UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Comparative genomics of *Bacillus anthracis* strains from anthrax outbreaks in Kruger National Park, South Africa

by

Sankwetea Prudent Mokgokong

(19386975)

Submitted in fulfilment of the requirements for the degree

Magister Scientia (Veterinary Tropical Diseases)

Faculty of Veterinary Science, University of Pretoria

Supervisor:

Prof. Henriette van Heerden

(University of Pretoria, South Africa)

Co supervisors:

Dr. Kgaugelo Edward Lekota

(North-West University, South Africa)

Dr. Wendy Turner

(Department of Forest and Wildlife Ecology

University of Wisconsin-Madison)

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Veterinary Science

Research Ethics Committee 01 April 2021

LETTER OF APPROVAL

Ethics Reference No	**REC066-20**
Protocol Title	**Comparative genomics of Bacillus anthracis strains isolated from animal anthrax outbreaks in Kruger National Park, South Africa**
Principal Investigator	**Sankwetea Mokgokong**
Supervisors	**Prof H van Heerden**

Dear Sankwetea Mokgokong,

We are pleased to inform you that your submission conforms to the requirements of the Faculty of Veterinary Sciences Research Ethics committee.

Please note the following about your ethics approval:

1. Please use your reference number (REC066-20) on any documents or correspondence with the Research Ethics Committee regarding your research.
2. Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, monitor the conduct of your research, or suspend or withdraw ethics approval.
3. Please note that ethical approval is granted for the duration of the research as stipulated in the original application (for Post graduate studies e.g. Honours studies: 1 year, Masters studies: two years, and PhD studies: three years) and should be extended when the approval period lapses.
4. The digital archiving of data is a requirement of the University of Pretoria. The data should be accessible in the event of an enquiry or further analysis of the data.

Ethics approval is subject to the following:

1. The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.
2. **Applications using Animals:** FVS ethics recommendation does not imply that AEC approval is granted. The application has been pre-screened and recommended for review by the AEC. Research may not proceed until AEC approval is granted.

We wish you the best with your research.

Yours sincerely

PROF M. OOSTHUIZEN
Chairperson: Research Ethics Committee

Declaration

I, Sankwetea Prudent Mokgokong, hereby declare that "Comparative genomics of *Bacillus anthracis* strains from anthrax outbreaks in Kruger National Park, South Africa" is my original work, that it has not been submitted to any other university for a degree or examination, and that all sources I have used have been acknowledged with full citations.

Sankwetea Prudent Mokgokong

November 2022

Onderstepoort, Pretoria, South Africa

Acknowledgements

Firstly, to God be the glory. Isaiah 60:22 “When the time is right, I, The Lord will make it happen. I would like to sincerely thank the following:

My supervisor, Prof. Henriette van Heerden for giving me the opportunity to work on the project, a privilege not afforded to many.

My co-supervisor, Dr Edward Lekota, who pushed me to limits I didn’t know I could get to. Thank you for believing in me, for serving beyond duty and for always willing to give me support.

Prof Wendy Turner for all the guidance and criticism. I value your expertise.

Dr Morne du Plessis for sharing your knowledge of bioinformatics, your patience and support as a mentor and a friend.

Dr Ayesha Hassim for the assistance and patience with the samples and data.

Colleagues at South African National Biodiversity Institute and North West University for the coffee and encouragement. Prof Desire Dalton, thank you for giving me some time off to work on my MSc. Sesi Betty and Mogolo Lucas for the constructive feedback. Indeed, we seek the light from those who have already been through it.

My family, for always supporting my dreams even when I didn’t believe in myself. I extend my gratitude to my son Omphile Letlotlo. You were my light and although it was hard you understood that I had to do what I had to do.

Finally, thanks to the Agri-SETA, National Research Foundation (NRF), National Science Foundation (NSF), and University of Pretoria, Department of Veterinary Tropical Diseases (UP-DVTD) for the funding.

Table of Contents

List of figures

List of tables

Dissertation summary

Anthrax is a zoonotic disease caused by a Gram-positive, rod-shaped, soil bacterium known as *Bacillus anthracis*. The global phylogenomic structure of *B. anthracis* consists of three major lineages namely clades A, B, and C. Anthrax is endemic in the northern part of Kruger National Park (KNP) in South Africa, which show cases of distinctive genetic diversity of *B. anthracis* A and B-clade strains. Until the 1990s, animals in the KNP were mainly infected by B-clade strains, but since the 1990s, only A-clade strains have been isolated from animals. In this study, whole genome sequencing (WGS) approach was employed to screen *B. anthracis* isolates (n=80) collected from animal carcasses and environmental samples in the KNP between 2012 and 2015. Whole genome single nucleotide polymorphism (wgSNP) and pan-genome analysis were employed to infer the population structure of *B. anthracis* A- and B-clades, as well as to understand the genetic differences amongst these clades, respectively. Furthermore, genetic characterization of the Bacillus collagen-like protein A (*bclA*), tryptophan operon and antibiotic resistance (AR) genes were investigated on A- and B-clades of the KNP *B. anthracis* genomes.

The population structure of the KNP *B. anthracis* consisted of diverse strains grouping in the prominent A.Br.005/006 (Ancient A) SNP lineage. The *B. anthracis* isolates from 2012-2015 are dispersed amongst minor sub-clades that present a non-stabilized genetic population within this SNP lineage. Pan-genomics showed a clear distinction between A- and B-clade *B. anthracis* genomes with 11,374 predicted clusters of protein coding sequences. B-clade genomes contain unique accessory genes that dissociates them phylogenetically from the A-clade genomes. These consisted of multiple copies of Heme-based aerotactic transducer (*hemAT)*, long-chain-fatty-acid--CoA ligase (*fadD13)* and murein hydrolases (*mepH*) which form part of the cell wall biosynthesis pathway of *B. anthracis*. The *mepH* gene, located on pXO2 in B-clade genomes is truncated and structurally differs from the A-clade genomes. AR profiles showed that the genomes of A and B lineages consist of *fosB1* and *fosB2* genes that encode for fosfomycin resistance. However, a gene copy of a multidrug gene (*mdtG_2)* was a unique feature of the B-clade

genomes. Furthermore, one of the KNP A.Br.005/006 strain 1298 had a unique antimicrobial resistance profile (*blaTEM-116_1*) which conferred resistance to amoxicillin, ampicillin, cephalothin, piperacillin and ticarcillin. Genes linked to prophage regions were also identified as a unique feature in A-clade genomes. Genetic differences in the tryptophan operon showed a truncated tryptophan synthase subunit alpha (*trpA*) gene in the B-clade strains. Sequence length variation of the BclA protein on the exosporium of the *B. anthracis* endospores and the copy number of the GXX amino acid repeats identified in the collagen like region (CLR) could be used for subtyping and differentiation of *B. anthracis* strains. The A.Br.005/006 genomes had a higher copy number (4:3) of variable number of tandem repeats (VNTR) in comparison to B-clade genomes. Comparative genomics was successfully used to infer genetic diversity of *B. anthracis* strains.

This study highlights the ubiquity of the A-clade, and its persistence in both enzootic and non-enzootic (areas south of the endemic region) regions of the KNP. This is the first study to report genetic variability in A- and B-clade *B. anthracis* strains based on the tryptophan operon and the *Bacillus* collagen-like protein A (*bclA*) in the KNP. The observed mutation in the *trpA* gene results into a pseudogene, inhibiting the termination step of the tryptophan biosynthesis pathway. Additionally, we were able to identify genes that contribute to the phylogenomic clustering of A and B KNP *B. anthracis* strains. Further studies on cell wall structure (genes and associated pathways) will provide insight on the persistence of A-clade *B. anthracis* genotypes over B-clade strains. This will inform adaptability and survival in the endemic regions. This study has revealed that WGS is a powerful tool that can be used to understand the evolution of *B. anthracis* for epidemiological surveillance and to trace anthrax outbreaks.

Chapter 1: Introduction and rationale

1.1 Introduction

Bacillus anthracis is an endospore–forming and soil-born pathogen that causes anthrax in mammals (Turnbull, 1998; WHO, 2008). The *B. anthracis* spores are highly resistant to desiccation in harsh environmental conditions and remain dormant for extended period (WHO, 2008). The spores are found in specific soil types and conditions, thriving in alkaline soils (Hugh-Jones and de Vos, 2002; WHO, 2008). When environmental conditions are favourable, the endospores can germinate into vegetative bacterial cell, causing disease (Hugh-Jones and de Vos, 2002; WHO, 2008) .

B. anthracis is enzootic to the Pafuri region of Kruger National Park (KNP) (de Vos, 1990) and the Ghaap region in the Northern Cape province (NCP) (Hugh-Jones and de Vos, 2002). KNP has a history of episodic outbreaks of anthrax recorded since the 1950s and certain geographical localities within it have been described as anthrax endemic areas (Smith, 1999). In the KNP, the endemic anthrax outbreaks are attributed to dry seasons after heavy rainfalls (de Vos, 1990; Steenkamp *et al.*, 2018) with the main host being kudu (*Tragelaphus strepsiceros)* (Hugh-Jones and de Vos, 2002). However, Basson *et al.* (2018) indicated that outbreaks since the 2000s occurred in the wet season infecting predominantly impala (*Aepyceros melampus*). Huang *et al.* (2022), showed that the dominant host turnover occurred in the 1990s, possibly attributed to the reduced population density of kudu after large outbreaks and drought or floods (Parsons *et al.*, 2005).

Based on the genetic structure, *B. anthracis,* three major clades A, B and C have been described (Keim *et al.*, 1999; Smith *et al.*, 2000; Van Ert *et al.*, 2007). These clades are well defined using canonical single nucleotide polymorphism (canSNP), MLVA (multi-loci variable number tandem repeat (VNTR) assay) and whole genome SNP (wgSNP) analysis (Smith *et al.*, 2000; Le Flèche *et al.*, 2001; Beyer *et al.*, 2012; Gachohi *et al.*, 2019; Lekota *et al.*, 2020). The *B. anthracis* genotypes were initially characterized using

eight VNTR markers for the assay (MLVA-8) (Keim *et al.*, 2000; Smith *et al.*, 2000) identifying both the A and B (KrugerB) clades in the KNP. In KNP, the rare B-clade (B.Br.001) isolates were dominant in the northern region of Pafuri while the A-clade isolates were dominant in the central region of the park (Smith *et al.*, 2000). However, since the 1990s, the A-clade isolates were the only clade present in the northern Pafuri region of the KNP (Smith *et al.*, 2000). There is a high genetic diversity of *B. anthracis* strains in the KNP consisting of the extinct B-clade (B.Br.001), various A-subclades such as A.Br.005/006 (Ancient A), and A.Br.002 (Sterne) compared to the diversity in Etosha National Park (ENP) in Namibia. The ENP consists mainly of the A.Br.Aust94 and A.Br.008/9 clades genotyped using MLVA-31 analysis over 50 years (Figure 1.1). The genotypes in ENP are mainly comprised of the A.Br.Aust94 subclade and are stable over time (Beyer *et al.*, 2012), while in the KNP the dominant genotypes changed over time from B to A (Figure 1.1).

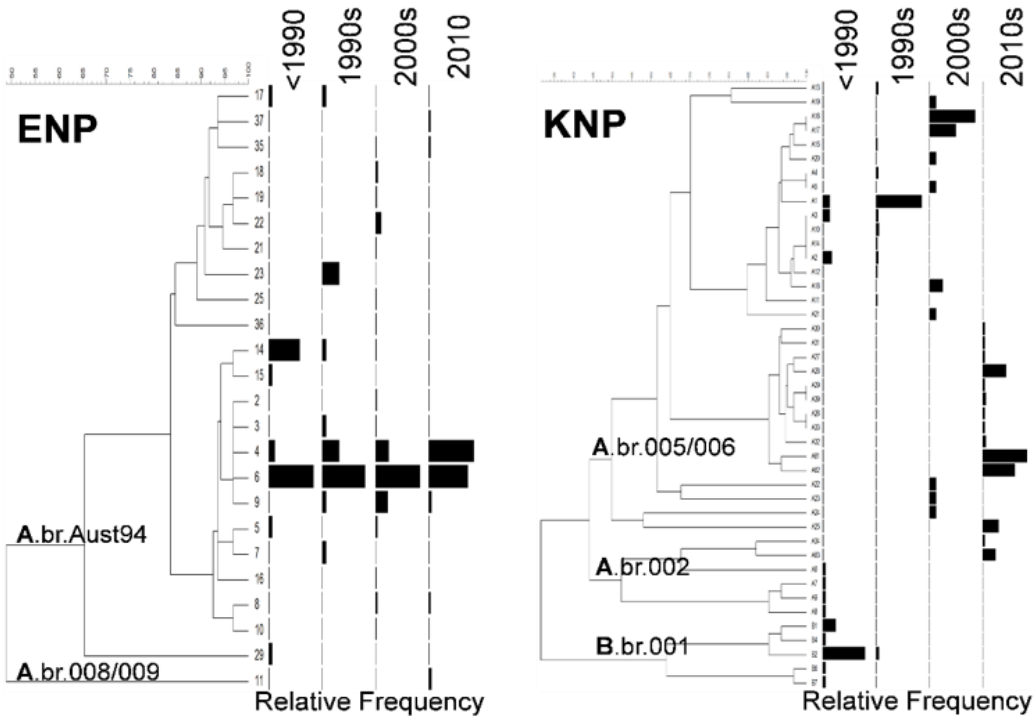


Figure 1.1: *Bacillus anthracis* strains diversity in Etosha National Park (ENP) (N=24) genotypes (Beyer *et al.*, 2012), and KNP (N=45 genotypes) (Hassim, 2016); Lekota *et al.*, 2018) (B clade)) over 50 years, based on 31-marker multi-locus-VNTR-analysis (MLVA). The histograms show the relative frequency among strains detected by decade. The dominant strains in ENP are relatively stable over time, while in the KNP show significant change over time.

This diversity is also evident when comparing the enzootic regions in South Africa. The *B. anthracis* strains from the NCP and the KNP groups occur in heterogenous genetic lineages that include A.Br.005/006 (Ancient A), A.Br.001/002 (Ames/Sterne), A.Br.0057 (V770), and B.Br.010 (B-branch) clades with a considerable geographic and genetic diversity of the A.Br.005/006 clade observed in the KNP (Lekota *et al.*, 2018) (Figure 1.2).

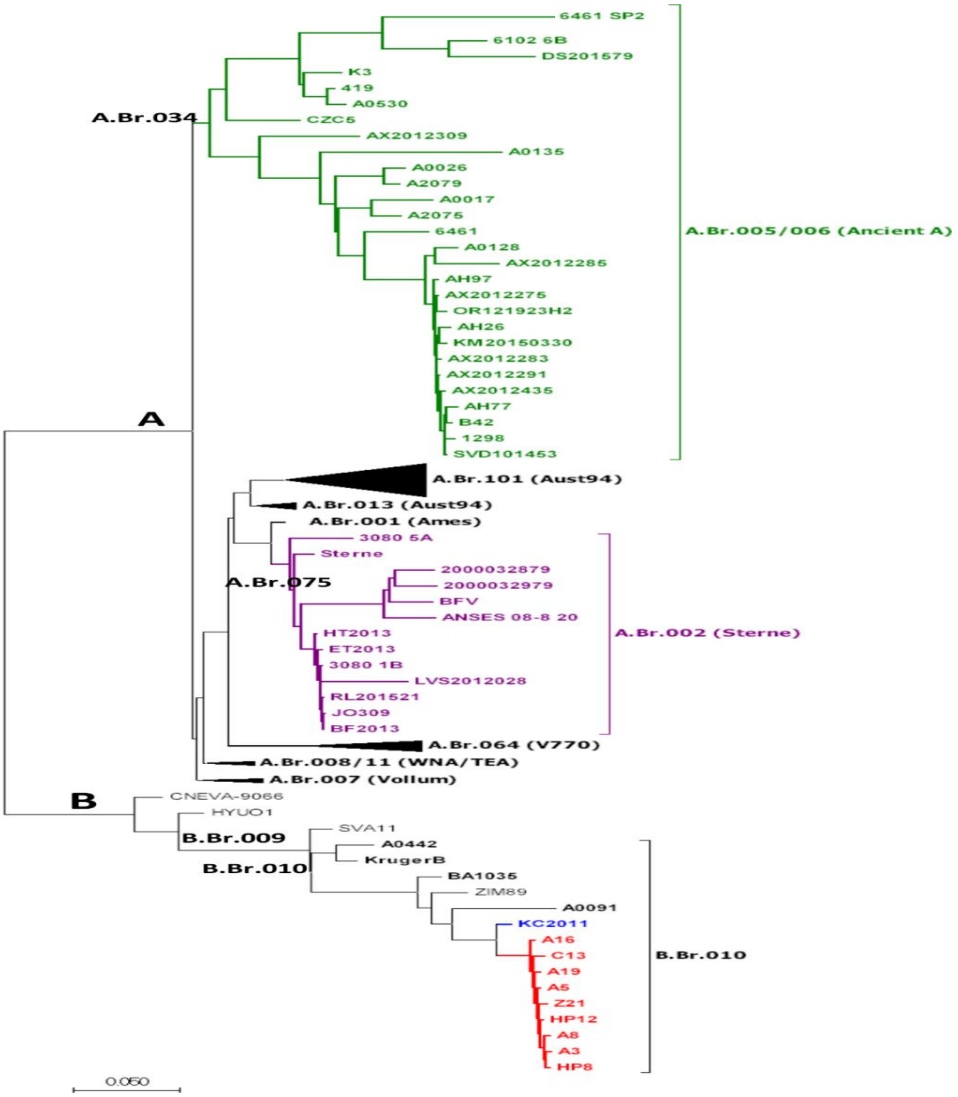


Figure 1.2: Whole genome phylogeny of *Bacillus anthracis* strains in Kruger National Park (KNP) showing the genetic diversity (Lekota *et al.*, 2018). The phylogenetic structure indicates the clustering of the strains into different sub-clades The South African strains grouped in the A.Br.005/006 (Ancient A) indicated in dark green, A.Br.001/2 (Sterne) in purple, and B.Br.10 (B-clade) in red. Image from Lekota *et al.*, (2018).

1.2. Rationale

Kruger National Park consists of heterologous strains that are mostly genotyped based on MLVA. The genetic diversity of *B. anthracis* is not fully exploited as MLVA results in homoplasticity (Achtman, 2008 ; Le Flèche *et al.*, 2001). This presents a challenge when assessing the evolution and population dynamics of monomorphic bacteria such as *B. anthracis*; therefore, understanding the anthrax outbreaks and tracing the dissemination of *B. anthracis* in the KNP is imperative. Whole genome sequencing exploits the complete genomic region and is a powerful tool to genotype monomorphic pathogenic bacteria such as *B. anthracis*. The population genetic structure of *B. anthracis* in the KNP shows a prominent A.Br.005/006 SNP branch that is heterogeneous and present in the enzootic and non-enzootic phylogeographies (Lekota *et al.*, 2016, 2018). This SNP branch is circulating in both the northern and central part of the KNP and is the most probable source of future outbreaks in the KNP; thus, understanding the genetic diversity will enable curbing of the disease in the park. Comparative genomics of A- and B-clade strains will also contribute towards understanding the ecology of *B. anthracis* in the endemic regions of the KNP.

1.3. Aims and objectives

The aim of the study was to determine and further delineate the genetic diversity of *B. anthracis* strains in the Kruger National Park (KNP) and to evaluate genetic variations among *B. anthracis* A- and B-clade strains.

Objectives

- To conduct whole genome sequencing of the *B. anthracis* strains in the KNP.
- To determine the genetic diversity of *B. anthracis* strains in the KNP using whole genome single nucleotide polymorphism (wgSNP) approach.
- To conduct comparative genomics on *B. anthracis* A- and B- clades, targeting the Bacillus collagen-like protein of anthracis (*bclA* gene), the tryptophan operon genes and antimicrobial resistance (AMR) genes.
- To identify unique genes by employing pan-genome analysis of A- and B-clade strains.

Chapter 2: Literature review

2.1 Anthrax and its history

Anthrax is an infectious zoonotic disease that primarily affects both wild herbivores and domestic animals (WHO, 2008). Infections have also been reported in carnivores and humans however, these hosts are resistant, and anthrax occurs less often in these hosts. This infectious zoonotic disease was coined from the Latin word (anthrax) meaning coal, which is characterized by the black lesions presented on humans infected with cutaneous anthrax (WHO, 2008). Anthrax is caused by a Gram-positive, endospore forming bacterium called *B. anthracis* (Sterne, 1937; Leppla, 1982). The causative agent was first identified in 1863 then isolated by Robert Koch in 1876. *Bacillus anthracis* is a member in the family Bacillaceae consisting of saprophytic, sporulating, non-motile rods (Turnbull and WHO, 1998).). The bacteria can exist in two forms in its life cycle namely as vegetative cells in hosts or as endospores. In the environment, the spores remain viable in soil for extended periods and are not prone to decay (Dragon and Rennie, 1995; Carlson *et al.*, 2019). The dormant endospores are tolerant to extreme conditions such as ultraviolet radiation (Nicholson and Galeano, 2003),, high temperatures and pH (WHO, 2008), humidity (Minett, 1950) and nutrient starved habitats (Dragon and Rennie, 1995). Upon entry of spores into susceptible host cells, the nutrient rich internal environment triggers germination and propagation of the spores into vegetative cells (WHO, 2008). The event results in the death of the infected animal and the release of vegetative cells into the surrounding environment. Sporulation occurs upon exposure to oxygen and the released spores remain in the environment until they are taken up by another host (Sterne, 1937; WHO, 2008).

Anthrax records date as far back as the first millennium BC with numerous scholars believing it may have caused the fifth and six plaques in Egypt in cattle and humans (approximately 1250 BC). It has additionally been described in the Siege of Troy (approximately 1200 BC), the writings of Homer (approximately 1000 BC), Hippocrates (approximately 400 BC), Varro (116-27 BC), Virgil (70–19 BC) and Galen (approximately 200 AD) (Turnbull and Shadomy, 2010).

Major outbreaks have been recorded from the 17-19th century in Europe and resulted in severe domestic livestock losses. The disease in animals was first described in 1780 by Chabert (Wistreich and Lechtman, 1973, Turnbull, 1996) and Barthelemy further demonstrated disease transmission via inoculation of blood from diseased animals to healthy animals (Wilson and Miles, 1975). In 1850, small filiform bodies were identified in a sheep with anthrax by Pierre Rayer and Casimir Joseph Davaine (Klemm and Klemm, 1959; Dirckx, 1981; Rouli *et al.*, 2014). Later Davaine described the presence of bacilli in the blood of a diseased animal and suggested that these microorganisms were causing the disease (Dirckx, 1981). In the 1870's, the disease was extensively investigated by Robert Koch and Louis Pasteur (Carter, 1988).

Koch described the life cycle of the *Bacillus* and identified spores that could survive for extended periods of time in the environment. Louis Pasteur developed the first animal anthrax vaccine in 1881 using attenuated live organisms (Encyclopaedia Britannica, 2018). Later, in the late 1930's the Sterne's livestock vaccine was used in South Africa and then in the rest of the world (Sterne, 1967). The vaccine in combination with other control programs resulted in a dramatic reduction in the disease in enzootic areas from 1940-1960, leading to the disease been reported as being of minor concern to both veterinary and public health (Riedel, 2005).

2.2 *Bacillus anthracis* pathogenesis

The *B. anthracis* genome comprises of an approximately 5.2 Mb single chromosome and two double stranded circular plasmids (Figure 2.1), pXO1 and pXO2, which carry the major virulence factors required for pathogenesis (Read *et al.*, 2003; Ravel *et al.*, 2009). The chromosome encodes over 5000 predicted proteins (Read *et al.*, 2003). The pXO1 (182 kb) encodes a tripartite protein exotoxin complex, while pXO2 (95 kb) encodes the *cap*BCAD operon responsible to produce a polyglutamate (PGA) capsule. The capsule protects the pathogen from phagocytosis, allowing invasion and evasion of the host immune response (Mock and Fouet, 2001).

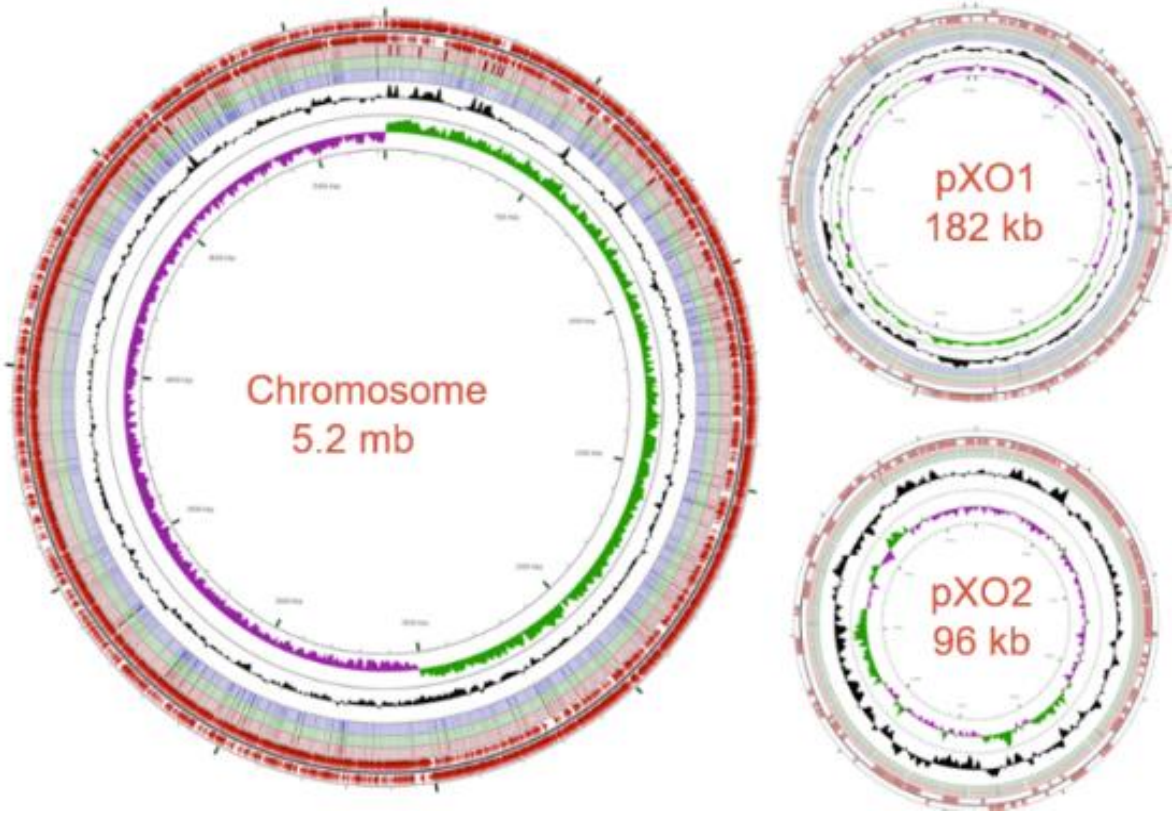


Figure 2.1: The genome structure of *Bacillus anthracis* indicating the double-stranded circular chromosome and the two plasmids (pXO1 and pXO2) which encode the main virulence genes and capsule genes, respectively (Lekota, 2018).

The two major virulence factors are formed by three components of the exotoxin, the protective antigen (PA), the lethal factor (LF) and the edema factor (EF). The lethal toxin (LT) comprises of the PA and the LF, whereas the edema toxin (ET) is composed of the PA and the EF (Pezard *et al.*, 1991; Mock and Fouet, 2001). The PA (molecular weight of 83 kDa) (PA83) attaches to the cellular receptors on target host cells and is proteolytically digested into PA63 (63 kDa) and PA20 (20 kDa) by furin or furin-like proteases (Petosa *et al.*, 2013; Hammamieh *et al.*, 2008; Froude *et al.*, 2011). The PA63 binds to the anthrax toxin receptor (ATRs) after PA20 dissociates and diffuses into the surrounding medium (anthrax toxin receptors) (Figure 2.2). The PA heptamerization causes ATR clustering, exposing the edema factor (EF) and the lethal factor (LF) binding domains to the receptor-bound PA63 that is carried into cells via membrane lipid rafts (Abrami *et al.*, 2003). The LF is a zinc-dependent metalloprotease with four domains, while the EF is a calmodulin-dependent adenylyl cyclase (Leppla, 1982; Abrami *et al.*, 2003; Han *et al.*, 2006; Pilo and Frey, 2018). The LF cleaves the N-terminal region of mitogen-activated protein kinase kinases (MAPKK).

This cleavage halts each of the MAPKKs from activating, and it has the potential to impair downstream signaling, required for a range of basic cell functions, such as immunological and stress response activation (Sherer *et al.*, 2007). The process results in cell lysis, inducing pro-inflammatory cytokines, leading to vascular collapse, shock and death (Figure 2.2). The LF causes cytolysis and increases macrophage death by altering membrane permeability, dissipating mitochondrial membrane, and fragmenting DNA (Leppla, 1982). In the host cells, the catalysis by the EF converts adenosine triphosphate to cyclic adenosine monophosphate (CAMP) (Han *et al.*, 2006), which de-regulates the cytokine response (Leppla, 1982; Chaudry *et al.*, 2002; Tang and Guo, 2009). The action of EF on host cytosolic target induces edema, necrosis, and hypoxia (Prince, 2003; Sherer *et al.*, 2007).

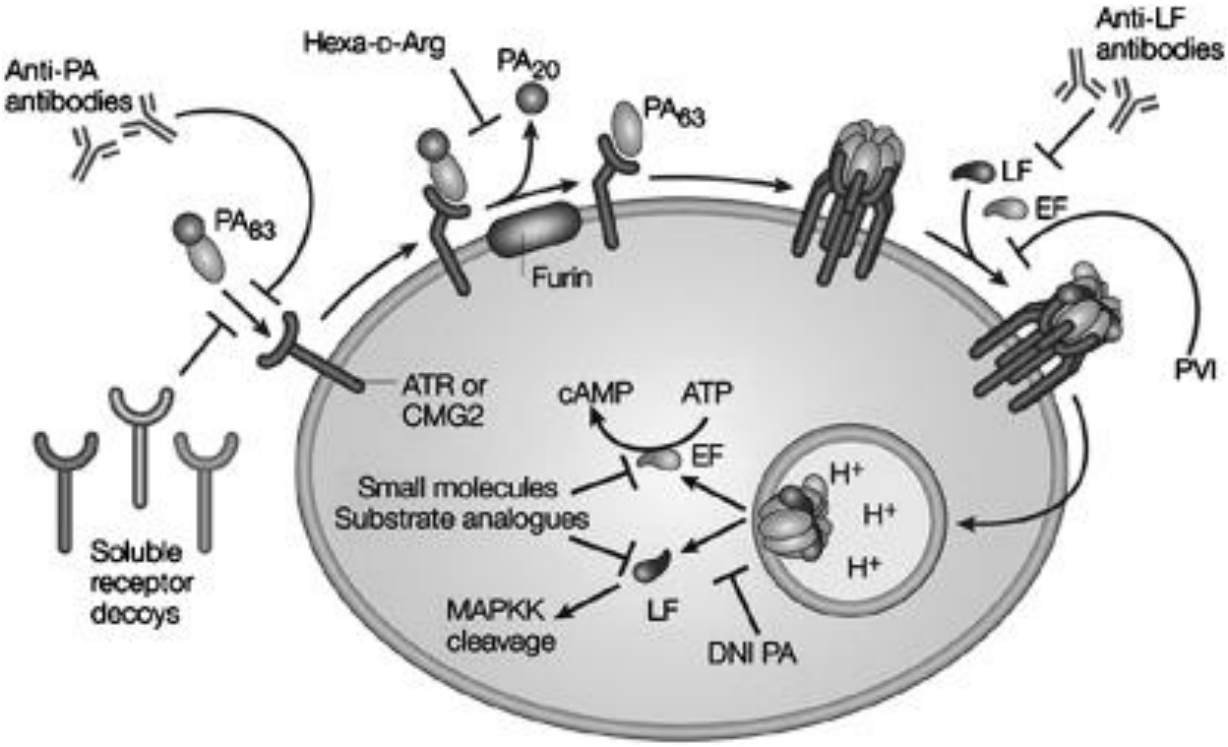


Figure 2.2: Schematic representation of uptake of *Bacillus anthracis* lethal and edema toxins by host cells. The toxin genes are encoded on the pXO1 plasmid. Abbreviations: ATR = anthrax toxin receptor (tumor endothelial marker-8); CMG2 = capillary morphogenesis gene 2; DNI = dominant-negative inhibitor; EF = edema factor; Hexa-d-Arg = hexa-d-arginine; LF = lethal factor; MAPKK = mitogen-activated protein kinase kinase; PA = protective antigen; PVI = polyvalent inhibitor. Image adapted from Sherer *et al.* (2007).

2.3 Diagnosis of anthrax

Microbiological and molecular techniques are used for the detection and diagnosis of *B. anthracis* from different samples. These samples can be isolated from biological fluids (e.g. blood), skin, tissue (e.g. carcass), soil and vegetation and examined using microscopy (Hammond and Diamond, 1984; Spencer, 2003; Riedel, 2005). The Gram-positive bacilli are present as short chains in infected blood or tissues, surrounded by the polypeptide capsule (Spencer, 2003). Blood smears are stained with Giemsa or MacFadyean stain, and visualized under a microscope (Hammond and Diamond, 1984) for the confirmation of *B. anthracis* capsules. This procedure was established in 1903 and is recommended as a highly reliable, rapid diagnostic test for anthrax identification. Anthrax colonies have to be grown on plate media containing 0.7% bicarbonate or 5% serum and incubated in a 5 to 10% carbon dioxide environment to induce the capsules (Riedel, 2005).

Phage lysis and antibiotic (specifically penicillin) susceptibility tests are also recommended for the diagnosis of anthrax, mainly for the confirmation of clinical cases (OIE, 2018). Gamma phage, which is regarded as the standard phage for anthrax diagnosis was initially reported in 1951 (Brown and Cherry, 1951). Even though *B. anthracis* is known to be susceptible to penicillin, not all strains are lysed by the gamma-phage, therefore confirmation with polymerase chain reaction (PCR) is recommended to screen for virulent factors namely PA, LF and the capsule genes (OIE, 2018).

2.4 Global prevalence and epidemiology

Anthrax occurs worldwide, however, reoccurring outbreaks occur in endemic regions where the soil type and environmental factors favour the *B. anthracis* endospore survival. Outbreaks in these endemic regions are associated with alkaline, nutrient (calcium and/or nitrogen) rich soils and changes in seasonal climatic factors such as rainfall and ambient temperatures (Turnbull, 2008; WHO, 2008). Anthrax remains prevalent in developing countries where control measures using vaccination of livestock is poorly implemented. These include central Asia, sub-Saharan Africa, south and central America, certain regions of China, and the Indian sub-continent, with sporadic cases in other areas/countries (WHO, 1996; Schmid and Kaufmann, 2002; Hugh-Jones and Blackburn, 2009). Despite control measures such as vaccination in Africa, occasional outbreaks occur in Botswana, Namibia, Zimbabwe, Uganda and Tanzania (WHO, 2008, Hampson *et al.*, 2011) with sporadic cases reported in South Africa (Viljoen, 1920; Hassim, 2016).

2.5 Anthrax epidemiology in South Africa

The anthrax prevalence peaked in 1923 in South Africa as a result of commercial farming, resulting in 60,000 wildlife and livestock dying from the infection (de Vos, 1990). As mentioned, the prevalence of anthrax was significantly reduced with the implementation of the Sterne vaccine for livestock in 1937 (Sterne, 1967). However, outbreaks and endemics still occur in natural environments such as Kruger National Park (KNP) and Northern Cape province (NCP) and affects wild ungulate herbivores (Hugh-Jones and De Vos, 2002).

In KNP, the spore survival in the northern parts is supported by neutral-to-alkaline soils and the increased calcium levels (Dragon and Rennie, 1995; Smith *et al.*, 2000). Anthrax outbreaks generally occur during dry seasons after extended average rainfall season with the main host being kudu (*Tragelaphus strepsiceros*) (Hugh-Jones and de Vos, 2002). However, recent outbreaks occurred in the wet seasons and predominantly impala (*Aepyceros melampus*) were infected (Basson *et al.*, 2018; Huang *et al.*, 2022). Various factors and vectors play a role in the transmission of *B.*

anthracis during anthrax outbreaks as *B. anthracis* is infectious but non-contagious and can thus not be spread from one animal to the other (WHO, 2008).

In contrast, vaccination of livestock can be used to control anthrax transmission in NCP, as the disease occurs on livestock farms with free roaming wildlife. Thus, the susceptible wildlife can transmit the disease to animals that are not vaccinated. Kudu (*Tragelaphus strepsiceros*) and Zebra (*Equus burchelli*) are the most affected species in NCP (Hugh-Jones and de Vos, 2002). Anthrax is spread by blowflies and louse flies (*Hippobosca rufipes*) in this region (Hugh-Jones and de Vos, 2002). Outbreaks in the Ghaap region of NCP occurred during summer and affected livestock (goats, sheep, horses and cattle), springbok (*Antidorcas marsupialis*), roan antelope (*Hippotragus equinus*) and gemsbok (*Oryx gazella*) (Henton and Briers, 1998). Other reported cases include the 2001 outbreak which affected eland (*Taurotragus oryx*) and the 2007/8 outbreak in antelope and equids which resulted in a great economic loss to game farmers (Hassim *et al.*, 2017).

2.6 Infection and transmission

Three forms of anthrax (cutaneous, gastrointestinal and inhalation) occur with infection (WHO, 2008). In animals, gastrointestinal anthrax which occurs through the ingestion of contaminated vegetation and inhalation of spores while grazing or while drinking at contaminated watering holes are the main forms of the disease in animals (WHO, 2008) (Figure 2.3).

Transmission of infectious spores to susceptible hosts can be through infected *B. anthracis* vegetation or by scavenger vectors from infected carcasses such as flies, vultures, jackal, hyena, etc. (McBride and Turnbull, 1998; Beyer and Turnbull, 2009). Grazing at carcass sites pose risk to *B. anthracis* spore exposure to herbivores. Turner *et al.* (2014) observed that carcass sites alter the nutrient composition of the soil, promoting vegetation growth, thus appealing to some herbivores. These dynamics influence disease transmission and species-specific exposure risks during different outbreaks (Turner *et al.*, 2014). Recent studies showed that transmission varies between carcass sites and water sources (Huang *et al.*, 2022). A high exposure to *B. anthracis* spores while grazing as compared to while drinking was observed (Turner

et al., 2016). The transmission of anthrax through inhalation of *B. anthracis* spores is unlikely as observed by Barandongo *et al.* (2018). The study showed that dust bathing behaviors do not pose a significant risk for herbivores to inhale anthrax (Barandongo *et al.*, 2018) and that it is unlikely that animals will engage in the behaviors at carcass sites.

In KNP, transmission of *B. anthracis* by vectors is well documented. Fly vectors include necrophagic and haemophagic flies such as *Tavanus* and *Stomoxya* biting flies (WHO, 2014) and non-biting flies such as *Chrysomya* (blow flies) (Braack and de Vos, 1990; Hugh-Jones and de Vos, 2002; WHO, 2008). Blowflies contaminate surrounding vegetation near carcass thus infecting animals. In KNP, *Chrysomya albiceps* and *Chrysomya marginalis* feed on infected carcasses and regurgitate *B. anthracis* spores on vegetation (Basson *et al.*, 2018). Susceptible hosts ingest spores while browsing on woody vegetation and this plays an important role in anthrax transmission (de Vos, 1990; Basson *et al.*, 2018) (Figure 2.2). Additionally, vultures are suspected of spreading anthrax from infected carcasses to water sources (dams, ponds, water troughs, and water holes) (Pienaar, 1967; Hugh-Jones and Blackburn, 2009).

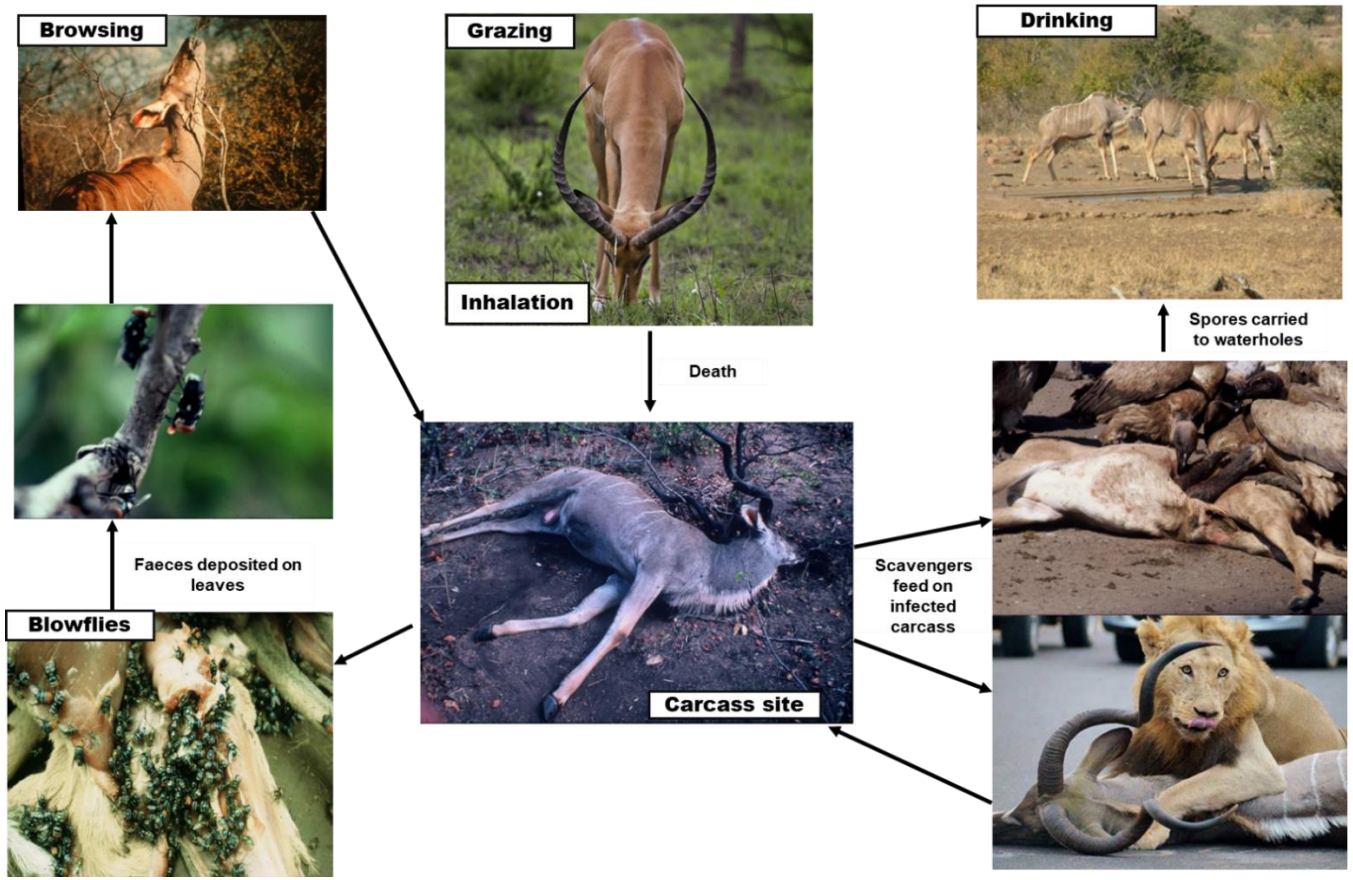


Figure 2.3: *Bacillus anthracis* transmission pathways indicating the source of infection and multiple vectors. The spores are inhaled or ingested by herbivores (impala, kudu etc.) while grazing/browsing and drinking. The infected animal dies at carcass site, releasing infectious spores into the soil. The spores are transmitted when animals (carnivores/scavengers) consume the infected meat. Scavengers (e.g., vultures) carry spores to drinking waterholes where new hosts may be exposed. Blowflies feeding on diseased carcasses contaminate surrounding vegetation near carcass site thus infecting browsing animals (de Vos, 1990).

2.7 *Bacillus anthracis* endospore and germination

The *B. anthracis* endospore is vital for infections in hosts. Endospores have a well-maintained structure and are metabolically inactive (Liu *et al.*, 2004). The structure ensures survival for extended periods in non-conducive environments. The endospore consists of 4 layers; the exosporium, the spore coat, the cortex and the spore core (Figure 2.4). The outermost layer of the spore is made up of lipids, polysaccharides and proteins encasing the spore coat. The exosporium is the primary contact surface between the spore and the environment (Beaman *et al.*, 1972; Turnbull, 1996, Liu *et al.*, 2008).

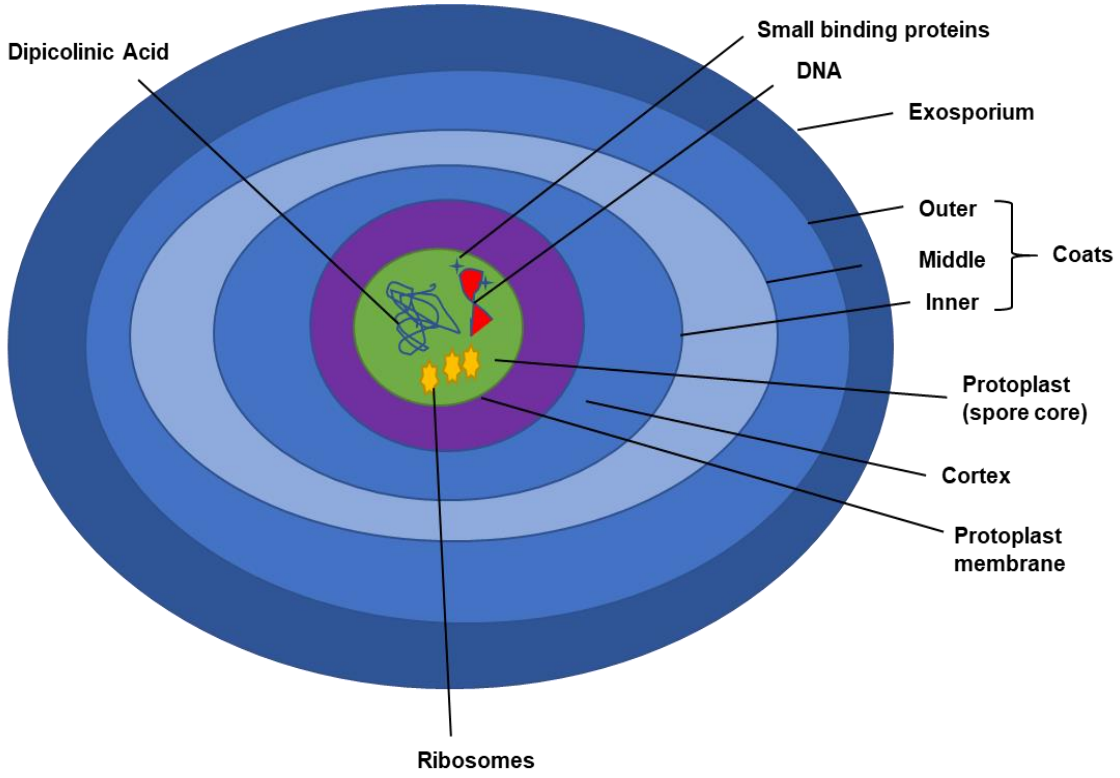


Figure 2.4: Cross section illustration of a bacterial endospore. The core contains the constituents (genomic material) of the future vegetative cell and dipicolinic acid essential for heat resistance. The core is surrounded by a murein peptidoglycan layer (cortex) which is vital for protection against heat and radiation. The wall surrounding the core forms the new vegetative cell during germination and the coats protect the spore from enzymatic and chemical degradation (Adapted from http://creativecommons.org/licenses/by-sa/4.0 and Turnbull, 1996).

The second layer is the double layered spore coat which mainly contains cysteine and hydrophobic amino acids that protect the spore from the harsh environmental conditions. The cortex comprises a loosely arranged peptidoglycan layer, a tightly arranged inner layer and a loosely arranged outer layer made of alanine, tetra-peptide and muramic lactum. The outer layer can be easily hydrolysed during germination. The innermost structure of the spore harbours the genomic material bound to small acid-soluble proteins (Driks, 2009; Swick *et al.*, 2016). The spore core contains excessive amounts of calcium and dipicolinic acid that ensures the spore is resistant to DNA damaging factors, maintaining the dormant phase. The surrounding membrane prevents water diffusion (Setlow, 2003; Knudsen *et al.*, 2016). The local environment provides signals for germination when nutrients are available.

Initiation of bacterial germination is dependent on proper germination or germination receptor interactions (Dixon *et al.*, 1999). For *B. anthracis*, amino acids, which are important sources of soil organics, and purine nucleosides (guanine and adenosine and its primary metabolite inosine) are major nutrient germinants and interact with their specific receptors located in the inner membrane of the spore (Ireland and Hanna, 2002; Weiner *et al.*, 2003; Fisher and Hanna, 2005). Alanine or inosine are typically the primary germinants, with a separate amino acid functioning as a co-germinant. When highly concentrated, L-alanine can trigger germination of spores by itself (Ireland and Hanna, 2002). In contrast, inosine can stimulate a greater germination response as a primary germinant at a diminished concentration as compared to L-alanine. Although inosine is a potent primary germinant, its independent action is not sufficient to germinate *B. anthracis* spores. However, it can be paired with a variety of L-amino acids, including L-histidine, L-serine, L-valine, L-tryptophan and L-methionine, as well as with the primary germinant L-alanine (Weiner *et al.*, 2003).

Germination events are preceded by receptor recognition to their specific germinants. For spore rehydration, dipicolinic acid and its associated calcium ions (Ca-DPA) reserves are released from the spore core, allowing for water to flow back into the core (Moir, 2003). Lytic enzymes activation hydrolysis the peptidoglycan in the spore cortex resulting in further spore rehydration (Setlow *et al.*, 2001). This is followed by the outgrowth phase when RNA, protein and DNA are synthesised, resulting in a

vegetative bacillus. The germination process is quick and occurs within a few minutes under favourable conditions.

When environmental conditions are not optimal, vegetative cells are released from diseased animals in a process called sporulation (Sterne,1967; Turnbull, 2008). This process is influenced by changes in temperature, humidity, pH, availability of nutrients and water (WHO, 2008). Although not well studied, sporulation of the vegetative cells is a sensitive phenomenon that impacts survival in the environment. In their study, Minett (1950) observed that sporulation is inhibited at humidity lower than 60%, however, temperature is the main factor when humidity levels are higher. Temperatures between 8˚C and 45˚C favour multiplication of vegetative cells and persistence of spores (Davies, 1960). Bacillus spores have the ability to survive in various environments. These spores have been isolated in temperature-controlled, low humidity spacecrafts (Venkateswaran *et al.*, 2001; La Duc *et al.*, 2004) and in deserts (Benardini *et al.*, 2003) where there is high ultraviolet radiation.

2.8 Molecular characterization of *B. anthracis*

Bacillus anthracis is a monomorphic organism that makes subtyping of the strains challenging (Achtman, 2008). Discriminatory molecular markers have long been used to characterize and differentiate the genetic lineages of *B. anthracis* (Keim *et al.*, 2000; Pearson *et al.*, 2004; Van Ert *et al.*, 2007a; Jung *et al.*, 2012). The genetic structure consists mainly of three lineages, the ubiquitous A-clade, the rare B-clade and the C-clade with their sub-clades (Figure 2.5). Majority of anthrax outbreaks (90%) are attributed to the A-clade, 10% to the B-clade and less than 1% cluster in the C-clade (Van Ert *et al.*, 2007a).

2.8.1 Multi-locus variable tandem repeats (VNTR) assay (MLVA)

The MLVA is a popular technique used for bacterial DNA genotyping (Keim *et al.*, 2007; Keim *et al.*, 2000; Van Ert *et al.*, 2007a). The analysis enables differentiation of rapidly evolving bacterial species that would otherwise have similar fingerprinting profiles using other techniques. The method relies on PCR amplification of VNTRs

found on the chromosome and plasmids. Variable number of tandem repeats (VNTR) allelic differences resulting from DNA replication generate variation in the number of observed VNTRs in strains of the same species. Amplification of the repeat regions can be analysed by electrophoresis to determine the relationship between bacterial strains. The MLVA however, requires optimized protocols and only common pathogens can be subtyped. The technique has been used successfully to cluster the genetic lineages of *B. anthracis* (Keim *et al.*, 2000; Van Ert *et al.*, 2007a). This assay has a high resolution and could be used to define the global population structure of *B. anthracis* into three clades (A, B and C) in 42 different countries (Van Ert *et al.*, 2007a).

Initial characterization of *B. anthracis* strains employed MLAV8 consisting of VNTR markers targeting variable chromosomal regions vrrA, vrrB1, vrrB2, vrrC1, vrrC2 and CG3 as well as two VNTRs markers located on *B. anthracis* plasmids pXO1 and pXO2 (Keim *et al.*, 1997; Keim *et al.*, 1999). These markers were used to reconstruct the evolutionary lineages, identifying various subpopulations among *B. anthracis* isolates. In their study, Keim *et al.* (2000) successfully used MLVA8 to characterize 426 worldwide *B. anthracis* collections into 89 distinct MLVA genotypes (Figure 2.5), revealing clustering of these strains into 6 dominant genetic clusters (A1, A2, A3, A4, B1 and B2).

The MLVA8 markers, however, could not distinguish strains from closely related geographic locations. Furthermore, 3 major clusters (A, B and C) were identified to resolve the evolutionary relationship of *B. anthracis* strains using MLVA15. Over 200 MLVA genotypes were identified, as compared to the analysis using MLVA8 (Van Ert *et al.*, 2007). MLVA15 provided a higher resolution when coupled with canonical single nucleotide polymorphisms (canSNPs) analysis. Classification of these genotypes using additional 7 VNTR markers revealed that the A lineage is predominant (89.6%) whereas the B and C are rare and are represented by fewer than 15% of the studied isolates (Van Ert *et al.*, 2007a).

Subsequent studies expanded on VNTR analysis to subtype *B. anthracis* isolates using MLVA25 analysis by capillary electrophoresis, leading to the identification of new genotypes (Lista *et al.*, 2006). The genotyping of French and Italian isolates using the 25 VNTR loci identified 69 genotypes. The method was reproducible, had an increased discriminatory power for characterization of *B. anthracis* strains and had the potential for epidemiological surveys (Lista *et al.*, 2006). The combination of MLVA15 and MLVA25 markers by Beyer *et al.,* (2012) to form MLVA31 has become the most powerful discriminatory method for differentiating *B. anthracis* genotypes. MLVA31 was used to characterize 384 *B. anthracis* isolates in the Etosha National Park, Namibia (Beyer and Turnbull, 2009). The study identified over 37 genotypes that clustered into 4 major SNP clusters.

The MVLA31 and SNP analysis provided a better understanding of the evolution of *B. anthracis*, revealing its history and present genotype circulating in the endemic area (Beyer *et al.*, 2012). In addition, sub-typing of French strains also revealed that MLVA31 can provide high resolution for genotyping of *B. anthracis* strains and be used for epidemiological surveillance of *B. anthracis* in different regions (Thierry *et al.*, 2014).

2.8.2 Canonical (can) SNPs

The use of canSNPs using real time PCR or melt analysis of mismatch amplification mutation assay (melt-MAMA assay based on conventional PCR) (Birdsell *et al.*, 2012) has been deployed as a high resolution genotyping method to discriminate whole genome phylogeny into major clonal lineages of *B. anthracis* clades (A, B and C) and sub-clades (Van Ert *et al.*, 2007b; Birdsell *et al.*, 2012). This method employed 12 canSNPs differentiating Ames strains from non-Ames strains (Van Ert *et al.*, 2007b). Although SNP analysis provides some resolution for genotyping, it doesn't have enough discriminatory power compared to MLVA (Lista *et al.*, 2006).

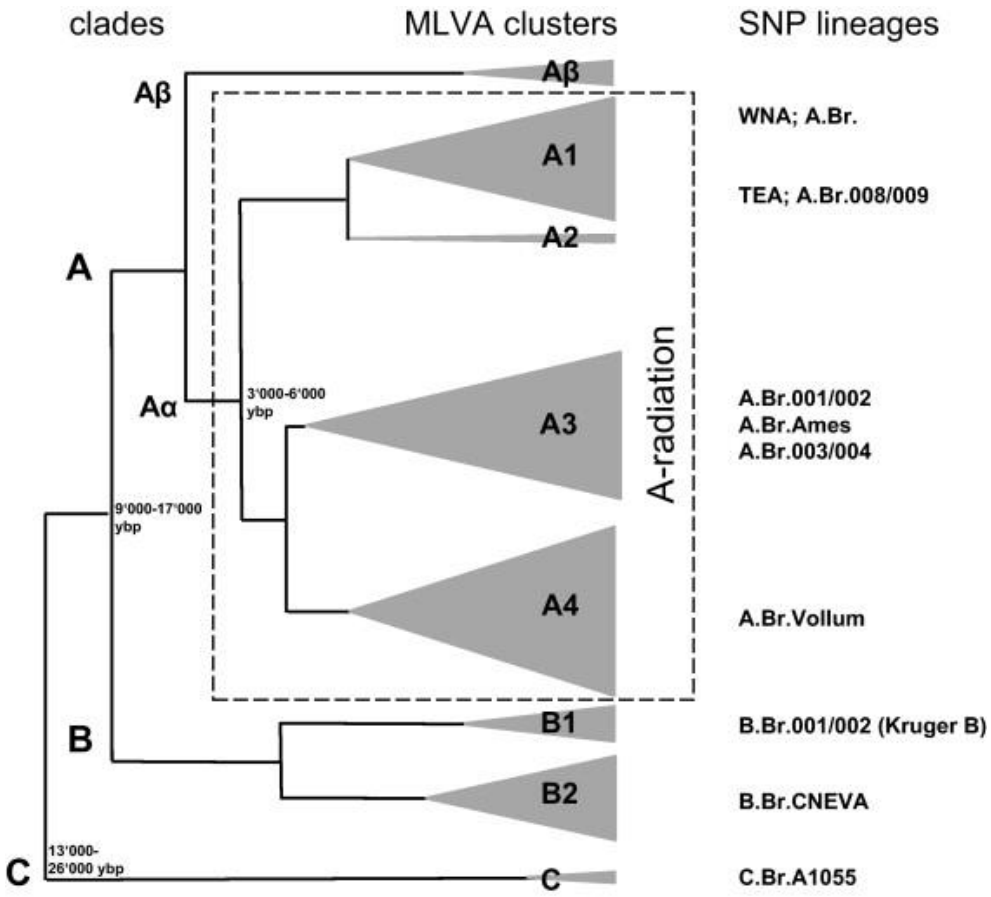


Figure 2.5: Genetic structure of *Bacillus anthracis* populations depicting clustering using multi-locus variable tandem repeats (VNTR) assay (MLVA) and canonical single nucleotide polymorphisms (canSNP) markers. The major subclusters (Aα1, Aα2, Aα3, Aα4, Aβ, B1, B2 and C) based on MLVA are denoted with triangles. The global population structure with collapsed major SNP lineages shown on the right, with the dominant A- clade shown with a dashed boarder. Adapted from (Pilo and Frey, 2011).

2.8.3 Whole genome single nucleotide polymorphism (wgSNP) analysis

Whole genome sequencing (WGS) has been deployed as a powerful tool for routine surveillance of pathogen transmission and for outbreaks tracing. The use of WGS as a genotyping technique has enabled discrimination of the phylogenomic structure of *B. anthracis* using SNP analysis (Pearson *et al.*, 2004; Sahl *et al.*, 2016). Whole genome sequencing is superior to conventional genotyping methods as it generates vast quantities of sequence data for comprehensive genome information (Rasko *et al.*, 2005).

The discovery and analysis of whole genome SNP (wgSNP) data has significantly improved the identification of novel genetic markers that enable studies on phylogenetic relationships among *B. anthracis* isolates. This is due to the availability and the increasing number of genomes deposited in GenBank (Keim *et al.*, 2000; Pearson *et al.*, 2004). The wgSNP analysis defines the global *B. anthracis* clades and sub-clades with higher resolution than canSNP and/or MLVA analysis (Van Ert *et al.*, 2007b; Derzelle *et al.*, 2015; Sahl *et al.*, 2016; Vergnaud *et al.*, 2016; Bruce *et al.*, 2020). .

2.9 South African genetic population structure of *B. anthracis*

South Africa has a high diversity of *B. anthracis* due to the presence of isolates from both A and B clusters (Keim *et al.*, 1999; Smith *et al.*, 1999, 2000; Keim *et al.*, 2000; Van Ert *et al.*, 2007b). Smith *et al.* (2000) showed that both the A- and B-clade strains occurred in the KNP during the 1970-1981 outbreaks, with the high mortality rates attributed to the B-clade strains. Using MLVA31, South African isolates cluster in the A- clade and B- clade. Historical anthrax outbreaks in the enzootic Pafuri region were attributed to both A and B lineages (Figure 1.1), however after the 1990s, only the A-clade strains could be isolated (Keim *et al.*, 2000; Smith *et al.*, 2000; Van Ert *et al.*, 2007b).

Cluster analysis showed that the KNP strains dominated the B1 cluster and are restricted to the region (Keim *et al.*, 2000). Keim *et al.* (2000) identified genotype 87 from the B-clade to be a significant contributor to current anthrax outbreaks in the KNP (Keim *et al.*, 2000). The study also showed that other KNP strains clustered within the A3 cluster which is widely distributed and represented by over 39 genotypes (Keim *et al.*, 2000). Recent studies have isolated the widely distributed A-clade in the KNP. Subtyping of 72 isolates from the 2012-2013 KNP outbreak using MLVA31 revealed clustering of the strains into 23 genotypes (Ledwaba, 2014). MLVA31 identified two dominant genotypes (A1, SNP subclade A.Br.005/009) as the dominant genotypes in the KNP isolated from 2014-2015 outbreaks (Venter, 2016).

Melt-MAMA SNP analysis using 19 markers on 99 smears from KNP isolates clustered the isolates into 2 clades, A.Br.005/006 and Aust94 (Hassim *et al.*, 2017). Genotyping of *B. anthracis* strains isolated from the 1975,1990,2010 and 2012-2015 outbreaks depicts clustering into the A-cluster, placing the strains in genotype 67 using MLVA-8, these isolates could be further separated into 8 different genotypes using MLVA-31 (Hassim *et al.*, 2017; Lekota *et al.*, 2018). The studies indicated diverse *B. anthracis* strains circulation in the KNP.

Whole genome sequencing has been successfully used to characterize the genetic structure of *B. anthracis* isolates from KNP and NCP. Comparative analysis of South Africa isolates to global genomes, had identified SNP branches unique to the South African strains (Lekota, 2018, Lekota *et al.*, 2020). Using isolates from 1990 anthrax outbreaks and previously isolated strains from the 1975 outbreak, phylogenetic analysis revealed clustering of the majority of the KNP isolates into the most prevalent genetic sub-group A.Br.005/006 (A-clade) and A.Br.002 (Sterne) subclades. The isolates from 1975 strains clustered within the B.Br.014 branch (B-clade). The A.Br.005/006 (A.Br.041) major SNP group could be resolved into minor subclades dispersed across the park, showing its wide distribution (Lekota, 2018). WGS-SNP analysis identified over 6000 informative SNPs that defined the phylogeny of South African *B. anthracis* A-clade strains. Novel SNPs that differentiate the B.Br.010 branch from other B-clade branches were also identified (Lekota, 2018).

In NCP, MLVA analysis assigned and defined the population structure of *B. anthracis* isolates into the major A.Br.014 subclade (Hassim, 2016). The phylogenetic structure defined by high throughput sequencing revealed clustering of the isolates into the A.Br.003/004 (A.Br.101) clade. The NCP isolates from the 1998 and 2008/9 outbreaks were placed into the distinct A.Br.101 subclade and clustered with isolates from Mozambique and Zimbabwe (Lekota, 2018; Lekota *et al.*, 2018). The branch is defined by two subbranches, A.Br.172 (1998 isolates) and A.Br.173 (2008/9 isolates) found at the top and bottom of the Ghaap plateau escarpment. The next generation sequencing (NGS) technology used in the study by Lekota *et al.* (2018) revealed higher resolution MLVA. Over time sequencing costs has reduced and resulted in a spike in the sequencing of different organisms (Koboldt *et al.*, 2013).

2.10 Next Generation Sequencing (NGS) and comparative genomics

Since the birth of genomics in 1977 (Sanger *et al.*, 1977) using Sanger sequencing, DNA sequencing approaches have drastically improved to parallel sequencing (Hutchison, 2007; Mardis, 2013) investigating the genome, transcriptome (Adams *et al.*, 1991), and epigenome (Sarda and Hannenhalli, 2014). These methods include WGS, whole exome sequencing (WES) and targeted sequencing. Different sequencing technologies (Table 2.1) can be employed for WGS, and the chosen technology is subject to individual requirements and the biological questions to be addressed. NGS technologies are robust and are constantly being improved. Similar to capillary electrophoresis, the NGS process involves sequential sequencing of nucleotide bases of a small, fragmented DNA/RNA, identifying signals emitted as each fragment is re-synthesised from the provided template strand (Shendure and Ji, 2008).

The basic principle behind NGS involves DNA or RNA fragmentation, adapter ligation, library sequencing and the assembling of genomic sequence data (Head *et al.*, 2014). The process is rapid, employs parallel sequencing, producing enormous high-throughput data (Koboldt *et al.*, 2013; Mardis, 2013). Sequencing using next generation technologies has not only revolutionized biological research but has also become more affordable and accessible (Buermans and den Dunnen, 2014; Head *et al.*, 2014). The produced data is stored in a particular format and analysed using bioinformatics software and optimized programs or workflows.

Table 2.1:Frequently used Next Generation Sequencing technologies

Sequencer	**454-GS FLX**	**Illumina MiSeq/ NextSeq**	**Ion Torrent**	**PacBio**	**Oxford Nanopore**
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Semiconductor sequencing	Single-molecule real-time (SMRT) sequencing	Sequencing by synthesis
Read length	700 bp	150-300bp Paired-end	200 bp	2 900 bp	500-2.3 mb
Output	0.7 Gb	15-600 Gb	25 Gb	100-150 Gb	50 Gb
Advantage	Read length, fast	High throughput	High throughput and accuracy	Long read length	Ultra-long reads
Disadvantage	Costly, low throughput	High substitution error rate	Short read length	Low throughput, low accuracy	High error rate
References	(Segerman, 2020)	(Su *et al.*, 2011; Lahens *et al.*, 2017)	(Lahens *et al.*, 2017)	(Rhoads and Au, 2015; Allali *et al.*, 2017)	(Segerman, 2020)

2.11 Genomic analysis of NGS data using bioinformatics tools

The analysis of NGS data can be intensive. With the advancement of sequencing technologies, analysis programs and workflows have been developed to handle the enormous data. For all NGS data analysis workflows, quality control is a pivotal starting point to ensure that the data is of good standard for downstream analysis. Various computer software are available for analysis of large NGS data. These include genome assembly and mapping software and programs desired for annotation of draft assembled genomes.

2.11.1 Genome mapping/assembly and annotation

Mapping refers to the alignment of sequence reads to a known or closely related genome. During mapping, short paired-end reads are assembled in comparison to a reference genome. The comparison can only be made between genomes that have a 90% similarity (Hardison, 2003; Li and Durbin, 2009). Reference mapping approaches are limited by sequence divergence between the genome being assembled and the reference (Li and Durbin, 2009). In contrast, *de novo* assembly is used when there is no reference genome available for comparison (Miller *et al.*, 2010). This type of assembly builds up genomes from short DNA fragments into contigs (Flicek and Birney, 2009; Miller *et al.*, 2010). The over-lapping contigs are then joined to generate scaffolds. These scaffolds are then combined to form a draft genome that can be annotated based on reference genomes (Abril and Castellano, 2019; Salzberg, 2019). Genome annotation employs gene prediction programs to identify related gene sequences. Automated annotation pipelines are available to perform the gene searches to predict arrangement and functional coding sequences (Overbeek *et al.*, 2005; Aziz *et al.*, 2008; Salzberg, 2019).

2.12 The Pan-genome analysis

An organism's pan-genome refers to all genes that make up a species (Rouli *et al.*, 2014). The pan-genome defines the structure of prokaryotic systems and includes the core genes and accessory genes that are variable among strains (Medini *et al.*, 2005; Tettelin *et al.*, 2005, Rouli *et al.*, 2014). The core genome is comprised of all genes responsible for the fundamental biology of a species and its major phenotypic characteristics. In contrast, the accessory genome represents all genes acquired through horizontal gene transfer that contribute to species diversity and the adaptation of some strains to occupy and thrive in specific niches (Medini *et al.*, 2005).

The ability of some species to acquire and exchange genetic material or inhabit variable niches means they have an open pan-genome whereas species that only colonize specific niches have closed pan-genomes (Medini *et al.*, 2005; Inglin *et al.*, 2018). Next generation sequencing has shed light on strain-to-strain variability, providing more information on the organism's evolutionary history. Instead of a single gene approach, pan-genomic analysis focuses on various regions of the genome, moreover on the evolutionary history of organisms' genomes.

The pan-genome of various bacterial species has been studied, defining their phylogenomic structure, virulence profiles and the presence or absence of orthologous gene groups (Zou *et al.*, 2014; Laing *et al.*, 2017; Luo *et al.*, 2018; Wu *et al.*, 2018; Blaustein *et al.*, 2019; Sinha *et al.*, 2021). Comparative analysis has shown that the pan-genome can identify strain specific genes and separate closely related bacterial lineages. The specific genes in *Streptococcus agalactiae* (Tettelin *et al.*, 2005), *Salmonella* spp. (Zou *et al.*, 2014; Laing *et al.*, 2017), *Escherichia coli* (Rasko *et al.*, 2008) provided insight into the evolution of the lineage and their functionality in pathogenesis.

2.13 Characterization of genetic differences in *Bacillus anthracis*

2.13.1 The tryptophan operon

Bacterial communities typically inhabit an ever-changing environment in which nutrient availability may increase or decrease dramatically. To respond to the fluctuations, bacteria modify their gene expression pattern, expressing different enzymes as required. The regulation of biosynthetic pathways is vital for the survival or propagation of bacteria; thus, control mechanisms are tightly controlled (Yanofsky, 2013). Variability in the organization of pathway genes within operons has been observed for groups of organisms, suggesting exposure to varying environmental demands. The ability to synthesize L-tryptophan is one of the requirements for survival and replication of most bacteria, however this process is energy taxing. A tryptophan biosynthesis pathway consists of five chemical reactions that are mostly conserved across microbial genomes, however, the genes of the pathway enzymes differ greatly in their arrangement, operon structure (gene fusion and splits, for example) and regulation (Merino *et al.*, 2008; Priya *et al.*, 2014; Eremenko *et al.*, 2020). Differences in gene and operon organization reflect evolutionary divergence and as well as adaptability in various environments (Eremenko *al.*, 2020). The alignment and phylogenetic analysis of different *B. anthracis* strains shows variability of single nucleotide polymorphism and InDels in tryptophan operon genes also demonstrated grouping to different clades. Phylogenetic reconstruction of tryptophan operon genes amongst *B. anthracis* strains showed that there were differences between A and B lineages.

The tryptophan synthase subunit alpha (*trpA*) gene codes for the a-subunit of tryptophan synthase, which converts indole-3-glycerol phosphate to tryptophan. The distinctions in *trpA* gene sequences in *B. anthracis* have been noted and distinguish the B-clade strains from lineages A and C (Eremenko *et al.*, 2020). Strains from lineage A have 19 additional amino acids VSLFFLFCVINVKIYRKYI at C-end of the polypeptide and a G258A substitution. In contrast, there is a mutation in the *trpA* gene in B lineage strains. The mutation results into a non-coding pseudogene which could account for the lineage B

strains' apparent tryptophan dependency (Eremenko *et al.*, 2020). The inability of the *B. anthracis* B-lineage strains to synthesize tryptophan, will have detrimental effects on their survival in the environment (Merino *et al.*, 2008). Studies on tryptophan dependence can provide information on intraspecific evolution of *B. anthracis*.

2.13.2 The *Bacillus* collagen-like protein A (BclA)

The *Bacillus* collagen-like protein A (BclA) is a dominant surface protein on the exosporium, the outermost layer on the surface of *B. anthracis* spores (Sylvestre *et al.,* 2003*)*. The BclA glycoprotein plays a vital structural role in the spore coat, and aids entry to host cells. The protein serves as the site of host recognition by the phagocytosis complement system and initiates carriage of spores across the epithelium (Le Flèche *et al.*, 2001; Sylvestre *et al.*, 2003). The structure and function of the BclA gene influences spore survival and persistence in the environment (Liu *et al.*, 2004).

The exosporium is the outermost surrounding of the *B. anthracis* endospore (Sylvestre *et al.*, 2002, 2003). The Bacillus collagen-like protein of anthracis (BclA*)* glycoprotein is the major component of the exosporium hair-like filaments. The BclA glycoprotein encode for internal collagen-like region (CLR) with a number of continuous GXX collagen-like repeats, including a large proportion of GTP triplets. The GXX repeats varies considerably between the stains and this variation is responsible for the length variation of the filament nap covering the outer layer of the exosporium (Sylvestre *et al.*, 2003). The Bams13 marker is located within the open reading frame that encodes BclA glycoprotein. The Bams13 marker in combination with other VNTR markers Bams15, Bams30 and Bams31 (Lista *et al.*, 2006) can be used for subtyping of *B. anthracis* strains.

Profiling of bacteria including *B. anthracis* using variable number tandem repeat (VNTR) has revealed the presence of the number of repeats and variability between species. Repeat motif vary in sizes and range between single to 100 repeats (Le Flèche *et al.*, 2001). When the VNTR spans across a coding region, it can influence the species' ability

to respond to environmental perturbations and evolution. The VNTR repeat size, however, has no effect on the translation open reading frame (ORF) but changes the amino acid sequence. In particular, Bams13 used for genotyping *B. anthracis* strains exhibits the greatest variability in fragment size among VNTRs and influence phenotypic presentation on the endospore surface (Sylvestre *et al.*, 2003). The Bams13 VNTR locus contained a region of the *bclA* gene, and repeat length is directly linked to length of filaments on the exosporium surface (Thompson *et al.*, 2007). Therefore, Bams13 is a highly polymorphic marker and contains the DNA encoding collage like region (CLR) of BclA (Lista *et al.,* 2006). Differences in the GXX amino acid repeats of the CLR region attribute to the variability in the *bclA* gene sizes between lineages. Assessment of VNTR variations carried on the *bclA* gene can shed light on the pathogen fitness and transmission (Sylvestre *et al.*, 2003).

Despite *bclA*'s crucial role in uptake and spore structure, this gene shows very high sequence diversity, even at local/regional scales (Sylvestre *et al.*, 2003). These phenotypic differences may alter rates of survival or virulence among genetic variants, thus affecting disease transmission rates and pathogen fitness. Variable Number Tandem Repeat (VNTR) analysis can be successfully used as a genotyping method to discriminate *B. anthracis* strains and may provide an understanding of spore survival in response to environmental changes (Sylvestre *et al.*, 2003).

Chapter 3: Materials and methods

3.1. *B. anthracis* culture collection

Bacillus anthracis isolates (n=83) collected during the 2012-2015 Kruger National Park (KNP) anthrax outbreak and sequences of strains (n=19) from the 1975-2011 outbreaks were used in this study (Table 3.1). The *B. anthracis* strains were isolated from soil (n=9),impala (*Aepyceros melampus,* n=42), African elephant (*Loxodonta africana,* n=5), hippopotamus (*Hippopotamus amphibious*, n=4), Burchell's zebra (*Equus quagga burchellii*, n=5), greater kudu (*Tragelaphus strepsiceros*, n=9), nyala (*Tragelaphus angasii*, n=4), roan antelope (*Hippotragus equinus*, n=11), buffalo (Syncerus caffer, n=3) vulture faeces (n=2), white rhinoceros (*Ceratotherium simum*, n=4), blue wildebeest (*Connochaetes taurinus*, n=1) carcasses as well as environmental samples (n=2) (Table 3.1).

DNA extraction and PCR confirmation of *B. anthracis*

B. anthracis strains were inoculated in 2 ml nutrient broth, incubated at 37 ^{o}C overnight and cells harvested by centrifugation. Genomic DNA of *B. anthracis* strains (n=80) were extracted from the harvested cells using DNA Blood Mini Kit (Qiagen) following manufacturer's protocol for Gram-positive bacteria with 20 mg/mL lysozyme (Sigma Aldrich). The DNA was quantified on the qubit fluorometric quantization systemusing the Broad Range assay kit (Invitrogen™). The quality of the DNA was analysed by electrophoresis on 0.8% agarose gel using ethidium bromide and visualized under UV-light.

All isolates were confirmed to contain *B. anthracis* virulence factors using fluorescence resonance energy transfer quantitative PCRs (FRET qPCRs) targeting the virulence genes on plasmids, namely *bapa* (*Bacillus anthracis* protective antigen) and *cap*C (capsule), along with chromosomal *sasp* (small acid soluble proteins) genes, as described by Ellerbrok *et al.*, (2002).

Table 3.1: Metadata of *Bacillus anthracis* strains from Kruger National Park used in this study

Strain	Location in the KNP	Source	Year
[1]A3 #	Pafuri	Soil	1975
[1]A5 #	Pafuri	Soil	1975
[1]A8 #	Pafuri	Soil	1975
[1]A11#	Pafuri	Soil	1975
[1]A16 #	Pafuri	Soil	1975
[1]A19 #	Pafuri	Soil	1975
[1]C13 #	Pafuri	Soil	1975
[1]HP8 #	Pafuri	Soil	1975
[1]HP12 #	Pafuri	Soil	1975
[1]Z21 #	Pafuri	Kudu	1975
B42	Pafuri	Buffalo	1990
[1]419 #	Pafuri	Buffalo	1990
[1]AH26 #	Punda Maria, Dzundzwini	Nyala	2010
AH97 #	Pafuri, Masengane Pan	Zebra	2010
[1]AH77 #	Punda Maria, Xitshova	Kudu	2010
JO309 (AH14)	Shingwedzi	Buffalo	2010
[1]AD201016B (KNP1) * #	Pafuri	Impala (Bone)	2010
[1]AD201016b (KNP3) * #	Pafuri	Impala (Blood)	2010
[1]KC2011 #	Hoedspruit	Cheetah	2011
[1]AX2012435 #	Pafuri	Elephant	2012
[1]AX2012291 #	Pafuri, Woodlands	White Rhinoceros	2012
[1]AX2012283 #	Houtboschrand	White Rhinoceros	2012
[1]AX2012275 #	Houtboschrand	White Rhinoceros	2012
AX2012309	Satara	White Rhinoceros	2012
AX2012285	Houtboschrand	Elephant	2012
OR121027H2	Sabie sands	Hippopotumus	2012
LVS2012028 #	Shingwedzi	Kudu	2012
LS030v (KNP2) *	Mooiplaas	Environmental (Blowfly at carcass)	2012
LVS2012007Ns (KNP25) *	Mooiplaas	Kudu	2012
10201209 (KNP26) *	Mooiplaas	Roan	2012
LVS201209 (KNP27) *	Mooiplaas	Nyala	2012
OR121026K2 (KNP29) *	Letaba	Kudu	2012
OR121923H2*	Sabie sands	Hippopotumus	2012
32201209 (KNP30) *	Mooiplaas	Roan	2012
3201209 (KNP32) *	Mooiplaas	Roan	2012
18201209 (KNP33) *	Mooiplaas	Roan	2012
4201209 (KNP35) *	Mooiplaas	Roan	2012
23201209(KNP38) *	Mooiplaas	Roan	2012
13201209 (KNP40) *	Mooiplaas	Roan	2012
22201209 (KNP41) *	Mooiplaas	Roan	2012
LVS2012015 (KNP42) *	Mooiplaas	Vulture faeces	2012
LVS2012017 (KNP43) *	Phalaborwa	Kudu	2012
20201209BTA (KNP46) *	Mooiplaas	Roan	2012
14201209 (KNP48) *	Mooiplaas	Roan	2012
LVS201229 (KNP49) *	Mooiplaas	Environmental (Grass at carcass site)	2012
LVS2012001 (KNP50) *	Mooiplaas	Kudu	2012
342012001 (KNP51) *	Mooiplaas	Roan	2012
LVS2012022 (KNP52) *	Mooiplaas	Kudu	2012
OH121029K (KNP53) *	Letaba	Kudu	2012
BF2013	Letaba	Hippopotumus	2013
ET2013	Olifants	Elephant tusk	2013

Strain	Location in the KNP	Source	Year
[1]**HT2013** [#]	Letaba	Hippopotumus tusk	2013
DS201334 (KNP28)	Pafuri	Burchell's zebra	2013
***DS201333* (KNP31) ***	Pafuri	Impala	2013
DS201326 (KNP34) *	Pafuri	Impala	2013
DS201328 (KNP36) *	Pafuri	Impala	2013
DS201317 (KNP37) *	Pafuri	Nyala	2013
DS201316 (KNP39) *	Pafuri	Impala	2013
DS201325 (KNP44) *	Pafuri	Impala	2013
DS201332 (KNP45) *	Pafuri	Impala	2013
DS201318 (KNP47) *	Pafuri	Nyala	2013
SVD201453*	Pafuri	Impala milk	2014
RL2014115RbM (KNP13) * [#]	Pafuri	Impala	2014
KM2015130315P (KNP9) *	Punda Maria	Elephant	2015
[1]**KM20150330 ***	Pafuri	Elephant tusk	2015
[1]**DS201579 *** [#]	Pafuri	Blue Wildebeest	2015
RL201521 *	Pafuri,	Vulture feaces	2015
RL201528 (KNP4) *	Pafuri	Impala	2015
[1]**DS201577 (KNP5) *** [#]	Pafuri	Impala	2015
[1]**1257032015UP (KNP6) *** [#]	Pafuri	Impala	2015
DS201513 (KNP7) *	Pafuri	Impala	2015
Rbk32 (KNP8) *	Pafuri	Impala	2015
[1]**1298*** [#]	Pafuri, Bobomeni	Impala	2015
DS20150962 (KNP10) *	Pafuri	Impala	2015
DS20150457 (KNP11) *	Pafuri	Impala	2015
[1]**13680315 (KNP12) *** [#]	Pafuri	Impala	2015
[1]**13030315 (KNP14) *** [#]	Pafuri	Impala	2015
13610315 (KNP15) *	Pafuri	Impala	2015
[1]**DS20150761 (KNP16) *** [#]	Pafuri	Impala	2015
13670315 (KNP17) *	Pafuri	Impala	2015
13020315 (KNP18) *	Pafuri	Impala	2015
[1]**DS20150663 (KNP20) *** [#]	Pafuri	Impala	2015
[1]**DS20152454 (KNP21) *** [#]	Pafuri	Impala	2015
1367PETA (KNP24) * [#]	Pafuri	Impala	2015
1249	Pafuri	Impala	2015
[1]**1365*** [#]	Pafuri	Impala	2015
[1]**1370*** [#]	Pafuri	Impala	2015
[1]**DS201505*** [#]	Pafuri	Impala	2015
[1]**DS201510*** [#]	Pafuri	Impala	2015
DS201513	Pafuri	Impala	2015
DS201515*	Pafuri	Impala	2015
[1]**DS201523*** [#]	Pafuri	Impala	2015
DS201525* [#]	Pafuri	Impala	2015
[1]**DS201578*** [#]	Pafuri	Impala	2015
DS201584* [#]	Pafuri	Impala	2015
[1]**DS201588*** [#]	Pafuri	Impala	2015
RL201528*	Pafuri	Impala	2015
RL201530*	Pafuri	Impala	2015
RL201531* [#]	Pafuri	Impala	2015
12950315 (KNP22) *	Pafuri	Burchell's zebra	2015
[1]**12460315 (KNP23) *** [#]	Pafuri	Burchell's zebra	2015
[1]**12610315 (KNP19) *** [#]	Pafuri	Burchell's zebra	2015

***Samples sequenced in this study, samples in italics were not *B. anthracis* and were removed after mapping. [#]Samples used for pan-genome analysis; [1]Samples used for tryptophan (trp) genes analysis.

3.2 High-throughput sequencing and quality assessment

DNA extracts of the *B. anthracis* strains (n=80) were processed and libraries were prepared using the Nextera XT DNA Sample Prep kit (Illumina). Sequencing of paired end library was performed on the Illumina MiSeq and HiSeq 2500 sequencer using the 200-cycle SBS (sequencing by synthesis) sequencing v4 kit (Illumina). Genome quality was assessed using the FastQC software 0:10.1 (Andrews, 2010) and the sequence adapters and ambiguous nucleotides were removed using Trimommatic (v0.32).

3.3 Read mapping and single nucleotide polymorphism (SNP) variant detection

The trimmed reads of the *B. anthracis* strains (n=83) were aligned to the *B. anthracis* Ames ancestor reference genome (GenBank: AE017334.2) using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). Sequence reads of *B. anthracis* global genomes were downloaded and retrieved using the SRA-tool kit (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software). These included *B. anthracis* complete and draft genomes (n= 109, Table 3.2) from different canonical clades available from NCBI Genbank (http://www.ncbi.nom.nih.gov). SAMtools v1.10 (Li *et al.*, 2009) was deployed to convert the aligned mapped reads to bam files and to sort and index the aligned sequenced reads. Picard-tools (http://picard.sourceforge.net/) was used to generate a sequence dictionary of the *B. anthracis* Ames ancestor, to mark duplicate reads and to build bam index of the mapped reads. SNPs were called using the UnifiedGenotyper method in GATK v4.0.12.0 (McKenna *et al.*, 2010; Depristo *et al.*, 2011; Van der Auwera *et al.*, 2013) and ambiguous variants were filtered prior to selecting variants.. SNPs positioning sets were deducted from the aligned genomes of *B. anthracis* Ames ancestor using MEGAX (molecular evolutionary genetics analysis software (Kumar, Stecher and Tamura, 2016). SNPs with informative sites (core SNPs) in all genome sequences were used for the phylogenetic tree construction using MEGAX (Kumar, Stecher and Tamura, 2016) and/or Bayesian Evolutionary Analysis by Sampling Trees (BEAST) v2 (Suchard *et al.*, 2018).

Table 3.2: *Bacillus anthracis* whole genomes from NCBI Genbank used in this study.

Strain	Country	Major canSNP	Bruce *et al.* 2020 cluster	Accession Number	Year
Ames ancestor	United States of America	A.Br.001 (Ames)	5.2	NC_007530.2	Unknown
[1]**Shikan-NIID**	Japan, Tokyo	A.Br.001 (Ames)		DRR150094	1928
[1]**Sterne** [#]	Unknown	A.Br.002 (Sterne)	5.2	AE017225.1	Unknown
Tokushima1 [#]	Japan	A.Br.002 (Sterne)		PRJDB3126; SAMD00115654	1984
3080_1B [#]	South Africa, NCP[a]	A.Br.002 (Sterne)		SRP227303; SAMN13151848; SRR10357977	2009
ANSES_08-8_20[#]	France	A.Br.002 (Sterne)	6.1	JHCB02000000.1	2008
2000032979	United States of America	A.Br.002 (Sterne)		JTAE01000000.1	Unknown
2000032879	United States of America	A.Br.002 (Sterne)		JSZY01000000.1	1962
[1]**BFV** [#]	Jamaica	A.Br.002 (Sterne)		CP007704.1	Unknown
A0252	Zimbabwe	A.Br.002 (Sterne)		PRJNA302749; SAMN04283840; SRS1185179	Unknown
[1]**Delta sterne**	Unknown	A.Br.002 (Sterne)		CP008752.1; PRJNA243519; SAMN02736981	1930
FT2012 [#]	South Africa, Letaba	A.Br.002 (Sterne)			2012
[1]**2000031075** [#]	South Africa			PRJNA264742; SAMN03165118	Unknown
52-G [#]	Country of Georgia	A.Br.003/004 (Aust94)		AZUF00000000.1	1998
8903_G [#]	Country of Georgia	A.Br.003/004 (Aust94)		AZUD00000000.1	1997
9080_G [#]	Country of Georgia	A.Br.003/004 (Aust94)		AZUE00000000.1	1998
[1]**A.Br.003** [#]	Scotland	A.Br.003/004 (Aust94)		JMPV00000000.1	Unknown
Aust94 [#]	Australia	A.Br.003/004 (Aust94)	5.3	GCA_000167335.1	1994
[1]**K1**	Namibia	A.Br.003/004 (Aust94)		PRJNA281298; SAMN03486970	2014
[1]**K2**	Namibia	A.Br.003/004 (Aust94)		SAMN03486971	2014
2110 [#]	South Africa, NCP	A.Br.003/004 (Aust94)	5.1	PRJNA510736; SAMN10614342	1998
[1]**V770 –NP-1R** [#]	United States of America	A.Br.003/004 (V770)	5.1	AZQO00000000	1951
2000031039	United States of America	A.Br.003/004 (V770)		JSZR01000000.1	1957
[1]**K8215** [#]	Argentina	A.Br.003/004 (V770)		LGIG01000000.1	1996
[1]**A1039**	Bolivia	A.Br.003/004 (V770)		LAKZ01000000.1	1999
ATCC14185	Israel	A.Br.003/004 (V770)		AZQO00000000.1	
A0094	South Africa, NCP	A.Br.003/004 (V770)	5.1	SRR2968135; PRJNA302749; SAMN04283813	1949
A0096	South Africa, NCP	A.Br.003/004 (V770)	5.1	SRR2968191; PRJNA302749; SAMN04283814	1939
A0097	South Africa, NCP	A.Br.003/004 (V770)	5.1	SRR2968192; PRJNA302749; SAMN04283815	1938

Strain	Country	Major canSNP	Bruce *et al.* 2020 cluster	Accession Number	Year
20SD [#]	South Africa, Mpumalanga	A.Br.003/004 (V770)		LGCD00000000	2011
SA047	South Africa	A.Br.003/004 (V770)		PRJNA309927; SAMN08812797	1999
2000031023	Unknown	A.Br.003/004 (V770)		PRJNA264742; SRR5811189	1957
BA1015	USA, Maryland	A.Br.003/004 (V770)	5.1	SRR2175366	1939
[1]**2002734065**	United Kingdom	A.Br.003/004 (V770)		PRJNA264742; SRR5811123	1975
[1]**K1285**	Namibia	A.Br.100 (Aust94)	5.4	SRR2071843	1996
A3716	Namibia	A.Br.014 (Aust94)		SRS1185190	2006
3631_1C [#]	South Africa, NCP	A.Br.101 (A.Br.014)		LGCC00000000	2009
[1]**2991_1B** [#]	South Africa, NCP	A.Br.101 (A.Br.014)		RXZV00000000	2009
[1]**3008_1B** [#]	South Africa, NCP	A.Br.101 (A.Br.014)		RXZU00000000	2009
[1]**3122_2B** [#]	South Africa, NCP	A.Br.101 (A.Br.014)		RXZT00000000	2009
[1]**3517_2C** [#]	South Africa, NCP	A.Br.101 (A.Br.014)		RXZP00000000	2009
3618_2D	South Africa, NCP	A.Br.014 (Aust94))		SAMN10614346	2009
[1]**3275_2D** [#]	South Africa, NCP	A.Br.101 (A.Br.014)		RXZR00000000	2009
[1]**3132_1B**	South Africa, NCP	A.Br.101 (A.Br.014)		RXZS00000000	2009
3631_8D [#]	South Africa, NCP	A.Br.003/004 (Aust94)	5.1	PRJNA510736; SAMN10614341	2009
[1]**3631_3D**	South Africa, NCP	A.Br.101 (A.Br.014)		PRJNA510736; SAMN10614340	2009
[1]**3631_4C** [#]	South Africa, NCP	A.Br.101 (A.Br.014)		RXZO00000000	2009
[1]**3631_8D** [#]	South Africa, NCP	A.Br.101 (A.Br.014)		RXZM00000000	2009
JB10	South Africa, NCP	A.Br.101 (A.Br.014)		RXZK00000000	2009
JB25	South Africa, NCP	A.Br.101 (A.Br.014)		SDEF00000000	2009
3631_7C	South Africa, NCP	A.Br.101 (A.Br.014)		SRP227303; SAMN13151842; SRR10357981	2009
5838	South Africa, NCP	A.Br.101 (A.Br.014)		SRP227303; SAMN13151843; SRR10357980	2009
[1]**2991_2B**	South Africa, NCP	A.Br.101 (A.Br.014)		SRP227303; SAMN13151844; SRR10357985	2009
3080_3B	South Africa, NCP	A.Br.101 (A.Br.014)		SRP227303; SAMN13151845; SRR10357983	2009
3079_1C	South Africa, NCP	A.Br.101 (A.Br.014)		SRP227303; SAMN13151846; SRR10357984	2009
3090_1B	South Africa, NCP	A.Br.101 (A.Br.014)		SRP228283; SAMN10614343; SRR10390628	2009
5838	South Africa, NCP	A.Br.104 (Aust94)		PRJNA580142; SAMN13151843	1998
BA_3154	Bulgaria	A.Br.009/011 (WNA/TEA)		ANFF00000000.1	1960

Strain	Country	Major canSNP	Bruce *et al.* 2020 cluster	Accession Number	Year
Sen2Col2	Senegal	A.Br.0011/009		CAVC000000000.1	2010
[1]H9401	South Korea	A.Br.H9401		NC_017729.1	1994
A0455	Mozambique			SRS1185174	Unknown
6461_SP2	South Africa, NCP	A.Br.005/006 (Ancient A)		SRP227303; SAMN13151840; SRR10357978	
6102_6B	Botswana	A.Br.005/006 (Ancient A)		SRP227303; SAMN13151841; SRR10357979	
A2075	Tanzania	A.Br.005/006 (Ancient A)	3.2	SRR2968187; PRJNA3-2749; SAMN04283799	1999
A2079	Tanzania	A.Br.005/006 (Ancient A)	3.1	SRR2968188; PRJNA302749, SAMN04283798	1999
A0530	Botswana	A.Br.005/006 (Ancient A)	3.3	SRR2968170; PRJNA302749; SAMN04283802	Unknown
A0135	Albania	A.Br.005/006 (Ancient A)	6.1	SRR2968140; PRJNA302749; SAMN04283796	Unknown
A0026	England	A.Br.005/006 (Ancient A)	3.1	SRR2968152; PRJNA302749; SAMN04283797	1992
A0017	Zambia	A.Br.005/006 (Ancient A)	3.2	SRR2968151; PRJNA302749; SAMN04283800	Unknown
A0533 (6461)	South Africa, NCP	A.Br.005/006 (Ancient A)		SRR2968171; PRJNA302749; SAMN04283801	Unknown
A0128	South Africa, NCP	A.Br.005/006 (Ancient A)		SRR2968156, PRJNA302749, SAMN04283804	Unknown
[1]CZC5	Zambia	A.Br.005/006 (Ancient A)	3.2	DRR014735, DRX013255, DRP002748	2011
[1]A0021	Zambia	A.Br.005/006 (Ancient A)		PRJNA257008; SAMN03862113	1997
Zambia_36	Zambia	A.Br.005/006 (Ancient A)	3.2	DRR125654	2012
Zambia_37	Zambia	A.Br.005/006 (Ancient A)	3.3	DRR125656	2013
Zambia_38	Zambia	A.Br.005/006 (Ancient A)		PRJDB1571; DRR014741	2013
Zambia_39	Zambia	A.Br.005/006 (Ancient A)	3.3	PRJDB1571; DRR014742	2013
Zambia_40	Zambia	A.Br.005/006 (Ancient A)	3.3	PRJDB1571; DRR125653	2012
Zambia_41	Zambia	A.Br.005/006 (Ancient A)	3.3	PRJDB1571; DRR125655	2013
[1]Vollum	United Kingdom	A.Br.007 (Vollum)	4	AAEP00000000.1	Unknown
CDC684	Unknown	A.Br.007 (Vollum)		NC_012581.1	1964
[1]Vollum	United Kingdom	A.Br.007 (Vollum)	4	AAEP00000000.1	Unknown
[1]BA0008	Italy, Sicily	A.Br.007 (Vollum)		PRJNA656733; SRR12435811	1997
[1]COVASU	India, Maharashtra	A.Br.007(Vollum)	4.2	SRR6037789	2017
A0442	South Africa, KNP	B.Br.001/002 (B-branch)		ABKG00000000.1	Unknown
[1]A0091	South Africa	B.Br.001/002	2.2	SRR2968134; PRJNA302749; SAMN04283795	1939

Strain	Country	Major canSNP	Bruce *et al.* 2020 cluster	Accession Number	Year
SVA11	Sweden	B.Br.001/002		CP006742.1	2011
BA1035	South Africa	B.Br.001/002		CP009698.1; CP009699.1	Unknown
Zim89	Zimbabwe	B.Br.001/002		JMPU01000000.1	Unknown
[1]KC2011	South Africa, Limpopo	B.Br.001/002		NJGK00000000	2011
ANSES_99-100	France	A.Br.009/011 (WNA/TEA)		PRJNA242332; SAMN02699415	1999
[1]Sen2Col2	Senegal	A.Br.0011/009		PRJEB1516; SAMEA2272511; ERS379929	2010
BA0004	Italy, Apulia	A.Br.009/011 (WNA/TEA)		PRJNA656733; SAMN15801198	1993
[1]BA0042	Italy, Apulia	A.Br.009/011 (WNA/TEA)		PRJNA656733; SRR12435845	1984
[1]Smith_1013	Unknown	A.Br.009/011 (WNA/TEA)		PRJNA243516; SAMN02732407	Unknown
2000031052	USA, Wyoming	A.Br.009/011 (WNA/TEA)	1.1	PRJNA264742; SRR1739967	1956
2000032823	Unknown	A.Br.009/011 (WNA/TEA)		PRJNA264742; SAMN07332931	Unknown
[1]BA0002	Italy, Apulia	A.Br.009/011 (WNA/TEA)		PRJNA656733; SAMN15801197	1993
ATCC 937 #	United States of America	A.Br.009/011 (WNA/TEA)		PRJNA561583	ATCC 937 #
ANSES_00-82	France	B. Br.006		PRJNA242332; SAMN02699416	2000
CNEVA-9066	France	B.Br.CNEVA		NZ_AAEN00000000.1	Unknown
KrugerB	South Africa, KNP	B.Br.010		AAEQ00000000.1	Unknown
HYU01	South Korea	B.Br.001/002		CP008846	2009
2002013094	United States of America	C.Br.002	1.2	SRR2164197	1956
[1]BA500 [#]	Unknown			PRJNA399155; SAMN07523108	Unknown
[1]2002013094 [#]	North America	C.Br.001		SRR2164197	1956
2000031052 [#]	Wyoming	C.Br.001		SRR1739967	1956
[1]2002013011 [#]	Unknown			SAMN04033054 SRS1063471	Unknown
[1]2002013007 [#]	Unknown			SAMN04033052 SRS1063449	Unknown

[a]NCP: North Cape province
[b] KNP: Kruger National Park
[1]Samples used for tryptophan (*trp*) genes analysis
[#]Samples used for pan-genome analysis

3.4 Genome assembly and annotation

The paired end trimmed reads of *B. anthracis* strains were *de novo* assembled using Shovill Faster SPAdes v1.1.0 pipeline (Seemann, 2017). The minimum contig length was set to 500 bp and kmer sizes 21, 33, 55, 77, 99, 127 were used for the assembly. CheckM (Parks *et al.*, 2015) was additionally used to assess the potential contaminants in individual assembled *B. anthracis* genomes. Genomes that did not meet the criteria of 98% completeness for *B. anthracis* were excluded for downstream analysis. Quast v 2.3 (Gurevich *et al.*, 2013) was used to evaluate the draft genome assemblies (n=121). The generated contigs were ordered using the moving contigs function in Mauve version 2.3.1. The assemblies were aligned to the reference *B. anthracis* Ames ancestor using the progressive Mauve alignment function (Darling *et al.*, 2004). Contigs not aligned to the reference genome were checked for sequence similarity using the nucleotide Basic Local Alignment Tool (BLASTn) (Altschul *et al.*, 1990) to confirm if they were *B. anthracis*. Draft assemblies were annotated using Rapid Annotation Using Subsystem Technology (RAST) annotation server (Overbeek *et al.*, 2005; Aziz *et al.*, 2008).

3.5 Pan-genome analysis

The *B. anthracis* draft genomes (n=111) that included KNP 2012-2015 strains (n=44, Table 3.1) and reference strains (n=67, Table 3.2) were used to determine pan-genome of *B. anthracis* A- and B-clade strains using Roary (Page *et al.*, 2015; Sitto and Battistuzzi, 2020). The *B. anthracis* genomes were annotated using Prokka v.1.14.0 (Seemann, 2014). Similarity searches between the coding domain sequences (CDS) of assembled genomes were conducted using pair-wise BLASTp (Altschul *et al.*, 1990) and Markov Cluster Algorithm (MCL).

Clusters were created, paralogs identified and the isolates were ordered by presence/absence of orthologs (Page *et al.*, 2015). Pan-genome clusters were defined as follows: Core-genes present in all isolates; soft core-genes present in at least 95% of isolates; shell-genes present between 15-95% of isolates; cloud-genes in less than 15%

of isolates (Tettelin *et al.*, 2005; Fischer *et al.*, 2011). BEAST2 (Suchard *et al.*, 2018b) was used to construct the phylogenetic tree of the aligned accessory-genes using default parameters. The phylogenetic tree was visualized using Figtree v1.16.6 (Rambaut, 2009) and ITOL (Letunic and Bork, 2019).

3.5.1 Sequence analysis of unique genes on *B. anthracis* clades

Binary-accessory genes that indicated the presence-absence of unique genes were extracted from A- and B-clade *B. anthracis* genomes using Roary (Page *et al.*, 2015). The unique genes were compared with annotated sequence data from Prokka (Seemann, 2014) and further compared and validated with output files annotated by RAST server (Aziz *et al.*, 2008). Multiple nucleotide sequence alignment for the unique accessory-binary genes were performed using Multiple Alignment using Fast Fourier Transform (MAFFT) (Katoh and Standley, 2013) and phylogenetic trees constructed using BEAST (Suchard *et al.*, 2018).

3.6 Genetic characterization of *B. anthracis* A- and B-clade strains

3.6.1 Tryptophan (*trp*) operon genes

Coding sequences linked to the tryptophan operon genes (*trpEDCBA*) were identified and extracted from annotated *B. anthracis* genomes (n=108, Table 3.1 and 3.2, indicated by the # sign) to determine variability in the clustering between A- and B- clade strains. Multiple sequence alignment for each *trp* gene was performed using MAFFT (Katoh and Standley, 2013) and phylogeny was inferred using the Maximum Parsimony method of evolution in MEGAX (Kumar *et al.*, 2018). Phylogenetic trees were visualized on FigTree v1.16.6 (Rambaut, 2009) and ITOL (Letunic and Bork, 2007).

3.6.2.1 *Bacillus* collagen-like protein of anthracis (*bclA* gene)

B. anthracis bclA gene coding sequences (n=109) were extracted from the A and B-clade annotated genomes and submitted to tandem repeat finder (TRF v 4.09) (Benson, 1999) to determine the copy number of repeats. The *bclA* sequences were translated into amino acid sequences using Molecular Evolutionary Genetics Analysis (MEGAX) (Kumar *et al.*, 2018), aligned using MAFFT (Katoh and Standley, 2013) and the phylogeny was inferred using the maximum parsimony method.

3.6.2.2 PCR amplification of *B. anthracis* VNTR regions

Bacillus anthracis isolates from KNP representing the A-clade (AX2012309, DS201330, DS201578, DS201588, 6461_SP1 and V54) and B- clade (A3, A5, C13 and Z21) were used to determine the VNTR copy numbers using MLVA markers Bams13, Bams15, Bams30 and Bams31. The 50 µl multiplex PCR reactions were prepared as described by Lista *et al.*., (2006) and consisted of 5 ng genomic DNA template, 0.2 µM forward primer, 0.2 µM reverse primer and 1X Ampliqon Red mastermix (Ampliqon). The PCR conditions included an initial enzyme activation and DNA denaturation at 98°C for 5 min followed by 36 cycles of 98°C for 20 sec, annealing at 60°C for 30 sec and extension at 65°C for 2 min. A final extension was done at 65°C for 5 min (Lista *et al.*, 2006). The PCR products were electrophoresed on a 1 % agarose gel and visualized with ethidium bromide. The amplicons were subjected to an enzymatic clean-up using Exonuclease I and Alkaline Phosphatephosphatase (Applied Biosystems), cycle sequencing and clean-up using the BigDye Terminator Kit v3.1 and the BigDye Xterminator kit (Applied Biosystems) respectively. The products were sequenced by capillary electrophoresis on the genetic analyzer 3500 (Applied Biosystems). The generated sequences were submitted to tandem repeat finder (TRF v 4.09) (Benson, 1999) to determine the copy number of repeats.

3.6.3 Antimicrobial resistance (AMR) profiles of *B. anthracis* genomes

Antimicrobial resistance (AMR) profiles were determined for the sequenced *B. anthracis* genomes (Table 3.1) and reference genomes (Table 3.2). The ABRicate pipeline (Seemann, 2014) was deployed to identify antimicrobial resistant (AMR) determinants in each assembled genome using the ResFinder database (–db ResFinder; accessed 23 January 2022) (Feldgarden *et al.*, 2019) with minimum identity and coverage thresholds of 75 (– minid 75) and 50% (–mincov 50), respectively (Zankari *et al.*, 2012; Bortolaia *et al.*, 2020).

Chapter 4: Results

4.1. Genomic features of the assemblies in *B. anthracis*

Sequence reads of *B. anthracis* strains were *de novo* assembled and resulted in draft genomes (n=120) of different sizes. The size of the draft genomes was approximately 5.5 Mb with contigs greater than 1000 bp (Table 4.1). The generated contigs for the 2012-2015 KNP isolates varied from 28 to 99. The minimum length of contigs in which half of the bases of the assembly are covered (N50) ranged from 9767 to 5227966. Assessment of potential contaminants in individual assemblies revealed that three isolates 1249, DS201313 and RL201530 were not classified as *B. anthracis* genomes and were excluded from downstream processing. Genome assembly statistics showed that the draft genome sizes of the sequenced *B. anthracis* genomes in this study were approximately 5.5 Mb and were comparable to the reference Ames ancestor (Table 4.1). The GC content of the draft genomes was 35%, similar to the GC content in other species in the *B. cereus* group. Alignment of the KNP 2012-2015 draft genomes revealed an exceptionally high level of homogeneity among the studied genomes, with over 90% of the reads aligning to the reference Ames ancestor chromosome and its plasmids.

The RAST annotation server annotated the genes in the draft assemblies and calculated over 5000 coding sequences and 300 subsystems for the Kruger A- and B-clades draft genomes. Approximately 27% of the genes were found in known subsystems, whereas 70% could not be assigned to known subsystems. The annotated *B. anthracis* strains revealed that the majority of the subsystem genes identified were responsible for basic cellular (e.g., cell wall synthesis) and metabolic functions (carbohydrates and amino acids metabolism). RAST also identified genes responsible for virulence, defense, dormancy and sporulation (Figure S1).

Table 4.1: Summary statistics of the *de novo* assembled *Bacillus anthracis* A- and B-clade Kruger National Park genomes and reference draft genomes from NCBI (n=120 used in this study.

Strain	# contigs (>= 0 bp)	# contigs (>= 1000 bp)	Total length (>= 0 bp)	Total length (>= 1000 bp)	# contigs	Largest contig	Total length	N50	N75
12460315	64	31	5471300	5458434	38	1768527	5463529	596566	238319
1249	34	25	3281671	3277601	28	609626	3279856	285718	189071
1257032015UP	57	29	5466458	5456893	33	1768527	5459929	565919	263595
12610315	81	44	5513698	5497688	55	1769458	5505748	596647	239924
1298	1319	38	5884225	5456351	111	923037	5503314	426149	216010
13020315	72	20	4675775	4655498	29	1461836	4661115	614414	309472
13030315	81	36	5492602	5473153	52	1768521	5484198	565838	263595
1365	82	35	5468585	5453061	42	1768981	5457919	1052322	150308
13680315	50	27	5465474	5456975	32	1603157	5460411	596643	239923
1370	37	33	5453394	5450390	37	1031227	5453394	404441	238303
2000031075	46	39	5353374	5348527	46	1853649	5353374	565655	203890
2000032823	48	42	5262802	5259650	45	702963	5261623	204490	132264
2002013007	123	103	5438681	5426465	118	356132	5436841	113112	53646
2002013011	152	125	5475088	5458459	143	254029	5471831	85796	49624
2002013094	62	51	5499927	5493153	59	702710	5498767	204277	110826
2002734065	37	32	5172725	5170397	34	1766407	5171701	419175	172624
20SD	52	48	5441630	5438435	51	867549	5441305	285522	188952
2949_1D	470	366	5416140	5345645	452	125072	5407937	28406	13810
2991_1B	378	253	5396612	5322561	321	185192	5373197	38630	19765
3008_1B	442	344	5418987	5351929	428	226189	5412702	34402	14068
3122_2B	431	325	5401847	5336815	393	175230	5385684	35340	16623
3132_1B	170	140	5350330	5337202	147	335825	5343030	98947	44920
3275_2D	751	614	5352180	5251890	751	89998	5352180	14738	6670
3517_1C	121	83	5416293	5396636	98	343375	5406625	203477	77569
3517_2C	1194	778	5266628	5007131	1070	55932	5212645	9767	4752
3618_2D	36	29	5434738	5430427	32	1637151	5433149	586769	237871
3631_1C	45	37	5356744	5353754	40	850822	5355516	271019	166369
3631_3D	598	426	5365636	5249771	566	246615	5350807	22854	11560
3631_4C	385	307	5402081	5347349	375	177852	5397440	35768	16345
3631_8D	882	631	5252949	5088181	822	98835	5225531	14279	6656
419	23	23	5341601	5341601	23	1161725	5341601	867965	239317
52-G	3	3	5504444	5504444	3	5227966	5504444	5227966	5227966
DS20150654	51	28	5465716	5457296	33	1603094	5460646	596642	238256
8903-G	3	3	5504767	5504767	3	5228314	5504767	5228314	5228314
9080-G	3	3	5508663	5508663	3	5232192	5508663	5232192	5232192
ABr003	3	3	5487716	5487716	3	5207816	5487716	5207816	5207816
A0021	32	25	5454873	5451308	29	1636680	5454115	1161585	238462
A0091	691	24	5754100	5442389	125	1766960	5497830	1160066	263873
A0193	60	60	5392880	5392880	60	536225	5392880	170206	80261
A0442	46	46	5374836	5374836	46	1040654	5374836	218283	92406
A1039	32	21	5458264	5452851	25	1768573	5455937	1198122	579970
A11	27	27	5424547	5424547	27	1637235	5424547	700661	217798
A16_KNP	38	38	5423735	5423735	38	630221	5423735	275881	117979
A19	34	34	5422453	5422453	34	630005	5422453	309999	165790
A3	36	36	5296518	5296518	36	630222	5296518	275881	142040
A5	31	31	5402392	5402392	31	630221	5402392	408653	159863
A8	72	72	5417873	5417873	72	489427	5417873	154041	73141
AH14	290	267	5322432	5304618	290	369679	5322432	48459	19137
AH26	132	125	5453982	5448651	132	407083	5453982	101000	51655
AH77	131	48	5474622	5450303	59	786780	5458233	212452	95244
AH97	54	47	5454988	5450060	54	917937	5454988	263375	144081
Ames_Ancestor	3	3	5503926	5503926	3	5227419	5503926	5227419	5227419
ANSES_00-82	28	28	5436014	5436014	28	1160487	5436014	868432	263671
ANSES_08-8_20	31	31	5440708	5440708	31	1114119	5440708	440320	239264
ANSES_99-100	30	30	5446472	5446472	30	1160908	5446472	868537	263516
ATCC_11966	32	23	5459259	5453841	30	1769259	5458391	936758	289144
ATCC_937	23	19	5278732	5276516	21	1770352	5277810	944336	289146
Aust94	49	48	5504562	5503918	49	1143198	5504562	240598	172362
AX2012275	117	38	5473629	5452296	45	803587	5457446	263118	179326
AX2012283	100	56	5468672	5450002	65	550281	5456434	204206	128615
AX2012291	105	49	5474062	5450993	61	786736	5459246	225720	137105
AX2012435	229	45	5533490	5463253	75	899819	5483514	365209	159925

Strain	# contigs (>= 0 bp)	# contigs (>= 1000 bp)	Total length (>= 0 bp)	Total length (>= 1000 bp)	# contigs	Largest contig	Total length	N50	N75
BA0002	53	53	5365101	5365101	53	1088458	5365101	299248	125307
BA0004	40	40	5273665	5273665	40	798953	5273665	238325	119366
BA0008	56	56	5444732	5444732	56	774634	5444732	222886	96586
BA0042	109	109	5223378	5223378	109	289709	5223378	99736	49299
BA1015	3	3	5487255	5487255	3	5210596	5487255	5210596	5210596
BA500	44	31	5358909	5352157	36	957178	5356276	729795	237869
BFV	3	3	5508355	5508355	3	5230053	5508355	5230053	5230053
C13	28	28	5423260	5423260	28	1637111	5423260	586769	217656
CNEVA-9066	30	29	5488676	5488337	29	1053408	5488337	277207	228247
COVASU	28	24	5461108	5458132	28	1213004	5461108	937089	291182
CZC5	25	22	5454337	5452788	23	1445599	5453735	590084	271413
Delta_sterne	1	1	5229650	5229650	1	5229650	5229650	5229650	5229650
DS201505	61	33	5462029	5451578	40	1162408	5456388	412401	214795
DS201510	36	31	5453456	5449803	36	1603075	5453456	483213	203905
DS201523	70	31	5471253	5453764	48	1640007	5465545	1162327	239939
DS201525	966	812	5508258	5400709	966	40577	5508258	10487	5198
DS201577	53	27	5466290	5456796	33	1769431	5460804	1197885	263595
DS201578	53	46	5455255	5450424	53	734067	5455255	267729	160352
DS201579	31	31	5450054	5450054	31	1768393	5450054	472345	239877
DS201584	33	27	5455976	5451767	33	1604240	5455976	1162214	239893
DS201588	35	28	5457146	5452624	35	1161971	5457146	703210	204674
FT2012	175	60	5387254	5352414	79	918487	5364655	266587	94848
H9401	3	3	5495471	5495471	3	5218947	5495471	5218947	5218947
HP12	36	36	5422775	5422775	36	630221	5422775	262421	127319
HT2013	753	541	5303035	5191833	651	131487	5268968	22555	7985
HYU01	3	3	5490124	5490124	3	5213498	5490124	5213498	5213498
K1	38	33	5458420	5454955	38	987529	5458420	401017	206875
K1285	36	24	5460829	5453632	32	1686950	5459104	1162096	288926
K2	38	35	5461141	5458890	38	579918	5461141	354382	237210
K8215	42	33	5458728	5453341	40	1162737	5458195	316575	169777
KC2011	1	1	5227419	5227419	1	5227419	5227419	5227419	5227419
KM20150330	229	45	5533490	5463253	75	899819	5483514	365209	159925
AD201016B	34	27	5360116	5355294	34	1768503	5360116	587282	238257
13680315	32	27	5460411	5456975	32	1603157	5460411	596643	239923
RL201415	38	30	5465476	5459732	38	1768503	5465476	1161938	263595
13030315	52	36	5484198	5473153	52	1768521	5484198	565838	263595
DS201506B	33	28	5454537	5451101	33	1769458	5454537	455808	238256
DS20152454b	33	28	5460646	5457296	33	1603094	5460646	596642	238256
12460315	38	31	5463529	5458434	38	1768527	5463529	596566	238319
1367PETA	33	27	5459637	5455360	33	1768502	5459637	1197885	239923
DS20133434	2205	293	5773096	5433550	320	201773	5452352	38647	18224
AD201016b	33	26	5459307	5454382	33	1162311	5459307	899933	264008
DS201577	33	27	5460804	5456796	33	1769431	5460804	1197885	263595
1257032015UP	33	29	5459929	5456893	33	1768527	5459929	565919	263595
KrugerB	64	62	5470007	5468381	64	1294190	5470007	286973	149281
LVS2012028	763	119	5458301	5366950	137	291142	5380085	112209	47314
RL2014115RbM	63	30	5473128	5459732	38	1768503	5465476	1161938	263595
RL201531	794	705	5530942	5463052	794	73086	5530942	12351	6579
SA047	35	28	5457902	5452496	35	1039789	5457902	867552	289076
Sen2Col2	44	27	5174112	5170218	30	1768306	5172027	482552	258148
Shikan-NIID	3	3	5504652	5504652	3	5228065	5504652	5228065	5228065
Smith1013	2	2	5286989	5286989	2	5189487	5286989	5189487	5189487
Sterne	1	1	5228663	5228663	1	5228663	5228663	5228663	5228663
SVA11	3	3	5487517	5487517	3	5210966	5487517	5210966	5210966
V770-NP1-R	36	33	5164523	5163724	33	1002900	5163724	403394	183429
Vollum	52	48	5488459	5485426	51	812727	5488148	382062	176056
Z21	41	41	5428551	5428551	41	585656	5428551	324409	127315
Zim89	3	3	5460923	5460923	3	5184202	5460923	5184202	5184202

4.2. Phylogenetic diversity of *B. anthracis* strains using wgSNP

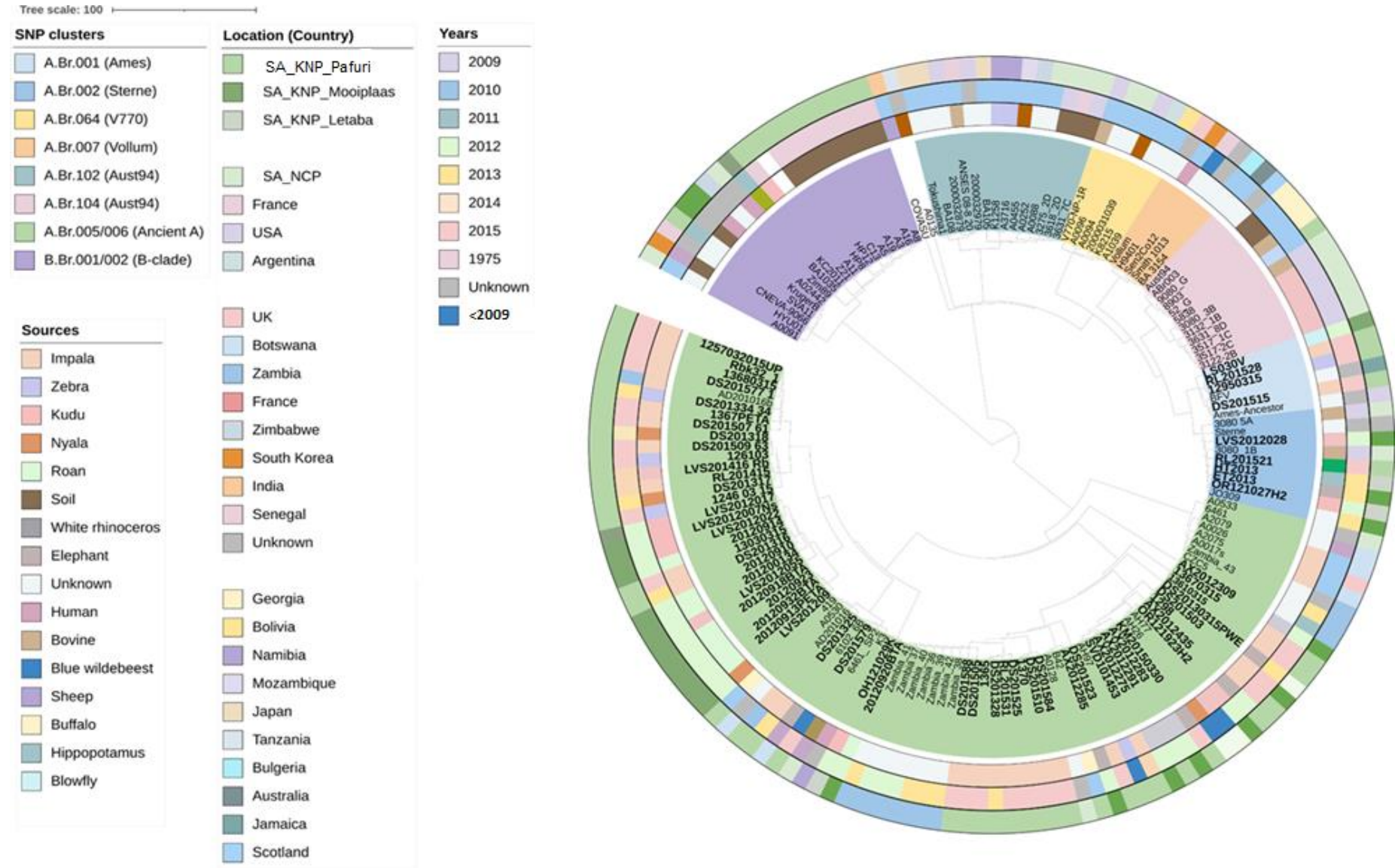
To infer phylogeny of the 2012-2015 KNP isolates (n=83), paired end sequencing was performed, producing over 1 million reads per isolate. Filter passed, trimmed (high quality) reads were mapped to the *B. anthracis* Ames ancestor as a reference genome. Whole genome single nucleotide polymorphisms (wgSNPs) were determined using *B. anthracis* strains (n=165) (Table 3.1 and Table 3.2) consisting of the sequenced 2012-2015 KNP genomes from this study, the 1975-2011 available sequenced genomes from KNP, and global genomes obtained from NCBI (Table 3.2). The wgSNP phylogeny was defined by 7713 parsimony informative SNPs that clustered the KNP and canonical global *B. anthracis* genomes (Figure 4.1A). All KNP isolates of the 2012-2015 anthrax outbreaks clustered in A.Br.005/006 (Ancient A) clade. Other major SNP clusters were not well represented during the KNP 2012-2015 outbreaks. None of the isolates clustered in the B.Br.001/002 clade (B-clade) (Figure4.1A), however seven isolates (12950315, DS201515, LS030V, LVS2012028, RL201521 RL201528 and OR121027H2) from Pafuri clustered in the A.Br.001/002 (Ames/Sterne) group (blue) with the previously sequenced (2013) genomes from Letaba (HT2013) and Olifants (ET2013). This clade was defined by 1125 parsimony informative sites.

The dominant A.Br.005/006 clade (2012-2015 isolates), could further be resolved into different sub-clades based on the year of outbreak, locality and the source of isolates (Figure 4.1A). The 2012-2015 isolates within A.Br.005/006 clade were divided into the 4 minor sub-clades (2012, 2012/13/15, 2013/14/15 and 2015). The 2012 anthrax outbreak is presented by LVS2012 (LVS2012007Ns, LVS2012017, LVS2012022, LVS2012055 and LVS201209) and AX2012 (AX2012275, AX2012283, AX2012291, AX2012435) genomes (Figure 4.1A, depicted in green). These genomes clustered separately due to different localities. The LVS2012- genomes were from isolates collected in Mooiplaas and Phalaborwa while the AX2012- genomes were isolated from strains collected in the Pafuri/Houtboschrand and Satara regions. Mooiplaas (dark green) and Pafuri (light green) appear to have heterologous genotypes defined by unique non-informative SNPs (Figure 4.1A). The AX2012- genomes clustered with the 2010 (Figure 4.1A) sequences (AH26, AH77 and AH97) and the 2012/13/15 genomes (AX2012309, KM20150330, DS20130315PWE, 13610315 and 13670315)

from Pafuri and Satara. The South African NCP (A0053, 6461), Tanzania (A2079) and Zambia (CZC5, A0017) strains grouped separately from the KNP genomes (Figure 4.1A).

Within the A.Br.005/006 major clade, the KNP isolates grouped in separate minor sub-clades, defined by the different regions on the park (Figure 4.1A). Within the A.Br.005/0006 (Figure 4.1B), the phylogeny was resolved into two distinct sub-clades (A1 and A2) that separated the isolates from the different localities. The A1 minor sub-clade consisted of isolates from Satara/Houtbouschrand and some 2015 impala genomes (1365,1370, DS201505/10/23/25/84 and RL201531) from Pafuri, which grouped with strains from Zambia (Zambia36/37/38/39/40/41/42), separate from other Pafuri 2015 isolates (Figure 4.1B). The A2 minor sub-clade consisted of the roan isolates from Mooiplaas (20120914, 20120910, 201200134, 20120918BTA, 20120923, 20120932BTA, 20120913PETA, RL2015) and other 2015 *B. anthracis* impala isolates (1257032015UP, Rbk32_1, DS201577, AD201016b, 1367PETA, DS20150761, DS20150963, RL201415, 13030315 and DS201316); nyala (DS201318, DS201317and LVS201209) and zebra (DS201334_34, 126103 and 12460315) (Figure 4.1B).

A



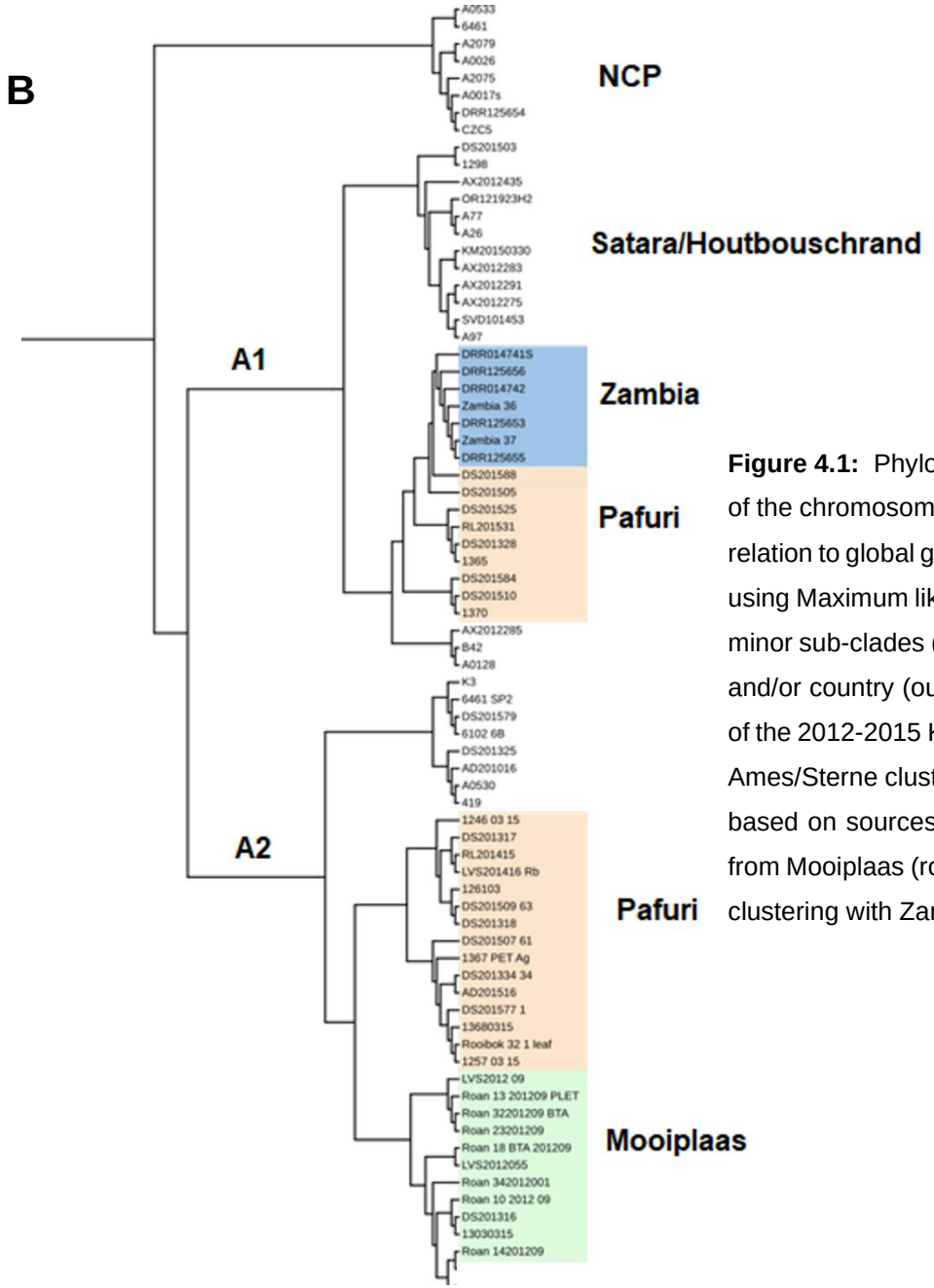


Figure 4.1: Phylogenetic relationship of *Bacillus anthracis* strains based on whole-genome SNP analysis of the chromosome indicating the clustering of 2012-2015 isolates from the Kruger National Park (KNP) in relation to global genomes. About 7713 parsimony informative SNPs were used for the phylogeny inference using Maximum likelihood method. **A**: Colour ranges indicate grouping of the isolates in different major and minor sub-clades (SNP clusters - inner ring), sources (second ring), outbreak year (third ring), and location and/or country (outermost ring). Strains sequenced in this study are indicated in a bold font. The majority of the 2012-2015 KNP isolates clustered in the A.Br.005/006 clade while a few grouped in the A.Br.001/002 Ames/Sterne cluster. **B**: Clustering of the KNP A.Br.005/006 isolates into two minor sub-clades A1 and A2 based on sources; roan (green) and impala (orange) and locality. The A1 clade is made up of samples from Mooiplaas (roan) and Pafuri (Impala) whereas sub-clade A2 consists of other Pafuri (impala samples) clustering with Zambia isolates and the Satara/Houtbouschrand (multiple species).

4.3 Pan-genome analysis and gene classification

Pan-genomic analysis was conducted on 121 genomes that included KNP strains from clades A (n=52) and B (n=9) and canonical global *B. anthracis* genomes strains (n=60). The pan-genome of *B. anthracis* was identified by 11 374 clusters of protein coding sequences (CDS) in this study. The core-genes in *B. anthracis* consisted of 3532 CDS, while the soft-core and shell genes were identified to be 1404 and 1038 respectively. About 5497 CDS were identified as cloud genes predicted amongst these genomes (Table 4.2). The cloud genes (binary accessory genes) and shell genes were used to investigate the placement of the South African sequenced genomes with global *B. anthracis* strains. Phylogenetic analysis revealed a clear distinction between A- and B-clade isolates, clustering the South African KNP genomes in their respective canSNPs (Figure 4.2). The 2012-2015 isolates clustered in the A.Br.005/006 primary clade (green). Within this clade, a monophyletic cluster of the South African genome was observed consisting of genomes of AX2012435, AX2012275, AX2012283, AX2012291, AH26, AH77, AH97, 1298, 13030315, the 2015 sub-clade presented by genomes 12460315, 12570315UP, 1365, 1367PETA, 1370, DS201505, DS201510, DS201523, DS201578, DS201579, DS20150761, DS20150663, DS201577, DS201584, DS201588, 13680315, DS20150654 and RL2014115RbM. Some of the South African genomes i.e., AH14, DS201525, DS20133434, HT2013 and RL201531 were identified having a significant number of cloud genes presented by long branches (Figure 4.2). The pan-genome analysis fairly clusters the A.Br.001/002 (Ames/Sterne) clade consisting of South African KNP genomes such as LVS2012028 from Shingwedzi and FT2012 (Figure 4.2).

Furthermore, pan-genome assigned the A.Br.104 (Aust94) that consisted of the South African NCP *B. anthracis* genomes (3517_2C, 3631_3B, 3631_8B e.tc.) into their own genetic sub-clade (Aust94, dark pink). This sub-clade also includes the Namibia (K1 and K2), Georgia (52_G, 8903_G and 9080_G) and Scotland (A.Br.003). None of the 2012-2015 KNP isolates clustered in the B-clade. However, pan-genome analysis revealed that the available sequenced South African B-clade (purple) genomes A3, A5, A8, A11, A16_KNP, A19, C13, HP8, KC2011 (Table 3.2) assigned as B.Br.010 by wgSNP analysis, grouped with the global B-clade genomes defined by B.Br.001/002 canSNP using wgSNP analysis (Figure 4.2).

Table 4.2: Summary statistics analysis of the *Bacillus anthracis* pan-genome used in this study listed in Table 3.1 and 3.2.

Core genes	(99% <= strains <= 100%)	3532
Soft core genes	(95% <= strains < 99%)	1404
Shell genes	(15% <= strains < 95%)	1038
Cloud genes	(0% <= strains < 15%)	5497
Total genes	(0% <= strains <= 100%)	11471

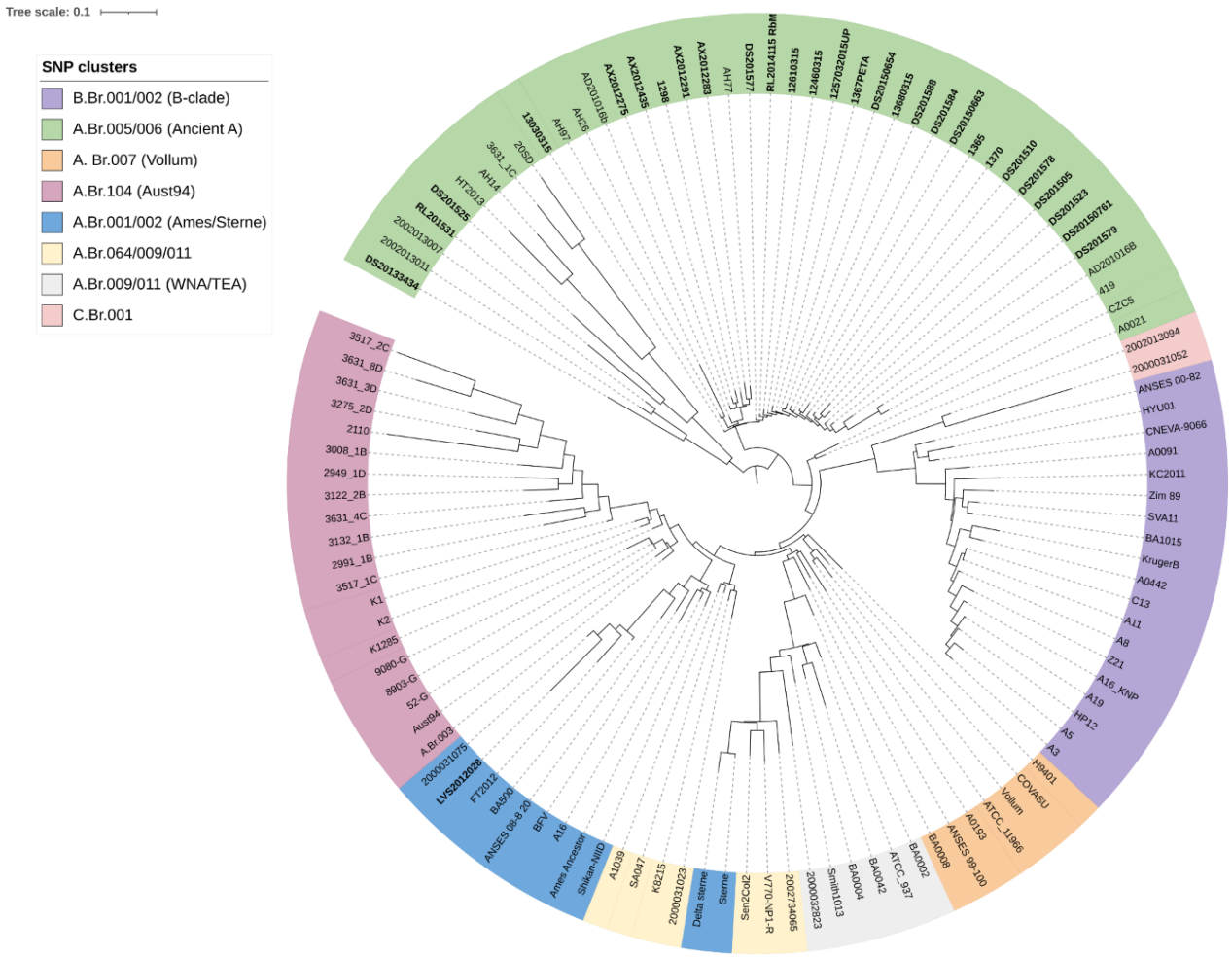


Figure 4.2: Pan-genome phylogeny of *Bacillus anthracis* based on accessory genes, showing the clustering of the South African, Kruger National Park (KNP) genomes (bold) in relation to global genomes. The phylogeny was inferred using Maximum likelihood method in BEAST. Colour ranges indicate the clustering into major sub-clades. Pan-genome analysis generated 8 primary clusters and grouped the 2012/2015 KNP genomes (bold) in the A.Br.005/006 major clade (green), separate from the Northern Cape Province genomes (pink) and B.Br.001/002 (B-clade) genomes (purple).

4.3.1 Assessment of the shell and cloud genes in A.Br.005/006 and B.Br.001/002

Pan-genome analysis of *B. anthracis* genomes allowed the identification of clade specific genes (Table 4.3). A total of 6535 shell and cloud genes were identified and assigned to A- and B-clade genomes. Ten genes were unique to the A.Br.005/006 branch, and included proteins linked to YeeV-YeeU toxin-antitoxin system, membrane associated lysozymes and inner membrane protein (two copies of *yohK_2*), putative prophage proteins and phage DNA replication proteins (Table 4.3). Pan-genomics also identified multiple hypothetical proteins only found in A.Br.005/006 genomes.

In contrast, B-clade genomes consisted of 15 genes that dissociates them from the A-clade genomes (Table 4.3). These were identified as hypothetical proteins (n=10), a multidrug resistance protein (*mdt*G_2) and chromosome encoded cell wall biosynthesis genes; long-chain-fatty-acid--CoA ligase *fad*D13 (4-5 copies), two copies of the Heme-based aerotactic transducer *(hem*AT or *hemZ*) and two copies of the RNA processing ribosomal RNA small subunit methyltransferase-J (Figure 4.3). Furthermore, this study identified a murein endopeptidase (*mepH*) that encodes the lipoprotein (NLP/P60 family) murein endopeptidase (Figure 4.3) in the chromosome and pXO2 (Figure 4.4A and B). The gene arrangement of *mepH* located on pXO2 of B-clade genomes is distinct from the gene in the A-clade genomes. In the B-clade, the *mepH* gene is composed of two subunits of 209 and 148 aa residues length respectively as compared to a complete 381 aa residues protein in A-clade strains (Figure 4.4B). The sequence alignment of the *mepH* gene 5' residue in B-clade is similar to the A-clade strains, however it has a stop codon at position 209 and results in the deletion of 127 amino acid residues (Figure 4.4B).

Table 4.3: Cloud genes of *Bacillus anthracis* strains in B.Br.001/002 and A.Br.005/006 identified in this study.

Gene name	Gene product name in full	Lineage/strain	Copy number
dprA	DNA processing protein A	B.Br.001/002	2
fadD13	*long-chain-fatty-acid--CoA ligase*	B.Br.001/002	4-5
rsmJ	*Ribosomal RNA small subunit methyltransferase J*	B.Br.001/002	2
mdtG	Multidrug resistance protein	B.Br.001/002	3
mepH	*Murein DD-endopeptidase*	B.Br.001/002	1
hemAT	*Heme-based aerotactic transducer*	B.Br.001/002	2
	antitoxin of the YeeV-YeeU toxin-antitoxin system	A.Br.005/006	1
spxA	Regulatory protein Spx	A.Br.005/006	3
yohK	inner membrane protein yohk	A.Br.005/006	2
	putative prophage protein	A.Br.005/006	2
	membrane-associated lysozyme; Qin prophage	A.Br.005/006	1
	Major tail protein V Phage	A.Br.005/006	1

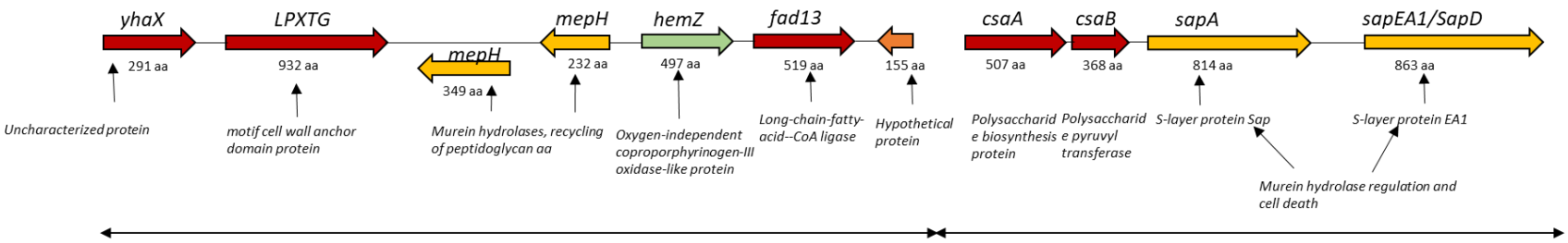


Figure 4.3: Schematic representation of *Bacillus anthracis* cell wall biosynthesis pathway gene cluster located on the chromosome. Cluster depicts organization and orientation of genes and the sizes of the coding regions for proteins responsible for cell wall anchoring (LPXTG), formation and recycling of peptides (*mepH, sapA, csaAB*), fatty acids metabolism (*fad13*) and oxidases (*hemZ*). The expression and interaction of genes result in the normal peptidoglycan synthesis or programmed cell death.

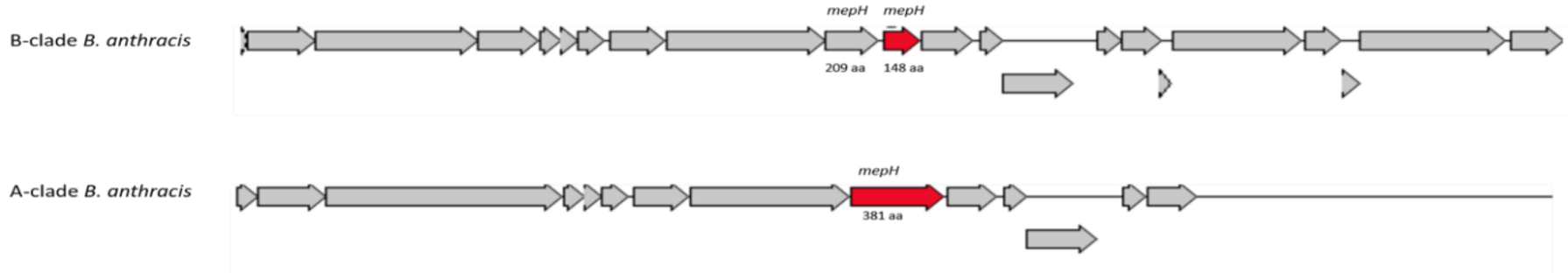


Figure 4.4A: Structural organization of the murein DD-endopeptidase hydrolase (*mepH*) protein in the pXO2 of *Bacillus anthracis* strains. The organization of the *mepH* in B-clade strains (top) consist of two proteins; 209 aa and 148 aa (top) compared to a single 381 aa protein in A- clade strains. Regions downstream and upstream the *mepH* coding region (indicated in grey) are similar in strains of both clades.

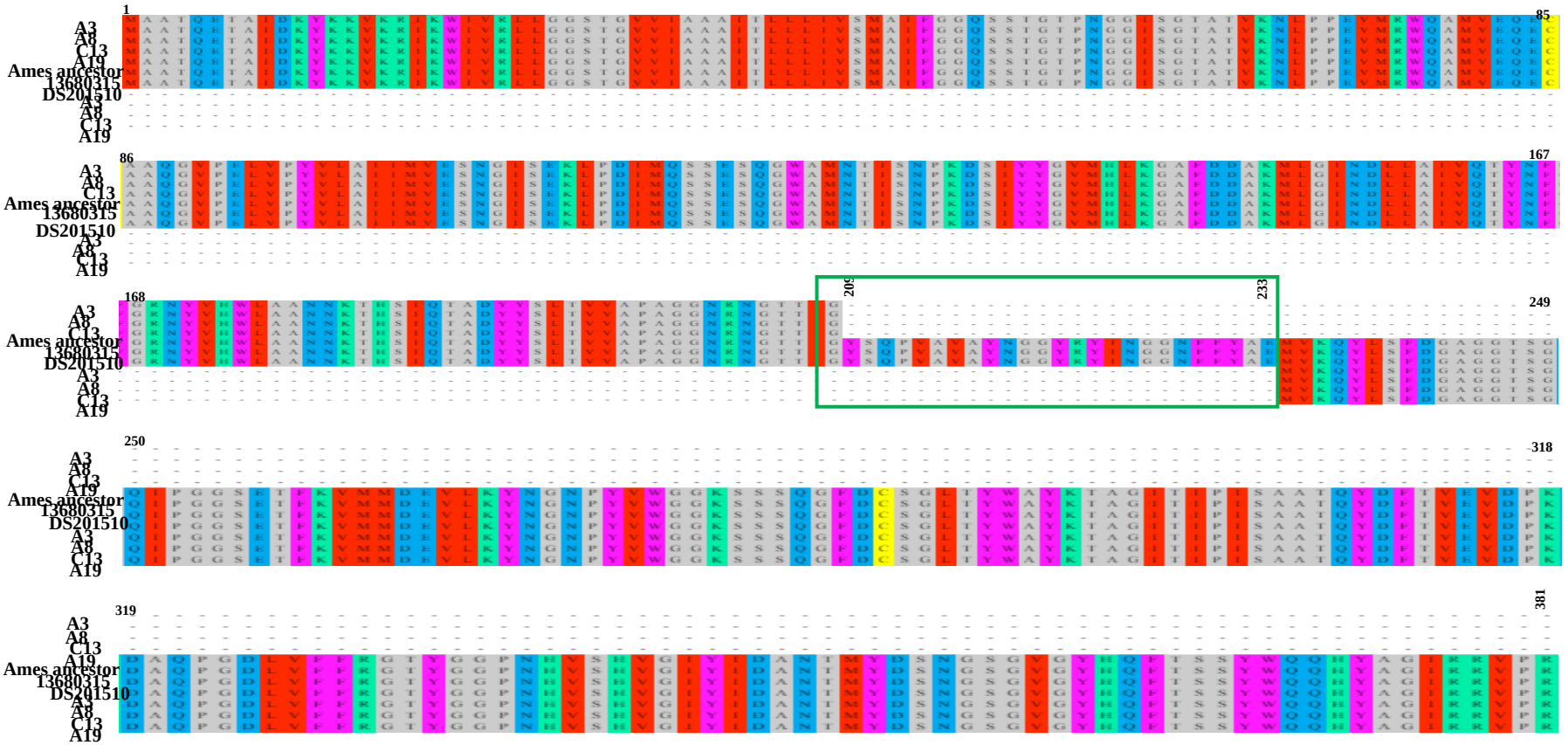


Figure 4.5B: Amino acid sequence alignment of *mepH* in *Bacillus anthracis* B-clade and A-clade genomes. The truncated protein at position 209 in B-clade compared to 308 amino acid residues in A-clade.

4.4 Genetic characterization of *B. anthracis* A and B strains in the KNP

4.4.1 Antimicrobial resistance (AMR) profiles of *B. anthracis* strains

The ResFinder database identified the AMR profiles of the assembled *B. anthracis* genomes. All *B. anthracis* genomes encode proteins related toto Fosfomycin resistance (Table 4.4). *B. anthracis* strain 1298 of the A-clade possesses a unique coding sequence for multidrug resistance towards amoxicillin, ampicillin, cephalothin, piperacillin and ticarcillin identified as *blaTEM*-116 present in contig 46 of the draft genome. All *B. anthracis* strains carry *mdtG_1,* a fosmoycin and deoxycholate resistant gene identified by pan-genome analysis. Pan-genome analysis revealed signatures of a second copy of the *mdtG_2* gene, found only in B-clade strains (Table 4.3).

Table 4.4: Antimicrobial resistance (AMR) profiles of the KNP *Bacillus anthracis* genomes used in this study.

Strain	Start	End	Strand	Gene	Coverage	% Coverage	% Identity	Resistance
12460315	872849	873268	-	fosB2_1	1-420/420	100	100	Fosfomycin
12460315	363821	364236	-	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
1257032015UP	872849	873268	-	fosB2_1	1-420/420	100	100	Fosfomycin
1257032015UP	363817	364232	-	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
1298	895133	895552	+	fosB2_1	1-420/420	100	100	Fosfomycin
1298	5	861	+	blaTEM-116_1	1-857/861	99.54	99.88	Amoxicillin;Ampicillin;Cephalothin;Piperacillin;Ticarcillin
1298	193416	193831	-	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
13030315	895260	895679	+	fosB2_1	1-420/420	100	100	Fosfomycin
13030315	232331	232746	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
1365	895492	895911	+	fosB2_1	1-420/420	100	100	Fosfomycin
1365	819562	819977	-	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
13680315	895235	895654	+	fosB2_1	1-420/420	100	100	Fosfomycin
13680315	232331	232746	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
1370	455094	455513	-	fosB2_1	1-420/420	100	100	Fosfomycin
1370	232322	232737	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
DS20150654	707416	707835	-	fosB2_1	1-420/420	100	100	Fosfomycin
DS20150654	232330	232745	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
DS20150761	872849	873268	-	fosB2_1	1-420/420	100	100	Fosfomycin
DS20150761	363898	364313	-	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
DS20150663	873780	874199	-	fosB2_1	1-420/420	100	100	Fosfomycin
DS20150663	232330	232745	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
A11	744065	744485	-	fosB2_1	1-420/420	100	99.76	Fosfomycin
A11	25043	25458	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
A16	1916543	1916958	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
A16	3778850	3779269	+	fosB2_1	1-420/420	100	100	Fosfomycin
A3	232255	232670	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
A3	262636	263056	+	fosB2_1	1-420/420	100	99.76	Fosfomycin
A5	307639	308054	-	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
A5	262636	263056	+	fosB2_1	1-420/420	100	99.76	Fosfomycin
A8	25098	25513	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
A8	247454	247874	+	fosB2_1	1-420/420	100	99.76	Fosfomycin
AH77	51933	52348	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
AH77	571372	571791	+	fosB2_1	1-420/420	100	100	Fosfomycin
AH97	545578	545997	+	fosB2_1	1-420/420	100	100	Fosfomycin
AH97	232307	232722	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
AX2012275	193430	193845	-	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
AX2012275	4267	4686	-	fosB2_1	1-420/420	100	100	Fosfomycin
AX2012283	4267	4686	-	fosB2_1	1-420/420	100	100	Fosfomycin
AX2012283	33824	34239	-	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
AX2012291	137084	137499	-	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
AX2012291	4267	4686	-	fosB2_1	1-420/420	100	100	Fosfomycin
AX2012435	66837	67252	+	fosB1_1	1-416/417	99.76	87.5	Fosfomycin
AX2012435	4267	4686	-	fosB2_1	1-420/420	100	100	Fosfomycin

4.4.2 Assessment of the tryptophan operon genes in *B. anthracis* genomes

The tryptophan operon structural gene genes (*trpE*, *trpD*, *trpC*, *trpB*, and *trpA*) sequences from the *B. anthracis* genomes (n=109) were aligned. The multiple sequence alignment of the *trpE*, *trpD*, *trpC*, *trpB* genes showed no nucleotide variation between A and B clade strains. In contrast, the alignment of the *trpA* gene depicts variation in gene size amongst A and B strains. The size of the *trpA* gene in A-clade strains is 777 bp compared to the 655 bp in B-clade strains (Figure 4.5). The 122 bp deletion at the 3' end on the B-clade strains results in a truncated gene. The presence of a G/T SNP at position 652 bp was observed on the aligned *trpA* gene and signals a stop codon during protein translation (Figure 4.5). There is a genetic distinction between A and B clade isolates as they cluster into two separate clades or branches (Figure 4.6). All isolates of the 2012-2015 anthrax outbreaks clustered in the A-clade with all available sequences of the KNP genomes.

Figure 4.6: Multiple sequence alignment of the *trpA* gene (region 643-777 bp) sequences of *Bacillus anthracis* showing the difference between the A and B strains. Blue box = A-clade strains, green box= B-clade strains. The *trpA* gene missing 122 bp in B-clade and G/T SNP at position 652 results in a stop codon, signalling termination of protein translation.

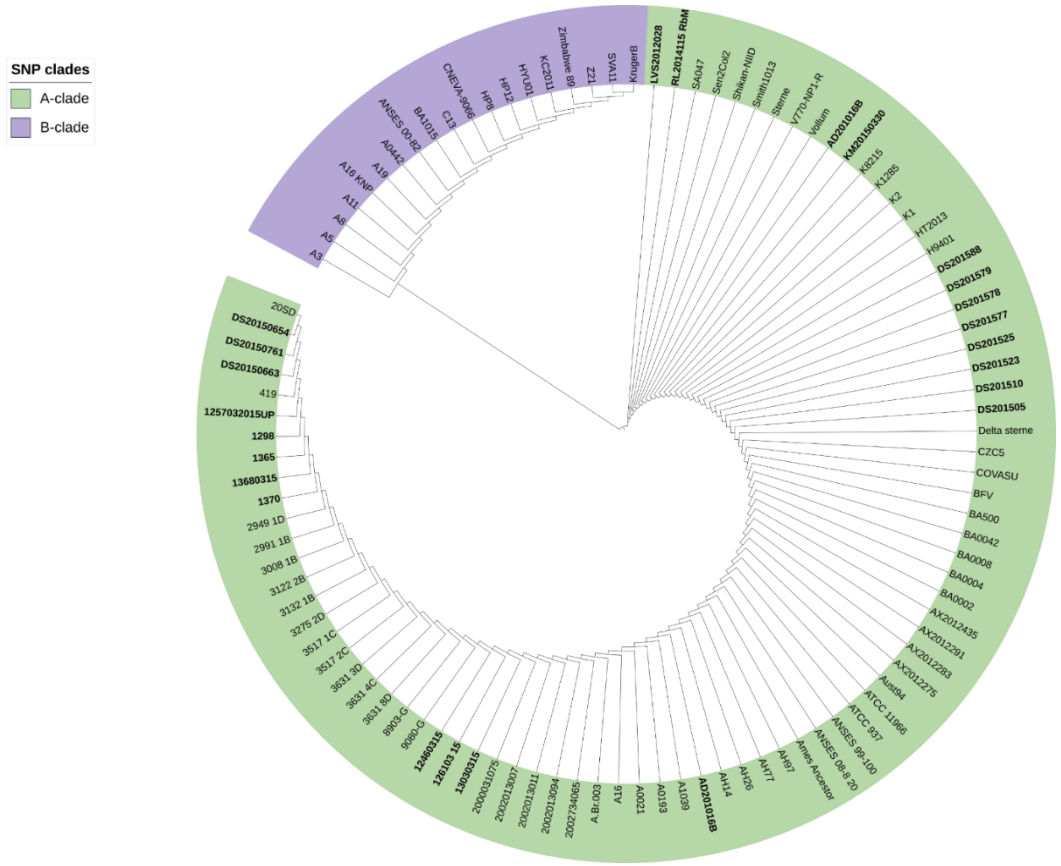


Figure 4.7: Phylogenetic relationship of *Bacillus anthracis* based on the tryptophan synthase alpha chain (*trpA*) gene of A and B clade *B. anthracis* isolates. The evolutionary history was inferred using the Maximum Parsimony method. The tree depicts two groups, a cluster for lineage B strains (purple) and another for lineage A strains (green). All isolates from the 2012-2015 (bold) Kruger National Park outbreaks clustered in the A-clade.

4.4.3 The *bclA* gene and VNTR copy numbers in *B. anthracis*

The multiple sequence alignment of the *B. anthracis* strains showed differences in the length of the *bclA* gene. The A and B clade strains consist of an 816 bp and 772 bp gene sequence encoding a protein of 270 and 257 amino acids respectively (Figure 4.7). The *B. anthracis* strains were differentiated by the number of VNTRs based on the length of the collagen-like regions (CLRs). Multiple sequence alignment of the BclA protein showed that the CLRs are composed of GXY (glycine and two additional residues) repeats varying from 51 to 273 amino acids and had a large proportion of GPT triplet repeat units (Table S2).

The copy number of the repeat units varied between lineages A and B and within strains in the same lineage. The 2012-2015 KNP A-clade (A.Br.005/006) strains had 4 to 6 *bclA* repeat units as compared to 3 units located in the B-clade strains (Table S2, in green). Sequence comparison showed that the copy number of repeats within the 2012-2015 A-clade from KNP (A.Br.005/006) group is 4, with the exception of strain 1365 (3 repeats) and strain DS201578 (5 repeats). Furthermore, variation in the number of VNTR within the A-clade genomes was also observed in the Punda Maria strains (AH26, AH77 and AH97) isolated during the 2010 outbreak, and 2012-2015 strains from the Pafuri region (Table S2 and Figure 4.7). Low VNTR copy numbers were observed in the B.Br.001/002 genomes (Table S2, light blue); the numbers varied between 2 and 3. The genomes in the A.Br.Sterne group had 3 to 5 repeats (Table S2, light purple), with the exception of the high copy number in strain Delta_sterne (6 repeats). Phylogenetic analysis on the *bclA* gene sequences showed that strains from A and B lineages cluster in two clades or two separate branches (Figure 4.8, Figure S2. A clear distinction can be observed between the A.Br.005/006, A.Br.064, A.Br.104 and B.Br.001/002 genomes. Some genomes in the B.Br.001/002 subclade (HYU01, CNEVA) consisted of 2 VNTR copies on the *bclA* gene, compared to the other studied genomes in the same subclade (Figure 4.8, Table S2). The KNP genomes in the B.Br.001/002; with the exception of the KrugerB genome (8 copies), all the other genomes have the same number of repeats (2) as the Zimbabwe89 genome. The nucleotide sequence alignment of the *bclA* gene and phylogenetic analysis fairly groups the strains into the major canonical SNPs of *B. anthracis*.

Confirmation of repeats units using MLVA (VNTR) PCR markers on the *B. anthracis* A-clade strains AX2012309, DS201330, DS201578, DS201588, 6461_SP1 and V54) and B- clade (A3, A5, C13 and Z21) strains showed that A-clade strains had longer fragments with high copy numbers of VNTR (Bams 15 and Bams 30) compared to B-clade isolates (Figure S3 and S4). No differences in fragment length were observed between A- and B-clade strains using MLVA markers Bams13 and Bams31 (Supplementary Figure S3 and S4). In addition, sanger sequencing of the PCR products and submission to the tandem repeat finder (TRF v 4.09) (Benson, 1999) to determined that the number of repeats were identical between whole genome sequence data and Sanger sequences of the same isolates.

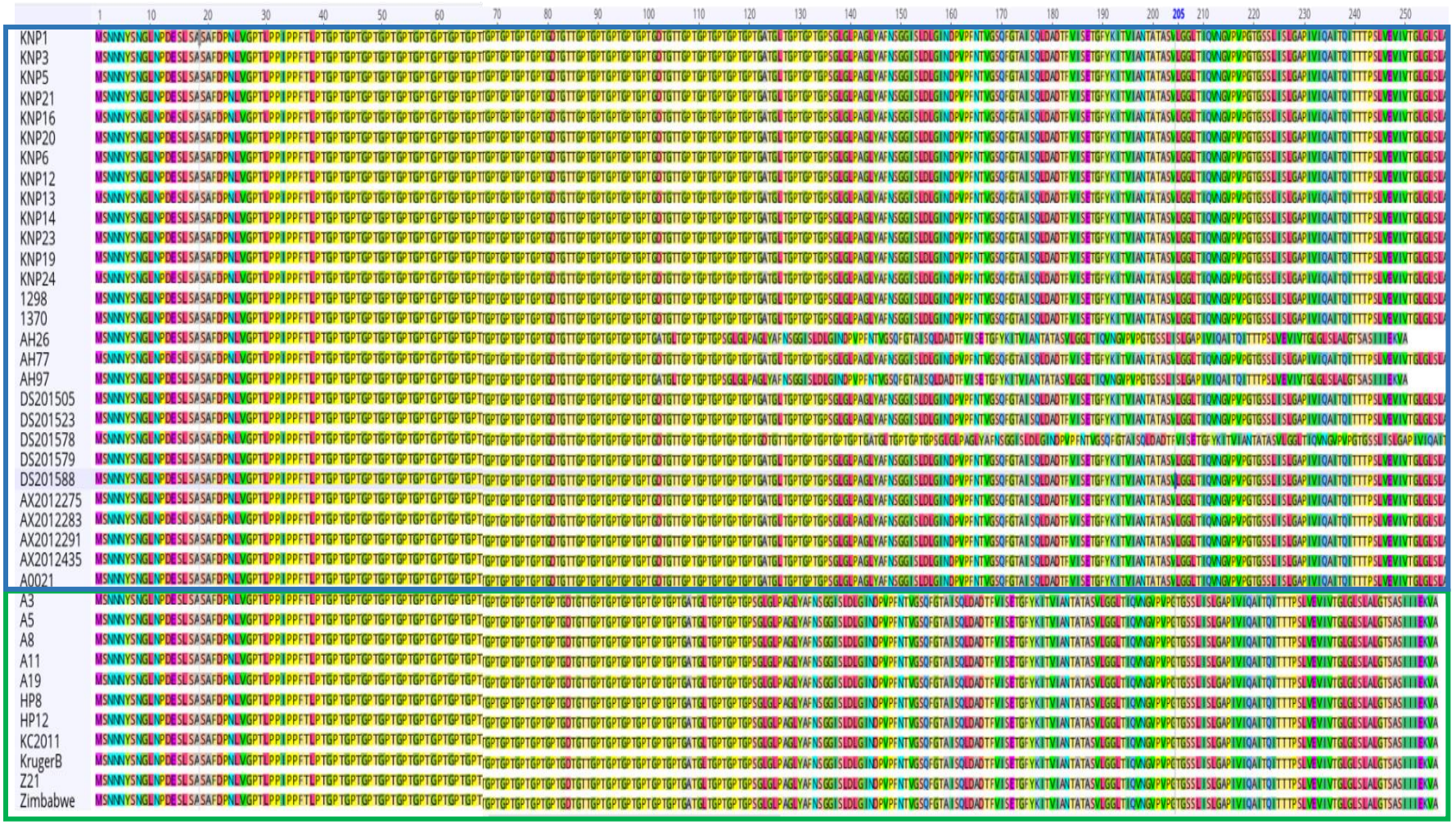


Figure 4.8: Multiple sequence alignment of the BclA protein sequences showing differences in the GPT repeat sequences in *Bacillus anthracis* lineage A and B strains from Kruger National Park. The blue box contains sequences from A-clade strains, and B-clade strain sequences in the green box. The GPT repeats start at position 27 and vary in length between A and B-clade isolates.

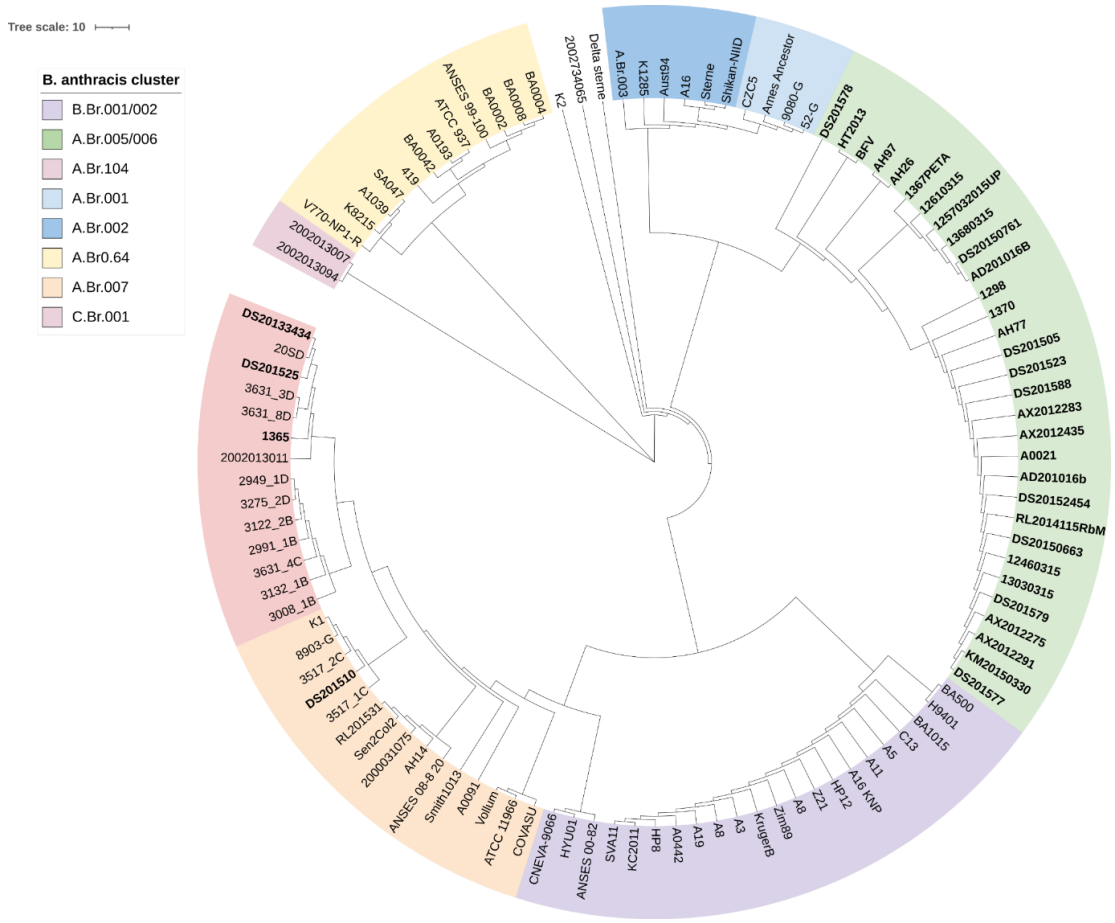


Figure 4.9: Phylogenetic tree of *Bacillus anthracis* based on the Bacillus collagen-like protein of anthracis (*bclA* gene) of A and B clade strains from Kruger National Park and global strains. Phylogeny was constructed using Maximum Parsimony method. The tree depicts 8 clusters, the dominant A.Br.005/006 (green), B.Br.001/002 (purple), A.Br.001/002 (blue), A.Br.002 (light blue), A.Br.104 (NCP) for lineage B strains (purple) and another for lineage A strains (green). All KNP isolates of the 2012-2015 (bold) anthrax outbreaks clustered in the A-clade.

Chapter 5: Discussion

In this study we combined whole genome sequencing (WGS) and pan-genome analysis to characterize differences between *B. anthracis* A- and B-clade strains that occur in the Kruger National Park (KNP). Whole genome SNP analysis provided a better understanding of *B. anthracis* genetic diversity in a natural environment such as KNP.

5.1 Genomic features of the assemblies in *B. anthracis*

Comparative genomics of the 2012-2015 KNP draft genomes (n=83) and available global genomes (41) revealed that the chromosome structure was conserved in the studied *B. anthracis* genomes. The genome sizes (~5.5 Mb) of the KNP A- and B-clade strains were comparable to the genome of the *B. anthracis* Ames ancestor, i.e., have minor large rearrangements and high regions of similarity (Keim *et al.*, 2009). The low GC content (35%) shows a true feature of the species as *B. anthracis* has a low genetic diversity with no evidence of gene transfer (Keim *et al.*, 2009).

5.2 Phylogenetic diversity of *B. anthracis* using wgSNP analysis

Bacillus anthracis is a monomorphic organism with low genetic diversity (Harrell *et al.*, 1995; Van Ert *et al.*, 2007) and traditional genotyping methods are limited in distinguishing *B. anthracis* isolates. The advent of new generation sequencing technologies has revolutionized microbial studies and have a high discriminatory power for characterization of complex micro-organisms. On a worldwide scale, *B. anthracis* genetic diversity has been well-described (Keim *et al.*, 2000; Smith *et al.*, 2000, Pearson *et al.*, 2004; Van Ert *et al.*, 2007; Sahl *et al.*, 2016; Hassim *et al.*, 2017; Lekota *et al.*, 2018).

The global structure classifies isolates into three primary clades (A, B, and C), with the A clade being the most common and globally prevalent (Van Ert *et al.*, 2007). Isolates from the *B. anthracis* B lineage have been detected in more limited or specific geographical locations (Smith *et al.*, 2000). Although lineages are endemic to particular regions such

as in the KNP, Zimbabwe and Mozambique and B " CNEVA" isolated in France, there is evidence of multiple variants co-circulating (Keim *et al.*, 2000; Smith *et al.*, 2000; Van Ert *et al.*, 2007). Anthrax is endemic in the northern part (Pafuri) of the KNP, where both A-clade and B-clade strains have been attributed to historical outbreaks before 1990 (Smith, 1999). After the 1990's, however, only the ubiquitous A-clade could be isolated from anthrax cases in the park (Smith *et al.*, 2000). On-going surveillance of anthrax outbreaks is in place in South Africa and different methods such as canSNPs, MLVA and wgSNP analysis have been employed to characterize *B. anthracis* phylogeny and have resolved the relatedness of the dominant A-clade and its minor sub-clades (Smith *et al.*, 2000; Keim *et al.*, 2009; Ledwaba, 2014; Lekota *et al.*, 2016, 2018, 2020; Hassim *et al.*, 2017).

The phylogenetic structure of *B. anthracis* strains in the KNP using wgSNP analysis showed that all the sequenced isolates from the 2012-2015 anthrax outbreaks grouped in the A-clade (A.Br.005/006 branch). This clade is well represented and prevalent among the 2012-2015 anthrax outbreak strains. The 2012-2015 KNP genomes clustered with other KNP strains from previous anthrax outbreaks and genomes from Southern Africa (Botswana and Zambia). The A.Br.005/006 SNP branch is characterized by diverse *B. anthracis* strains and presents an unstable genetic population supported by non-informative SNPs that are dispersed among the minor sub-clades. The short distance phylogenetic branches showed that the genotypes within the prevalent A-clade are closely related but are heterologous. In this study, wgSNP analysis revealed unique genotypes for the different regions of the KNP. Within the A.Br.005/006 sub-clade, wgSNP analysis could distinguish *B. anthracis* strains isolated from different KNP localities, clustering Mooiplaas (2012) and Pafuri isolates (2013 to 2015), as well as various animal species (roan and impala) and outbreak years. Although both Pafuri and Mooiplaas are in the northern part of the KNP, they present distinct dominant genotypes (Ledwaba, 2014). This shows that there may be other contributing factors to the diversification of *B. anthracis* in the KNP and requires further investigation. Additionally, isolates from KNP were clearly distinguished from the closest global strains from Zambia based on unique SNPs.

The dominance of A-clade genotypes indicates reproductive fitness as well as significant long-distance dispersal. KNP A-clade strains survive in diverse environmental conditions and over 90% of cases reported globally have been attributed to the A-clade (Van Ert *et al.*, 2007). These genotypes are highly diverse as compared to the regionally restricted B genotypes. The absence of B-clade strains in the recent KNP outbreaks (2012-2015) is not peculiar as the lineage has declined over several decades (Smith *et al.*, 2000). The lack of adaptability and diminished ability to survive has resulted in B-clade strains vanishing from the local enzootic environments. This can be associated to the propagation of B-clade strains only in areas with elevated soil pH and high calcium levels (Smith *et al.*, 2000).

Comparative analysis of A- and B-clade isolates from KNP based on wgSNPs revealed that A strains were more adaptable to a wider range of habitats than B strains, which were limited to a narrower range. This restricted or specific dispersal is characteristic of the B and C lineages (Dragon and Rennie, 1995; Smith *et al.*, 2000; Blackburn *et al.*, 2007, 2015; Van Ert *et al.*, 2007b; Keim *et al.*, 2009; Lekota *et al.*, 2016, 2018, 2020; Bruce *et al.*, 2020). The A-lineage has played an important role in the evolution of anthrax as evidenced by the success of the A branch and its clonal offspring, the involvement of A genotypes in the majority of recent anthrax outbreaks worldwide, and the short phylogenetic branch lengths within this group. This study was able to define the phylogenetic structure of *B. anthracis* in the enzootic KNP.

5.3 Pan-genome comparative analysis and gene classification

Accessory genes contribute to an organism's lifestyle and adaptation qualities to its environment (Yichao *et al.*, 2018). These genes are expressed to aid proliferation in different niches and are thus hypothesized to play a role in selection and contribute to the evolution of bacterial species (Whelan *et al.*, 2021). Pan-genome comparative analyses of the KNP *B. anthracis* strains from the 2012-2015 outbreak identified that conserved gene clusters between the A- and B- clade strains. Over 40% of the gene clusters were

strain specific. Pan-genome analysis of *B. anthracis* identified hypothetical proteins and prophage proteins unique to A-clade strains). Further investigation on the sequence diversity of the phages found in A- and B- clade strains is required to understand their influence on *B. anthracis* evolution and persistence of A-clade strains worldwide. These prophages may contain several genes (acquired horizontally) that offer stress tolerance, antibiotic resistance, virulence, and metabolic functions linked to survival (Ramisetty and Sudhakari, 2019).

Bacteriophages play a vital role in the ecology and evolution of bacteria by interacting with host and phage genomes in various ways (Roossinck, 2011; Nasir, Kim and Caetano-Anollés, 2017). The co-evolution of phages and their hosts affect their respective genomes' survival, persistence, and evolution (Buckling and Brockhurst, 2012; Koskella and Brockhurst, 2014; Nasir, Kim and Caetano-Anollés, 2017; Ramisetty and Sudhakari, 2019). *Bacillus anthracis* strains contain unique phages that can assist in bacterial identification (Hassim *et al. 2016*; Lekota *et al.*, 2018). Prophage signatures identified in isolates from A-clade genomes may offer an added survival advantage and persistence over B-clade strains in the environment.

Additionally, coding sequences linked to antitoxin of the YeeV-YeeU toxin-antitoxin (TA) system were identified in genomes from the A. Br.005/006 clade (n=45). Toxin-antitoxins are not required for normal cell growth; however, their abundance is evident on bacterial plasmids and chromosomes (Fozo *et al.*, 2010). It's been proposed that TAs are vital in cell survival in nature for transitioning to a latent, drug-resistant state to enable tolerance to excessive amounts of antibiotic stress (Page and Peti, 2016). The function of TAs is comparable to that of antibiotics, they inhibit cell growth by targeting different critical cellular processes including DNA replication, transcription, and cell wall construction (Davies, 1996; Yang and Walsh, 2017). Toxin-antitoxins influence virulence evolution (due to enhanced antibiotic resistance) and persistence of A lineage strains in the environment.

Sequence comparison of cell wall biosynthesis proteins identified multiple copies and variants of long-chain-fatty-acid--CoA ligases between strains from A and B lineages. The B-clade genomes consisted of 4-5 copies of this *FadD13* compared to 2 copies in A.Br.005/006 genomes. These ligases are essential in metabolic and regulatory processes, such as lipid biosynthesis and fatty acid degradation, intermediary metabolism and gene expression (Færgeman *et al.*, 1997; Black and DiRusso, 2003; Dong *et al.*, 2012). The result of this study suggests that the survival of A-clade strains in different niches is dependent on the cell wall structure. The protein composition and the interaction of the cell wall proteins aid in maintaining cell wall integrity and inhibits desiccation of spores (Færgeman *et al.*, 1997; Black and DiRusso, 2003; Dong *et al.*, 2012). Higher copy number of this gene might influence the ability of different lineages to survive in harsh or unsuitable environments.

Variability in terms of the size and structure of the *mepH* gene was observed. The *mepH* gene on pXO2 encodes two lipoproteins in B-clade genomes as compared to only one in A-clade genomes. The lipoproteins (*mepH*) are generally found in the chromosome responsible for cell-wall biosynthesis (Vermassen *et al.*, 2019). However, this gene was also found in the pXO2 plasmid of *B. anthracis* genomes. Amino acid sequence comparison of the *mepH* depicts a truncated protein in B-clade strains as compared to A-clade strains. It is not clear if the truncated proteins (in the B-clade genomes) observed in this study have any functional significance and there is currently limited information on the interactions of chromosomal cell wall structure proteins linked with the *mepH* protein of pXO2. This study suggests that there is a cross-link between the *mepH* of pXO2 with chromosomal cell wall structure proteins. Further experimental investigation can establish any links between the presence of the truncated proteins and the stability of the cell wall on the survival of the strains. Although there were some similarities in the cell wall proteins, changes or mutation of proteins (e.g murein hydrolases) can affect cell wall structure, and thus influence the spore survival in various environments (Vermassen *et al.*, 2019). Any differences may contribute to cell wall biosynthesis and degradation since murein hydrolases are involved in the recycling of peptidoglycan, maintaining cell wall

integrity, and programmed cell death (Vermassen *et al.*, 2019). The structure of the cell wall, the protein composition and their functions/interactions require further investigation. We hypothesize that the difference in spore cell wall structure influences the dominance of A-clade strains over the B-clade strains dominated before the 1990s but since has become almost extinct.

5.4 Genetic characterization of *B. anthracis* A- and B-clade strains in the KNP

5.4.1 Antimicrobial resistance (AMR) genes

Penicillins are effective against most naturally occurring *B. anthracis* strains (Odendaal *et al.*, 1991; Mohammed *et al.*, 2002; Athamna *et al.*, 2004; Luna *et al.*, 2007). Antibiotics such as penicillin and oxytetracycline can be used to treat natural anthrax in large animals (such as cattle, sheep, goats, swine, and horses) in the early stages of the disease (Jensen and Mackey, 1979; Odendaal *et al.*, 1991; WHO, 2008). Since disease progression is quick , antibiotics must be administered as soon as possible. Ampicillin, penicillin and methicillin susceptibility has been observed in *B. anthracis* strains isolated from KNP (Odendaal *et al.*, 1991). This study revealed that all isolates from lineage A and B likely have resistance to fosfomycin, with additional multidrug resistance (MDR) signatures identified in B-clade genomes. Fosfomycin is effective for most Gram-negative bacteria and inhibits cell wall synthesis (Falagas *et al.*, 2019), however, there has been an increase in fosfomycin resistance in bacterial species such as *Escherichia coli* (Kim *et al.*, 1996; Skarzynski *et al.*, 1996). Both endogenous and phage-encoded genes are hypothesized to contribute to the increased resistance of *B. anthracis* to fosfomycin (Sambrook *et al.*, 1989, Kedar *et al.*, 2008).

Studies have shown that the conferred fosfomycin resistance is linked to mutation in UDP-N-Acetylglucosamine enolpyruvyl transferase (MurA) responsible for peptidoglycan (PG) biosynthesis of the bacterial cell wall (Falagas *et al.*, 2019; Skarzynski *et al.*, 1996). The mutated enzyme contributes to the integrity of the cell wall structure and prevents fosfomycin from entering bacterial cells (Falagas *et al.*, 2019). Fosfomycin resistance and

the presence of additional MDR genes enable propagation of bacterial species in different environments (Schuch and Fischetti, 2006). Antimicrobial resistance (AMR) profiles showed presence of fosfomycin resistance in both A- and B-clade strains. However further studies are needed to understand whether the genes are expressed in both lineages and if the genes contribute to the persistence of A-clade isolates over B-clade strains. The presence of the *blaTEM*-116 gene in one of the A-clade strains, isolate 1298 from Pafuri, is peculiar as it has never been reported in the KNP. However, it is not present in other 2015 isolates from the same area. This AMR gene has been previously isolated from strains from Zambia (Bruce *et al.*, 2021), however it was not found in the strains used in this study. The *blaTEM*-116 gene confers resistance to penicillins and broad-spectrum cephalosporins (Weldhagen *et al.*, 2003). Additional studies on the *blaTEM*-116 will provide insight on the relationship between this gene and the overall evolution of *B. anthracis* strains in the KNP.

5.4.2 *B. anthracis* tryptophan operon variability

The amino acid tryptophan requires a lot of energy to synthesise in the cell, thus its metabolic pathway is expected to be tightly controlled. The tryptophan biosynthetic pathway consists of highly conserved multiple enzymatic processes in various microbial genomes. The tryptophan pathway genes vary in arrangements, operon structure and are regulated differently (Priya *et al.*, 2014). Five structural genes (*trpE, trpD, trpC, trpB,* and *trpA*) encode enzymes needed for tryptophan biosynthesis. These genes form a single transcriptional unit, namely, the *trp* operon which is tightly regulated (Yanofsky, 2013). Differences in gene and operon organization reflect evolutionary divergence, adaptability in various environments (Eremenko *et al.*, 2020). Multiple sequence alignment of the *trpA* gene on the KNP *B. anthracis* strains presented a SNP mutation on B-clade strains. The G/T SNP at position 652 bp signals termination of the tryptophan operon in the B-clade strains and thus results in no synthesis of tryptophan. Phylogenetic analysis of this gene clearly separate A and B lineages. This study is supported by previous observations by Emerenko *et al.* (2020), showing variability of SNPs and InDels in *trp* operon genes.

Studies on tryptophan dependence can provide information on intraspecies evolution of *B. anthracis*.

5.4.3 Variation of the *bclA* gene and VNTR copy numbers in *B. anthracis* genomes

The *BclA* glycoprotein is expressed on the outermost layer of *B. anthracis* spores (Thompson *et al.*, 2007; Thompson and Stewart, 2008). It is a key structural component of the spore coat and aids in the recognition and binding of spores to hosts (Wang *et al.*, 2016). Spore binding triggers phagocytosis and the movement of spores across the epithelium (Thompson *et al.*, 2007). Despite its critical function in absorption and spore shape, the *bclA* gene size varies in different strains. Sequence comparison of the BclA protein in this study depicts variability in the CLR of A- and B-clade strains. The length of the GPT repeat sequence in the polymorphic CLR region is responsible for the variation in the spore filament length (Sylvestre *et al.*, 2003).

The phylogenetic analysis based on the *bclA* gene does not easily distinguish *B. anthracis* strains into separate major SNP clades and thus cannot be used independently for the clustering of isolates in A- and B-clades. The number of tandem repeats using MLVA Bams markers, however, can be used to differentiate *B. anthracis* lineages. These markers can distinguish subpopulations with high discriminatory power (Pearson *et al*, 2004; Lista *et al.*, 2006). The results in the current study were supported by Sanger sequencing of VNTR markers related to the *bclA* gene (Bam13 and Bams30) which showed a distinction between different canonical SNP branches of A- and B- clades. Since the exosporium filaments form the spore's outermost structure, variations in BclA protein may influence the spore's characteristics in response to environmental perturbations (Sylvestre *et al.*, 2003). Bacteria are known to use the genetic repertoire to invade and evade host immune response, The variation in the VNTR within a surface structural gene shows that *B. anthracis*, like other pathogens, may employ the variation to regulate its ability for disease transmission and pathogen fitness, thus altering rates of survival or virulence among genetic variants (Shields *et al.*, 1995; van Belkum *et al.*, 1998). It is still unclear how the exosporium structure contribute to the survival of A-clade strains compared to the B-clade. Additionally, further investigation is required to

determine the germination dynamics of *B. anthracis* spores when conditions are favourable.

5.5 Conclusion, recommendations and study limitations

Whole genome sequencing has demonstrated high discriminatory power for *B. anthracis* surveillance in the KNP. Whole genome single nucleotide polymorphism and pan-genomics were successfully employed to define the phylogenetic structure of *B. anthracis* strains from different outbreaks as well as determined their genomic differences. In this study, the ubiquity of the ancient A- clade strains present in both enzootic and non-enzootic regions of the park was revealed. Comparative genomics provided more insight into the predominance of the A.Br.005/006 and the genetic diversity in the KNP. Pan-genomics revealed genetic differences that contribute to the occurrence and persistence of some strains over others that are linked with cell wall biosynthetic genes. The presence of multiple hypothetical proteins and multiple gene copies (and gene variants) in the accessory genome between *B. anthracis* lineages A- and B-clade genomes require further investigation. Characterization of the core and accessory genes will shed light on environmental survival strategies of A- and B- clade lineages. Although the aims of the study were achieved, sequencing of additional *B. anthracis* strains and further analysis of the genome will provide more understanding on anthrax evolution in South Africa.

The study “Comparative genomics of *Bacillus anthracis* strains from anthrax outbreaks in Kruger National Park, South Africa” is part of a bigger study on *B. anthracis* evolution. The work on *bclA* gene variations requires experimental evidence to determine the link to disease transmission and pathogen transmission. Additionally, functional analysis to evaluate the effect of the truncated *trpA* gene on tryptophan synthesis should be done. The use of third generation sequencing technologies for genome wide studies will compliment this study. Further investigation on *B. anthracis* pan-genome, specifically the difference in cell wall biosynthesis pathway will aid in resolving the evolutionary relationship and inform pathogen evolution and survival world-wide**.**

5.6 References

1. Abrami L. Liu S. Cosson P. Leppla S.H. Van der Goot F.G. 2003. Anthrax toxin triggers endocytosis of its receptor via a lipid raft-mediated clathrin-dependent process. Journal of Cell Biology 160 321–328. https://doi.org/10.1083/jcb.200211018

2. Abril J.F. Castellano S. 2019. Genome Annotation in: Ranganathan S. Gribskov M. Nakai K. Schönbach C. (Eds.) Encyclopedia of bioinformatics and computational biology. Academic Press Oxford pp. 195–209. https://doi.org/10.1016/B978-0-12-809633-8.20226-4

3. Achtman M. 2008. Evolution Population Structure and Phylogeography of Genetically Monomorphic Bacterial Pathogens. Annu. Rev. Microbiol. 62 53–70. Beyer W. Bellan S. Eberle G. Ganz H.H. Getz W.M. Haumacher R. Hilss K.A. Kilian W. Lazak J. Turner W.C. Turnbull P.C.B. 2012. Distribution and molecular evolution of *Bacillus anthracis* genotypes in Namibia. PLoS Neglected Tropical Diseases 6. https://doi.org/10.1371/journal.pntd.0001534

4. Adams M.D. Kelley J.M. Gocayne J.D. Dubnick M.A.K. Polymeropoulos M.H. Xiao H. Merril C.R. Wu A. Olde B. Moreno R.F. Kerlavage A.R. McCombie W.R. Venter J.C. 1991. Complementary DNA sequencing: Expressed sequence tags and human genome project. Science 252 1651–1656. https://doi.org/10.1126/science.2047873

5. Allali I. Arnold J.W. Roach J. Cadenas M.B. Butz N. Hassan H.M. Koci M. Ballou A. Mendoza M. Ali R. Azcarate-Peril M.A. 2017. A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. BMC Microbiol 17 194. https://doi.org/10.1186/s12866-017-1101-8

6. Altschul S.F. Gish W. Miller W. Myers E.W. Lipman D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology 215 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

7. Athamna A. Athamna M. Abu-Rashed N. Medlej B. Bast D.J. Rubinstein E. 2004. Selection of *Bacillus anthracis* isolates resistant to antibiotics. Journal of Antimicrobial Chemotherapy 54 424–428. https://doi.org/10.1093/jac/dkh258

8. Aziz R.K. Bartels D. Best 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. 2008. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics 9 1–15. https://doi.org/10.1186/1471-2164-9-75

9. Barandongo Z.R. Mfune J.K.E. Turner W.C. 2018. Dust-Bathing Behaviours Of African herbivores and the potential risk of inhalational anthrax. Journal of Wildlife Diseases 54 34–44. https://doi.org/10.7589/2017-04-069

10. Basson L. Hassim A. Dekker A. Gilbert A. Beyer W. Rossouw J. van Heerden H. 2018. Blowflies as vectors of *Bacillus anthracis* in the Kruger National Park. Koedoe 60 1–6. https://doi.org/10.4102/koedoe.v60i1.1468

11. Beaman T.C. Pankratz H.S. and Gerhardt P.1972. Ultrastructure of the exosporium and underlying inclusions in spores of Bacillus megaterium strains. Journal of Bacteriology 109(3) pp. 1198–1209. Available at: https://doi.org/10.1128/jb.109.3.1198-1209.1972.

12. Benardini J.N. Sawyer J. Venkateswaran K. Nicholson W.L. 2003. Spore UV and acceleration resistance of endolithic *Bacillus pumilus* and *Bacillus subtilis* isolates obtained from Sonoran desert basalt: implications for lithopanspermia. Astrobiology 3 709–717. https://doi.org/10.1089/153110703322736033

13. Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 27 573–580. https://doi.org/10.1093/nar/27.2.573

14. Beyer W. and Turnbull P.C.B. 2009. Anthrax in animals. Molecular Aspects of Medicine 30: 481–489.

15. Beyer W. Bellan S. Eberle G. Ganz H.H. Getz W.M. Haumacher R. Hilss K.A. Kilian W. Lazak J. Turner W.C. Turnbull P.C.B. 2012. Distribution and molecular evolution of

Bacillus anthracis genotypes in Namibia. PLoS Neglected Tropical Diseases 6. https://doi.org/10.1371/journal.pntd.0001534

16. Birdsell D.N. Pearson T. Price E.P. Hornstra H.M. Nera R.D. Stone N. Gruendike J. Kaufman E.L. Pettus A.H. Hurbon A.N. Buchhagen J.L. Harms N.J. Chanturia G. Gyuranecz M. Wagner D.M. Keim P.S. 2012. Melt analysis of mismatch amplification mutation assays (melt-MAMA): A functional study of a cost-effective snp genotyping assay in bacterial models. PLoS ONE 7. https://doi.org/10.1371/journal.pone.0032866

17. Black P.N. DiRusso C.C. 2003. Transmembrane movement of exogenous long-chain fatty acids: proteins enzymes and vectorial esterification. Microbiology and Molecular Biology Reviews 67 454–472. https://doi.org/10.1128/mmbr.67.3.454-472.2003

18. Blackburn J.K. McNyset K.M. Curtis A. Hugh-Jones M.E. 2007. Modeling the Geographic Distribution of *Bacillus anthracis* the Causative Agent of Anthrax Disease for the Contiguous United States using Predictive Ecologic Niche Modeling. The American Journal of Tropical Medicine and Hygiene Am J Trop Med Hyg 77 1103–1110. https://doi.org/10.4269/ajtmh.2007.77.1103

19. Blaustein R.A. McFarland A.G. Maamar S.B. Lopez A. Castro-Wallace S. Hartmann E.M. 2019. Pangenomic Approach To Understanding Microbial Adaptations within a Model Built Environment the International Space Station Relative to Human Hosts and Soil. mSystems 4 e00281-18. https://doi.org/10.1128/mSystems.00281-18

20. Bortolaia V. Kaas R.S. Ruppe E. Roberts M.C. Schwarz S. Cattoir V. Philippon A. Allesoe R.L. Rebelo A.R. Florensa A.F. Fagelhauer L. Chakraborty T. Neumann B. Werner G. Bender J.K. Stingl K. Nguyen M. Coppens J. Xavier B.B. Malhotra-Kumar S. Westh H. Pinholt M. Anjum M.F. Duggett N.A. Kempf I. Nykäsenoja S. Olkkola S. Wieczorek K. Amaro A. Clemente L. Mossong J. Losch S. Ragimbeau C. Lund O. Aarestrup F.M. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. Journal of Antimicrobial Chemotherapy 75 3491–3500. https://doi.org/10.1093/jac/dkaa345

21. Braack L.E.O. De Vos V. 1990. Relation to anthrax transmission in the Kruger national park South Africa 142 141–142.

22. Britannica T. Editors of Encyclopaedia (2018 April 2). bacterial disease. Encyclopedia Britannica. https://www.britannica.com/science/bacterial-disease.

23. Brown E.R. and Cherry W.B. 1951. Specific identification of *Bacillus anthracis* by means of a variant bacteriophage pp. 34–39.

24. Bruce S.A. Huang Y.-H. Kamath P.L. van Heerden H. Turner W.C. 2021. The roles of antimicrobial resistance phage diversity isolation source and selection in shaping the genomic architecture of *Bacillus anthracis*. Microbial Genomics. https://doi.org/10.1099/mgen.0.000616

25. Bruce S.A. Schiraldi N.J. Kamath P.L. Easterday W.R. Turner W.C. 2020. A classification framework for *Bacillus anthracis* defined by global genomic structure. Evolutionary Applications 13 935–944. https://doi.org/10.1111/eva.12911

26. Bruce S.A. Schiraldi N.J. Kamath P.L. Easterday W.R. Turner W.C. 2020. A classification framework for *Bacillus anthracis* defined by global genomic structure. Evolutionary Applications 13 935–944. https://doi.org/10.1111/eva.12911Darling A.C.E. Mau B. Blattner F.R. Perna N.T. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Research 14 1394–1403. https://doi.org/10.1101/gr.2289704

27. Buckling A. Brockhurst M. 2012. Bacteria–Virus Coevolution 347–370. https://doi.org/10.1007/978-1-4614-3567-9

28. Buermans H.P.J. den Dunnen J.T. 2014. Next generation sequencing technology: Advances and applications. Biochimica et Biophysica Acta - Molecular Basis of Disease 1842 1932–1941. https://doi.org/10.1016/j.bbadis.2014.06.015

29. Carlson C.J. Kracalik I.T. Ross N. Alexander K.A. Hugh-Jones M.E. Fegan M. Elkin B.T. Epp T. Shury T.K. Zhang W. Bagirova M. Getz W.M. Blackburn J.K. 2019. The global distribution of *Bacillus anthracis* and associated anthrax risk to humans, livestock and wildlife. Nature Microbiology 4 1337–1343. https://doi.org/10.1038/s41564-019-0435-4

30. Carter K.C. 1988. The Koch-Pasteur dispute on establishing the cause of anthrax. Bulletin of the History of Medicine 62(1) pp.42-57.

31. Chaudry G.J. Moayeri M. Liu S. Leppla S.H. 2002. Quickening the pace of anthrax research: Three advances point towards possible therapies. Trends in Microbiology 10 58–62. https://doi.org/10.1016/S0966-842X(01)02294-6

32. Davies D.G. 1960. The influence of temperature and humidity on spore formation and germination in *Bacillus anthracis*' Journal of Hygiene 58(2) pp. 177–186. Available at: https://doi.org/10.1017/S0022172400038250

33. Davies J. 1996. Origins and evolution of antibiotic resistance. Microbiología (Madrid Spain) 12 9–16. https://doi.org/10.1128/mmbr.00016-10

34. de Vos V. 1990. The Epidemiology / Ecology of anthrax Eco-historical perspective. https://repository.up.ac.za/bitstream/handle/2263/74462/2020_WAHVM_DeVosValerius.pdf?

35. de Vos V. 1990. The Epidemiology/ Ecology of Anthrax. Eco-historical perspective.

36. de Vos V. Turnbull P.C.B. 2004. Anthrax in J.A.W. Coetzer & R.C. Tustin (eds.) Infectious Diseases of Livestock 3rd edn. vol. 3 pp. 1788–1818 Oxford University Press Cape Town South Africa.

37. Depristo M.A. Banks E. Poplin R. Garimella K.V. Maguire J.R. Hartl C. Philippakis A.A. Del Angel G. Rivas M.A. Hanna M. McKenna A. Fennell T.J. Kernytsky A.M. Sivachenko A.Y. Cibulskis K. Gabriel S.B. Altshuler D. Daly M.J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43 491–501. https://doi.org/10.1038/ng.806

38. Derzelle S. Girault G. Kokotovic B. Angen Ø. 2015. Whole genome-sequencing and phylogenetic analysis of a historical collection of *Bacillus anthracis* strains from Danish cattle. PLoS ONE 10 1–11. https://doi.org/10.1371/journal.pone.0134699

39. Dirckx J.H. 1981. Virgil on anthrax. The American Journal of Dermatopathology 3.

40. Dixon T.C. Meselson M. Guillemin J. Hanna P.C. 1999. Anthrax. New England Journal of Medicine 341 815–826. https://doi.org/10.1056/NEJM199909093411107

41. Dong Y. Du H. Gao C. Ma T. Feng L. 2012. Characterization of two long-chain fatty acid CoA ligases in the Gram-positive bacterium *Geobacillus thermodenitrificans* NG80-2. Microbiological Research 167 602–607. https://doi.org/10.1016/j.micres.2012.05.001

42. Dragon D.C. Rennie R.P. 1995. The ecology of anthrax spores: Tough but not invincible Can Vet J.

43. Driks A. 2009. Proteins of the spore core and coat 2002. In: Sonenshein AL Hoch JA Losick R editors. Bacillus subtilis and its closest relatives. pp. 527–536 American Society for Microbiology; Washington D.C.

44. Ellerbrok H. Nattermann H. Ozel M. Beutin L. Appel B. Pauli G. 2002. Rapid and sensitive identification of pathogenic and apathogenic *Bacillus anthracis* by real-time PCR. FEMS microbiology letters 214 1 51–9.

45. Eremenko E.I. Bobrysheva O. Pisarenko S. Ryazanova A. Semenova O. Aksenova L. Kovalev D. Pechkovskii G. Evchenko A. Kulichenko A. 2020. Variability of the *Bacillus anthracis* tryptophan operon 1–19. https://doi.org/10.21203/rs.3.rs-21866/v1

46. Færgeman N.J. DiRusso C.C. Elberger A. Knudsen J. Black P.N. 1997. Disruption of the Saccharomyces cerevisiae homologue to the murine fatty acid transport protein impairs uptake and growth on long-chain fatty acids. Journal of Biological Chemistry 272 8531–8538. https://doi.org/10.1074/jbc.272.13.8531

47. Falagas M.E. Athanasaki F. Voulgaris G.L. Triarides N.A. Vardakas K.Z. 2019. Resistance to fosfomycin: mechanisms frequency and clinical consequences. Int J Antimicrob Agents 53 22–28. https://doi.org/10.1016/j.ijantimicag.2018.09.013

48. Feldgarden M. Brover V. Haft D.H. Prasad A.B. Slotta D.J. Tolstoy I. Tyson G.H. Zhao S. Hsu C.-H. McDermott P.F. Tadesse D.A. Morales C. Simmons M. Tillman G. Wasilenko J. Folster J.P. Klimke W. 2019. Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. Antimicrob Agents Chemother 63. https://doi.org/10.1128/AAC.00483-19

49. Fischer S. Brunk B.P. Chen F. Gao X. Harb O.S. Iodice J.B. Shanmugam D. Roos D.S. Stoeckert C.J.J. 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr Protoc Bioinformatics Chapter 6 6.12.1-6.12.19. https://doi.org/10.1002/0471250953.bi0612s35

50. Fisher N. Hanna P. 2005. Characterization of *Bacillus anthracis* germinant receptors in vitro. J Bacteriol 187 8055–8062. https://doi.org/10.1128/JB.187.23.8055-8062.2005

51. Flicek P. Birney E. 2009. Sense from sequence reads: methods for alignment and assembly. Nature Methods 6 S6–S12. https://doi.org/10.1038/nmeth.1376

52. Fozo E.M. Makarova K.S. Shabalina S.A. Yutin N. Koonin E.V. Storz G. 2010. Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. Nucleic Acids Res 38 3743–3759. https://doi.org/10.1093/nar/gkq054

53. Froude J.W. Thullier P. Pelat T. 2011. Antibodies against anthrax: Mechanisms of action and clinical applications. Toxins 3 1433–1452. https://doi.org/10.3390/TOXINS3111433

54. Gachohi J.M. Gakuya F. Lekolool I. Osoro E. Nderitu L. Munyua P. Ngere I. Kemunto N. Bett B. Otieno F. Muturi M. Mwatondo A. Widdowson M.A. Kariuki Njenga M. 2019. Temporal and spatial distribution of anthrax outbreaks among Kenyan wildlife 1999–2017. Epidemiology and Infection 147. https://doi.org/10.1017/S0950268819001304

55. Gurevich A. Saveliev V. Vyahhi N. Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

56. Hammamieh R. Ribot W.J. Abshire T.G. Jett M. Ezzell J. 2008. Activity of the *Bacillus anthracis* 20 kDa protective antigen component. BMC Infectious Diseases 8. https://doi.org/10.1186/1471-2334-8-124

57. Hammond K.A. Diamond J. 1984. in Humans and Animals.

58. Hampson K. Lembo T. Bessell P. Auty H. Packer C. Halliday J. Beesley C.A. Fyumagwa R. Hoare R. Ernest E. Mentzel C. Metzger K.L. Mlengeya T. Stamey K.

Roberts K. Wilkins P.P. Cleaveland S. 2011. Predictability of anthrax infection in the Serengeti Tanzania. Journal of Applied Ecology 48 1333–1344. https://doi.org/10.1111/j.1365-2664.2011.02030.x

59. Han C.S. Xie G. Challacombe J.F. Altherr M.R. Bhotika S.S. Brown N. Bruce D. Campbell C.S. Campbell M.L. Chen J. Chertkov O. Cleland C. Dimitrijevic M. Doggett N.A. Fawcett J.J. Glavina T. Goodwin L.A. Green L.D. Hill K.K. Hitchcock P. Jackson P.J. Keim P. Kewalramani A.R. Longmire J. Lucas S. Malfatti S. McMurry K. Meincke L.J. Misra M. Moseman B.L. Mundt M. Munk A.C. Okinaka R.T. Parson-Quintana B. Reilly L.P. Richardson P. Robinson D.L. Rubin E. Saunders E. Tapia R. Tesmer J.G. Thayer N. Thompson L.S. Tice H. Ticknor L.O. Wills P.L. Brettin T.S. Gilna P. 2006. Erratum: Pathogenomic sequence analysis of Bacillus cereus and Bacillus thuringiensis isolates closely related to *Bacillus anthracis* (Journal of Bacteriology (2006) 188 9 (3382-3390)). Journal of Bacteriology 188 7711. https://doi.org/10.1128/JB.01430-06

60. Hardison R.C. 2003. Comparative genomics. PLoS Biol 1 E58. https://doi.org/10.1371/journal.pbio.0000058

61. Harrell L.J. Andersen G.L. Wilson K.H. 1995. Genetic variability of *Bacillus anthracis* and related species. J Clin Microbiol 33 1847–1850. https://doi.org/10.1128/jcm.33.7.1847-1850.1995

62. Hassim A. 2016. Distribution and molecular characterization of South African *Bacillus anthracis* strains and their associated bacteriophages. PhD University of Pretoria.

63. Hassim A. Dekker E.H. Byaruhanga C. Reardon T. van Heerden H. 2017. A retrospective study of anthrax on the Ghaap Plateau Northern Cape province of South Africa with special reference to the 2007–2008 outbreaks. Onderstepoort Journal of Veterinary Research 84. https://doi.org/10.4102/ojvr.v84i1.1414

64. Head S.R. Kiyomi Komori H. LaMere S.A. Whisenant T. Van Nieuwerburgh F. Salomon D.R. Ordoukhanian P. 2014. Library construction for next-generation sequencing: Overviews and challenges. BioTechniques 56 61–77. https://doi.org/10.2144/000114133

65. Henton M. Briers G. Anthrax spores isolated above burning carcases. OIE International Congress with WHO-Cosponsorship on Anthrax Brucellosis CBPP Clostrial and Mycobacterial diseases Berg-en-Dal (South Africa) 9-15 Aug 1998 1998. Onderstepoort Veterinary Inst

66. Huang Y.-H. Kausrud K. Hassim A. Ochai S.O. van Schalkwyk O.L. Dekker E.H. Buyantuev A. Cloete C.C. Kilian J.W. Mfune J.K.E. Kamath P.L. van Heerden H. Turner W.C. 2022. Environmental drivers of biseasonal anthrax outbreak dynamics in two multihost savanna systems. Ecological Monographs n/a e1526. https://doi.org/10.1002/ecm.1526

67. Hugh-Jones M. Blackburn J. 2009. The ecology of *Bacillus anthracis*. Molecular Aspects of Medicine 30 356–367. https://doi.org/10.1016/j.mam.2009.08.003

68. Hugh-Jones M.E. de Vos V. 2002. Anthrax and wildlife. Revue Scientifique et Technique de l'OIE 21 359–383. https://doi.org/10.20506/rst.21.2.1336

69. Hutchison C.A. 2007. DNA sequencing: Bench to bedside and beyond. Nucleic Acids Research 35 6227–6237. https://doi.org/10.1093/nar/gkm688

70. Inglin R.C. Meile L. Stevens M.J.A. 2018. Clustering of Pan- and Core-genome of Lactobacillus provides Novel Evolutionary Insights for Differentiation. BMC Genomics 19 284. https://doi.org/10.1186/s12864-018-4601-5

71. Ireland J.A.W. Hanna P.C. 2002. Amino acid- and purine ribonucleoside-induced germination of *Bacillus anthracis* ΔSterne endospores: gerS mediates responses to aromatic ring structures. Journal of Bacteriology 184 1296–1303. https://doi.org/10.1128/JB.184.5.1296-1303.2002

72. Jensen R. & Mackey D.R. 1979. Diseases of feedlot cattle. 3rd ed. Lea & Febiger Publ. Philadelphia. pp. 59-65; 75-84; 89-95; 142-146; 167-173.

73. Jung K.H. Kim S.H. Kim S.K. Cho S.Y. Chai J.C. Lee Y.S. Kim J.C. Kim S.J. Oh H.B. Chai Y.G. 2012. Genetic populations of *Bacillus anthracis* isolates from Korea. Journal of Veterinary Science 13 385–393. https://doi.org/10.4142/jvs.2012.13.4.385

74. Katoh K. Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution 30 772–780. https://doi.org/10.1093/molbev/mst010

75. Kedar G.C. Brown-Driver V. Reyes D.R. Hilgers M.T. Stidham M.A. Shaw K.J. Finn J. Haselbeck R.J. 2008. Comparison of the essential cellular functions of the two murA genes of *Bacillus anthracis*. Antimicrob Agents Chemother 52 2009–2013. https://doi.org/10.1128/AAC.01594-07

76. Keim P. Gruendike J.M. Klevytska A.M. Schupp J.M. Challacombe J. Okinaka R. 2009. The genome and variation of *Bacillus anthracis*. Molecular Aspects of Medicine 30 397–405. https://doi.org/10.1016/j.mam.2009.08.005

77. Keim P. Kalif A. Schupp J. Hill K. Travis S.E. Richmond K. Adair D.M. Hugh-Jones M. Kuske C.R. Jackson P. 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. Journal of Bacteriology 179 818–824. https://doi.org/10.1128/jb.179.3.818-824.1997

78. Keim P. Klevytska A.M. Price L.B. Schupp J.M. Zinser G. Smith K.L. Hugh-Jones M.E. Okinaka R. Hill K.K. Jackson P.J. 1999. Molecular diversity in *Bacillus anthracis*. J Appl Microbiol 87 215–217. https://doi.org/10.1046/j.1365-2672.1999.00873.x

79. Keim P. Price L.B. Klevytska A.M. Smith K.L. Schupp J.M. Okinaka R. Jackson P.J. Hugh-Jones M.E. 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. Journal of Bacteriology 182 2928–2936. https://doi.org/10.1128/JB.182.10.2928-2936.2000

80. Kim D.H. Lees W.J. Kempsell K.E. Lane W.S. Duncan K. Walsh C.T. 1996. Characterization of a Cys115 to Asp substitution in the *Escherichia coli* cell wall biosynthetic enzyme UDP-GlcNAc enolpyruvyl transferase (*MurA*) that confers resistance to inactivation by the antibiotic fosfomycin. Biochemistry 35 4923–4928. https://doi.org/10.1021/bi952937w

81. Klemm D.M. Klemm W.R. 1959. A history of anthrax. J Am Vet Med Assoc 135 458–462.

82. Knudsen S.M. Cermak N. Delgado F.F. Setlow B. Setlow P. Manalis S.R. 2016. Water and small-molecule permeation of dormant *Bacillus subtilis* spores. Journal of Bacteriology 198 168–177. https://doi.org/10.1128/JB.00435-15

83. Koboldt D.C. Steinberg K.M. Larson D.E. Wilson R.K. Mardis E.R. 2013. The next-generation sequencing revolution and its impact on genomics. Cell 155 27–38. https://doi.org/10.1016/j.cell.2013.09.006

84. Koskella B. Brockhurst M.A. 2014. Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS Microbiol Rev 38 916–931. https://doi.org/10.1111/1574-6976.12072

85. Kumar S. Stecher G. Li M. Knyaz C. Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35 1547–1549. https://doi.org/10.1093/molbev/msy096

86. Kumar S. Stecher G. Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 33 1870–1874. https://doi.org/10.1093/molbev/msw054

87. La Duc M.T. Kern R. Venkateswaran K. 2004. Microbial monitoring of spacecraft and associated environments. Microb Ecol 47 150–158. https://doi.org/10.1007/s00248-003-1012-0

88. Lahens N.F. Ricciotti E. Smirnova O. Toorens E. Kim E.J. Baruzzo G. Hayer K.E. Ganguly T. Schug J. Grant G.R. 2017. A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. BMC Genomics 18 602. https://doi.org/10.1186/s12864-017-4011-0

89. Laing C. Buchanan C. Taboada E.N. Zhang Y. Kropinski A. Villegas A. Thomas J.E. 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. https://doi.org/10.1186/1471-2105-11-461

90. Le Flèche P. Hauck Y. Onteniente L. Prieur A. Denoeud F. Ramisse V. Sylvestre P. Benson G. Ramisse F. Vergnaud G. 2001. A tandem repeats database for bacterial

genomes: Application to the genotyping of Yersinia pestis and *Bacillus anthracis*. BMC Microbiology 1 1–14. https://doi.org/10.1186/1471-2180-1-2

91. Ledwaba M.B. 2014. Molecular characterization of *Bacillus anthracis* and brucella species from southern Africa. MSc Dissertation University of Pretoria. viewed 20220604. <http://hdl.handle.net/2263/79267>

92. Lekota K.E. 2018. Genomic study of *Bacillus anthracis* and Bacillus species isolated from anthrax cases in South Africa. PhD University of Pretoria.

93. Lekota K.E. Hassim A. Madoroba E. Hefer C.A. van Heerden H. 2020. Phylogenomic structure of *Bacillus anthracis* isolates in the Northern Cape Province South Africa revealed novel single nucleotide polymorphisms. Infection Genetics and Evolution 80 104146. https://doi.org/10.1016/j.meegid.2019.104146

94. Lekota K.E. Hassim A. Mafofo J. Rees J. Muchadeyi F.C. van Heerden H. Madoroba E. 2016. Polyphasic characterization of Bacillus species from anthrax outbreaks in animals from South Africa and Lesotho. Journal of Infection in Developing Countries 10 814–823. https://doi.org/10.3855/jidc.7798

95. Lekota K.E. Hassim A. Rogers P. Dekker E.H. Last R. de Klerk-Lorist L. van Heerden H. 2018. The reporting of a *Bacillus anthracis* B-clade strain in South Africa after more than 20 years. BMC Res Notes 11 264. https://doi.org/10.1186/s13104-018-3366-x

96. Leppla S.H. 1982. Anthrax toxin edema factor: A bacterial adenylate cyclase that increases cyclic AMP concentrations in eukaryotic cells. Proceedings of the National Academy of Sciences of the United States of America 79 3162–3166. https://doi.org/10.1073/pnas.79.10.3162

97. Letunic I. Bork P. 2019. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Research 47 256–259. https://doi.org/10.1093/nar/gkz239

98. Li H. Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

99. Li H. Handsaker B. Wysoker A. Fennell T. Ruan J. Homer N. Marth G. Abecasis G. Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

100. Lista F. Faggioni G. Valjevac S. Ciammaruconi A. Vaissaire J. Le Doujet C. Gorgé O. De Santis R. Carattoli A. Ciervo A. Fasanella A. Orsini F. D'Amelio R. Pourcel C. Cassone A. Vergnaud G. 2006. Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci Multiple Locus Variable-Number Tandem Repeats Analysis. BMC Microbiology 6. https://doi.org/10.1186/1471-2180-6-33

101. Liu H. Bergman N.H. Thomason B. Shallom S. Hazen A. Crossno J. Rasko D.A. Ravel J. Read T.D. Peterson S.N. Yates J. Hanna P.C. 2004. Formation and composition of the *Bacillus anthracis* endospore. Journal of Bacteriology 186 164–178. https://doi.org/10.1128/JB.186.1.164-178.2004

102. Lobo I. 2008. Basic Local Alignment Search Tool (BLAST). Nature Education 1(1):215

103. Luna V.A. King D.S. Gulledge J. Cannons A.C. Amuso P.T. Cattani J. 2007. Susceptibility of *Bacillus anthracis* Bacillus cereus Bacillus mycoides Bacillus pseudomycoides and Bacillus thuringiensis to 24 antimicrobials using Sensititre® automated microbroth dilution and Etest® agar gradient diffusion methods. Journal of Antimicrobial Chemotherapy 60 555–567. https://doi.org/10.1093/jac/dkm213

104. Luo H. Ribeiro A. Folador C. Singhal N. Wu Y. Zaiden N. Cao B. 2018. The Core- and Pan-Genomic Analyses of the Genus Comamonas: From Environmental Adaptation to Potential Virulence. https://doi.org/10.3389/fmicb.2018.03096Mardis E.R. 2013. Next-generation sequencing platforms. Annual Review of Analytical Chemistry 6 287–303. https://doi.org/10.1146/annurev-anchem-062012-092628

105. McBride B.W. Turnbull P.C.B. 1998. Bacillus Infection and Immunity in: Delves P.J. (Ed.) Encyclopedia of Immunology (Second Edition). Elsevier Oxford pp. 311–315. https://doi.org/10.1006/rwei.1999.0080

106. McKenna A. Hanna M. Banks E. Sivachenko A. Cibulskis K. Kernytsky A. Garimella K. Altshuler D. Gabriel S. Daly M. DePristo M.A. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20 1297–1303. https://doi.org/10.1101/gr.107524.110

107. Medini D. Donati C. Tettelin H. Masignani V. Rappuoli R. 2005. The microbial pan-genome. Current Opinion in Genetics & Development 15 589–594. https://doi.org/10.1016/j.gde.2005.09.006

108. Merino E. Jensen R.A. Yanofsky C. 2008. Evolution of bacterial trp operons and their regulation. Current Opinion in Microbiology 11 1–13.

109. Miller J.R. Koren S. Sutton G. 2010. Assembly algorithms for next-generation sequencing data. Genomics 95 315–327. https://doi.org/10.1016/j.ygeno.2010.03.001

110. Minett F.C. 1950. Sporulation and viability of *B. anthracis* in relation to environmental temperature and humidity. Journal of Comparative Pathology and Therapeutics 60 161–176. https://doi.org/10.1016/S0368-1742(50)80016-4

111. Mock M. Fouet A. 2001. Anthrax. Annual Reviews in Microbiology 55 647–671.

112. Mohammed M.J. Marston C.K. Popovic T. Weyant R.S. Tenover F.C. 2002. Antimicrobial susceptibility testing of *Bacillus anthracis*: Comparison of results obtained by using the National Committee for Clinical Laboratory Standards broth microdilution reference and etest agar gradient diffusion methods. Journal of Clinical Microbiology 40 1902–1907. https://doi.org/10.1128/JCM.40.6.1902-1907.2002

113. Moir A. 2003. Bacterial spore germination and protein mobility. Trends in Microbiology 11 452–454. https://doi.org/10.1016/j.tim.2003.08.001

114. Nasir A. Kim K.M. Caetano-Anollés G. 2017. Long-term evolution of viruses: A Janus-faced balance. BioEssays 39. https://doi.org/10.1002/bies.201700026

115. Nicholson W.L. Galeano B. 2003. UV resistance of *Bacillus anthracis* spores revisited: validation of *Bacillus subtilis* spores as UV surrogates for spores of *B. anthracis* Sterne. Appl Environ Microbiol 69 1327–1330. https://doi.org/10.1128/AEM.69.2.1327-1330

116. Odendaal M.W. Pieterson P.M. de Vos V. Botha A.D. 1991. The antibiotic sensitivity patterns of *Bacillus anthracis* isolated from the Kruger National Park. The Onderstepoort journal of veterinary research 58 17–19.

117. Overbeek R. Begley T. Butler R.M. Choudhuri J.V. Chuang H.Y. Cohoon M. de Crécy-Lagard V. Diaz N. Disz T. Edwards R. Fonstein M. Frank E.D. Gerdes S. Glass E.M. Goesmann A. Hanson A. Iwata-Reuyl D. Jensen R. Jamshidi N. Krause L. Kubal M. Larsen N. Linke B. McHardy A.C. Meyer F. Neuweger H. Olsen G. Olson R. Osterman A. Portnoy V. Pusch G.D. Rodionov D.A. Rül;ckert C. Steiner J. Stevens R. Thiele I. Vassieva O. Ye Y. Zagnitko O. Vonstein V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Research 33 5691–5702. https://doi.org/10.1093/nar/gki866

118. 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. 2008. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics 9 1–15.

119. Page A.J. Cummins C.A. Hunt M. Wong V.K. Reuter S. Holden M.T.G. Fookes M. Falush D. Keane J.A. Parkhill J. 2015. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics 31 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

120. Page R. Peti W. 2016. Toxin-antitoxin systems in bacterial growth arrest and persistence. Nat Chem Biol 12 208–214. https://doi.org/10.1038/nchembio.2044

121. Parks D.H. Imelfort M. Skennerton C.T. Hugenholtz P. Tyson G.W. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates single cells and metagenomes. https://doi.org/10.1101/gr.186072.114

122. Parsons M. McLoughlin C.A. Kotschy K.A. Rogers K.H. Rountree M.W. 2005. The effects of extreme floods on the biophysical heterogeneity of river landscapes. Frontiers in Ecology and the Environment 3 487–494. https://doi.org/10.1890/1540-9295(2005)003[0487:TEOEFO]2.0.CO;2

123. Pearson T. Busch J.D. Ravel J. Read T.D. Rhoton S.D. U'Ren J.M. Simonson T.S. Kachur S.M. Leadem R.R. Cardon M.L. Van Ert M.N. Huynh L.Y. Fraser C.M. Keim P.

2004. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. Proceedings of the National Academy of Sciences of the United States of America 101 13536–13541. https://doi.org/10.1073/pnas.0403844101

124. Petosa C. Collier R.J. Klimpel K.R. Leppla S.H. Liddington R.C. 2013. Crystal structure of the anthrax toxin protective antigen. https://doi.org/10.1038/385833a0

125. Pezard C. Berche P. Mock M. 1991. Contribution of individual toxin components to virulence of *Bacillus anthracis* Infection and Immunity. https://doi.org/10.1128/iai.59.10.3472-3477.1991

126. Pienaar U.D. 1967. Epidemiology of anthrax in wild animals and the control of anthrax epizootics in the Kruger National Park South Africa. Federation proceedings 26 1496–1502.

127. Pilo P. Frey J. 2011. *Bacillus anthracis*: Molecular taxonomy population genetics phylogeny and patho-evolution. Infection Genetics and Evolution 11 1218–1224. https://doi.org/10.1016/j.meegid.2011.05.013

128. Pilo P. Frey J. 2018. Pathogenicity population genetics and dissemination of *Bacillus anthracis*. Infection Genetics and Evolution 64 115–125. https://doi.org/10.1016/j.meegid.2018.06.024

129. Prince A.S. 2003. The host response to anthrax lethal toxin: Unexpected observations. Journal of Clinical Investigation 112 656–658. https://doi.org/10.1172/JCI200319581

130. Priya V.K. Sarkar S. Sinha S. 2014. Evolution of tryptophan biosynthetic pathway in microbial genomes: A comparative genetic study. Systems and Synthetic Biology 8 59–72. https://doi.org/10.1007/s11693-013-9127-1

131. Rambaut A. 2009. FigTree v1.2.2. <http://tree.bio.ed.ac.uk/software/figtree/>

132. Ramisetty B.C.M. Sudhakari P.A. 2019. Bacterial 'Grounded' Prophages: Hotspots for Genetic Renovation and Innovation. Front. Genet. 10 65. https://doi.org/10.3389/fgene.2019.00065

133. Rasko D.A. Altherr M.R. Han C.S. Ravel J. 2005. Genomics of the Bacillus cereus group of organisms. FEMS Microbiology Reviews 29 303–329. https://doi.org/10.1016/j.femsre.2004.12.005

134. Rasko D.A. Rosovitz M.J. Myers G.S.A. Mongodin E.F. Fricke W.F. Gajer P. Crabtree J. Sebaihia M. Thomson N.R. Chaudhuri R. Henderson I.R. Sperandio V. Ravel J. 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. J Bacteriol 190 6881–6893. https://doi.org/10.1128/JB.00619-08

135. Ravel J. Jiang L. Stanley S.T. Wilson M.R. Decker R.S. Read T.D. Worsham P. Keim P.S. Salzberg S.L. Fraser-Liggett C.M. Rasko D.A. 2009. The complete genome sequence of *Bacillus anthracis* Ames "Ancestor." Journal of Bacteriology 91 445–446. https://doi.org/10.1128/JB.01347-08

136. Read T.D. Peterson S.N. Tourasse N. Baillie L.W. Paulsen I.T. Nelson K.E. Tettelin H. Fouts D.E. Eisen J.A. Gill S.R. Holtzapple E.K. Økstad O.A. Helgason E. Rilstone J. Wu M. Kolonay J.F. Beanan M.J. Dodson R.J. Brinkac L.M. Gwinn M. DeBoy R.T. Madpu R. Daugherty S.C. Durkin A.S. Haft D.H. Nelson W.C. Peterson J.D. Pop M. Khouri H.M. Radune D. Benton J.L. Mahamoud Y. Jiang L. Hance I.R. Weidman J.F. Berry K.J. Plaut R.D. Wolf A.M. Watkins K.L. Nierman W.C. Hazen A. Cline R. Redmond C. Thwaite J.E. White O. Salzberg S.L. Thomason B. Friedlander A.M. Koehler T.M. Hanna P.C. Kolstø A.B. Fraser C.M. 2003. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature 423 81–86. https://doi.org/10.1038/nature01586

137. Rhoads A. Au K.F. 2015. PacBio Sequencing and its applications. Genomics Proteomics & Bioinformatics 13 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

138. Riedel S. 2005. Anthrax: A Continuing Concern in the Era of Bioterrorism. Baylor University Medical Center Proceedings 18 234–243. https://doi.org/10.1080/08998280.2005.11928074

139. Roossinck M.J. 2011. The good viruses: viral mutualistic symbioses. Nat Rev Microbiol 9 99–108. https://doi.org/10.1038/nrmicro2491

140. Rouli L. MBengue M. Robert C. Ndiaye M. La Scola B. Raoult D. 2014. Genomic analysis of three African strains of *Bacillus anthracis* demonstrates that they are part of the clonal expansion of an exclusively pathogenic bacterium. New Microbes and New Infections 2 161–169. https://doi.org/10.1002/nmi2.62

141. Sahl J.W. Pearson T. Okinaka R. Schupp J.M. Gillece J.D. Heaton H. Birdsell D. Hepp C. Fofanov V. Noseda R. Fasanella A. Hoffmaster A. Wagner D.M. Keim P. 2016. A *Bacillus anthracis* genome sequence from the Sverdlovsk 1979 autopsy specimens. mBio 7. https://doi.org/10.1128/mBio.01501-16

142. Salzberg S.L. 2019. Next-generation genome annotation: we still struggle to get it right. Genome Biology 20 92. https://doi.org/10.1186/s13059-019-1715-2

143. Sambrook K. E. Fritsch and T. Maniatis. 1989. Molecular cloning: a laboratory manual 2nd ed. Cold Spring Harbor Laboratory Press Cold Spring Harbor NY.

144. Sanger F. Air G. M. Barrell B. G. Brown N. L. Coulson A. R. Fiddes J. C. Hutchison C. A. Slocombe P. M. Smith M. 1977. Nucleotide sequence of bacteriophage phi X174 DNA. Nature 265.

145. Sarda S. Hannenhalli S. 2014. Next-generation sequencing and epigenomics research: A Hammer in Search of Nails. Genomics & Informatics 12 2. https://doi.org/10.5808/gi.2014.12.1.2

146. Schmid G. Kaufmann A. 2002. Anthrax in Europe: Its epidemiology clinical characteristics and role in bioterrorism. Clinical Microbiology and Infection 8 479–488. https://doi.org/10.1046/j.1469-0691.2002.00500.x

147. Schuch R. Fischetti V.A. 2006. Detailed genomic analysis of the Wbeta and gamma phages infecting *Bacillus anthracis*: implications for evolution of environmental fitness and antibiotic resistance. J Bacteriol 188 3037–3051. https://doi.org/10.1128/JB.188.8.3037-3051.2006

148. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

149. Seemann T. 2017. Shovill: Faster SPAdes assembly of Illumina reads.

150. Segerman B. 2020. The most frequently used sequencing technologies and assembly methods in different time segments of the bacterial surveillance and RefSeq genome databases. Front. Cell. Infect. Microbiol. 10 527102. https://doi.org/10.3389/fcimb.2020.527102

151. Setlow P. 2003. Spore germination. Current opinion in microbiology 6 6 550–6.

152. Shendure J. Ji H. 2008. Next-generation DNA sequencing. Nature Biotechnology 26 1135–1145. https://doi.org/10.1038/nbt1486

153. Sherer K. Li Y. Cui X. Eichacker P.Q. 2007. Lethal and edema toxins in the pathogenesis of *Bacillus anthracis* septic shock: Implications for therapy. American Journal of Respiratory and Critical Care Medicine 175 211–221. https://doi.org/10.1164/rccm.200608-1239CP

154. Shields D.C. McDevitt D. Foster T.J. 1995. Evidence against concerted evolution in a tandem array in the clumping factor gene of Staphylococcus aureus. Mol Biol Evol 12 963–965. https://doi.org/10.1093/oxfordjournals.molbev.a040275

155. Sinha D. Sun X. Khare M. Drancourt M. Raoult D. Fournier P.-E. 2021. Pangenome analysis and virulence profiling of Streptococcus intermedius. BMC Genomics 22 522. https://doi.org/10.1186/s12864-021-07829-2

156. Sitto F. Battistuzzi F.U. 2020. Estimating Pangenomes with Roary. Molecular Biology and Evolution 37 933–939. https://doi.org/10.1093/molbev/msz284

157. Skarzynski T. Mistry A. Wonacott A. Hutchinson S.E. Kelly V.A. Duncan K. 1996. Structure of UDP-N-acetylglucosamine enolpyruvyl transferase an enzyme essential for the synthesis of bacterial peptidoglycan complexed with substrate UDP-N-acetylglucosamine and the drug fosfomycin. Structure 4 1465–1474. https://doi.org/10.1016/s0969-2126(96)00153-0

158. Smith K.L. 1999. Epidemiology of Anthrax in the Kruger National Park South Africa: Genetic Diversity and Environment. LSU Historical Dissertations and Theses. 6962. https://digitalcommons.lsu.edu/gradschool_disstheses/6962

159. Smith K.L. DeVos V. Bryden H. Price L.B. Hugh-Jones M.E. Keim P. 2000. *Bacillus anthracis* diversity in Kruger National Park. Journal of Clinical Microbiology 38 3780–3784. https://doi.org/10.1128/jcm.38.10.3780-3784.2000

160. Spencer R.C. 2003. *Bacillus anthracis*. Journal of Clinical Pathology 56 182–187. https://doi.org/10.1136/jcp.56.3.182

161. Steenkamp P.J. Van Heerden H. Van Schalkwyk L. 2018. Ecological suitability modeling for anthrax in the Kruger National Park South Africa. https://doi.org/10.1371/journal.pone.0191704

162. Sterne M. 1937. Variation in *Bacillus anthracis*. Onderstepoort Journal of Veterinary Science and Animal Industry 8 271–348.

163. Sterne M. 1967. Distribution and economic importance of anthrax. Federation proceedings 26 1493-1495.Su Z. Ning B. Fang H. Hong H. Perkins R. Tong W. Shi L. 2011. Next-generation sequencing and its applications in molecular diagnostics. Expert Review of Molecular Diagnostics 11 333–343. https://doi.org/10.1586/erm.11.3

164. Suchard M.A. Lemey P. Baele G. Ayres D.L. Drummond A.J. Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evolution 4 1–5. https://doi.org/10.1093/ve/vey016

165. Swick M.C. Koehler T.M. and Driks A.2016 Surviving between hosts: Sporulation and transmission' virulence mechanisms of bacterial pathogens 4(4) pp. 567–591. Available at: https://doi.org/10.1128/9781555819286.ch20.

166. Sylvestre P. Couture-Tosi E. Mock M. 2003. Polymorphism in the collagen-like region of the *Bacillus anthracis* BclA protein leads to variation in exosporium filament length. Journal of Bacteriology 185 1555–1563. https://doi.org/10.1128/JB.185.5.1555-1563.2003

167. Tang W.J. Guo Q. 2009. The adenylyl cyclase activity of anthrax edema factor. Molecular Aspects of Medicine 30 423–430. https://doi.org/10.1016/j.mam.2009.06.001

168. 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. Fraser C.M. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A 102 13950–13955. https://doi.org/10.1073/pnas.0506758102

169. Thierry S. Tourterel C. Le Flèche P. Derzelle S. Dekhil N. Mendy C. Colaneri C. Vergnaud G. Madani N. 2014. Genotyping of French *Bacillus anthracis* strains based on 31-loci multi locus VNTR analysis: epidemiology marker evaluation and update of the internet genotype database. PLoS One 9 e95131. https://doi.org/10.1371/journal.pone.0095131

170. Thompson B.M. Stewart G.C. 2008. Targeting of the BclA and BclB proteins to the *Bacillus anthracis* spore surface. Mol Microbiol 70 421–434. https://doi.org/10.1111/j.1365-2958.2008.06420.x

171. Thompson B.M. Waller L.N. Fox K.F. Fox A. Stewart G.C. 2007. The BclB glycoprotein of *Bacillus anthracis* is involved in exosporium integrity. Journal of Bacteriology 189 6704–6713. https://doi.org/10.1128/JB.00762-07

172. Turnbull P.C.B. 1996. Bacillus. In: Baron S editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; Chapter 15.

173. Turnbull P. C. B. 2008. Anthrax in humans and animals World Health Organization.

174. Turnbull P.C.B WHO 1998. Guidelines for the surveillance and control of anthrax in humans and animals.

175. Turnbull P.C.B. Shadomy S.V. 2010. Anthrax from 5000BC to AD 2010. In: *Bacillus anthracis* and Anthrax. NH Bergman ed. Hoboken NJ: Wiley-Blackwell.

176. Turner W.C. Kausrud K.L. Beyer W. Easterday W.R. Barandongo Z.R. Blaschke E. Cloete C.C. Lazak J. Van Ert M.N. Ganz H.H. Turnbull P.C.B. Stenseth N.Chr. Getz W.M.

2016. Lethal exposure: An integrated approach to pathogen transmission via environmental reservoirs. Scientific Reports 6 27311. https://doi.org/10.1038/srep27311

177. Turner W.C. Kausrud K.L. Krishnappa Y.S. Cromsigt J.P.G.M. Ganz H.H. Mapaure I. Cloete C.C. Havarua Z. Küsters M. Getz W.M. Stenseth N.Chr. 2014. Fatal attraction: vegetation responses to nutrient inputs attract herbivores to infectious anthrax carcass sites. Proceedings of the Royal Society B: Biological Sciences 281 20141785. https://doi.org/10.1098/rspb.2014.1785

178. van Belkum A. Scherer S. van Alphen L. Verbrugh H. 1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiol Mol Biol Rev 62 275–293. https://doi.org/10.1128/MMBR.62.2.275-293.1998

179. Van der Auwera G.A. Carneiro M.O. Hartl C. Poplin R. del Angel G. Levy-Moonshine A. Jordan T. Shakir K. Roazen D. Thibault J. Banks E. Garimella K.V. Altshuler D. Gabriel S. DePristo M.A. 2013. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline Current Protocols in Bioinformatics. https://doi.org/10.1002/0471250953.bi1110s43

180. Van Ert M.N. Easterday W.R. Huynh L.Y. Okinaka R.T. Hugh-Jones M.E. Ravel J. Zanecki S.R. Pearson T. Simonson T.S. U'Ren J.M. Kachur S.M. Leadem-Dougherty R.R. Rhoton S.D. Zinser G. Farlow J. Coker P.R. Smith K.L. Wang B. Kenefic L.J. Fraser-Liggett C.M. Wagner D.M. Keim P. 2007a. Global genetic population structure of *Bacillus anthracis*. PLOS ONE 2 1–10. https://doi.org/10.1371/journal.pone.0000461

181. Van Ert M.N. Easterday W.R. Simonson T.S. U'Ren J.M. Pearson T. Kenefic L.J. Busch J.D. Huynh L.Y. Dukerich M. Trim C.B. Beaudry J. Welty-Bernard A. Read T. Fraser C.M. Ravel J. Keim P. 2007b. Strain-specific single-nucleotide polymorphism assays for the *Bacillus anthracis* Ames strain. J Clin Microbiol 45 47–53. https://doi.org/10.1128/JCM.01233-06

182. Venkateswaran K. Satomi M. Chung S. Kern R. Koukol R. Basic C. White D. 2001. Molecular microbial diversity of a spacecraft assembly facility. Syst Appl Microbiol 24 311–320. https://doi.org/10.1078/0723-2020-00018

183. Venter F. 2016. Genotypic diversity of *Bacillus anthracis* from 2014 to 2015 in the Kruger National Park. MSc University of Pretoria.

184. Vergnaud G. Girault G. Thierry S. Pourcel C. Madani N. Blouin Y. 2016. Comparison of French and worldwide *Bacillus anthracis* strains favors a recent post-columbian origin of the predominant North-American clade. PLoS ONE 11. https://doi.org/10.1371/journal.pone.0146216

185. Vermassen A. Leroy S. Talon R. Provot C. Popowska M. Desvaux M. 2019. Cell wall hydrolases in bacteria: Insight on the diversity of cell wall amidases glycosidases and peptidases toward peptidoglycan. Frontiers in Microbiology 10. https://doi.org/10.3389/fmicb.2019.00331

186. Viljoen PR. 1920. Anthrax in South Africa. Journal of Comparative Pathology and Therapeutics 33 137–152. https://doi.org/10.1016/s0368-1742(20)80025-3

187. Wang Y. Jenkins S.A. Gu C. Shree A. Martinez-Moczygemba M. Herold J. Botto M. Wetsel R.A. Xu Y. 2016. *Bacillus anthracis* spore surface protein BclA mediates complement factor H binding to spores and promotes spore persistence. PLoS Pathog 12 e1005678. https://doi.org/10.1371/journal.ppat.1005678

188. Weiner M.A. Read T.D. Hanna P.C. 2003. Identification and characterization of the gerH operon of *Bacillus anthracis* endospores: a differential role for purine nucleosides in germination. J Bacteriol 185 1462–1464. https://doi.org/10.1128/JB.185.4.1462-1464.2003

189. Weldhagen G.F. Poirel L. Nordmann P. 2003. Ambler class A extended-spectrum beta-lactamases in *Pseudomonas aeruginosa*: novel developments and clinical impact. Antimicrob Agents Chemother 47 2385–2392. https://doi.org/10.1128/AAC.47.8.2385-2392.2003

190. Whelan FJ Hall RJ McInerney JO. Evidence for selection in the abundant accessory gene content of a prokaryote pangenome. Mol Biol Evol. 2021 Aug 23;38(9):3697-3708. doi: 10.1093/molbev/msab139. PMID: 33963386; PMCID: PMC8382901

191. WHO 1996. Anthrax: memorandum from a WHO meeting. 1996. Bulletin of the World Health Organization.

192. WHO 2008. Anthrax in humans and animals 4th ed World Health Organization. World Health Organization.

193. Wilson G.S. and Miles A.A. 1975. Topley and Wilson's principles of bacteriology virology and immunity (Vol. 2 No. 6th edition).

194. Wistreich G. and Lechtman M.1973. Microbiology and Human Disease. New York United States of America: Glencoe Press.

195. Wu Y. Zaiden N. Cao B. 2018. The core- and pan-genomic analyses of the genus comamonas: from environmental adaptation to potential virulence. Front Microbiol 9 3096. https://doi.org/10.3389/fmicb.2018.03096

196. Yang Q.E. Walsh T.R. 2017. Toxin-antitoxin systems and their role in disseminating and maintaining antimicrobial resistance. FEMS Microbiology Reviews 41 343–353. https://doi.org/10.1093/femsre/fux006

197. Yanofsky C. 2013. Tryptophan operon of *Escherichia coli*. Brenner's Encyclopedia of Genetics: Second Edition 221–223. https://doi.org/10.1016/B978-0-12-374984-0.01676-4

198. Yichao W. Zaiden N. Cao B. 2018. The core- and pan-genomic analyses of genus comamonas: From Environmental Adaptation to Potential Virulence. https://doi.org/10.3389/fmicb.2018.03096

199. Zankari E. Hasman H. Cosentino S. Vestergaard M. Rasmussen S. Lund O. Aarestrup F.M. Larsen M.V. 2012. Identification of acquired antimicrobial resistance genes. Journal of Antimicrobial Chemotherapy 67 2640–2644. https://doi.org/10.1093/jac/dks

200. Zou Q.H. Li R.Q. Liu G.R. Liu S.L. 2014. Comparative genomic analysis between typhoidal and non-typhoidal *Salmonella serovars* reveals typhoid-specific protein families. Infection Genetics and Evolution 26 295–302. https://doi.org/10.1016/j.meegid.2014.06.008

5.7 Supplementary data

Table S1: Quality control for mapping of selected (n=9) Kruger National Park 2015 *Bacillus anthracis* sequence reads to the *B. anthracis* Ames ancestor reference genome

Sample	Number of reads	%mapped paired reads	Mean Coverage	Mean mapping quality
DS201584	2047362	82.15	53.34	41.38
DS201505	192276	84.62	48.08	40.8
1370	2002196	81.32	51.65	42
DS201523	2725806	83.67	67.79	41.08
1365	2179530	78.89	56.01	41.77
DS201588	1298748	82.49	39.43	46.42
DS201525	2228848	74.76	43.55	44.64
DS201578	2738308	89.5	60.33	38.33
DS201510	2160756	74.85	41.93	41.54

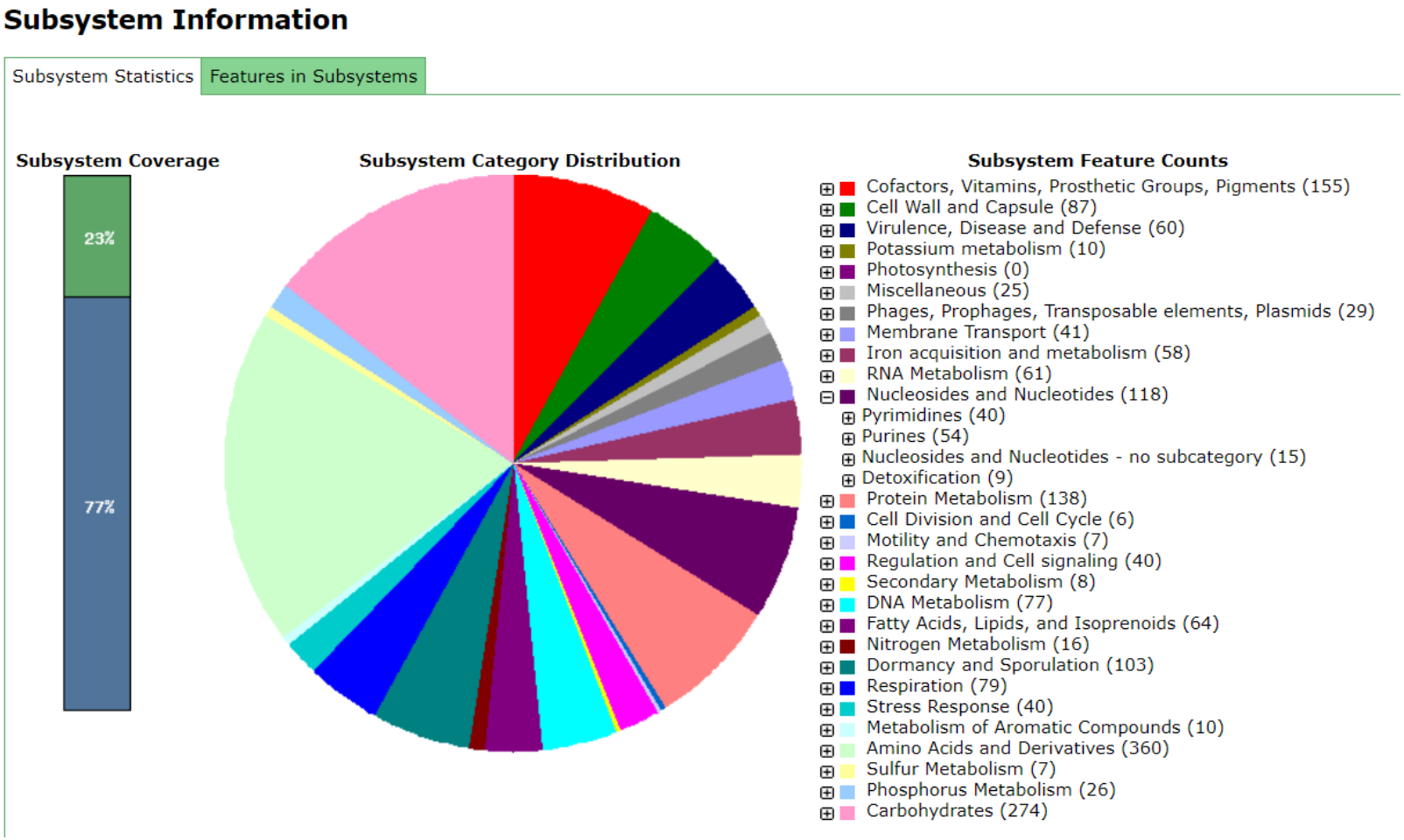


Figure S1: RAST annotation server subsystem analysis of *B. anthracis genomes*. The genes represented in a known subsystem are indicated by green bar (23%). The blue bar (77%) represents the genes that are not part of any subsystem. The pie chart represents the subsystem distribution. Image from RAST webserver (www.rast.nmpdr.org)

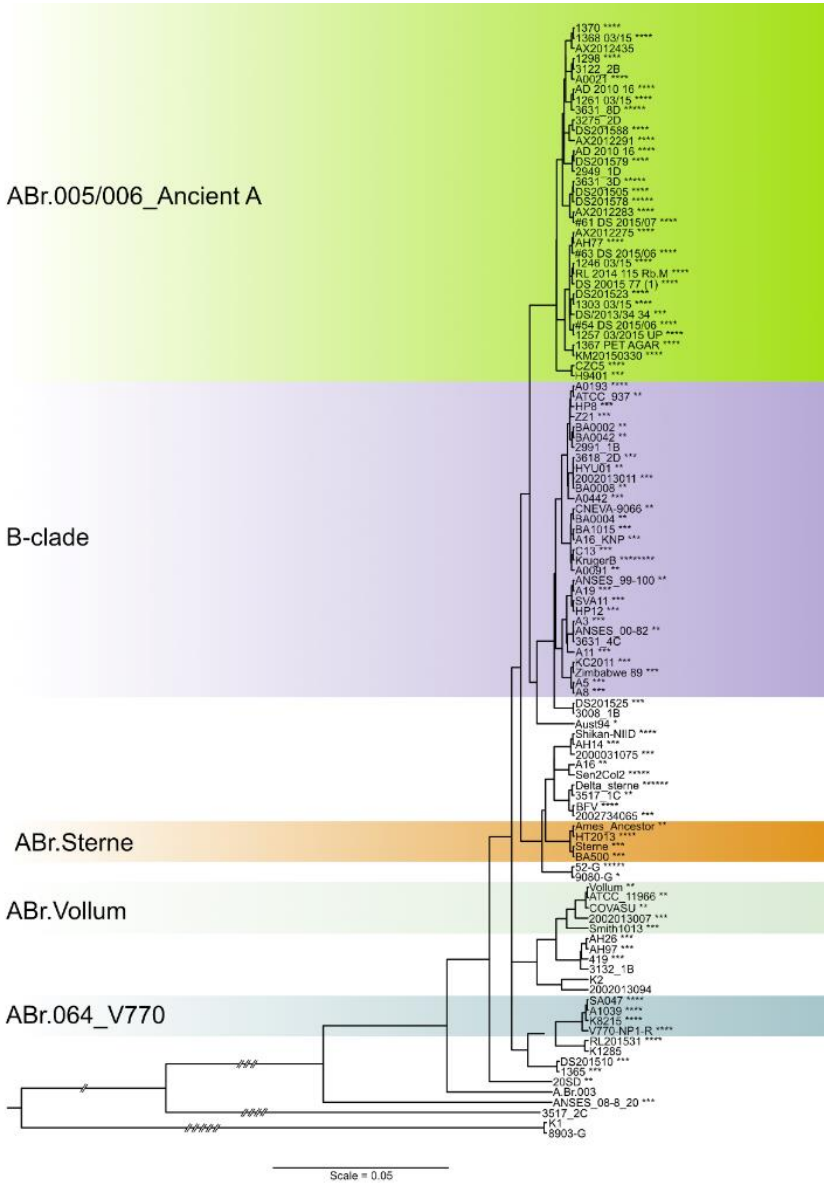


Figure S2: The variable number of tandem repeat (VNTR) profiles of *Bacillus anthracis* based on the multiple sequence alignment of *Bacillus* collagen-like protein of anthracis. A maximum likelihood method was used to infer the phylogeny of *B. anthracis* strains from Kruger National Park. Copy numbers were determined in Tandem Repeat Finder. Number of asterisk (*) represent the number of repeats.

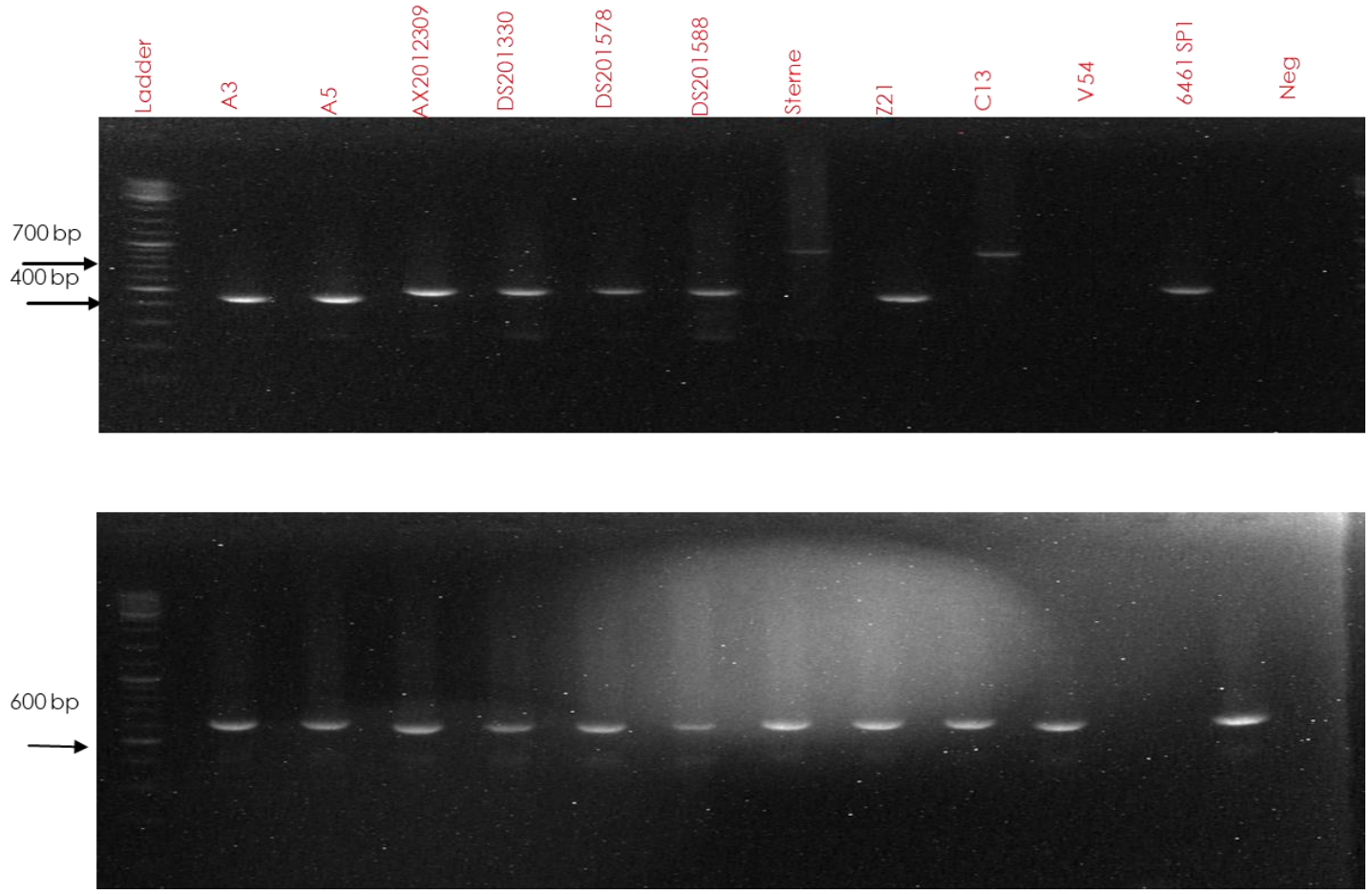


Figure S3: The variable number of tandem repeat (VNTR) profiles of *Bacillus anthracis* A-clade strains (AX2012309, DS201330, DS201578, DS201588, 6461_SP1 and V54) and B- clade strains (A3, A5, C13 and Z21) using Bams13 (Top) and Bams15 (Bottom) markers.

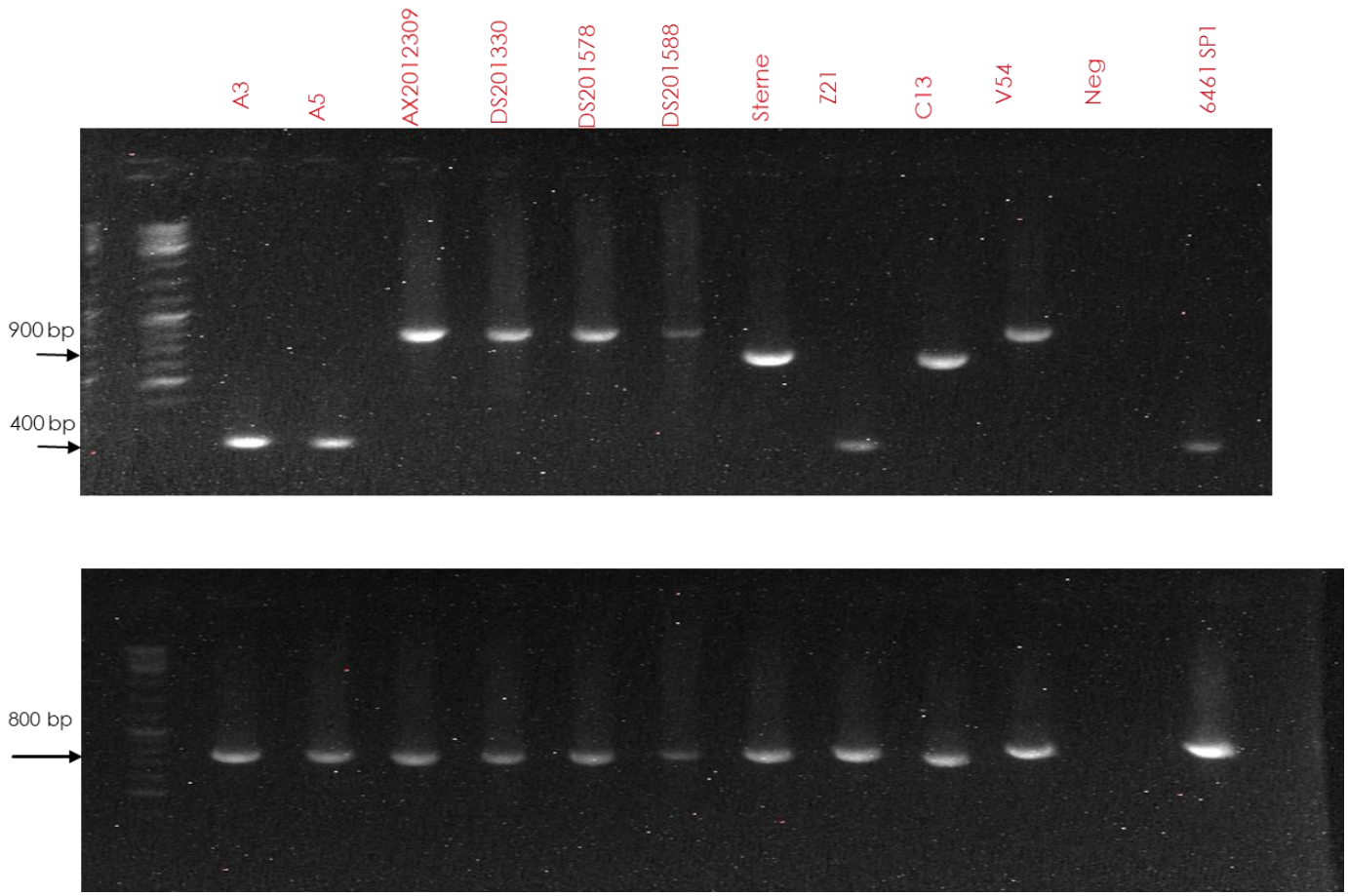


Figure S 4: The variable number of tandem repeat (VNTR) profiles of *Bacillus anthracis* A-clade strains (AX2012309, DS201330, DS201578, DS201588, 6461_SP1 and V54) and B- clade strains (A3, A5, C13 and Z21) using Bams30 (Top) and Bams31 (Bottom) markers. PCR amplicons depicts difference in length of VNTR for A- and B-clade strains.

Table S2: Analysis of variable number of tandem repeats (VNTR) copy numbers on the *bclA* gene for *Bacillus anthracis* strains from clades A and B lineages. Strains are defined by the primary SNP lineages.

Sample ID	Number of Repeats	Lineage/Strain	Sample ID	Number of Repeats	Lineage/Strain	Sample ID	Number of Repeats	Lineage/Strain	Sample ID	Number of Repeats	Lineage/Strain
Ames-ancestor	2	Ames	KNP28_3	3	A.Br.005/006	1298	4	A.Br.005/006	A16	2	B.Br.001/002
Sen2Col2	5	A.Br.0011/009	AH14	3	A.Br.005/006	124603_15_2	4	A.Br.005/006	A0091_2	2	B.Br.001/002
Smith1013_2	3	A.Br.009/011	AH26_2	3	A.Br.005/006	130303_15_2	4	A.Br.005/006	A0193	2	B.Br.001/002
ANSES_08-8_20_2	3	A.Br.002 (Sterne)	AH97	3	A.Br.005/006	1370_2	4	A.Br.005/006	ANSES_00-82	2	B.Br.001/002
BA500	3	A.Br.002 (Sterne)	2002013011_2	3	A.Br.005/006	KNP16	4	A.Br.005/006	A0442_2	3	B.Br.001/002
2000031075_3	3	A.Br.002 (Sterne)	419_2	3	A.Br.005/006	KNP20	4	A.Br.005/006	C13_2	3	B.Br.001/002
Sterne	3	A.Br.002 (Sterne)	DS201510_2	3	A.Br.005/006	KNP20_2	4	A.Br.005/006	SVA11	3	B.Br.001/002
2002734065_2	3	A.Br.002 (Sterne)	DS201525	3	A.Br.005/006	KNP21_2	4	A.Br.005/006	A0021	4	B.Br.001/002
BFV_2	4	A.Br.002 (Sterne)	AX2012275_2	4	A.Br.005/006	KNP23_2	4	A.Br.005/006	KrugerB	8	B.Br.Kruger
HT2013	4	A.Br.002 (Sterne)	AX2012283_2	4	A.Br.005/006	KNP24_2	4	A.Br.005/006	CNEVA	2	B.Br.CNEVA
Delta_sterne_2	6	A.Br.002 (Sterne)	AX2012291_2	4	A.Br.005/006	AD201016_2	4	A.Br.005/006	ANSES_99-100	2	A.Br.007 (Vollum)
9080-G	1	A.Br.003/004 (Aust94)	AX2012435	4	A.Br.005/006	AH77_2	4	A.Br.005/006	ATCC_11966_2	2	A.Br.007 (Vollum)
Aust94	1	A.Br.003/004 (Aust94)	CZC5	4	A.Br.005/006	KM20150330	4	A.Br.005/006	COVASU	2	A.Br.007 (Vollum)
20SD_2	2	A.Br.003/004 (Aust94)	DS201505_2	4	A.Br.005/006	RL201531_2	4	A.Br.005/006	Vollum_2	2	A.Br.007 (Vollum)
A.Br.003_2	2	A.Br.003/004 (Aust94)	DS201523_2	4	A.Br.005/006	DS201578	5	A.Br.005/006	BA0008	2	A.Br.007 (Vollum)
SA047	4	A.Br.003/004 (Aust94)	DS201577_2	4	A.Br.005/006	H9401	3	A.Br.H9401	BA1015_2	3	A.Br.007 (Vollum)
Shikan-NIID_2	4	A.Br.003/004 (Aust94)	DS201579	4	A.Br.005/006	HYU01	2	B.Br.001/002	3517_1C	2	A.Br.101 (A.Br.014)
52-G	5	A.Br.003/004 (Aust94)	DS201588_2	4	A.Br.005/006	HP12	3	B.Br.001/002	3618_2D_2	3	A.Br.101 (A.Br.014)
BA0002	2	A.Br.003/004 (V770)	KNP1_2	4	A.Br.005/006	KC2011	3	B.Br.001/002	3631_3D	5	A.Br.101 (A.Br.014)
BA0004	2	A.Br.003/004 (V770)	KNP12_2	4	A.Br.005/006	HP8_2	3	B.Br.001/002	3631_8D	5	A.Br.101 (A.Br.014)
BA0042	2	A.Br.003/004 (V770)	KNP19	4	A.Br.005/006	A3	3	B.Br.001/002	3631_4C	No repeats	A.Br.101 (A.Br.014)
K1825-2	2	A.Br.003/004 (V770)	KNP3_2	4	A.Br.005/006	A5_2	3	B.Br.001/002	3517_2C	No repeats	A.Br.101 (A.Br.014)
K8215_2	4	A.Br.003/004 (V770)	KNP5_2	4	A.Br.005/006	A8	3	B.Br.001/002	2949_1D	No repeats	A.Br.101 (A.Br.014)
V770-NP1-R	4	A.Br.003/004 (V770)	KNP6_2	4	A.Br.005/006	A11_2	3	B.Br.001/002	2991_1B	No repeats	A.Br.101 (A.Br.014)
A1039	4	A.Br.003/004 (V770)	KNP13_2	4	A.Br.005/006	A16_KNP	3	B.Br.001/002	3008_1B	No repeats	A.Br.101 (A.Br.014)
ATCC_937_2	2	A.Br.005/006	KNP14_2	4	A.Br.005/006	A19_2	3	B.Br.001/002	3122_2B	No repeats	A.Br.101 (A.Br.014)
1365_2	3	A.Br.005/006	54_DS201506_2	4	A.Br.005/006	Z21	3	B.Br.001/002	3275_2D	No repeats	A.Br.101 (A.Br.014)
2002013007_2	3	A.Br.005/006	RL201115_RbM_2	4	A.Br.005/006	Zimbabwe 89	3	B.Br.001/002	3132_1B	No repeats	A.Br.101 (A.Br.014)