyk, A.O., Berendsen, E.M., Eijlander, R.T., de Jong, A., Wells-Bennik, M.H.J., and Kuipers, O.P., 2015. Draft Genome Sequences of Four Bacillus thermoamylovorans Strains Isolated from Milk and Acacia Gum, a Food Ingredient. *Genome announcements*, 3 (2), 2–3.

Krawczyk, A.O., de Jong, A., Eijlander, R.T., Berendsen, E.M., Holsappel, S., Wells-Bennik, M.H.J., and Kuipers, O.P., 2015. Next-Generation Whole-Genome Sequencing of Eight Strains of Bacillus cereus, Isolated from Food. *Genome announcements*, 3 (6), 3–5.

Krawczyk, A.O., de Jong, A., Holsappel, S., Eijlander, R.T., van Heel, A., Berendsen, E.M., Wells-Bennik, M.H.J., and Kuipers, O.P., 2016. Genome Sequences of 12 Spore-Forming Bacillus Species, Comprising Bacillus coagulans, Bacillus licheniformis, Bacillus amyloliquefaciens, Bacillus sporothermodurans, and Bacillus vallismortis, Isolated from Foods. *Genome announcements*, 4 (3), e00103-16.

Kuhnigk, T., Borst, E.-M., Breunig, A., Ko$nig, H., Collins, M.D., Hutson, R.A., and Ka$mpfer, P., 1995. Bacillus oleronius sp. nov., a member of the hindgut flora of the termite Reticulitermes santonensis (Feytaud). *Can J Microbiol*, (41), 699–706.

Kuhnigk, T., Borst, E.-M., Breunig, A., König, H., Collins, M.D., Hutson, R.A., and Kämpfer, P., 1995a. Bacillus oleronius sp.nov., a member of the hindgut flora of the termite Reticulitermes santonensis (Feytaud). *Canadian Journal of Microbiology*, 41 (8), 699–706.

Kuhnigk, T., Borst, E.M., Breunig, A., König, H., Collins, M.D., Hutson, R.A., and Kämpfer, P., 1995b. Bacillus oleronius sp.nov., a member of the hindgut flora of the termite Reticulitermes santonensis (Feytaud). *Canadian journal of microbiology*, 41 (8), 699–706.

Kung, V.L., Ozer, E.A., and Hauser, A.R., 2010. The Accessory Genome of Pseudomonas aeruginosa. *Microbiology and Molecular Biology Reviews*, 74 (4), 621–641.

Kuo, C.H. and Ochman, H., 2009. The fate of new bacterial genes. *FEMS Microbiology Reviews*, 33 (1), 38–43.

LaBreck, C.J., May, S., Viola, M.G., Conti, J., and Camberg, J.L., 2017. The Protein Chaperone ClpX Targets Native and Non-native Aggregated Substrates for Remodeling,

Disassembly, and Degradation with ClpP. *Frontiers in Molecular Biosciences*, 4 (May), 1–13.

Lacey, N., Delaney, S., Kavanagh, K., and Powell, F.C.C., 2007. Mite-related bacterial antigens stimulate inflammatory cells in rosacea. *British Journal of Dermatology*, 157 (3), 474–481.

Laconi, S., Cailhol, L., Pourcel, L., Thalamas, C., Lapeyre-Mestre, M., and Chabrol, H., 2015. Relation entre mécanismes de défense et alliance thérapeutique. *Encephale*, 41 (5), 429–434.

Ladenstein, R. and Ren, B., 2006. Protein disulfides and protein disulfide oxidoreductases in hyperthermophiles. *FEBS Journal*, 273 (18), 4170–4185.

Laing, C., Buchanan, C., Taboada, E.N., Zhang, Y., Kropinski, A., Villegas, A., Thomas, J.E., and Gannon, V.P.J., 2010. Pan-genome sequence analysis using Panseq : an online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics*, 11 (461), 1–14.

Land, M., Hauser, L., Jun, S.R., Nookaew, I., Leuze, M.R., Ahn, T.H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., and Ussery, D.W., 2015. Insights from 20 years of bacterial genome sequencing. *Functional and Integrative Genomics*, 15 (2), 141–161.

Ledala, N., Sengupta, M., Muthaiyan, A., Wilkinson, B.J., and Jayaswal, R.K., 2010. Transcriptomic Response of Listeria monocytogenes to Iron Limitation and fur Mutation □ †, 76 (2), 406–416.

Lembke, F., 1995. Highly heat resistant spores in UHT milk. *Bulletin of the International Dairy Federation*, 302, 60–61.

Liberek, K. and Georgopoulos, C., 1993. Autoregulation of the *Escherichia coli* heat shock response by the DnaK and DnaJ heat shock proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 90 (23), 11019–11023.

Liu, J., Huang, C., Shin, D.-H., Yokota, H., Jancarik, J., Kim, J.-S., Adams, P.D., Kim, R., and Kim, S.-H., 2005. Crystal Structure of a Heat-inducible Transcriptional Repressor HrcA from Thermotoga maritima: Structural Insight into DNA Binding and Dimerization. *Journal of Molecular Biology*, 350 (5), 987–996.

Lovell, S., Goryshin, I.Y., Reznikoff, W.R., and Rayment, I., 2002. Two-metal active site binding of a Tn5 transposase synaptic complex. *Nature Structural Biology*, 9 (4), 278–281.

Lu, H., Giordano, F., and Ning, Z., 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, 14 (5), 265–279.

Lücking, G., Stoeckel, M., Atamer, Z., Hinrichs, J., and Ehling-Schulz, M., 2013. Characterization of aerobic spore-forming bacteria associated with industrial dairy processing environments and product spoilage. *International Journal of Food Microbiology*, 166 (2), 270–279.

Magge, A., Granger, A.C., Wahome, P.G., Setlow, B., Vepachedu, V.R., Loshon, C.A., Peng, L., Chen, D., Li, Y.Q., and Setlow, P., 2008. Role of dipicolinic acid in the germination, stability, and viability of spores of Bacillus subtilis. *Journal of Bacteriology*, 190 (14), 4798–4807.

Magno- erez-Bryan, M.C., Mar ınez-Gar ıa, P.M., Hierrezuelo, J., Rod ıguez-Palenzuela, P., Arrebola, E., Ramos, C., de Vicente, A., erez-Gar ıa, A., and Romero, D., 2015.

Comparative Genomics Within the Bacillus Genus Reveal the Singularities of Two Robust Bacillus amyloliquefaciens Biocontrol Strains. *Mpmi*, 28 (10), 1102–1116.

Mahadevan, R., Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V, Eiglmeier, K., Gas, S., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., Mclean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., and Barrell, B.G., 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393 (November), 537–544.

Mathys, a., Chapman, B., Bull, M., Heinz, V., and Knorr, D., 2007. Flow cytometric assessment of Bacillus spore response to high pressure and heat. *Innovative Food Science and Emerging Technologies*, 8 (4), 519–527.

Maughan, H. and Van der Auwera, G., 2011. Bacillus taxonomy in the genomic era finds phenotypes to be essential though often misleading. *Infection, Genetics and Evolution*, 11 (5), 789–797.

McGillivray, S.M., Ebrahimi, C.M., Fisher, N., Sabet, M., Zhang, D.X., Chen, Y., Haste, N.M., Aroian, R. V., Gallo, R.L., Guiney, D.G., Friedlander, A.M., Koehler, T.M., and Nizet, V., 2009. ClpX contributes to innate defense peptide resistance and virulence phenotypes of Bacillus anthracis. *Journal of Innate Immunity*, 1 (5), 494–506.

Mcinerney, J.O., Mcnally, A., and Connell, M.J.O., 2017a. Why prokaryotes have pangenomes. *Nature Publishing Group*, 2 (March), 1–5.

Mcinerney, J.O., Mcnally, A., and Connell, M.J.O., 2017b. Reply to ' The population genetics of pangenomes '. *Nature Microbiology*, 2 (December), 41564.

McNally, A., Oren, Y., Kelly, D., Pascoe, B., Dunn, S., Sreecharan, T., Vehkala, M., Välimäki, N., Prentice, M.B., Ashour, A., Avram, O., Pupko, T., Dobrindt, U., Literak, I., Guenther, S., Schaufler, K., Wieler, L.H., Zhiyong, Z., Sheppard, S.K., McInerney, J.O., and Corander, J., 2016. Combined Analysis of Variation in Core, Accessory and Regulatory Genome Regions Provides a Super-Resolution View into the Evolution of Bacterial Populations. *PLoS Genetics*, 12 (9), 1–16.

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R., 2005. The microbial pan-genome. *Current Opinion in Genetics and Development*, 15 (6), 589–594.

Meintanis, C., Chalkou, K.I., Kormas, K.A., Lymperopoulou, D.S., Katsifas, E. a, Hatzinikolaou, D.G., and Karagouni, a D., 2008. Application of rpoB sequence similarity analysis, REP-PCR and BOX-PCR for the differentiation of species within the genus Geobacillus. *Letters in applied microbiology*, 46 (3), 395–401.

Metzker, M.L., 2010. Sequencing technologies the next generation. *Nature Reviews Genetics*, 11 (1), 31–46.

Miller, W.J. and Capy, P., 2006. Applying mobile genetic elements for genome analysis and evolution. *Molecular biotechnology*, 33 (2), 161–174.

Mohapatra, B.R. and Mazumder, A., 2008. Comparative efficacy of five different rep-PCR methods to discriminate Escherichia coli populations in aquatic environments. *Water Science and Technology*, 58 (3), 537–547.

Montanari, G., Borsari, a, Chiavari, C., Ferri, G., Zambonelli, C., and Grazia, L., 2004.

Morphological and phenotypical characterization of Bacillus sporothermodurans. *Journal of applied microbiology*, 97 (4), 802–9.

Montanhini, M.T.M., Colombo, M., Nero, L. a, and Bersot, L.S., 2013. Short communication: Presence of neutral metallopeptidase (npr) gene and proteolytic activity of Bacillus cereus isolated from dairy products. *Journal of dairy science*, 96 (9), 5641–3.

Mosquera-Rendón, J., Rada-Bravo, A.M., Cárdenas-Brito, S., Corredor, M., Restrepo-Pineda, E., and Benítez-Páez, A., 2016. Pangenome-wide and molecular evolution analyses of the Pseudomonas aeruginosa species. *BMC Genomics*, 17 (1), 1–14.

Mugadza, D.T., 2018. Bacillus and Paenibacillus species associated with extended shelf life milk during processing and storage, 71 (2).

Mugadza, D.T., Owusu-Darko, R., and Buys, E.M., 2018. Short communication: Source tracking Bacillus cereus in an extended shelf-life milk processing plant using partial sequencing of rpoB and multilocus sequence typing. *Journal of Dairy Science*, 1–5.

Nadon, C., 2016. Whole Genome Sequencing : How significant is it for food safety. *Technical Meeting on the impact of Whole Genome Sequencing on food safety management within a One Health framework*, (May).

Nallapareddy, S.R., Duh, R., Singh, K. V, and Murray, B.E., 2002. Molecular Typing of Selected Enterococcus faecalis Isolates : Pilot Study Using Multilocus Sequence Typing and Pulsed-Field Gel Electrophoresis, 40 (3), 868–876.

Ndejjo, R., Mukama, T., Musinguzi, G., Halage, A.A., Ssempebwa, J.C., and Musoke, D., 2017. Women's intention to screen and willingness to vaccinate their daughters against cervical cancer - A cross sectional study in eastern Uganda. *BMC Public Health*, 17 (1),

1–12.

Neter, J., Kutner, M., Nachtsheim, C., and Wasserman, W., 1996. *Applied Linear Statistical Models*. 4th Editio. New York: WCB McGraw-Hill.

Nicholson, W.L., Munakata, N., Horneck, G., Melosh, H.J., and Setlow, P., 2000. Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments. *Microbiology and molecular biology reviews : MMBR*, 64 (3), 548–72.

Olivier, S. a., Bull, M.K., and Chapman, B., 2012. Bacillus spp. spores produced at lower temperatures are more resistant to high pressure thermal processes but mineralization does not predict relative resistance. *Innovative Food Science and Emerging Technologies*, 16, 96–101.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, 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 (D1), D206–D214.

Owusu-Darko, R., Allam, M., Mtshali, S., Ismail, A., and Buys, E.M., 2017. Draft genome sequence of Bacillus oleronius DSM 9356 isolated from the termite Reticulitermes santonensis. *Genomics Data*, 12, 76–78.

Ozer, E.A., 2018. ClustAGE : a tool for clustering and distribution analysis of bacterial accessory genomic elements. *BMC Bioinformatics*, 19 (150), 1–12.

Ozer, E.A., Allen, J.P., and Hauser, A.R., 2014. Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt. *BMC Genomics*, 15 (737), 1–17.

Patel, S., 2016. Drivers of bacterial genomes plasticity and roles they play in pathogen virulence, persistence and drug resistance. *Infection, Genetics and Evolution*, 45, 151–164.

Pecora, N.D., Li, N., Allard, M., Li, C., Albano, E., Delaney, M., Dubois, A., Onderdonk, A.B., and Bry, L., 2015. Genomically Informed Surveillance for Carbapenem-Resistant Enterobacteriaceae in a Health Care System, 6 (4), 1–11.

Peleg, M. and Cole, M.B., 1998. Reinterpretation of microbial survival curves. *Crit Rev Food Sci Nutr*, 38 (5), 353–380.

Peleg, M., Normand, M.D., and Corradini, M.G., 2005. Generating microbial survival curves during thermal processing in real time. *Journal of Applied Microbiology*, 98 (2), 406–417.

Peleg, M., Normand, M.D., and Corradini, M.G., 2008. Interactive software for estimating the efficacy of non-isothermal heat preservation processes. *International Journal of Food Microbiology*, 126 (1–2), 250–257.

Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D., and Yeates, T.O., 1999. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles (genomic bioinformatics metabolic pathways structural complexes). *Biochemistry*, 96 (April), 4285–4288.

Periago, P.M., Van Zuijlen, a., Fernandez, P.S., Klapwijk, P.M., Ter Steeg, P.F., Corradini, M.G., and Peleg, M., 2004. Estimation of the non-isothermal inactivation patterns of Bacillus sporothermodurans IC4 spores in soups from their isothermal survival data. *International Journal of Food Microbiology*, 95 (2), 205–218.

Peters, J.E., Makarova, K.S., Shmakov, S., and Koonin, E. V., 2017. Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proceedings of the National Academy of Sciences*, 201709035.

Pettersson, B., Lembke, F., Hammer, P., Stackebrandt, E., and Priest, F.G., 1996. Bacillus sporothemodurans, a New Species Producing Highly Heat-Resistant Endospores. *International Journal of Systematic Bacteriology*, 46 (3), 759–764.

Pinto, C.L.O., Souza, L. V., Meloni, V.A.S., Batista, C.S., Silva, R., Martins, E.M.F., Cruz, A.G., and Martins, M.L., 2018. Microbiological quality of Brazilian UHT milk: Identification and spoilage potential of spore-forming bacteria. *International Journal of Dairy Technology*, 71 (1), 20–26.

Poptsova, M.S. and Gogarten, J.P., 2010. Using comparative genome analysis to identify problems in annotated microbial genomes. *Microbiology*, 156 (7), 1909–1917.

Postollec, F., Mathot, A.G., Bernard, M., Divanac'h, M.L., Pavan, S., and Sohier, D., 2012. Tracking spore-forming bacteria in food: From natural biodiversity to selection by processes. *International Journal of Food Microbiology*, 158 (1), 1–8.

Pritchard, L., Glover, R.H., Humphris, S., Elphinstone, J.G., and Toth, I.K., 2016. Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Analytical Methods*.

Qin, Y., Polacek, N., Vesper, O., Staub, E., Einfeldt, E., Wilson, D.N., and Nierhaus, K.H., 2006. The Highly Conserved LepA Is a Ribosomal Elongation Factor that Back-Translocates the Ribosome. *Cell*, 127 (4), 721–733.

Rambaut, A., 2016. FigTree v1.4.3. Tree figure drawing tool [online]. *Available: http://*

tree.bio.ed.ac. uk/software/figtre.

Raskin, D.M., Seshadri, R., Pukatzki, S.U., and Mekalanos, J.J., 2006. Bacterial genomics and pathogen evolution. *Cell*, 124 (4), 703–714.

Rea, R.B., Gahan, C.G.M., and Hill, C., 2004. Disruption of Putative Regulatory Loci in Listeria monocytogenes Demonstrates a Significant Role for Fur and PerR in Virulence. *Infection and Immunity*, 72 (2), 717–727.

Reding-Roman, C., Hewlett, M., Duxbury, S., Gori, F., Gudelj, I., and Beardmore, R., 2017. The unconstrained evolution of fast and efficient antibiotic-resistant bacterial genomes. *Nature Ecology & Evolution*, 1 (January), 0050.

Di Resta, C. and Ferrari, M., 2018. Next Generation Sequencing: From Research Area to Clinical Practice. *Ejifcc*, 29 (3), 215–220.

Reznick, D. and King, K., 2017. Antibiotic resistance: Evolution without trade-offs. *Nature Ecology & Evolution*, 1 (January), 0066.

Rhoads, A. and Au, K.F., 2015. PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics*, 13 (5), 278–289.

Ritota, M., Di Costanzo, M.G., Mattera, M., and Manzi, P., 2017. New Trends for the Evaluation of Heat Treatments of Milk. *Journal of Analytical Methods in Chemistry*, 2017, 1–12.

Roberts, A.P. and Mullany, P., 2011. Tn916-like genetic elements: A diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiology Reviews*, 35 (5), 856–871.

Roberts, R.J., Carneiro, M.O., and Schatz, M.C., 2013. The advantages of SMRT sequencing. *Genome Biology*, 14 (6), 405.

Rocha, E.P.C., 2008. The Organization of the Bacterial Genome. *Annual Review of Genetics*, 42 (1), 211–233.

Rouli, L., Merhej, V., Fournier, P., and Raoult, D., 2015. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes and New Infections*, 7, 72–85.

Rowland, S.J. and Dyke, K.., 2013. Tn552 , a novel transposable element from Staphylococcus aureus. *Mol Microbiol.*, 4.

Santos, A.R., Barbosa, E., Fiaux, K., and Chaitankar, V., 2013. PANNOTATOR : an automated tool for annotation of pan-genomes. *Genetics and Molecular Research*, 12 (3), 2982–2989.

Van Schaik, W., Château, A., Dillies, M.A., Coppée, J.Y., Sonenshein, A.L., and Fouet, A., 2009. The global regulator CodY regulates toxin gene expression in Bacillus anthracis and is required for full virulence. *Infection and Immunity*, 77 (10), 4437–4445.

Schauer, K., Geginat, G., Liang, C., Goebel, W., Dandekar, T., and Fuchs, T.M., 2010. Deciphering the intracellular metabolism of Listeria monocytogenes by mutant screening and modelling. *BMC Genomics*, 11 (1), 573.

Scheldeman, P., Goossens, K., Rodriguez-Diaz, M., Pil, A., Goris, J., Herman, L., De Vos, P., Logan, N. a, and Heyndrickx, M., 2004. Paenibacillus lactis sp. nov., isolated from raw and heat-treated milk. *International journal of systematic and evolutionary microbiology*, 54 (Pt 3), 885–91.

Scheldeman, P., Herman, L., Foster, S., and Heyndrickx, M., 2006. Bacillus sporothermodurans and other highly heat-resistant spore formers in milk. *Journal of Applied Microbiology*, 101 (3), 542–555.

Scheldeman, P., Herman, L., Goris, J., Vos, P. De, Heyndrickx, M., De Vos, P., and Heyndrickx, M., 2002. Polymerase chain reaction identification of Bacillus sporothermodurans from dairy sources. *Journal of Applied Microbiology*, 92 (5), 983–991.

Scheldeman, P., Pil, A., Herman, L., De Vos, P., and Heyndrickx, M., 2005. Incidence and diversity of potentially highly heat-resistant spores isolated at dairy farms. *Applied and Environmental Microbiology*, 71 (3), 1480–1494.

Schroder, H., Langer, T., Hartl, F., and Bukaul, B., 1993. DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage, 12 (1), 4137–4144.

Schulz, A. and Schumann, W., 1996. hrcA , the first gene of the Bacillus subtilis dnaK operon encodes a negative regulator of class I heat shock genes . These include : hrcA , the First Gene of the Bacillus subtilis dnaK Operon Encodes a Negative Regulator of Class I Heat Shock Genes. *Journal of Bacteriology*, 178 (4), 1088–1093.

Schürch, A.C., Arredondo-Alonso, S., Willems, R.J.L., and Goering, R. V., 2018. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. *Clinical Microbiology and Infection*.

Setlow, P., 2006. Spores of Bacillus subtilis: Their resistance to and killing by radiation, heat

and chemicals. *Journal of Applied Microbiology*, 101 (3), 514–525.

Shafique, M., 2012. Isolation and screening of protease producing thermophilic Bacillus strains from different soil types of Pakistan. *African Journal of Microbiology Research*, 5 (31).

Sharma, A. and Satyanarayana, T., 2013. Comparative Genomics of Bacillus species and its Relevance in Industrial Microbiology. *Genomics insights*, 6, 25–36.

Sheppard, S., Didelot, X., Jolley, K., Darling, A., Pascoe, B., and Meric, G., 2013. Progressive genome-wide introgression in agricultural Campylobacter coli. *Mol Ecol*, 22, 1051–64.

Sheppard, S.K., Didelot, X., Jolley, K.A., Darling, A.E., Pascoe, B., Meric, G., Kelly, D.J., Cody, A., Colles, F.M., Strachan, N.J.C., Ogden, I.D., Forbes, K., French, N.P., Carter, P., Miller, W.G., McCarthy, N.D., Owen, R., Litrup, E., Egholm, M., Affourtit, J.P., Bentley, S.D., Parkhill, J., Maiden, M.C.J., and Falush, D., 2013. Progressive genome-wide introgression in agricultural Campylobacter coli. *Molecular Ecology*, 22 (4), 1051–1064.

Siciliano, R., Rega, B., Amoresano, A., Pucci, P., Internazionale, C., Ii, F., and Organica, C., 2000. Modern Mass Spectrometric Methodologies in Monitoring Milk Quality, 72 (2), 408–415.

Siefert, J.L., Larios-Sanz, M., Nakamura, L.K., Slepecky, R. a., Paul, J.H., Moore, E.R.B., Fox, G.E., and Jurtshuk, Jr., P., 2000. Phylogeny of Marine Bacillus Isolates from the Gulf of Mexico. *Current Microbiology*, 41 (2), 84–88.

Siguier, P., Filee, J., and Chandler, M., 2006. Insertion sequences in prokaryotic genomes.

Current Opinion in Microbiology, 9, 526–531.

de Silva, S., Petterson, B., de Muro, M.A., and Priest, F.G., 1998. A DNA Probe for the Detection and Identification of Bacillus sporothermodurans Using the 16S-23S rDNA Spacer Region and Phylogenetic Analysis of some Field Isolates of Bacillus which form Highly Heat Resistant Spores. *Systematic and Applied Microbiology*, 21 (3), 398–407.

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

Stamatakis, A., 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, 30 (9), 33–34.

Stern, J. and Proctor, B., 1954. A micro-method and apparatus for the multiple determination of rates of destruction of bacteria and bacterial spores subjected to heat. *Food Technology*, (8), 139–143.

Stoecklin, G. and Bukau, B., 2013. Telling right from wrong in life — cellular quality control. *Nature Reviews Molecular Cell Biology*, 14, 613.

Stokes, H.W. and Gillings, M.R., 2011. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiology Reviews*, 35 (5), 790–819.

Stothard, P., Grant, J.R., and Domselaar, G. Van, 2018. Visualizing and comparing circular genomes using the CGView family of tools, (August), 1–7.

Swerdlow, H.P., Smith, M.E., Harris, S.R., Bertoni, A., Quail, M., Otto, T.D., Coupland, P., Gu, Y., and Connor, T.R., 2012. A tale of three next generation sequencing platforms:

comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics*, 13 (1), 341.

Tabit, F.T., 2018. Prevalence and growth characteristics of Bacillus sporothermodurans in UHT milk. *British Food Journal*, 120 (10), 2250–2260.

Tabit, F.T. and Buys, E., 2010. The effects of wet heat treatment on the structural and chemical components of Bacillus sporothermodurans spores. *International Journal of Food Microbiology*, 140 (2–3), 207–213.

Tabit, F.T. and Buys, E.M., 2011. Incidence and survival of Bacillus sporothermodurans during processing of UHT milk. *British Food Journal*, 113 (4), 505–518.

Tagini, F. and Greub, G., 2017. Bacterial genome sequencing in clinical microbiology: a pathogen-oriented review. *European Journal of Clinical Microbiology and Infectious Diseases*, 36 (11), 2007–2020.

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., and Ostell, J., 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, 44 (14), 6614–6624.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S. V., Crabtree, J., Jones, A.L., Durkin, A.S., DeBoy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J.B., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels,

M.R., Rappuoli, R., and Fraser, C.M., 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial 'pan-genome'. *Proceedings of the National Academy of Sciences*, 102 (39), 13950–13955.

Tettelin, H., Riley, D., Cattuto, C., and Medini, D., 2008. Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, 11 (5), 472–477.

Tyler, A.D., Mataseje, L., Urfano, C.J., Schmidt, L., Antonation, K.S., Mulvey, M.R., and Corbett, C.R., 2018. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Scientific Reports*, 8 (1), 1–12.

Upadrasta, A., Pitta, S., and Madempudi, R.S., 2016. Draft Genome Sequence of the Spore-Forming Probiotic Strain *Bacillus coagulans* Unique IS-2. *Genome Announcements*, 4 (2), e00225-16.

Vaerewijck, M.J.M.J.M., De Vos, P., Lebbe, L., Scheldeman, P., Hoste, B., and Heyndrickx, M., 2001. Occurrence of Bacillus sporothermodurans and other aerobic spore-forming species in feed concentrate for dairy cattle. *Journal of Applied Microbiology*, 91 (6), 1074–1084.

Valero, E., Villamiel, M., Miralles, B., Sanz, J., and Martõânez-castro, I., 2001. Changes in flavour and volatile components during storage of whole and skimmed UHT milk, 72.

Verma, A., Pal, Y., Khatri, I., Ojha, A.K., Gruber-Vodicka, H., Schumann, P., Dastager, S., Subramanian, S., Mayilraj, S., and Krishnamurthi, S., 2017. Examination into the taxonomic position of Bacillus thermotolerans Yang et al., 2013, proposal for its reclassification into a new genus and species Quasibacillus thermotolerans gen. nov., comb. nov. and reclassification of B. encimensis Dastager et al., . *Systematic and*

Applied Microbiology, 40 (7), 411–422.

Větrovský, T. and Baldrian, P., 2013. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS ONE*, 8 (2), 1–10.

Vettone, A., Perugino, G., Rossi, M., Valenti, A., and Ciaramella, M., 2014. Genome stability: recent insights in the topoisomerase reverse gyrase and thermophilic DNA alkyltransferase. *Extremophiles*, 18 (5), 895–904.

Vos, M. and Eyre-walker, A., 2017. Are pangenomes adaptive or not? *Nature Microbiology*, 2 (12), 1576–1576.

Walsh, D. a, Bapteste, E., Kamekura, M., and Doolittle, W.F., 2004. Evolution of the RNA polymerase B' subunit gene (rpoB') in Halobacteriales: a complementary molecular marker to the SSU rRNA gene. *Molecular biology and evolution*, 21 (12), 2340–51.

Wang, J., Liu, B., Liu, G., Ge, C., Xiao, R., Zheng, X., and Shi, H., 2016. Draft Genome Sequence of *Bacillus farraginis* R-6540 ^{T} (DSM 16013), a Spore-Forming Bacterium Isolated at Dairy Farms. *Genome Announcements*, 4 (3), e00562-16.

Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., Machi, D., Mao, C., Nordberg, E.K., Olson, R., Overbeek, R., Pusch, G.D., Shukla, M., Schulman, J., Stevens, R.L., Sullivan, D.E., Vonstein, V., Warren, A., Will, R., Wilson, M.J.C., Yoo, H.S., Zhang, C., Zhang, Y., and Sobral, B.W., 2014. PATRIC , the bacterial bioinformatics database and analysis resource, 42 (November 2013), 581–591.

Wells-Bennik, M.H., Driehuis, F., and van Hijum, S.A., 2016. Prospects for improved control

of dairy-relevant sporeformers using -omics technologies. *Current Opinion in Food Science*.

Whitehead, S., Holroyd, S., Jagels, K., Skelton, J., Bentley, S.D., Baker, S., Karlyshev, A. V., Wren, B.W., Brooks, K., Davis, P., Dougan, G., James, K.D., Leather, S., Titball, R.W., Davies, R.M., Oyston, P.C.F., Parkhill, J., Prentice, M.B., Churcher, C., Barrell, B.G., Chillingworth, T., Thomson, N.R., Cerdeño-Tárraga, A.M., Sebaihia, M., Moule, S., Holden, M.T.G., Simmonds, M., Hamlin, N., Cronin, A., Stevens, K., Mungall, K.L., Quail, M., Basham, D., Feltwell, T., and Rutherford, K., 2002. Genome sequence of Yersinia pestis, the causative agent of plague. *Nature*, 413 (6855), 523–527.

Wickner, S., Maurizi, M.R., and Gottesman, S., 1999. Posttranslational Quality Control: Folding, Refolding, and Degrading Proteins. *Science*, 286 (5446), 1888 LP – 1893.

Wiedenbeck, J. and Cohan, F.M., 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews*, 35 (5), 957–976.

Willenbacher, J., Mohr, T., Henkel, M., Gebhard, S., Mascher, T., Syldatk, C., and Hausmann, R., 2016. Substitution of the native srfA promoter by constitutive Pveg in two B. Subtilis strains and evaluation of the effect on Surfactin production. *Journal of Biotechnology*, 224, 14–17.

Wolf, R., Ophir, J., Avigad, J., Lengy, J., and A., K., 1988. The hair follicle mites (Demodex spp.). Could they be vectors of pathogenic microorganisms? *Acta Derm Venereol.*, 68 (6), 535–7.

Wolf, Y.I., Rogozin, I.B., Kondrashov, A.S., and Koonin, E. V, 2001. Genome Alignment ,

Evolution of Prokaryotic Genome Organization , and Prediction of Gene Function Using Genomic Context Genome Alignment , Evolution of Prokaryotic Genome Organization , and Prediction of Gene Function Using Genomic Context. *Genome Research*, 11, 356–372.

Woo, P.C.Y., Lau, S.K.P., Teng, J.L.L., Tse, H., and Yuen, K.Y., 2008. Then and now: Use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clinical Microbiology and Infection*, 14 (10), 908–934.

Wozniak, M., Wong, L., and Tiuryn, J., 2011. CAMBer: an approach to support comparative analysis of multiple bacterial strains. *BMC Genomics*, 12 Suppl 2 (Suppl 2), S6.

Wozniak, M., Wong, L., and Tiuryn, J., 2014. ECAMBer: Efficient support for large-scale comparative analysis of multiple bacterial strains. *BMC Bioinformatics*, 15 (1).

Wu, X.-Y., Walker, M.J., Hornitzky, M., and Chin, J., 2006. Development of a group-specific PCR combined with ARDRA for the identification of Bacillus species of environmental significance. *Journal of microbiological methods*, 64 (1), 107–19.

Xiao, Jingfa; Zhang, Zhewen; Wu, Jiayan; Yu, J., 2015. RESOURCE REVIEW A Brief Review of Software Tools for Pangenomics. *Genomics, Proteomics & Bioinformatics*, 13 (1), 73–76.

Xu, S., Labuza, T.P., and Diez-Gonzalez, F., 2006. Thermal inactivation of Bacillus anthracis spores in cow's milk. *Applied and Environmental Microbiology*, 72 (6), 4479–4483.

Yang, G., Zhou, X., Zhou, S., Yang, D., Wang, Y., and Wang, D., 2013. Bacillus thermotolerans sp. nov., a thermophilic bacterium capable of reducing humus.

International Journal of Systematic and Evolutionary Microbiology, 63 (PART10), 3672–3678.

Yuan, D.D., Liu, G.C., Ren, D.Y., Zhang, D., Zhao, L., Kan, C.P., Yang, Y.Z., Ma, W., Li, Y., and Zhang, L.B., 2012. A survey on occurrence of thermophilic bacilli in commercial milk powders in China. *Food Control*, 25 (2), 752–757.

Zallen, D.T., 2003. Despite Franklin's work, Wilkins earned his Nobel. *Nature*, 425 (15).

Zhao, Y., Jia, X., Yang, J., Ling, Y., Zhang, Z., Yu, J., Wu, J., and Xiao, J., 2014. PanGP: A tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics*, 30 (9), 1297–1299.

Zhao, Y., Sun, C., Zhao, D., Zhang, Y., You, Y., Jia, X., Yang, J., Wang, L., Wang, J., Fu, H., Kang, Y., Chen, F., Yu, J., Wu, J., and Xiao, J., 2018. PGAP-X : extension on pan-genome analysis pipeline, 19 (Suppl 1).

Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., and Yu, J., 2012. PGAP : pan-genomes analysis pipeline. *Bioinformatics*, 28 (3), 416–418.

Zheng, Y., Xi, X., Xu, H., Hou, Q., Bian, Y., Yu, Z., Kwok, L.Y., Zhang, W., Sun, Z., and Zhang, H., 2016. Using PacBio Long-Read High-Throughput Microbial Gene Amplicon Sequencing to Evaluate Infant Formula Safety. *Journal of Agricultural and Food Chemistry*, 64 (37), 6993–7001.

Zick, A., Onn, I., Bezalel, R., Margalit, H., and Shlomai, J., 2005. Assigning functions to genes: Identification of S-phase expressed genes in Leishmania major based on post-transcriptional control elements. *Nucleic Acids Research*, 33 (13), 4235–4242.

van Zuijlen, A., Periago, P.M., Amézquita, A., Palop, A., Brul, S., and Fernández, P.S., 2010. Characterization of Bacillus sporothermodurans IC4 spores; putative indicator microorganism for optimisation of thermal processes in food sterilisation. *Food Research International*, 43 (7), 1895–1901.

Publications and presentations from this work

Scientific papers

Owusu-Darko, R., Allam, M., Mtshali, S., Ismail, A. and Buys, E. M. 2017. Draft genome sequence of *Bacillus oleronius* DSM 9356^{T} isolated from the termite *Reticulitermes santonensis*. *Genomics Data* 12, 76–78.

Owusu-Darko, R., Allam, M., de Oliveira, S. D., Ferreira, C. A. S., Grover, S., Mtshali, S., Ismail, A., Mallappa, R. H., Tabit, F. and Buys, E. M. 2019. Genome sequences of *Bacillus sporothermodurans* strains isolated from Ultra-High-Temperature milk. *Microbiology Resource Anouncements* In press.

Conference presentations

Oral presentations

Owusu-Darko, R., Dias de Oliveira, S. and Buys, E. M. 2017. *Bacillus sporothermodurans* in UHT milk: Use of Next Generation Sequencing to characterise heat resistance with the aim of improving milk safety and quality. South African Society of Dairy Technology Dairy Students Evening. Pretoria, SA. 31st August, 2017.

Owusu-Darko, R., Allam, M., Mtshali, S., Ismail, A. and Buys, E. M. 2017. Thermal death kinetics of *Bacillus sporothermodurans* spores isolated from UHT milk. International Association for Food Protection (IAFP) Annual meeting, Florida, US. 9th – 12th July 2017.

Owusu-Darko, R. and Buys, E. M. 2015. Use of molecular sequencing in enhancing thermal inactivation protocols. SAAFoST Biennial Congress, Durban, SA. 6th – 9th September, 2015.

Poster presentations

Owusu-Darko, R., Allam, M., Mtshali, S., Ismail, A. and Buys, E. M 2017. Draft genome sequence of *Bacillus oleronius* DSM 9356^{T}: from the hindgut of a termite to possible milk contaminant. Society for Applied Microbiology, Newcastle, UK. 3rd – 6th July 2017.

Genetic characterisation of *Bacillus sporothermodurans* in UHT milk. International Association for Food Protection (IAFP) Annual meeting, Oregon, US. 25th – 28th July 2015.