Polyphasic and genome characterization of *Bacillus* species from anthrax outbreaks in animals in South Africa and Lesotho

by

Kgaugelo Edward Lekota

(11124459)

Submitted in fulfillment of the requirements for the degree

Masters of Science (Veterinary Tropical Diseases)

Faculty of Veterinary Science, University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Promoter	: Dr. H. van Heerden
Co-promoters	: Dr. E. Madoroba
	Dr. F.C. Muchadeyi
	Dr. J. Mafofo

DECLARATION

I declare that “Polyphasic and genome characterization of *Bacillus* species from anthrax outbreaks in animals in South Africa and Lesotho” is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full name**:** **Kgaugelo Edward Lekota**

Date: November, 2013

Signed: .....

ACKNOWLEDGEMENTS

Give thanks to God, the creator, for being my leader, pillar of my strength and also for guidance He gave me through my years of study to overcome my challenges

I offer my sincere gratitude to thank my supervisor Dr. Henriette van Heerden for her guidance, confidence and willingness to give me the opportunity to explore the realms of Veterinary Microbiology.

I would like to extend a word of gratitude to thank my promoters Dr. E. Madoroba, F.C. Muchadeyi and J. Mafofo. They went the extra mile throughout my research and gave constructive discussion and encouragement towards the completion of this work.

Special thanks to Dr. J. Rees for the opportunity he gave me to further my studies at the Agricultural Research council-Biotechnology Platform (ARC-BTP).

This project would not have been possible without the funding provided by the National Research Foundation (NRF), ARC-Onderstepoort Veterinary Institute (ARC-OVI), ARC-BTP, and Department of Agriculture and Forestry (DAFF) grands. The expertise and facilities provided by Department of Veterinary Tropical Diseases (DVTD), University of Pretoria, ARC-OVI, and ARC-BTP.

I extend my appreciation to the Dr. A. Gelaw, Mrs R.N. Maluleka, M.J. Kekena of the ARC-OVI, Bacteriology for their training in Bacteriology.

Special thanks to Dr. J. Mafofo of the ARC-BTP for the time and effort he helped with bioinformatics work.

I would like to thank Ms. A Hassim at DVTD, University of Pretoria for training in MLVA and, Ms M. DeCastro at ARC-BTP for the training in sequencing and CLC-bio software.

Kind thanks and appreciations to my friends Molemi, Pabalelo, Ayesha, Mzamose, Thuli, Charles and to all the postgrads at the ARC-BTP and DVTD, University of Pretoria for their constant support and encouragements throughout my studies.

Finally I would like to send my sincere and grateful appreciation to my family. To my parents Selinah Makhegu and Jackson Lekota, for giving me the opportunity to further my studies, for believing in me, and helping build my future, and atchive my goals. To the only sisters Abigail, Mmapelo, and Millicent; for their motivation, inspiration, giving me hope and being there for me.

SCIENTIFIC OUTPUTS

Oral presentations:

K.E. Lekota, F.C. Muchadeyi, J. Mafofo, J. Rees, H. van Heerden, E. Madoroba. 2012. The potential role of *Bacillus* species from Southern Africa in the manifestation of anthrax. Annual Reviews, Agricultural Research Council-OVI. Onderstepoort, South Africa.

K.E. Lekota, A. Hassim, J. Mafofo',J. Rees, F.C. Muchadeyi, H. van Heerden, E. Madoroba. 2013. Polyphasic characterization of *Bacillus* species from anthrax outbreaks in animals from Southern Africa. International *Bacillus* ACT conference, Victoria, Canada.
[Manuscript submitted to BMC Microbiology]

K.E. Lekota, J. Mafofo', E. Madoroba, J. Rees, H. van Heerden, F.C. Muchadeyi, 2013. Whole genome sequencing and genetic variant analysis of two South African *Bacillus anthracis* strains. International *Bacillus* ACT conference, Victoria, Canada.
[Manuscript in preparation]

Poster presenations:

- Wildlife ranching South Africa (WRSA), Annual congress and expo, Nylstroom, South Africa.
- Faculty day, University of Pretoria Veterinary Science, Onderstepoort, South Africa
- Genomic research institute (GRI) seminar day, University of Pretoria, Hatfield, South Africa.

TABLE OF CONTENT PAGE

INDEX OF TABLES

INDEX OF FIGURES

LIST OF ABBREVIATIONS

ARC	Agricultural Research council
BASys	Bacterial annotation systems
cAMP	Cyclic adenosine monophosphate
canSNP	Canonical single nucleotide polymorphism
CDC	Centres for disease control
CDS	Coding sequence
DNA	Deoxyribonucleic acid
EF	Oedema factor
INDELs	Insertion and deletions
KNP	Kruger National Park
LF	Lethal factor
MAPK	Mitogen activated protein kinase
MELT-MAMA	Melt analysis of mismatch amplification mutation assays
MLVA	Multiple locus Variable number tandem repeats analysis
MLST	Multiple locus sequence typing
NCP	Northern Cape Province
NGS	Next generation sequencing
OIE	Office International des Epizooties
ORF	Open reading frame
PHAST	Phage search tool
PCR	Polymerase chain reaction
PA	Protective antigen
PlcR	Pleotropic transcriptional regulator
RAST	Rapid annotation using subsystem technology
rRNA	Ribosomal ribonucleic acid
SNP	Single nucleotide polymorphism
UPGMA	Unweighted pair group method using arithmetic averages
VNTR	Variable number tandem repeats
WGS	Whole genome sequencing
WHO	World Health Organisation

Thesis Summary

Polyphasic and genome characterization of *Bacillus* species from anthrax outbreaks in animals in South Africa and Lesotho

Promoter: Henriette van Heerden
Co-promoters: Evelyn Madoroba, Farai Catherine Muchadeyi, Joseph Mafofo
Department: Veterinary Tropical Diseases, University of Pretoria
Degree: MSc (Veterinary Science)

Summary

Bacillus anthracis, B. cereus, *B. mycoides, B. pseudomycoides, B. thuringiensis* and *B. weihenstephanensis* belong to the *B. cereus* group (*B. cereus sensu lato*). *Bacillus anthracis* is a causative agent of anthrax that primarily affects herbivores. In South Africa, anthrax occurs sporadic and is endemic or epidemic in the Northern Cape Province (NCP) and Kruger National Park (KNP). The Agricultural Research Council-Onderstepoort Veterinary (ARC-OVI) Institute is a national reference laboratory for diagnosis of anthrax. The ARC-OVI reference laboratory received isolates that were reported as anthrax cases and other *Bacillus* strains that were found to have caused anthrax-like symptoms, but showed inconclusive results with the standard *B. anthracis* biochemical tests. It was hypothesis that these isolates were unique strains of the *B. cereus* group, which previously could have been distinguished from *B. cereus* group using a few of the standard biochemical tests. However, the biochemical tests did not seem to be sensitive enough to give unequivocal results for these atypical *B. anthracis* strains that caused anthrax-like symptoms. Molecular based methods were therefore employed to further investigate the strains.

The study therefore characterized *B. anthracis* and *Bacillus* species that were isolated during anthrax outbreaks in the Northern Cape, Limpopo, Mpumalanga Provinces and Lesotho, using a polyphasic approach that included phenotypic and

molecular techniques. For this purpose, 3 *B. anthracis* and 10 *Bacillus* isolates were subjected to microbiology tests, Biolog OmniLog identification system (Biolog), 16S ribosomal RNA (rRNA) sequence analysis, PCR detection of protective antigen (*pag)* and capsule (*cap*) regions, real-time PCR using hybridization probes targeting the chromosome, *pag* and *cap*C genes, and multi-loci variable number of tandem repeat (VNTR) analysis (MLVA). The classical microbiological tests indicated that the *Bacillus* species could be differentiated from the typical *B. anthracis* only by the use of gamma phage resistant and the morphology of *B. anthracis*. Phenotypic and genotypic, some of the strains showed the presence of capsules that might be important for virulence or environmental survival. Some of the *Bacillus species* produced weak PCR amplicons of the *pag*A (pXO1) and different *cap*BCA (pXO2) regions. The Biolog OmniLog system and 16S rRNA gene sequencing identified the isolates as *B. endophyticus, B. thuringiensis* and *Brevibacterium frigotoleransis* with 16S rRNA sequencing being more reliable than Biolog Omnilog. The 16S rRNA sequencing identified the *B. anthracis* strains that clustered separately from the *Bacillus* species. The API 50 CHB system was rendered irrelevant for diagnosis of *Bacillus* species (*B. anthracis*-like) isolates in this study as it lacked correlation with the 16S rRNA sequences and Biolog Omnilog. The real-time PCR identified the *B. anthracis* strains. MLVA-26 differentiated the *Bacillus* species isolates from the typical *B. anthracis*. Some of the VNTR markers produced non-specific amplicons, which indicated differences in the genomes as the markers are specifically designed for *B. anthracis*. In this study we determined that *B. anthracis* can be differentiated from other *Bacillus* species using some of the confirmatory microbiological that includes gamma-phage sensitivity, 16S rRNA sequencing, real-time PCR and MLVA. Most of the *Bacillus* species isolates were identified by 16S rRNA sequencing and Biolog OmniLog as the endospore-forming *B. endophyticus*, which was first isolated from the inner tissue of healthy cotton plants and their role in anthrax-associated animal deaths could not be established. This was the first time that virulence genes of *B. anthracis* were detected in *B. endophyticus* bacteria that do not belong to the *B. cereus* group. However PCR analysis provided limited information about the genetic basis of these observations since it relies on predetermined, known genetic sequences. Whole genome sequencing was therefore considered to further investigate and resolve the variability within species and sub-species groups that are closely related amongst *B. cereus*/*subtilis* group. It was decided, as a first step, to

analyse typical South African *Bacillus anthracis* strains using whole genomes sequence data and build an understanding of the genomic structure and level of diversity in these isolates.

Bacillus anthracis strains (20SD and 3631-1C) were characterized by whole genome sequencing. Sequencing of the *B. anthracis* strains showed the presence of the chromosome and plasmids pXO1 and pXO2 replicons on strain 20SD, whereas isolate 3631-1C lacked plasmid pXO2. Genome comparison of the *B. anthracis* strains showed similar and highly conserved regions relative to *B. anthracis* Ames ancestor, with differences in the number of single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs). High proportion of SNPs were observed in the chromosome with few genetic rearrangements. Approximately 2% of the SNPs in both strains were found on the pXO1 and pXO2 plasmids when compared to *B. anthracis* Ames ancestor.

Four-lambda prophages (the latent form of a bacteriophage in which the viral genes are incorporated into the bacterial chromosome) has been reported to be unique to *B. anthracis.* The presence of the four putative prophage regions in *B. anthracis* was investigated as these regions may be useful for differentiating the *B. anthracis* chromosome from that of its neighbors. In the study the prophages identified on the two strains were common to the *B. anthracis* Ames ancestor with slight variations in the sizes and number of prophages. *B. anthracis* 20SD and 3631-1C had phage minor proteins regions that are similar to the *B. anthracis* H901 and CDC strains. Sequence data was also compared to the only available South African *B anthracis* whole genome sequences of *B. anthracis* KrugerB and A0442 that also exhibited high number of prophages. The four-lambda prophages indicated to be characteristic of *B. anthracis* were shared amongst the *B. anthracis* strains i.e Ames ancestor, KrugerB, A0442, 3631-1C and 20SD. Results from this study suggested that regions that have capsid or mobile phage elements were likely to change the structural genomic complex of the strain resulting in inverted regions. The genomic complexity of strain 3631-1C was highly similar to *B. anthracis* Ames ancestor, while 20SD has a few genetic rearrangements of inversions. The results from this study agreed with and explained why prophages like conjugative transposons, insertion sequences, introns and other elements that make up a mobile portion of bacterial genomes often

accounts for large-scale genomic rearrangements, insertion and deletions in bacterial chromosome. Overall the study showed the importance of using polyphasic approach to characterize the *Bacillus* isolates from *B. anthracis* and better diagnostic method should be developed to identify *B. anthracis* and *Bacillus* sepcies containing anthrax virulence genes. The study also provided the importance of using whole genome wide characterization for comparative genomics.

Chapter 1

1. GENERAL INTRODUCTION

Bacillus anthracis is the causative agent of anthrax that is endemic in the Northern Cape Province (NCP) and Kruger National Park (KNP) of South Africa. Anthrax affects mammalian species, but mortalities are most common in herbivores. *Bacillus anthracis* is a soil borne, Gram-positive, rod-shaped, facultative anaerobic bacteria that is non-motile and belong to the spore-forming (endospores) group. The endospores can survive in the soil and remain stable for decades in extreme conditions (Dragon and Rennie, 1995; Beyer and Turnbull, 2009). This spore-forming feature makes *B. anthracis* a suitable bioterrorism agent (Inglesby *et al*., 1999; 2002).

Bacillus anthracis, B. cereus, *B. mycoides, B. pseudomycoides, B. thuringiensis* and *B. weihenstephanensis* belong to the *B. cereus* group (*B. cereus sensu lato*) (Kaneko *et al.,* 1978; Helgason *et al*., 2000). *Bacillus cereus* (*sensu stricto*)*, B. anthracis* and *B. thuringiensis* share highly conserved chromosome but differ in pathogenicity that is mostly plasmid–encoded (Helgason *et al*., 2000). *Bacillus cereus* is a food borne pathogen, whereas *B. thuringiensis* produces insect toxins and *B. anthracis* has plasmid-encoding toxin and capsule proteins (Kotiranta *et al*., 2000; Schnepf *et al*., 1998; Stenfors Arnesen *et al*., 2008). Plasmid-encoded as well as chromosomal virulence factors are usually used to differentiate *B. cereus sensu stricto* species. However, *B. cereus* isolates with pXO1 and/or pXO2 have recently been found in people, chimpanzees or gorillas (Hoffmaster *et al.,* 2004; Klee *et al.,* 2006; Avashia *et al*., 2007; Klee *et al.,* 2010; Oh *et al*., 2011). Therefore, the presence of pXO1 and pXO2 is considered to be no longer sensitive enough to differentiate *B. anthracis* from the *B. cereus* group.

Morhological and biochemical characteristics that include motility, antibiotic susceptibility testing, staining of capsule and lysis of colonies by gamma phage are used to discriminate *B. cereus* group species (Logan *et al*., 1985; Odendaal *et al*., 1991). At the molecular level, attempts have been made to distinguish *B. cereus* group species despite the high degree of genetic relatedness using techiniques that includes whole-genome DNA-DNA hybridization (Read *et al*., 2003), 16S rRNA sequencing (Ash *et al*., 1991; Ash and Collins, 1992), variable number of tandem repeats (VNTR) using multiple loci (multiple-loci variable number tandem repeats

analysis (MLVA)) (Keim *et al.*, 1999; Le Flech *et al*., 2001; Van Ert *et al*., 2007; Beyer *et al.,* 2012) and multilocus sequence typing (MLST) (Helgason *et al*., 2000; 2004). These diverse range of methods have often failed to give consistent differences. For example MLVA and single nucleotide polymorphisms (SNPs) are known to reveal the genetic relationships within *B. anthracis* (Keim *et al.,* 2000; Van Ert *et al*., 2007; Marston *et al*., 2011), but fails to distinguish species belonging to *B. cereus sensu lato.* Differentiation of *B. cereus sensu lato* should therefore be reviewed to include molecular based and biochemical assays that will allow for accurate identification.

The genome size of *B. anthracis* Ames is 5 227 293 base pairs (bp) (5.23 Mb) with 5 508 predicted protein coding regions (Read *et al.,* 2003; Ravel *et al.*, 2009). The principal virulence factors of this bacterium are encoded on two plasmids that are circular, extrachromosomal and double stranded DNA. The plasmid pXO1 (182 kb) encodes a tripartite protein exotoxin complex while plasmid pXO2 (96 kb) encodes the polypeptide capsule genes (Read *et al*., 2003). The members of the *B. cereus* group have a pleiotropic regulator (PlcR) that controls transcription of secreted virulence factors. *Bacillus anthracis* carries a specific mutation in the *plc*R gene whereas in *B. cereus* and *B. thuringiensis* it encodes functional proteins (Agaisse *et al*., 1999; Kolsto *et al*., 2009). However, there are some *B. cereus* and *B. thurigiensis* strains that have a defective *plc*R gene and thus exhibit a *B. anthracis*-like phenotype (Slamti and Lereclus 2002; Kolsto *et al*., 2009). Kolsto *et al.* (2009) indicated that *B. anthracis* can be differentiate from its nearest neighbours based on (i) a separate and unique evolutionary lineage within the *B. cereus* group; (ii) a mutation on the *plc*R gene; and (iii) presence of a full complement of four prophages (lambda01-04) within the chromosome.

2. JUSTIFICATION

2.1. Motivation of the study

Surveys on anthrax outbreaks have revealed the existence of isolates that are not the same as the typical *B. anthracis,* but possess some similar features and cause anthrax-like symptoms in the affected animals. Some of these i Beesley solates, were found to be to be closely related to *B. anthracis*, which is closely related to *B. cereus* and *B. thuringiensis*. The plasmid encoding anthrax toxins and capsule genes have been found in some *B. cereus* isolates that caused anthrax-like clinical symptoms in mammals (Hoffmaster *et al*. 2004; Avashia *et al*., 2007; Klee *et al*., 2010), and in some *B. megaterium* strains (Beesley *et al*., 2010).

In South Africa, isolates were found that caused anthrax-like symptoms but showed inconclusive results with the standard biochemical tests used for diagnosing *B. anthracis*. It could be that these isolates are unique strains of the *B. cereus* group. Previously *B. anthracis* was easily distinguished from *B. cereus sensu stricto* using few biochemical tests (Klee *et al*., 2006). However, with atypical *B. anthracis* that cause anthrax-like symptoms, biochemical tests do not seem to be sensitive enough to give unequivocal results. As a result, hence molecular based methods are becoming increasingly important. For these reasons, it is paramount to characterise these microorganisms using a combination of molecular methods and classical microbiological tests (polyphasic approach) is now being considered to be comprehensive and paramount for a conclusive characterization of these microorganisms. In the next generation genomics era, whole genome sequencing (WGS) provides more promise in diagnostic microbiological practice, by providing inter- and intra- bacterial species and subspecies variability (Segerman *et al*., 2011; Koser *et al*., 2012). The use of next generation sequencing (NGS) technology has now made it possible to analyze all the genes of the genomes simultaneously. Genome wide characterization of strains of *B. anthracis* is imperative for epidemiological investigations of natural outbreaks. The use of whole genome sequencing for genome wide characterization is becoming more relevant in diagnostics monitoring and control of anthrax outbreaks. WGS facilitates the

identification of the infecting strains and enable differentiation of *B. anthracis* from closely related *B. cereus* group members.

A survey on anthrax outbreaks in South Africa revealed the existence of atypical *B. anthracis* (*Bacillus* species), which caused anthrax-like symptoms in the affected animals (Figure 1.1). *Bacillus* species strains from anthrax cases were therefore screened from the bacterial culture collection of South African anthrax reference laboratory at the Agricultural Research Council-Onderstepoort Veterinary Institute (ARC-OVI), Onderstepoort. Typical *B. anthracis* and atypical *Bacilllus* species were identified. In this study, we applied the use a polyphasic approach to characterize the typical and atypical strains, and characterized the genomes of two *B. anthracis* strains by whole genome sequencing approach.

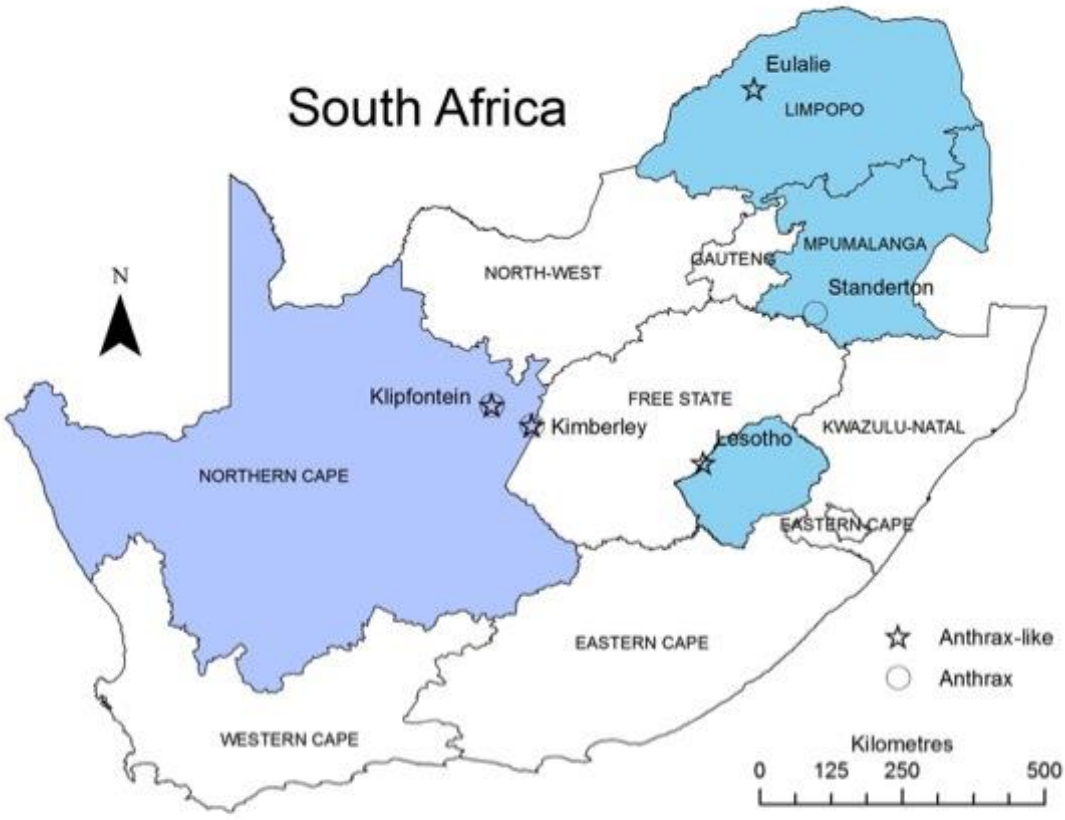


Figure 1.1. South African map indicating the locations of *Bacillus anthracis* and *Bacillus* species (atypical *B. anthracis*) used in this study.

2.2. Aim of the study

The aims of the study were to characterize atypical *Bacillus* isolates alongside *B. anthracis* obtained from anthrax outbreaks in South Africa and Lesotho using a polyphasic approach that combines the use of classical microbiological methods and DNA-based methods. In addition, this study strived to determine the genetic variations of two South African *B. anthracis* isolates (obtained from anthrax outbreaks and characterized using the polyphasic approach) relative to *B. anthracis* Ames ancestor using WGS.

2.4 Objectives

Polyphasic approach:

- To characterize the *B. anthracis* and *Bacillus* species using classical microbiological methods.
- To determine the virulence genes of *B. anthracis* and *Bacillus* species using species-specific PCR.
- To characterize the *B. anthracis*-like isolates using metabolic fingerprinting (Biolog OmniLog system).
- To characterize the *B. anthracis* and *Bacillus* species using 16S ribosomal RNA gene sequencing.
- To characterize the *B. anthracis* and *Bacillus* species using MLVA assay.

Whole genome sequencing:

- To sequence the genomes of two South African *B. anthracis* isolates obtained during anthrax outbreaks.
- To determine genetic variation between the South African strains relative to *B. anthracis* Ames ancestor.

The study will investigate *Bacillus* species isolated during anthrax outbreaks and determine whether the current methods allow for reliable identification. Sequencing the whole genome of *B. anthracis* is expected to show a broader view of the gene functions of the strains, genetic variable regions, and possibly determine novel genes that also contribute to pathogenicity or changes on the physiological property of *B. anthracis*. The study aimed to address whether *B. anthracis* strains from South Africa had typical *B. anthracis* sequences compared to the international reference strain.

2.5. Reference

Agaisse, H., Gominet, M., Okstad, O.A., Kolsto, A., Lereclus, D., 1999. PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. Mol. Microbiol. 32, 1043-1053.

Ash, C., Collins, M.D., 1992. Comparative analysis of 23S ribosomal RNA gene sequences of *Bacillus anthracis* and emetic *Bacillus cereus* determined by PCR-direct sequencing. FEMS Microbiol. Lett. 94, 75-80.

Ash, C., Farrow, J., Wallbanks, S., Collins, M., 1991. Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small subunit ribosomal RNA sequences. Lett. Appl. Microbiol. 13, 202-206.

Avashia, S.B., Riggins, W., Lindley, C., Hoffmaster, A., Drumgoole, R., Nekomoto, T., Jackson, P.J., Hill, K.K., Williams, K., Lehman, L., 2007. Fatal pneumonia among metalworkers due to inhalation exposure to *Bacillus cereus* containing *Bacillus anthracis* toxin genes. Clin. Infect. Dis. 44, 414-416.

Beesley, C.A., Vanner, C.L., Helsel, L.O., Gee, J.E., Hoffmaster, A.R., 2010. Identification and characterization of clinical *Bacillus* spp. isolates phenotypically similar to *Bacillus anthracis*. FEMS Microbiol. Lett. 313, 47-53.

Beyer, W., Bellan, S., Eberle, G., Ganz, H.H., Getz, W.M., Haumacher, R., Hilss, K.A., Kilian, W., Lazak, J., Turner, W.C., 2012. Distribution and molecular evolution of *Bacillus anthracis* genotypes in Namibia. PLoS Negl. Trop. Dis. 6, e1534.

Beyer, W., Turnbull, P., 2009. Anthrax in animals. Mol. Aspects Med. 30, 481-489.

Dragon, D.C., Rennie, R.P., 1995. The ecology of anthrax spores: tough but not invincible. The Can. Vet. J. 36, 295-301.

Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Friedlander, A.M., Hauer, J., McDade, J., Osterholm, M.T., O'Toole, T., 1999. Anthrax as a biological weapon. JAMA. 81, 1735-1745.

Inglesby, T.V., O'Toole, T., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Friedlander, A.M., Gerberding, J., Hauer, J., Hughes, J., 2002. Anthrax as a biological weapon, 2002. JAMA 287, 2236-2252.

Helgason, E., Okstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I., Kolsto, A., 2000. *Bacillus anthracis, Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. Appl. Environ. Microbiol. 66, 2627-2630.

Hoffmaster, A.R., Ravel, J., Rasko, D.A., Chapman, G.D., Chute, M.D., Marston, C.K., De, B.K., Sacchi, C.T., Fitzgerald, C., Mayer, L.W., 2004. Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. Proc. Natl. Acad. Sci. U. S. A. 101, 8449-8454.

Kaneko, T., Nozaki, R., Aizawa, K., 1978. Deoxyribonucleic acid relatedness between *Bacillus anthracis, Bacillus cereus* and *Bacillus thuringiensis*. Microbiol. Immunol. 22, 639-641.

Keim, P., Klevytska, A., Price, L., Schupp, J., Zinser, G., Smith, K., Hugh–Jones, M., Okinaka, R., Hill, K., Jackson, P., 1999. Molecular diversity in *Bacillus anthracis*. J. Appl. Microbiol. 87, 215-217.

Klee, S.R., Brzuszkiewicz, E.B., Nattermann, H., Bruggemann, H., Dupke, S., Wollherr, A., Franz, T., Pauli, G., Appel, B., Liebl, W., 2010. The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids. PLoS One 5, e10986.

Klee, S.R., Ozel, M., Appel, B., Boesch, C., Ellerbrok, H., Jacob, D., Holland, G., Leendertz, F.H., Pauli, G., Grunow, R., 2006. Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. J. Bacteriol. 188, 5333-5344.

Kolsto, A.B., Tourasse, N.J., Okstad, O.A., 2009. What sets *Bacillus anthracis* apart from other *Bacillus species*? Annu. Rev. Microbiol. 63, 451-476.

Koser, C.U., Ellington, M.J., Cartwright, E.J., Gillespie, S.H., Brown, N.M., Farrington, M., Holden, M.T., Dougan, G., Bentley, S.D., Parkhill, J., 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathogens 8, e1002824.

Kotiranta, A., Lounatmaa, K., Haapasalo, M., 2000. Epidemiology and pathogenesis of *Bacillus cereus* infections. Microb. Infect. 2, 189-198.

Le Fleche, 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, 2.

Logan, N., Carman, J., Melling, J., Berkeley, R., 1985. Identification of *Bacillus anthracis* by API tests. J. Med. Microbiol. 20, 75-85.

Marston, C.K., Allen, C.A., Beaudry, J., Price, E.P., Wolken, S.R., Pearson, T., Keim, P., Hoffmaster, A.R., 2011. Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States. PloS One 6, e28274.

Odendaal, M., Pieterson, P., De Vos, V., Botha, A., 1991. The antibiotic sensitivity patterns of *Bacillus anthracis* isolated from the Kruger National Park. Onderstepoort J. Vet. Res. 58, 17-19.

Oh, S., Budzik, J.M., Garufi, G., Schneewind, O., 2011. Two capsular polysaccharides enable *Bacillus cereus* G9241 to cause anthrax–like disease. Mol. Microbiol. 80, 455-470.

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., 2009. The complete genome sequence of *Bacillus anthracis* Ames "Ancestor". J. Bacteriol. 191, 445-446.

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., 2003. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature 423, 81-86.

Segerman, B., De Medici, D., Ehling Schulz, M., Fach, P., Fenicia, L., Fricker, M., Wielinga, P., Van Rotterdam, B., Knutsson, R., 2011. Bioinformatic tools for using whole genome sequencing as a rapid high resolution diagnostic typing tool when tracing bioterror organisms in the food and feed chain. Int. J. Food Microbiol. 145, S167-S176.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D., Dean, D., 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. 62, 775-806.

Slamti, L., Lereclus, D., 2002. A cell–cell signaling peptide activates the PlcR virulence regulon in bacteria of the *Bacillus cereus* group. EMBO J. 21, 4550-4559.

Stenfors Arnesen, L.P., Fagerlund, A., Granum, P.E., 2008. From soil to gut: *Bacillus cereus* and its food poisoning toxins. FEMS Microbiol. Rev. 32, 579-606.

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., 2007. Strain-specific single-nucleotide polymorphism assays for the *Bacillus anthracis* Ames strain. J. Clin. Microbiol. 45, 47-53.

Chapter 2

2. LITERATURE REVIEW

2.1. Anthrax and its importance

Anthrax is a zoonotic disease of warm-blooded animals that affects most mammals including domesticated and wild animals as well as humans (WHO, 2008). It is primarily a disease of herbivores such as sheep, cattle, horses, goats, while humans and carnivores are less susceptible. Humans become infected incidentally by contact with infected animals or contaminated products such as hides wool, hair, and bone meal (Watson and Keir, 1994; Kamal *et al*., 2011). In the early 1990s human cases of inhalation anthrax were reported in the United States of America (USA) among the workers in textile and tanning industries processing goat hair, or wools (Jernigan *et al*., 2001). From 1932 to 1945, bacteriologists of the Japanese Amy Unit 731 tested it as a biological weapon on Manchurian prisoners (Inglesby *et al*., 1999; Hudson *et al*., 2008).

2.2. Epidemiology and Global distribution of anthrax

Anthrax is a global, naturally occurring disease. The status of anthrax is considered to be endemic, hyper-endemic, epidemic, sporadic, probably free, free and unknown depending on the country's history (WHO, 2003; WHO, 2008). Sporadic is when anthrax cases occur occasionally or in random instances, while endemic indicates continually presense in a specific area. Epidemic is due to widespread occurrence in a community at a particular time. Sporadic cases or outbreaks in animals occur in most countries due to contaminated products of animal origin from endemic countries and the spread of *B. anthracis* from these areas to indigenous livestock via effluent from premises where such animal products are processed. Endemic regions are mostly associated with warmer temperatures that allow *B. anthracis* to sporulate (WHO, 2008). Anthrax cases in cold countries are thought to have been through the importation of contaminated animal products (Beyer and Turbull, 2009).

Anthrax is still regarded as endemic in many parts of the world such as central Asia, southern Africa, south and north-South America, certain regions of China, Indian sub-continent, and Australia (Dragon and Rennie, 1995; Beyer and Turnbull, 2009).

Anthrax is also hyper-endemic or endemic to most of the Middle East, Mexico and central America, some Mediterranean countries and other parts of the countries indicated by red colour on the map in Figure 2.1. In Africa the most common outbreaks were reported in Botswana, Namibia, Zimbabwe, Uganda and Tanzania. Zimbabwe and Zambia are considered as hyper-endemic due to the high annual anthrax-related humans deaths in these countries (WHO, 2008). In other countries like southern and eastern India, outbreaks were reported in both animals and humans as the disease is poorly controlled (WHO, 2008). Sporadic cases occur frequently in southern Europe and less frequency in northern Europe, certain regions of the southern and northern America. However, in the western region of Europe it is rare due to soil acidity (Beyer and Turnbull, 2009).

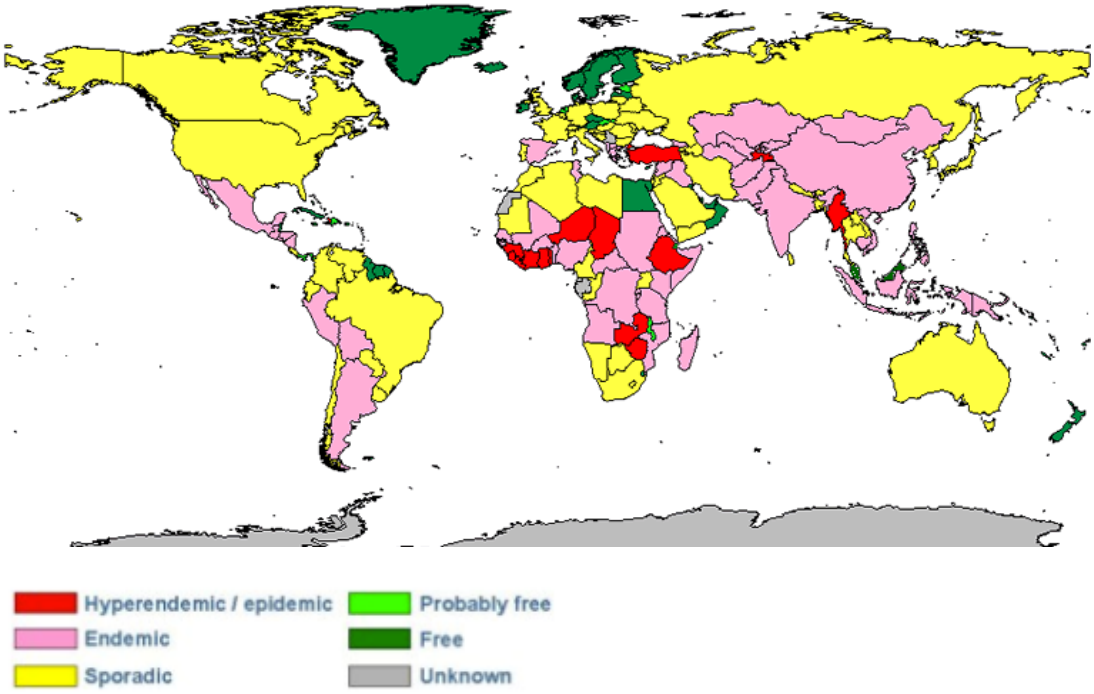


Figure 2.1. Global distribution of anthrax in animals and humans. Courtesy of world anthrax site; http://www.vetmed.1su.edu/whocc/mp_world.htm

2.2.1 Epidemiology and distribution of anthrax in South Africa

Distribution and spread of anthrax in South Africa shows a major peak in 1923 when an estimated 30 000 to 60 000 animals died of anthrax (Sterne, 1967; Fasanella *et al*., 2010). With the development and use of the Sterne attenuated live spore vaccine in 1937, anthrax cases where significantly reduced in South Africa. Sporadic cases still occur throughout the sourthern Africa, but the disease is now regarded as endemic in two main areas of Kruger National Park (KNP) where national parks occur, and Vaalbos and Kgalagadi national parks in the Northern Cape Province (NCP).

South African anthrax deaths have been recorded in at least 52 species and herbivores were the most affetced, while suids, carnivores and ostriches seem to be less susceptible although outbreaks do occur in these species (Hugh-Jones and de Vos, 2002). In KNP, kudu (*Tragelaphus strepsiceros*) is the principal host accounting >50% of all recorded anthrax cases followed by zebras falling into relatively small group of other affected species (de Vos 1990). In Kimberly district (Ghaap region), NCP anthrax outbreaks were also reported during summer of 1995-1996 affecting goats, cattle, sheep, horses, roan antelope, gemsbok (*Oryx gazella*), and springbok (*Antidorcas marsupialis*) (Henton and Briers, 1998). The outbreak has since spread from the initial focus in a south westerly direction over a distance of 270 km (40-50 km wide). This spread of anthrax was considered to be due to windborne spores emanating from incineration sites (Henton and Bries, 1998; WHO, 2008). Reports from the Department of Agriculture Forestry and Fishers (DAFF) in South Africa show a high prevalence of 53 anthrax cases that affected eland (*Taurotragus oryx*) during 2001 in NCP, Barkly West area. Another outbreak affected the area in 2008/9. Anthrax incidences in South Africa are now considered sporadic probably as a result of effective control measures and management of outbreaks in the endemic regions.

2.3. Life cycle of *Bacillus anthracis*

Bacillus anthracis life cycle infection occurs when the vegetative cells are released from infected hosts whereby the bacterium remains in dormant for long periods as an environmental spore until the disease is re-established (Figure 2.2). Terminal

hemorrhage and scavenger action release blood into the environment. Little is still known regarding the completion of the bacterial life cycle in the environment between outbreaks. Shafazand *et al.* (1999) suggested that sporulation and starvation are the only options for the completion of the life cycle. However, processes like biofilm and rizhosphere- interaction have been indicated as alternatives to sporulation (Saile and Koehler, 2006; Lee *et al*., 2007). Van Ness (1971) suggested that incubation areas exist in the environment where certain soil conditions may favor the vegetative growth prior to outbreaks. The most intriguing genetic factor investigated is DNA exchange amongst the bacilli species that could occur due to horizontal gene transfer in the soil amongst closely related species of the *B. cereus* group (Rasko *et al*., 2005; Vilas-Boas *et al*., 2007). The ecological and genetic factor that contributes towards the survival of *B. anthracis* in the environment mentioned above is still unclear (Figure 2.2).

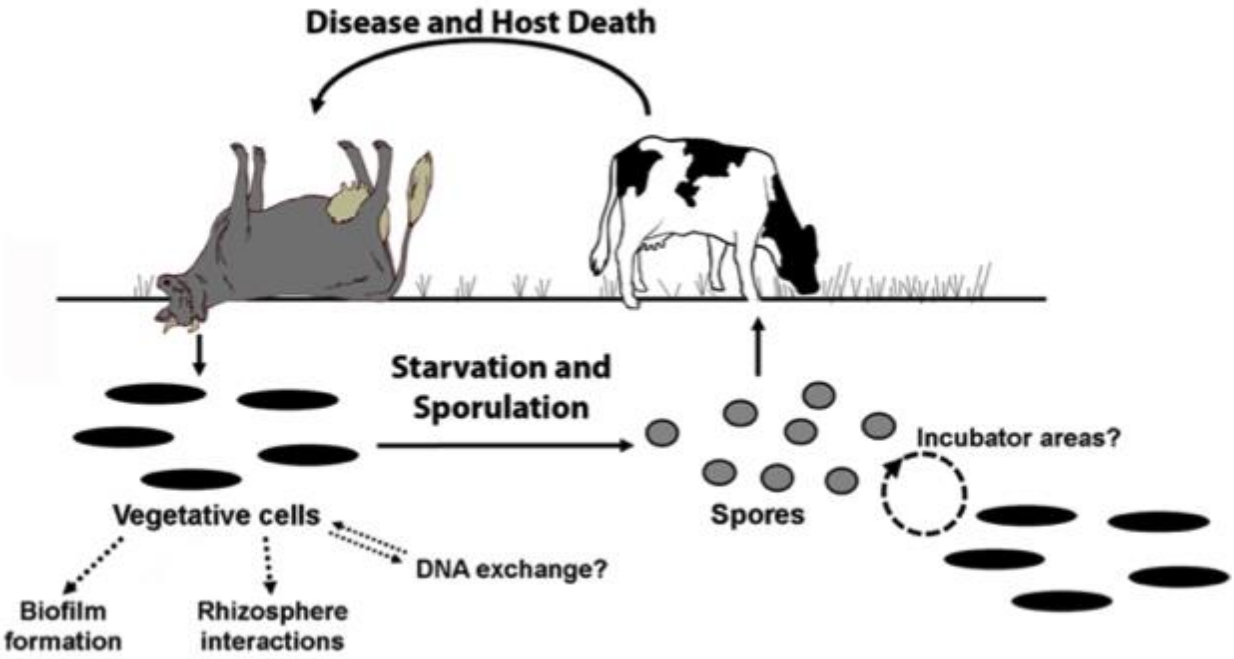


Figure 2.2. The life cycle of *Bacillus anthracis* where solid arrows trace the disease acquisition of the host by dormant spores that germinate and produce the vegetative cells. Dot arrows trace the ecological and genetic factors that occur in the environmental soil (Schuch and Fischetti, 2009).

2.4. Transmission

Anthrax occurs in three major clinical forms namely pulmonary (inhalation), gastrointestinal and cutaneous depending on the route of entry spores germinate in the body. Pulmonary anthrax is through inhalation of spore to the lungs, whilst gastrointestinal anthrax is by ingestion of contaminated meat, and cutaneous anthrax is by transmission of the spores through lesion in the skin (Turnbull, 1998; Mock and Fouet, 2001).

2.4.1. Disease acquisition on animals

Animals normally acquire the disease at grazing sites and waterholes where previous victims of the disease had died or at other infected sources (Smith *et al*., 2000; Hugh-Jones and de Vos, 2002; Beyer and Turnbull, 2009; Hugh-Jones and Blackburn, 2009, Fasanella *et al*., 2010). Ingestion of sufficient numbers of spores in soil or plants in pastures usually affects most herbivores (WHO, 2008). Animals at grazing sites acquire the disease through ingestion of spores, which require a lesion to initiate the mode of infection as *B. anthracis* is considered non-invasive. Ingestion of dry grass that is contaminated with anthrax-ridden soil can cause oro-gastrointestinal lesions enabling infection. Animals that graze over dry, dusty, contaminated soil inhale spores leading to infection by inhalation, which has been reported in cows that died of pulmonary anthrax (Bell and Laing, 1977).

2.4.2. Vectors

2.4.2.1. Insects

Insects or flies are strongly suspected of being important in the transmission of anthrax and may cause outbreaks (Beyer and Turnbull, 2009). The insects that are known to transmit the disease are necrophilic and haemophagic flies. Anthrax outbreaks in KNP, mostly affect browsing kudus as factors such as vector ecology and feeding preferences, land cover, and host species feeding-habits make then the most likely victims (Hugh-Jones and Blackburn, 2009). Biting flies such as *Tabanus*

or *Stomoxys* species have been reported in livestock and wildlife outbreaks in Africa, India, and USA, while non-biting files such as *Chrysoma* (blowfly) species were reported in outbreaks where browsing herbivores in southeastern Africa were affected (OIE, 2008). In KNP, blowflies have been incriminated as the principal vector of anthrax in herbivores (Braack *et al*., 1990; de Vos, 1990; de Vos and Bryden, 1996) although *B. anthracis* is not detected from the digestive tract of blowflies within two weeks of feeding on the carcass (WHO, 2008; Beyer and Turnbull, 2009). The other insect of great importance in spreading the disease is a tabanid fly where *B. anthracis* have not been found to multiply in the gut (Hugh-Jones and Blackburn, 2009) indicating that these insects might only be vectors (but further investigation is needed).

2.4.2.2. Vultures and Scavengers

In rural areas, scavenger birds, such as ravens (*Corpus corax*) and vultures spread the disease from infected carcass to nearby water hole, water trough, pond, or dam (Pienaar, 1967; Hugh-Jones and Blackburn, 2009; Fasanella *et al*., 2010). Anthrax spores were recovered from faeces of jackals, vultures, and hyenas (Lindeque and Turnbull, 1994).

2.4.2.3. Trade and transmission of anthrax

In the agricultural sectors, feedstuffs with contaminated ingredients are traditionally the primary source of infection (WHO, 2008). Sporadic cases that have been reported around the countries are mainly through the trade of infected animals or meat hides, hair, wool or bones in the near areas or at a distance. Products with contaminated spores find their way to infect livestock in the importing countries (Turnbull, 1998; Beyer and Turnbull, 2009).

2.5. Clinical manifestation

The clinical signs of anthrax vary with species affected and form of exposure. In animals, clinical signs range from peracute to subacute (OIE, 2008). The peracute

form is characterized by sudden death that occurs within few hours after clinical signs such as fever, anorexia, diarrhoea, declining milk production, and sometimes abortion. Other symptoms include failure of blood to clot, bloody discharges from body orifices, rapid bloating, and incomplete rigor mortis. The acute form of the disease will result in death within 24 hours to few days after clinical signs that are characterized by excitement followed by depression, convulsions, abortion, staggering, death, lack of rigor mortis, and a bloody discharge from body orifices. Subcutaneous is more prolonged and lasts for several days, and result in clinical forms such as, subcutaneous oedema of the ventral parts of the body (Turnbull, 1998; Hugh-Jones and Blackburn, 2009; Fasanella *et al*., 2010).

2.6. Pathogenicity

Bacillus anthracis causes disease when the spores enter the body, germinate, multiply, and release toxins. The spores can be transmitted in three primary forms already mentioned, namely: pulmonary anthrax through inhalation of *B. anthracis* spores into the lungs; cutaneous anthrax through a lesion in the skin; and through consumption of contaminated meat and water resulting in gastrointestinal form of anthrax (Turnbull, 1998; Mock and Fouet, 2001; Vilas-Boas *et al*., 2007). Whatever the route of infection, the *B. anthracis* spores are transported by macrophages to draining lymph nodes and then enter the blood stream and multiply rapidly. The pathogenicity of *B. anthracis* depends on the quality of the capsular coat, the amount of toxins produced (Coker *et al*., 2002; Shoop *et al*., 2005; Fasenella *et al*., 2010) and the sensitivity of the host species (Smith, 1973). To herbivores such as cattle, sheep and equines, it is an acute form and death occurs within a 48-72 hours (Smith, 1973).

2.7. Vaccine and control

Anthrax was used as the first bacterial vaccine by Pasteur in 1881 (Leppla *et al.,* 2002). The heat attenuated *B. anthracis* cultures resulted in the loss of the toxin-encoding plasmid pXO1 therefore were capable of forming capsules, but not toxins in the Pasteur vaccine (Farrar, 1994). The Pasteur vaccine provides a lower level of protective immunity than toxigenic vaccine strains. Pasteur vaccine was displaced by the vaccine of Sterne utilising a toxin positive, capsule negative strain 34F2 (Sterne,

1939; Sterne, 1946). Widespread vaccination began in the 1930's, using the Sterne attenuated strain, which virtually abolished anthrax in industrialized countries. By 1939, the Sterne vaccine had proven itself to be safe and effective for use in domesticated animals. It seemed to completely protect them against challenge with highly virulent strains of *B. anthracis*. Live attenuated strains are still undesirable for human vaccine (Hambleton *et al.,* 1984). While vaccination programs and good farming practices have enable control of anthrax in most countries, it still remains a serious problem in many developing countries (Farrar, 1994).

2.8. Bacteriology: *Bacillus anthracis*

Taxonomy

Kingdom : Bacteria
Phylum : Firmicutes
Class : Bacilli
Order : Bacillales
Family : Bacillaceae
Genus : *Bacillus*
Species : *anthracis*

2.8.1. The cell and colony morphology

Bacillus anthracis is a Gram-positive and rod shape with endospores commonly found in the soil. The vegetative cell is about 1-1.5 µm in width and 3-10 µm in length, which survive poorly in environments such as water and soil (Turnbull, 2002; Saile and Koehler, 2006). The arrangements of the rods are square ends. During the exponential phase of batch culture the organism form long serpentine chains, and single cells or short chains are observed in infected tissues.

The colonies are fairly flat, rough, large, and greyish white with medusa heads on sheep blood agar (SBA) (Fasanella *et al*., 2010). The colonies have irregular edges (rugby ball shape) and a rough "ground glass" appearance when incubated under conditions that are not conducive for capsule formation. Colonies that are grown on

sodium bicarbonate for induction of capsule synthesis appears extremely mucoid that can be more than 3 µm in thickness (Drysdale *et al*., 2004).

Bacillus anthracis has only one flagellin gene with four proteins containing point mutation and frameshifts. Therefore the flagellum is non-functional and lacks motility (Rasko *et al*., 2005). Non-motility differentiates *B. anthracis* from closely related *B. cereus* group species. The cell structure of this organism contains extensive peptidoglycan layer, lipoteichoic acids, and crystalline cell surface proteins (S-layer protein). It differs from the other Gram-positive bacteria in terms of not containing the teichoic acids and while the S-layer proteins are not glycosylated (Bradley *et al*., 2001).

2.8.2. The spore

Bacillus anthracis can form a spore (endospore) in response to starvation. The spore is largely metabolically inactive and can resist a wide range of stresses found in nature, where it remains dormant in the soil until it sense the presence of nutrient and rapidly return to vegetative growth (Dragon and Rennie, 1995; Driks, 2009). The spores are regarded as biological weapon since they are fatal in pulmonary form and are resistant to extreme heat, cold, pH, desiccation, chemicals disinfectants, and ultraviolet radiation. These features allow the organism to persist for prolonged intervals within the environment (Dragon and Rennie, 1995; Driks, 2009).

Different layers constitute the structure of the spore (Figure 2.3). Spores can survive for decades in the environment during nutrient deprivation (unfavorable conditions). Recovery of the *B. anthracis* from bones estimated to be approximately 200 years old in KNP indicate its ability to survive for prolonged periods in the environment (Dragon and Rennie, 1995). When conditions are favorable, anthrax spores germinate and multiply when they enter an environment rich in amino acids, nucleosides, and glucose, in the blood or tissues of animal and human by killing the host through toxins (Figure 2.2; Inglesby *et al*., 1999).

2.8.3. Anatomy of the spore

The spore coat (core) represents approximately 50% of the volume spore. The spore coat houses the chromosome that is tightly complexed with proteins called the small acid-soluble proteins (SASPs) (Dirks and Setlow, 2000). It plays an important role in resistance to chemicals and physical disruption due to interactions between the DNA and SASPs together with high levels of calcium dipicolinic acid and other ions in the core. The spore cortex is essential for spore dehydration (which comprise of 10 to 30% of the water content of the vegetative cell) and for heat resistance (Figure 2.3; Nicholson and Galeano, 2003).

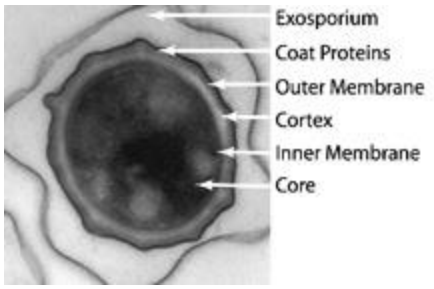


Figure 2.3. Electron micrograph of *Bacillus anthracis* spore structures listing different components (Liu *et al*., 2004).

The outermost surface of the *B. anthracis* spore is a protein called the exosporium, which is present to most species of *B. cereus* group. The exosporium (surrounds the coat) is the outermost layer of about 25 nm to 40 nm thick that contains important proteins and carbohydrate markers that are recognized by the antibodies (Fox *et al*., 2003; Williams *et al*., 2003). The proteins have received considerable attentions in recent years since they are candidates for vaccines and ligands for spore detection (Fox *et al*., 2003; Tournier *et al*., 2009).

2.8.4. Resistance

Due to spore formation, the bacterium is resistance to unfavorable conditions such as heat, cold, pH, desiccation, chemicals, irradiation and other adverse conditions. The spores are sensitive to certain chemicals such as 2-3% formaldehyde solutions at 40 °C for 20 minutes, or 0.25% at 60 °C for 6 hours or at 4% for at least 2 hours (Turnbull, 1998). The spores can be destroyed only after ten minutes at boiling temperatures, but preferable they are destroyed after 30 minutes in an autoclave set

at 121 °C. It was observed that they can withstand boiling temperature of 98 °C for 10 minutes (Fasanella *et al*., 2010).

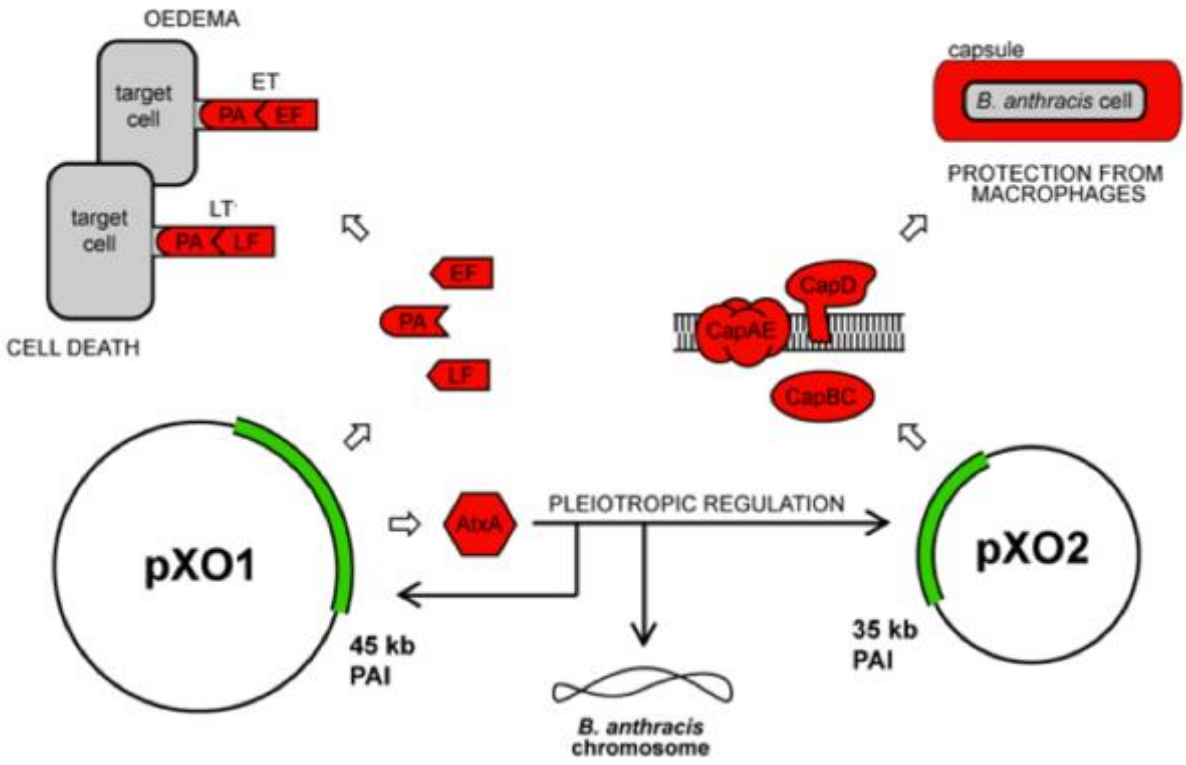
2.9. Distinguishing with its close relatives (phenotypes)

Bacillus anthracis is distinguishable from its closely relatives by its ability to synthesize virulence plasmids pXO1 (182 kb) which are composed of structural genes for anthrax toxin proteins and pXO2 (96 kb) which is important for the production of the poly-D-glutamic acid capsule (Makino *et al*., 1989; Okinaka *et al*., 1999; Candela and Fouet, 2005). The virulence genes of *B. anthracis* are produced in a number of complex (in the host) and defined media. The toxins are synthesized at optimal conditions under 37 °C in defined media containing nutrients with glucose as a carbon source and bicarbonate for induction (Leppla, 1982; Drysdale *et al*., 2004; Koehler, 2009). The production of capsule formation is dependent on the presence of bicarbonate in the media. The bacterium is differentiated by its absence of beta-hemolysin on sheep blood agar, lack of motility, absence of phospholipase C activity, susceptibility to penicillin, and sensitivity of the gamma bacteriophage (Marston *et al*., 2006; Koehler, 2009). *Bacillus anthracis* is classified as a species on its own and from its nearest neighbours based on (i) a separate and unique evolutionary lineage within the *B. cereus* group; (ii) a nonsense mutation in the regulatory *plc*R gene; and (iii) presence of a full complement of four prophages (lambda01-04) within the chromosome.

2.10. Genome of *Bacillus anthracis* and pathogenicity islands

The genome of *B. anthracis* is a tripartite organism that comprises of a single chromosome and two circular extra-chromosomal plasmids, pXO1 and pXO2 that are the primary virulence factors of anthrax. The chromosome DNA molecule size is about 5 227 293 bp encoding 5 508 predicted protein coding sequences (Read *et al*., 2003). The genome composition is highly biased towards adenine and thymine, with low GC (guanine and cytosine) content of approximately 35%. The plasmids code for many genes of which the most important are the toxin and capsule determinants (Okinaka *et al*., 1999; Pannucci *et al*., 2002). The pXO1 (182 kb) encodes the toxin complex gene while the pXO2 (96 kb) encodes the capsule genes (Figure 2.4). The

toxin complex is composed of three components namely, lethal toxin (LT) that composes of protective antigen (PA) and lethal factor (LF), and edema toxin (ET) which composed of PA and edema factor (EF) (Mock and Fouet, 2001; Mock and Mignot, 2003; Young and Collier, 2007). The PA controls the two toxin proteins in the cell by allowing the proteins to pass through the membrane via a special toxin delivery system (Rasko *et al*., 2005). The capsule gene consists of a five-gene operon (*cap*ABCDE) that interact at the bacterial cell membrane to produce poly-gamma-D-gluatamic acid which enable host immune system evasion by protecting the vegetative cells from being phagocytosed by macrophages (Green *et al*., 1985; Drysdale *et al*., 2004).



PAI= Pathogenicity island

Figure 2.4. The primary virulence factors of *Bacillus anthracis* are controlled by the anthrax toxin activator (AtxA). AtxA from the pXO1 controls the expression of the toxins components and the capsular formation components from the pXO2 (http://www.broadinstitute.org/files/shared/genomebio/B_cereusWP_10_22_09_pdf).

2. 11. Regulation of toxin and capsule by Anthrax toxin activator (Atxa)

The plasmid pXO1 encodes the three components (*pag*, *lef*, *cya*) that are all located within the 45 kb region PAI (pathogenicity island) (Okinaka *et al*., 1999). The plasmid also encodes the transcriptional activator *Atx*A (known as anthrax toxin activator,) that increases the rate of transcription. The activator is known to control the expression of the anthrax toxins components, the capsule formation components, and as well as expression of chromosomal genes. The expression of chromosomal genes such as those coding for the protein constituting the bacterial S-layer is also due to the activator *Atx*A (Saile and Koehler, 2002). The *pagR* regulates the expression of one or more toxin genes by binding to the operator (Fouet and Mock, 2006). The capsule expression on the pXO2 plasmid is activated by the transcriptional regulators (*Acp*A and *Acp*B), which are under the control of *Atx*A from pXO1, and the capsule operon and its regulators are part of the 35 kb PAI on pXO2 (Figure 2.4) (van der Auwera *et al*., 2005; Fouet and Mock, 2006). Therefore there is a cross-talk occurrence between the two-virulence plasmids pXO1 and pXO2. The induction of capsule formation and toxin expression is dependent on the presence of bicarbonate/CO_2 in the atmosphere (Drysdale *et al*., 2004).

2.12. Molecular mechanisms of pathogenicity and host-pathogen interactions

The virulence factors that contribute towards pathogenicity have been discussed in session 2.10. The LT is known as a zinc metalloprotease protein, while the ET is a calmodulin (CaM)-dependent adenylate cyclase (Leppla, 1982; Collier and Young, 2003; Pilo and Frey, 2011). The two proteins share the same adhesion subunit called PA that transports these toxins to the cytosol (Figure 2.5), which bind to anthrax toxin receptors (ATR) on the host cell surface (Bradley *et al*., 2001). PA (83 kDa) is cleaved by furin (cellular proteases) into two fragments, namely PA20 (20 kDa) and PA63 (63 kDa). PA20 dissociates and diffuses into the surrounding medium, leaving PA63 binding to the ATRs. Heptamerization of PA induces clustering of ATRs, that expose EF and LF binding domains to the receptor-bound PA63, and transported into the cells via membrane lipid rafts (Abrami *et al*., 2003). In the cell, LF acts

proteolytically to cleave N-terminal portion of mitogen-activated protein kinase kinase (MAPKK) resulting in cell lysis or induction of proinflammatory cytokines leading to vascular collapse, shock and death (Moayeri *et al*., 2003). LF is also known to cause cytolysis where it promotes macrophage apoptosis through change in membrane permeability dissipation of mitochondrial membrane and DNA fragmentation (Leppla, 1982; Gnade *et al*., 2010).

When EF is inside the host cell, it catalyses the conversion of adenosine triphosphate (ATP) to cyclic Adenosine monophosphate (cAMP), which therefore deregulates the cytokine response in the host (Figure 2.5; Leppla, 1982; Tang and Guo, 2009). EF and LF act on host cytosolic target to induce edema, necrosis and hypoxia (Prince, 2003). The presence of the capsule of *B. anthracis* from the pXO2 synthesize poly-D-glutamic acid capsule that imparts a highly negative charge to the organism that contributes pathogenesis by avoiding killing through phagocytosis (Figure 2.5).

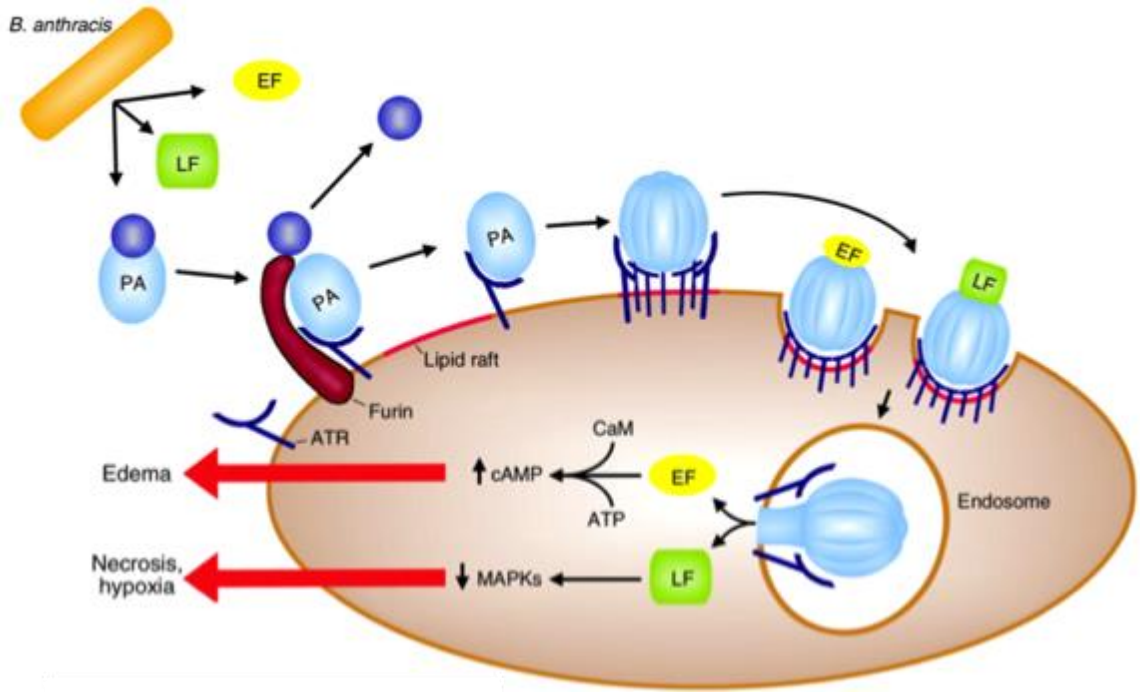


Figure 2.5. The molecular mechanism of *Bacillus anthracis* virulence factors and host-pathogen interactions of the cell. *Bacillus anthracis* cell synthesizing four major virulence genes: LF (lethal factor), EF (oedema factor), PA (protective antigen), all toxins are encoded on pXO1, and cap (capsule) encoded at plasmid XO2 (Prince, 2003).

2.13. The *Bacillus cereus* group

Members of the *B. cereus* group (*B. cereus sensu lato*) include six species: *B. thuringiensis*, *B. anthracis*, *B. cereus*, *B. mycoides*, *B. pseudomycoides*, and *B. weihenstephanensis*. Members of this group are all Gram-positive spore formers that belong to the Firmicutes with low GC content. The members of the *Bacillus* group are found in diverse habitats including various types of soils, the gut of soil-dwelling invertebrates, the plant rhizosphere, and in food production factories (Jensen *et al*., 2003; Kolsto *et al*., 2009). *Bacillus cereus sensu lato* consist of *B. cereus, B. anthracis* and *B. thuringiensis* that shares highly conserved chromosome but differ in pathogenicity that is mostly plasmid–encoded. The group has been suggested to be reclassified as a single species (*B. cereus*) due to their high genetic relatedness, and that the subspecies could then be assigned to various specialized groups within the species (Helgason *et al*., 2000) but this suggestion was never implemented.

2.14. *Bacillus anthracis*, *B. cereus* and *B. thuringiensis*

The three species *B. anthracis*, *B. cereus*, and *B. thuringiensis* have been studied within the other members of the *B. cereus* group. The description of *B. anthracis* has already been discussed in session 2.8.

Bacillus thuringiensis is commonly referred to as B.t., an insect pathogen that harbors plasmids encoding insecticidal proteins with toxicity against insects of the orders *Diptera*, *Lepidoptera*, and *Coleoptera* and nematodes (Whiteley and Schnepf, 1986; Schnepf *et al*., 1998). *Bacillus thuringiensis* is commercially used as a bio-pesticide worldwide. *Bacillus thuringiensis* produce the insecticidal proteinaceous toxin crystals (Cry and Cyt proteins), which are about 130-140 kDa (Aronson, 2002; Rasko *et al*., 2005). The crystals usually solubilize in the mid-gut of an insect where it is cleaved by a gut protease to produce endotoxins. When the toxin binds to the mid-gut epithelial cells by creating pores in the cell membrane, the gut is therefore immobilized and epithelial cells lyse as a results the insect will stop feeding and die (Schnepf *et al*., 1998).

Bacillus cereus causes is a food-borne spoilage pathogen that is associated with two forms of human food namely; emetic and diarrheal syndromes. The emetic syndrome is due to synthesis of the cereulide (a small heat-stable non-ribosomally synthesized dodecadepsipeptide) toxin. The emetic syndrome it is characterized by abdominal distress or nausea and vomiting. Diarrheal syndrome is due to production of enterotoxins that causes diarrhea by disruption. It is also responsible for pneumonia and endophthalmitis due to emetic toxin, enterotoxins, and degradative enzymes (Kotiranta *et al*., 2000; Stenfors Arnesen *et al*., 2008). Other virulence factors include degradative enzymes such as secreted phospholipases, haemolysins, and proteases, which are under the control of the pleiotropic transcriptional regulator *plcR* (Agaisse *et al*., 1999).

2.15. Horizontal gene transfer

Horizontal transfer of certain genes that integrate within the chromosome through transduction via bacteriophages or exchange of plasmids through transformation amongst the closely related species of the *B. cereus* group has been reported (Hoffmaster *et al*., 2004; Hu *et al*., 2009; Kolsto *et al*., 2009). The *B. cereus sensu lato* group contains bacteriophages which may integrate in the chromosome as prophages or which may replicate as independent linear elements (Rasko *et al*., 2005; Lapidus *et al*., 2008; Klee *et al*., 2010). Phages serve as vectors transferring certain virulence factors to the bacteria giving them the advantage of acquiring antibiotic resistance, changing the virulence factors and/or pathogenicity. Phages have been isolates from a variety of hosts. The plasmids, pXO1 and pXO2 found in *B. anthracis* are shared by closely related species such as *B. cereus* and *B. thuringiensis* (Hoffmaster *et al*., 2004; Klee *et al*., 2006; 2010; Leendertz *et al*., 2004) suggesting horizontal transfer amongst the closely related species.

2.16. *Bacillus* isolates resembling Anthrax-like disease

In the years 2001-2002, *Bacillus* isolates have been reported from chimpanzees in the Tai National Park, Ivory Coast that died from anthrax-like symptoms. During 2004/2005, *Bacillus* isolates were obtained from nine chimpanzees and one gorilla in

Cameroon (CA) and Cote d'Ivore (CI) (Klee *et al*., 2006; Leendertz *et al*., 2004; 2006). The isolates had *pagA* and *CapC* genes suggesting the presence of the pXO1 and pXO2-like sequences (Leendertz *et al*., 2006). These *B. anthracis*-like isolates were motile, resistant to gamma phage and the CA isolates were also resistance to penicillin (Leendertz *et al*., 2004; Klee *et al*., 2010). These features are in contrast to classic *B. anthracis,* but also had *B. anthracis* characteristics such as the *pagA* and *CapC* genes and were not heamolytic. Further genotyping, microbiological and functional characterization, as well as multilocus sequence typing (MLST) based phylogenetic analysis revealed that they are distinct from classical *B. anthracis* (Klee *et al*., 2006; 2010; Kolsto *et al*., 2009; Leendertz *et al*., 2006). MLST revealed a closely relationship with *B. anthracis* and *B. thuringiensis* serovar Konkukian strain (reported from human necrosis case), and *B. cereus* E33L which was reported from suspected anthrax case from a zebra (Klee *et al*., 2010).

The *pclR* gene of the CI and CA strains showed the lack of the non-sense mutation and they were motile. The specific *B. anthracis* prophage regions were absent on their chromosomes indicating that the strains were not *B. anthracis* (Okinaka *et al*., 2006; Klee *et al*., 2010). These isolates were later characterized as *B. cereus* var. *anthracis* since they had features of *B. cereus* and *B. anthracis*. This anthrax-like cases makes diagnosis difficult or are reported as non-anthrax (not *B. anthracis*) and are therefore overlooked by diagnostic laboratories.

2.17. *Bacillus cereus* group plasmids similar to *B. anthracis*

Bacillus cereus and *B. thuringiensis* isolates have been reported with *B. anthracis* pXO1 and/or pXO2-like sequences indicated in Table 2.1. However the pXO1-like plasmids are exclusively restricted to *B. cereus*. The extensive homology around the replication origin and in the putative replication protein (*RepX*) on pXO1-like, therefore form a coherent plasmid family (Pannucci *et al*., 2002; Hoffmaster *et al*., 2004; Kolsto *et al*., 2009; Klee *et al*., 2010). Few pXO2-like plasmids have been identified similar to pXO2 as compared to pXO1, and are mostly found on *B. thuringiensis* (Table 2.1).

Table 2.1. Plasmids content of close relatives to *Bacillus anthracis.*

Bacillus Species	Strain	Plasmid name	Size (bp)	Similar to pXO1/pXO2	GenBank Accession
B. anthracis	Ames Ancestor	pXO1	181677	_	AE017336
B. anthracis	Ames Ancestor	pXO2	94830	_	AE017335
B. cereus	ATCC10987	pBc1098	208369	pXO1-like	AE017195
B. cereus	AH187	pCER270	270082	pXO1-like	DQ889676
B. cereus	G9241	pBCXO1	209385	pXO1-like	DQ889679
B. cereus	AH818	pPER272	272145	pXO1-like	DQ889678
B. cereus	03BB108	p03BB108_239	238933	pXO1-like	ABDM02000063
B. cereus	03BB108	p03BB108_282	282009	pXO1-like	ABDM02000062
B. cereus	AH187; F4810/72	pAH187_272/ pCER270	270082	pXO1-like	CP001179
B. cereus biovar. anthracis	CI	pCI-XO1	181907	pXO1-like	CP001747
B. cereus biovar. anthracis	CI	pCI-XO2	94469	pXO2-like	CP001748
B. thuringiensis	HD73	pAW63	71777	pXO2-like	DQ025752
B. thuringiensis	97–27	pBT9727	77112	pXO2-like	CP000047
B. thuringiensis	_	pBtoxis	127923	pXO2-like	AL731825

2.18. Diagnosis of anthrax

2.18.1. Initial identification of anthrax

Diagnosis of anthrax consists of isolating *B. anthracis* from clinical samples like blood, organs, fluid samples from skin lesions, and aspirates of lymph or spleen. Care should be undertaken when collecting samples from dead animals suspected of anthrax to prevent human infection and environmental contamination due to bacterial sporulation upon expose to oxygen. A blood smear is made to assess the blood from animal parts such as limb, mammary, ear or other peripheral veins without opening the carcass (Hugh-Jones and de Vos, 2002; WHO, 2008). Blood can also be collected from the nasal cavity if difficulties are experienced on collection. The blood smear is examined for *B. anthracis* morphology using microscopy.

It is also known that the haemorrhagic nasal, buccal or anal exudate carry large number of the bacteria that can be cultured from swabs (WHO, 2008). In culture isolation, Gram stained *B. anthracis* is large, Gram-positive rod that occur singly, in pairs, or in chains (Figure 2.6A). From decaying carcass, the capsule is not always visible on a Gram stain that might result in mistaken diagnosis. Preferable, polychrome methylene blue known as MacFadyean stain (Figure 2.6B) is used to stain blood smear from decaying carcass. Smears are considered positive when the bacilli cells appear blue surrounded by pinkish-red capsule. The cells are arranged in pairs or short chains, and they are square ended rods. In Giemsa stain the capsule is stained reddish-Mauve (Figure 2.6C). The capsule of the *B. anthracis* is not observed when cultured on media like nutrient agar or broth, but rather on nutrient agar containing 0.7% of sodium bicarbonate and incubated in the presence of carbon dioxide (Fassanella *et al*., 2010).

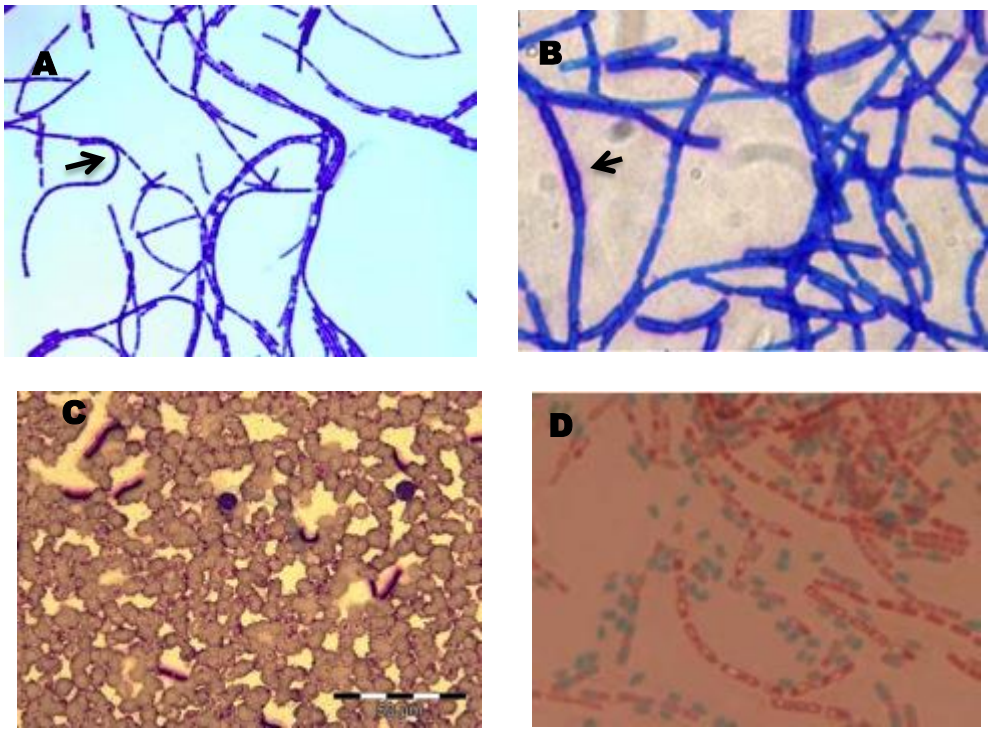


Figure 2.6. *Bacillus anthracis* cell morphology. (A) Gram staining of *B. anthracis* Gram-positive rods occurring in single, pairs or chains with bacilli cells appearing as box-car shaped with loop appearance at the end of long chain (indicated by an arrow); (B) *Bacillus anthracis* stained with polymethylene blue (M'Fadyean stain adapted from Owen *et al*., 2013) with bacilli cells appearing blue surrounded by pinkish red capsule (indicated by the arrow); (C) Bacilli cells of *B. anthracis* in the blood smear stained with Giemsa stain and; (D) Malachite green stained spores and red vegetative cells (picture adapted from WHO, 2008).

2.18.2. Culture

Bacillus anthracis grows readily on artificial media. Nutrient agar can be used for isolation of *B. anthracis* from uncontaminated fresh specimen, but best results are usually obtained on media containing serum or blood (Turnbull, 1998; Tomaso *et al*., 2006). In case of closely related *Bacillus* species, it is important to use selective media for isolation. Selective medium for *B. anthracis* like polymyxin-lysozyme-EDTA-thallous acetate (PLET) agar is recommended (Knisely, 1966). PLET agar plates are incubated for up to 48 hours at 37 °C, while cultures on serum, blood or nutrient agar are incubated for 12 to 24 hours at 37 °C. *Bacillus anthracis* can also be

easily be distinguished from closely related organisms by using cereus ident agar (CEI) as it grows white to cream coloured colonies compared to turquoise coloration by other *B. cereus* species (Figure 2.7A and B) (Peng *et al*., 2001; Tomaso *et al*., 2006). On blood agar *B. anthracis* forms large white to grey non-hemolytic colonies with irregularly tapered outgrowths (medusa's head appearance) (Figure 2.7F) while *B. cereus* usually exhibits beta-haemolysis.

Identification of *Bacillus anthracis* at reference laboratories generally includes examination for colony morphology, capsulation, lack of haemolysis activity, penicillin sensitivity and bacteriophage sensitivity (Turnbull, 1998; Fasenella *et al*., 2010). However some of *B. anthracis* isolates can demonstrate bacteriophage or penicillin resistance, and hemolytic (Ondendaal *et al*., 1991; Coker *et al*. 2002; Drysdale *et al*., 2004; Gierczynski *et al*., 2006). Therefore confirmation test by PCR for the capsular and toxins genes is commonly practice for diagnosis of anthrax.

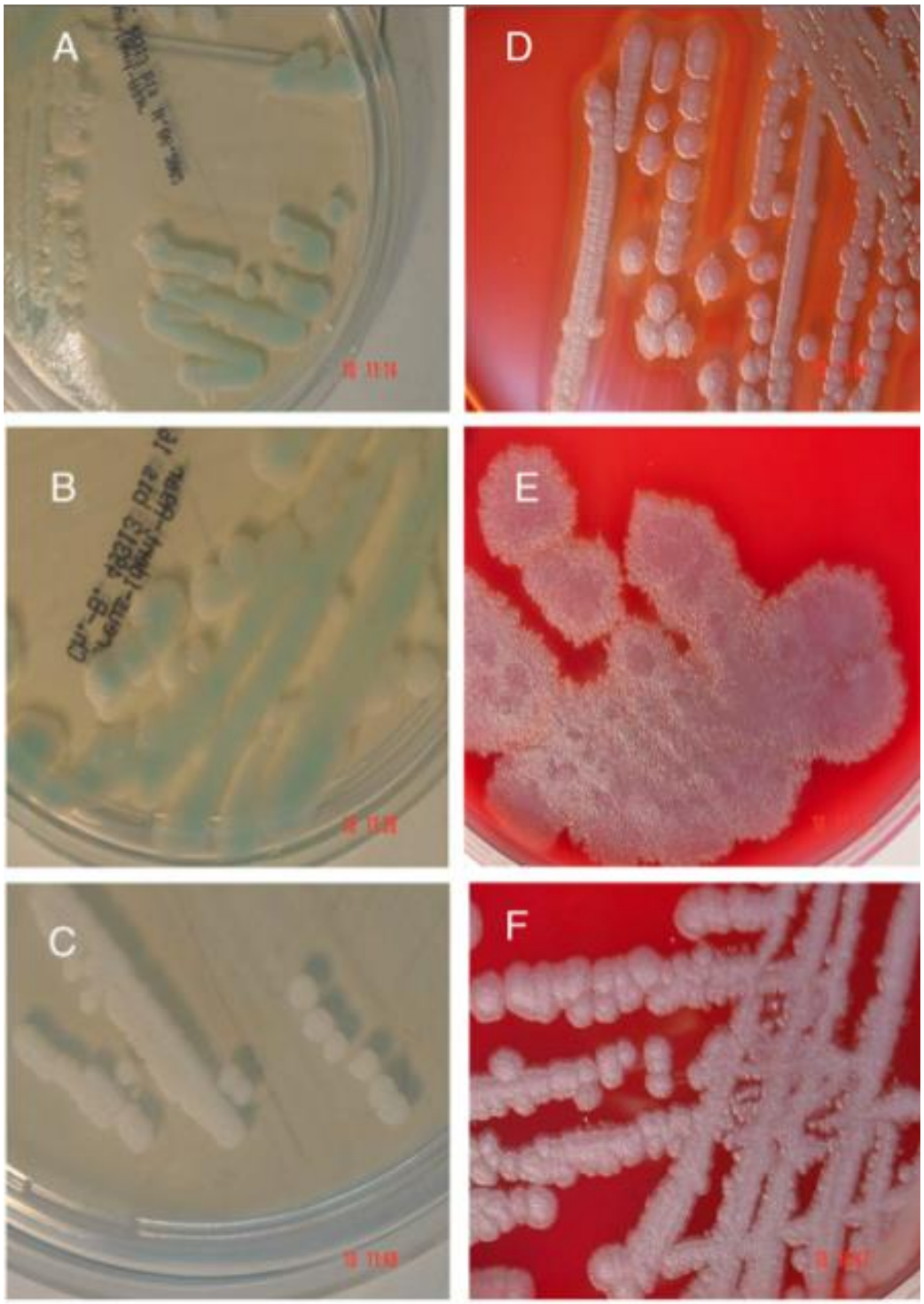


Figure 2.7. Colony morphology of *Bacillus cereus* group on different media: On cereus ident agar colonies morphology of *B. cereus* (A) and *B. thuringiensis* (B) acquire a turquoise coloration while *B. anthracis* (C) colonies are white to cream coloured. On blood agar *B. cereus* (D) and *B. thuringiensis* (E) usually produce haemolysis, but may grow with *B. anthracis*-like non-hemolytic colonies compared to of *B. anthracis* (F) (Tomaso *et al*., 2006).

2.18.3. Biochemical tests

Different biochemical tests are commercially available for identification of *Bacillus* species that include biochemical characteristics such as oxidase, catalase, indole, citrate, urease, voges-proskauer, nitrate etc. The API 50CHB system that is based on acidification of 49 carbohydrates is also available and used to identify *Bacillus* species. However the identification of *B. cereus* apart of being non-hemylotic, non-motile and pathogenic using biochemical tests, exhibits the same characteristics as *B. anthracis* (WHO, 2008). The Biolog system is useful for identification of *Bacillus* species and for characterization data profile when comparing with *B. anthracis* (Baillie *et al*., 1995).

2.18.4. Polymerase Chain Reaction (PCR)

PCR is used to identify virulence factors (capsule and toxins) present on the plasmids genes and also the chromosomal gene markers (Patra *et al*., 1996; Turnbull, 1998). The protective antigen (*pag* gene), edema factor (*cya*), lethal factor (*let*), and as well as different capsule genes, which include *cap*B, *cap*C, and *Cap*A are used to confirm *B. anthracis* (Beyer *et al*., 1995; Ramisse *et al*., 1996; Makino *et al*., 1993; WHO, 2008). However diagnostic primers of the capsule genes have been reported to react unspecifically with various *Bacillus* species (Beyer and Bohm, 2003).

The chromosomal markers include *Ba813* (Ramisie *et al*., 1996), *saspB* (Hoffmaster *et al*., 2002), *rpoB* (Qi *et al*., 2001), *gyrA* (Hurtle *et al*., 2004), *plcR* (Easterday *et al*., 2005), *adk* (Olsen *et al*., 2007) and other four genes identified by Radnedge *et al*. (2003). However the marker *rpoB*, *Ba813* and S-layer chromosomal proteins genes very often lead to inconsistence false positive results from environmental samples (Ramisse *et al*., 1999; Papaparaskevas *et al*., 2004; Kolsto *et al*., 2009). The *plcR* marker successfully differentiates *B. anthracis* from *B. cereus* and *B. thuriengiensis* (Agassie *et al*., 1999; Easterday *et al*., 2005).

2.19. Molecular characterization of members of the *Bacillus cereus* group

2.19.1. 16S rRNA analysis

Identification and differentiation of the *B. cereus sensu stricto* group have been investigated using molecular techniques (Ash and Collins, 1992; Rasko *et al*., 2005). In prokaryotes, the ribosomal RNA genes are transcribed from the ribosomal operon as 30S rRNA precursor molecules, and then RNase III cleave the molecules into 16S, 23S, and 5S rRNA molecules (Ash *et al*., 1991; Rajendhran and Gunasekaran, 2011). The 16S rRNA is the conserved amongst the three rRNAs sequences and therefore commonly used for identification in bacteria. It has been proposed as an "evolutionary clock", which has led to the reconstruction of the tree of life (Woese, 1987). The 16S rRNA sequences of this group revealed a 99-100% similarity within the *B. cereus* group. It was therefore suggested by Bayvkin *et al*. (2004) that the 16S rRNA sequences are not useful for discrimination of the *B. cereus sensu stricto* group but can be used to discriminate the *B. cereus* group from other *Bacillus* species.

2.19.2. Multilocus sequence typing (MLST)

MLST is based on sequencing of defined segments of various housekeeping genes that are present in the *B. cereus* group (Helgason *et al*., 2004; Priest *et al*., 2004). The MLST housekeeping genes including *adk*, *ccp*A, *fts*A, *glp*T, *pyr*E, *rec*F, *suc*C (Helgason *et al*., 2004), and additional markers *rpo*B, *lef*, *cap*, *plc*R and Ba5510 (Olsen *et al*., 2007) were designed to distinguish the members of the *B. cereus* group species. MLST has insufficient discriminatory power in differentiating *B. anthracis* isolates (intraspecies) due to very few differences (Helgason *et al*., 2004; Kim *et al*., 2005). However, it is not routinely used to diagnose *B. anthracis*.

2.19.3. Multiple-locus variable number tandem repeats (VNTR) assay (MLVA)

MLVA is a genotyping method used to determine genetic diversity within *B anthracis*. MLVA is based on amplifying variable number tandem repeats (VNTRs) found in the chromosome and plasmids locations that exhibit allelic differences/similarities within global *B. anthracis* population (Keim *et al*., 2000; 2004; 2009). The discriminatory power of each marker can be estimated by the number of alleles it detects and by its diversity. Different number of markers can be analysed, but initially only 8 chromosomal markers namely: variable repeat region (*vrr*A), *vrr*B1, *vrr*B2, *vrr*C1, *vrr*C2, CG3 were used (Keim *et al*., 2000). The other VNTRs markers are from *B. anthracis* plasmids, pXO1 and pXO2 (Keim *et al*., 1999). The MLVA-8 genotyping method was able to identify different subpopulations (89 genotypes around the world) and precise identification of particular strains of *B. anthracis* (Keim *et al*., 1999; 2000; Hoffmaster *et al*., 2002). In the previous studies, MLVA-8 lacked the discrimination power to differentiate between two geographical isolates from a natural outbreak (Keim *et al*., 2004; Lista *et al*., 2006), but differentiated isolates into major clonal lineages of the A, B and C clades.

The canonical single nucleotide polymorphisms (canSNPs) were discovered through whole genome sequencing and later used together with MLVA. The canSNP method identified 12 SNPs that define the major clonal lineages (clade A, B, C; Figure 2.8) (Pearson *et al*., 2004). Additional VNTR markers were added to MLVA-8 to make up the MLVA-15. The MLVA-15 in combination with the canSNPs were used by van Ert *et al*. (2007) to characterize the global distribution of genotypes of more than 1000 isolates around the world. With the MLVA-15 a high resolution within the major clades of *B. anthracis* (van Ert *et al*., 2007; Marston *et al*., 2011) is possible (Figure 2.8). The MLVA-25 was introduced by a European group to discriminate *B. anthracis* strains using agarose electrophoresis (Le Fleche *et al*., 2001; Lista *et al*., 2006; Ciammaruconi *et al*., 2008). The MLVA-15 and MLVA-25 methods were combined to form 31 VNTRs loci (MLVA-31) that show a superior discriminating power with a high resolution (Beyer *et al*., 2012).

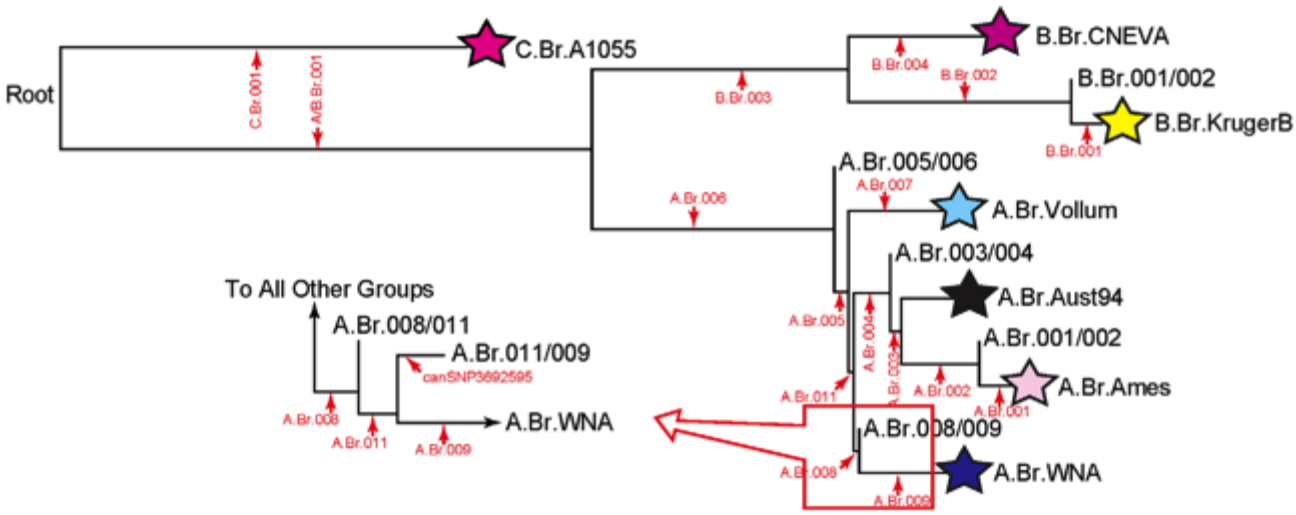


Figure 2.8. Phylogenetic structure of *Bacillus anthtacis* showing the relationship of three major lineages based on the canonical single nucleotide polymorphisms (canSNPs) and multiple locus Variable number tandem repeats assay (MLVA)-15 (Van Ert *et al*., 2007; Marston *et al*., 2011).

2.19.4. Whole genome sequencing (WGS)

The MLST and MLVA methods made it possible to separate closely related lineages intraspecies. However, using a single or a few genes can give unexpected results, as the selected genes are not fully representative for the genome. Using the high throughput sequencing technology, it is now possible to sequence all genes of the genome simultaneously (Rasko *et al*., 2005; Segerman *et al*., 2011). The complete genome sequencing of *B. anthracis* and *Bacillus* species enabled accurate comparative analyses, to identify genes that are conserved amongst closely related species, and genes unique for each species (Ivanova *et al*., 2003).

Genome sequencing projects of *B. anthracis* shows that up to date, about 29 genomes are available at genebank of national centre for biotechnology information (NCBI). However more than one hundred *B. anthracis* strains worldwide have been sequenced for genetic population studies (Pearson *et al*., 2004; Van Ert *et al*., 2007) and for comparative genomics (Rasko *et al*., 2005; 2011).

2.20. Next generation sequencing (NGS)

The development of next generation sequencing technology has led to the production of sequencing machines with lower costs and higher throughput than previously available (Pop and Salzberg, 2008) making whole genome sequence faster, less expensive and more feasible to implement for a number of projects (Shendure and Ji, 2008; Ansorge, 2009). Sequencers such as 454 Life sciences/Roche, Solexa/Illumina, Applied Biosystems (SOLID technology) and Helicos are already available. Sequencing enables determining whole genome variability within species and subspecies groups as well as between them.

2.20.1. Next Generation platforms

The NGS platforms that are commercially available include the 454 Life sciences/Roche, Solexa/Illumina Genome Analyzer and Applied Biosystems (SOLID technology). The basic principles of these platforms are illustrated on Figure 2.9. The Polonator and Helicos/HeliScope platforms have been introduced to the market in 2008 but are not widely used (Eid *et al*., 2009; Pushkarev *et al*., 2009). Technologies that are currently being developed and with promise to offer faster sequencing speed and more affordable prices are listed in Table 2.2. These will most likely available in the market within the next few years (Zhang *et al*., 2011).

2.20.2 Roche /454 pyrosequencing

The 454-genome sequencer (now own by Roche) was the first next generation sequencing technology commercially available on the market in 2005 (http://www.454.com). The sequencing by synthesis approach used by this technology method is based on emulsion PCR and pyrophosphate detection (Figure 2.9). A library of DNA templates is constructed by a highly efficient DNA amplification method known as emulsion PCR, where sheared DNA fragments are ligated to specific oligonucleotide adapters, resulting in the binding of each DNA fragment to a

fragment-carrying bead (Margulies *et al*., 2005). The beads that are used have sulphyryase and luciferase attached to them that are loaded into the wells. The beads function as amplification reactors to produce approximately 10 million clonal copies of DNA template that are needed for sufficient light signal intensities (Margulies *et al*., 2005; Shendure and Ji, 2008). During each cycle of a pyrosequencing reaction, a single species of unlabeled nucleotide is supplied to the reaction mixture to all beads on the chip, enabling syntheses of complementary strand of the DNA template. With the incorporation of each base in the growing strand, an inorganic pyrophosphate (PPi) group is released and converted to ATP that is then used by the luciferase enzyme to emit photons (Ronaghi *et al.,* 1996). Initially, the read lengths of 454 sequence were 100 bp but are now able to generate an average length of 400 bp with 600 bp as maximum capacity. The 454 sequencer has the longest short reads among all the NGS platforms and generates ~ 400-600 Mb of sequence reads per run (Zhang *et al*., 2011).

Table 2.2. Detailed information of next generation sequencing (NGS) platforms (Metzker *et al*., 2009; Zhang *et al*., 2011)

Technology	Amplification	Read length	Throughput	Sequence by synthesis	Run time (days)	Prons	Cons
Currently available							
Roche/GS-FLX Titanium	Emulsion PCR	400-600 bp	500Mbp/run	Pyrosequencing	0.35	Longer reads improve mapping in repetitive regions; fast run times	High reagent cost; high error rates in homo- polymer repeats
Illumina/HiSeq 2000, HiScan, MiSeq	Bridge PCR (Cluster PCR)	2 × 100 bp/ 2 × 250 bp	200Gbp/run	Reversible terminators	1.5-8	Currently the most widely used platform in the field	Low multiplexing capability of samples
ABI/SOLiD 5500xl	Emulsion PCR	50-100 bp	>100Gbp/run	Sequencing-by-ligation	7-14	Two-base encoding provides inherent error correction	Long run times
Polonator/G.007	Emulsion PCR	26 bp	8-10Gbp/run	Sequencing-by-ligation	5	Least expensive platform; open source to adapt alternative NGS chemistries	Users are required to maintain and quality control reagents; shortest NGS read lengths
Helicos/HeliScope	No	35 (25-55)bp	21-37 Gbp/run	True single-molecule sequencing (tSMS)	8	Non-bias representation of templates for genome and seq-based applications	High error rates compared with other reversible terminator chemistries

In development							
Pacific BioSciences/RS	No	1000 bp	N/A	Single-molecule real-time (SMRT)	N/A	N/A	N/A
Visigen Biotechnologies	No	>100 Kbp	N/A	Base-specific FRET	N/A	N/A	N/A
U.S. Genomics	No	N/A	N/A	Single-molecule mapping	N/A	N/A	N/A
Genovoxx	No	N/A	N/A	Single-molecule sequencing by synthesis	N/A	N/A	N/A
Oxford Nanopore Technologies	No	35 bp	N/A	Nanopores/exonuclease-coupled	N/A	N/A	N/A
NABsys	No	N/A	N/A	Nanopores	N/A	N/A	N/A
Electronic BioSciences	No	N/A	N/A	Nanopores	N/A	N/A	N/A
BioNanomatrix/nano Analyzer	No	400 Kbp	N/A	Nanochannel arrays	N/A	N/A	N/A
GE Global Research	No	N/A	N/A	Closed complex/ nanoparticle	N/A	N/A	N/A
IBM	No	N/A	N/A	Nanopores	N/A	N/A	N/A
LingVitae	No	N/A	N/A	Nanopores	N/A	N/A	N/A
Complete Genomics	No	70 bp	N/A	DNA nanoball arrays	N/A	N/A	N/A

N/A = Not available

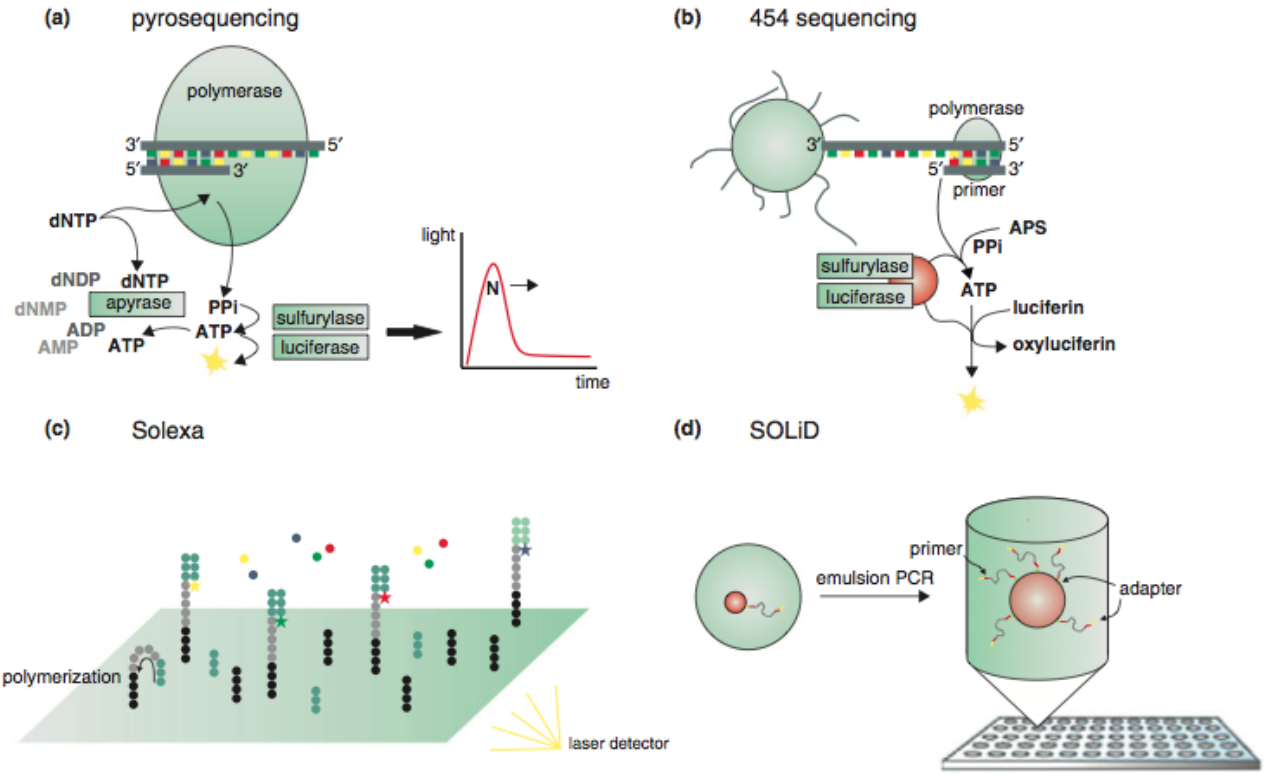


Figure 2.9. Basic principles of next generation sequencing (NGS) techniques. (a) Pyrosequencing with the incorporation of a new nucleotide generates detectable light. (b) The 454 sequencing where nucleotide incorporation is associated with the release of pyrophosphate resulting in a light signal. (c) Solexa consisting of DNA fragments building double-stranded bridges and after the addition of the labeled terminators the sequencing cycle starts. (d) SOLiD where adapters are bound, emulsion PCR is carried out to generate so-called bead clones (Mutz *et al*., 2012).

2.20.3. SOLID platform (Applied Biosystems)

The SOLID (Supported Oligonucleotide Ligation and Detection) platform uses a hybridization-ligation chemistry. Adapters coupled to microparticles are linked to DNA fragments and applied to an emulsion PCR system with small magnetic beads (Shendure *et al*., 2008) (Figure 2.8). This platform uses smaller beads than with 454 sequencing. The read length of SOLID analyzer is up to 50 bp (2 × 50 mate pairs) and can produce 80-100 Gbp of sequence per run. The latest 5500xl SOLID system can generate over 2.4 billion reads per run with a raw base accuracy of 99.94% (Zhang *et al*., 2011). This system provide the best quality data as a result of its sequencing by ligation approach. The limiting factor of SOLID is that it runs slower when compared to other systems and there is a degradation of template strand over time (Hossain *et al*., 2009; Mutz *et al*., 2012). Another

limiting factor of this system is the tedious and time consuming preparation of the DNA library (Zhang *et al*., 2011).

2.20.4. Illumina/ Solexa Genome Analyzer

After the introduction of the 454 technology, the Illumina/Solexa genome analyzer was the second platform to reach the market (Bentley, 2006) and currently is the most widely used system. An outline of library preparation to sequencing is represented below (Figure 2.10). Illumina platform uses a mixture of four fluorescently unique reversible dye terminators that are simultaneously introduced into oligo-primed cluster fragments in the flow cell channels along with the DNA polymerase. Sequence by synthesis is coupled by incorporating complimentary bases into individual fragment clusters, and by recordeding the bases specific moieties linked to the nucleotide bases and the 3' deoxyribose sugar position (Voelkerding *et al*., 2010). Illumina is widely used in most sequencing projects due to the superior data quality and proper read lengths. To date, the sequencing platform is still recognized as the most adaptable and easiest to use, and the productivity of the generated data is massive i.e. it generates about 200 Gbp of short sequence per run on pair-end reads (Zang *et al*., 2011). With the recently added sequencing Illumina machine (Miseq) that can cover read length of 2 × 250 bps (http://www.illumina.com).

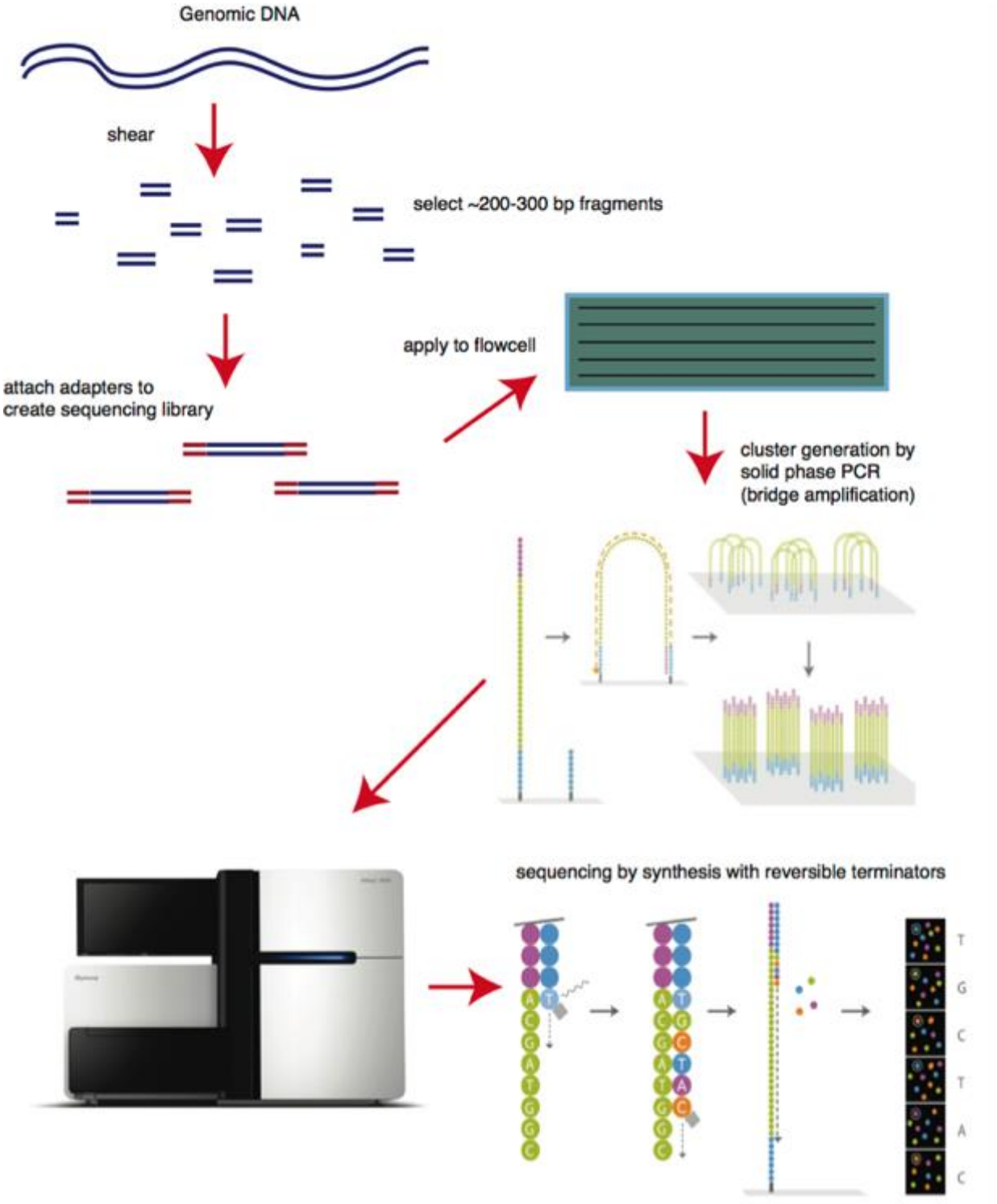


Figure 2.10. Schematic representation of the Illumina genome analyzer platform illustrating the three steps of sequencing. The DNA sample is prepared into a sequencing library by fragmentation into pieces each around 100-300 bases long. Adapters are added to each end and the library is flowed across a solid surface (flow cell). Following solid bridge amplification PCR process (cluster generation), the sequencing by synthesis with reversible terminators (http://www.illumina.com).

2.21. Bioinformatics for next generation sequencing

2.21.1. Bacterial genome assembly

DNA sequencing technologies cannot obtain the whole genome sequences, only small part of the DNA known as a read can be obtained. In order to obtain the original genome sequences, specific programs called genome assemblers are used to assemble or piece the genome sequence from the reads. This is essential for characterization of the organism's genome and for comparison studies. Many genome assemblers have been reported depending on the chosen-sequencing platform such as Burrows Wheeler Alignment (BWA) (Li and Durbin, 2009), Samtools (Li *et al*., 2009), EDENA (Hernandez *et al*., 2008), ALLPATHS (Butler *et al*., 2008), SOAPdenovo, SHARCGS (Dohm *et al*., 2007), Velvet (Zerbino and Birney, 2008), and Abyss (Simpson *et al*., 2009). Several genome assemblers and utility bioinformatics tools reported by Zhang *et al*. (2011) are based on reference-based and/or *de-novo* assembly.

2.21.2. Referenced-based assembly

Once the sequencing platforms have generated enough data for analysis, the most important part is to assembly the short reads to a reference genome (Flicek and Birney, 2009). Some challenges when using reference mapping are with the development of new algorithms to handle some uncertainly reads to the reference genome or lack of accuracy during alignment (Li and Homer, 2010; Zhang *et al*., 2011).

2.21.3. *De-novo* assembly

De-novo approach is important when dealing with unknown organism genome (with no reference genome sequence available). The approach involves the step of combining all short reads sequences into so-called contigs (contiguous sequences). *De-novo* assembly is essential to give the most suitable identification of the genome and identify unique sequences. One of the major drawbacks of *de-novo* assembly when compared to reference-based assembly is the short read lengths that makes assembly of the genome challenging. The majority of the *de-novo* assemblers have been created for prokaryotes rather than eukaryotes due to small bacterial genome size. Currently *de-novo* assembly with NGS data is generally limited (Wooley *et al*., 2010; Zhang *et al*. 2011).

2.21.4. End user packages

Convenient end user packages are user-friendly with easy data input and output formats. The packages include multiple computing programs compiled into one software package. CLC-Bio Genomic workbench is the mostly widely used end user package (Table 2.3). NextGENe from softGenetics is used for candidate gene resequencing projects but is not suitable for large genome sequencing projects (Zhang *et al*., 2011).

Table 2.3. End users software packages for next generation sequencing (NGS) with multi-task functions.

Software packages	Website
Genomic workbench/CLCbio	http://www.clcbio.com/index.php?id=1331
NextGENe/SoftGenetics	http://softgenetics.com/NextGENe.html
Genomatix Genome Analyzer	http://www.genomatix.de/genome_analyzer.html
Zoom	http://www.bioinformaticssolutions.com/products/zoom/index.php
SeqMan Ngen/DNASTAR	http://www.dnastar.com/t-products-seqman-ngen.aspx
JMP Genomics	http://www.jmp.com/software/genomics/index.shtml
RTG/Real-time Genomics	http://www.realtimegenomics.com/RTG-Software
PASS	http://pass.cribi.unipd.it/cgi-bin/pass.pl?action=Download
CASAVA	http://www.illumina.com/software/
Geneus/GenoLogics	http://www.genologics.com/solutions/research-informatics/
Roche Analysis tools	http://454.com/products-solutions/analysis-tools/index.asp
VSRAP	http://sourceforge.net/apps/mediawiki/vancouvershortr/
BING	http://www.dinulab.org/bing
PaCGeE/PGI	http://personalgenomicsinstitute.org/index.php/
GATK	http://www.broadinstitute.org/gsa/wiki/index.php/
Geneious Pro	http://www.geneious.com/default,1246,NGS%2520Assembly.sm
Partek GS/Partek	http://www.partek.com/partekgs
Bioscope	https://products.appliedbiosystems.com/ab/en/US/adirect/

2.21.5. Annotation of bacterial genome

Once the genome sequence of a bacterium is completed, the next step is to interpret the genome by using gene prediction programs that scan the sequence for coding proteins and functional RNA products. Identified genes are compared with existing databases of DNA or protein sequences to identify related sequences. Information and function of the related sequence is transferred to the new sequence. Annotation pipelines have been created and most employ gene prediction software such as GLIMMER (Delcher *et al*., 1999), which uses reference set of sequences to train the model enabling to predict coding regions in the genome. Various bacterial annotation pipelines have been published and

the best online automated annotation systems are RAST (Rapid Annotation using subsystem technology), IMG (Integrated Microbial Genomes), BASys (Bacterial Annotation Systems), JCVI (J. Craig Venter Institute), PGAAP (Prokaryotic Genomes Automatic Annotation Pipeline at NCBI (Kisand and Lettieri, 2013) and systems that are locally installed include AGeS, DIYA, PIPA etc. (Stothard and Wishart, 2006; Richardson and Watson, 2012). RAST system allows browsing of the annotated genomes and supports the use of external comparative tools for analyses (Aziz *et al*., 2008; Kisand and Lettieri, 2013). Program such as MICheck that checks annotated sequences for syntactic errors and frameshifts is also available (Cruveiller *et al*., 2005).

2.21. 6. Comparative genome analysis

Comparative genome analyses enable search of unique genes and shared genes amongst genomes. This is accomplished by genome alignment tools that compares the genome of interest with the reference or compared genome. Comparative genomics is used for identification of specific genes with important functions such as virulence genes or drug resistance determinants (Edward and Holt, 2013). The software tools that are commonly used include BRIG (BLAST Ring Image Generator), Mauve, ACT (Artemis Comparative Tool), MUMmer, and CGView or BLAST. Mauve is a java-based graphical-interface program that provides one way to visualize multiple alignments of whole genomes, with a built in viewer and the option to export comparative genomic information in various forms (Darling *et al.,* 2010; Edwards and Holt, 2013).

2.22. Reference

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. J. Cell. Biol.160, 321-328.

Agaisse, H., Gominet, M., Okstad, O.A., Kolsto, A., Lereclus, D., 1999. PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. Mol. Microbiol. 32, 1043-1053.

Ansorge, W.J., 2009. Next-generation DNA sequencing techniques. New biotechnology 25, 195-203.

Aronson, A., 2002. Sporulation and δ-endotoxin synthesis by *Bacillus thuringiensis*. Cell. Mol. Life. Sci. 59, 417-425.

Ash, C., Farrow, J., Wallbanks, S., Collins, M., 1991. Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small subunit ribosomal RNA sequences. Lett. Appl. Microbiol. 13, 202-206.

Ash, C., Collins, M.D., 1992. Comparative analysis of 23S ribosomal RNA gene sequences of *Bacillus anthracis* and emetic *Bacillus cereus* determined by PCR-direct sequencing. FEMS Microbiol. Lett. 94, 75-80.

Avashia, S.B., Riggins, W., Lindley, C., Hoffmaster, A., Drumgoole, R., Nekomoto, T., Jackson, P.J., Hill, K.K., Williams, K., Lehman, L., 2007. Fatal pneumonia among metalworkers due to inhalation exposure to *Bacillus cereus* containing *Bacillus anthracis* toxin genes. Clin. Infect. Dis. 44, 414-416.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75.

Baillie, L., Jones, M., Turnbull, P., Manchee, R., 1995. Evaluation of the Biolog system for the identification of *Bacillus anthracis*. Lett. Appl. Microbiol. 20, 209-211.

Bayvkin, S., Lysov, Y., Zakhariev, V., Kelly, J., Jackman, J., Stahl, D., Cherni, A., 2004. Use of 16S rRNA, 23S rRNA, and gyrB gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group. J. Clin. Microbiol. 42.

Beesley, C.A., Vanner, C.L., Helsel, L.O., Gee, J.E., Hoffmaster, A.R., 2010. Identification and characterization of clinical *Bacillus* spp. isolates phenotypically similar to *Bacillus anthracis*. FEMS Microbiol. Lett. 313, 47-53.

Bell, W., Laing, P., 1977. Pulmonary anthrax in cattle. Vet. Rec. 100, 573-574.

Bentley, D.R., 2006. Whole-genome re-sequencing. Curr. Opin. Genet. Dev. 16, 545-552.

Beyer, W., Bohm, R., 2003. Pitfalls in the diagnosis of *Bacillus anthracis* from environmental samples-Lessons from investigations of suspicious specimens. ISAH, Mexico. Htttp://www.isahsoc.org/documents/speakers/S45%20BeyerGermany.doc

Beyer, W., Turnbull, P., 2009. Anthrax in animals. Mol. Aspects Med. 30, 481-489.

Beyer, W., Pocivalsek, S., 1999. Polymerase chain reaction-ELISA to detect *Bacillus anthracis* from soil samples—limitations of present published primers. J. Appl. Microbiol. 87, 229-236.

Beyer, W., Bellan, S., Eberle, G., Ganz, H.H., Getz, W.M., Haumacher, R., Hilss, K.A., Kilian, W., Lazak, J., Turner, W.C., 2012. Distribution and molecular evolution of *Bacillus anthracis* genotypes in Namibia. PLoS Neglect. Trop. Dis. 6, e1534.

Beyer, W., Glockner, P., Otto, J., Bohm, R., 1995. A nested PCR method for the detection of *Bacillus anthracis* in environmental samples collected from former tannery sites. Microbiol. Res. 150, 179-186.

Birdsell, D.N., Antwerpen, M., Keim, P., Hanczaruk, M., Foster, J.T., Sahl, J.W., Wagner, D.M., Grass, G., 2013. Draft genome sequences of two Bulgarian *Bacillus anthracis* strains. Genome announc. 1 e00152-13.

Braack, L., De Vos, V., 1990. Feeding habits and flight range of blow-flies (Chrysomyia spp.) in relation to anthrax transmission in the Kruger National Park, South Africa. Onderstepoort J. Vet. Res. 57, 141-142.

Bradley, K.A., Mogridge, J., Mourez, M., Collier, R.J., Young, J.A.T., 2001. Identification of the cellular receptor for anthrax toxin. Nature 414, 225-229.

Buck, C., Anacker, R., Newman, F., Eisenstark, A., 1963. Phage isolated from lysogenic *Bacillus anthracis*. J. Bacteriol. 85, 1423-1430.

Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I.A., Belmonte, M.K., Lander, E.S., Nusbaum, C., Jaffe, D.B., 2008. ALLPATHS: de novo assembly of whole-genome shotgun microreads. Genome Res. 18, 810-820.

Candela, T., Fouet, A., 2005. *Bacillus anthracis* CapD, belonging to the γ-glutamyltranspeptidase family, is required for the covalent anchoring of capsule to peptidoglycan. Mol. Microbiol. 57, 717-726.

Candela, T., Mock, M., Fouet, A., 2005. CapE, a 47-amino-acid peptide, is necessary for *Bacillus anthracis* polyglutamate capsule synthesis. J. Bacteriol. 187, 7765-7772.

Ciammaruconi, A., Grassi, S., De Santis, R., Faggioni, G., Pittiglio, V., D'Amelio, R., Carattoli, A., Cassone, A., Vergnaud, G., Lista, F., 2008. Fieldable genotyping of *Bacillus anthracis* and Yersinia pestis based on 25-loci Multi Locus VNTR Analysis. BMC Microbiol. 8, 21.

Coker, P.R., Smith, K.L., Hugh-Jones, M.E., 2002. Antimicrobial susceptibilities of diverse *Bacillus anthracis* isolates. Antimicrob. Agents Chemother. 46, 3843-3845.

Collier, R.J., Young, J.A., 2003. Anthrax toxin. Annu. Rev. Cell Dev. Biol. 19, 45-70.

Cruveiller, S., Le Saux, J., Vallenet, D., Lajus, A., Bocs, S., Medigue, C., 2005. MICheck: a web tool for fast checking of syntactic annotations of bacterial genomes. Nucleic Acids Res. 33, W471-W479.

Darling, A.E., Mau, B., Perna, N.T., 2010. Progressive Mauve: multiple genome alignment with gene gain, loss and rearrangement. PloS One 5, e11147.

De Vos, V., 1990. The ecology of anthrax in the Kruger National Park, South Africa. Salisbury Med. Bull Suppl. 68, 19-23.

De Vos, V., Bryden, H., 1996. Anthrax in the Kruger National Park: temporal and spatial patterns of disease occurrence. Salisbury Med. Bull Suppl. 87, 26-30.

Delcher, A.L., Harmon, D., Kasif, S., White, O., Salzberg, S.L., 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27, 4636-4641.

Dohm, J.C., Lottaz, C., Borodina, T., Himmelbauer, H., 2007. SHARCGS, a fast and highly accurate short-read assembly algorithm for de novo genomic sequencing. Genome Res. 17, 1697-1706.

Dragon, D.C., Rennie, R.P., 1995. The ecology of anthrax spores: tough but not invincible. Can. Vet. J. 36, 295-301.

Driks, A., 2009. The *Bacillus anthracis* spore. Mol. Aspects Med. 30, 368-373.

Driks, A., Setlow, P., 2000. Morphogenesis and properties of the bacterial spore. In: Brun, Y.V., Shimkets, L.J. (Eds), Prokaryotic development. American Society for Microbiology, Washington, DC, 191-218.

Drysdale, M., Bourgogne, A., Hilsenbeck, S.G., Koehler, T.M., 2004. AtxA controls *Bacillus anthracis* capsule synthesis via acpA and a newly discovered regulator, acpB. J. Bacteriol. 186, 307-315.

Drysdale, M., Heninger, S., Hutt, J., Chen, Y., Lyons, C.R., Koehler, T.M., 2004. Capsule synthesis by *Bacillus anthracis* is required for dissemination in murine inhalation anthrax. EMBO J. 24, 221-227.

Easterday, W.R., Van Ert, M.N., Simonson, T.S., Wagner, D.M., Kenefic, L.J., Allender, C.J., Keim, P., 2005. Use of single nucleotide polymorphisms in the plcR gene for specific identification of *Bacillus anthracis*. J. Clin. Microbiol. 43, 1995-1997.

Edwards, D.J., Holt, K.E., 2013. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. Microb. Inform. Exp. 3, 2.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., 2009. Real-time DNA sequencing from single polymerase molecules. Science 323, 133-138.

Farrar, W.E., 1994. Anthrax: virulence and vaccines. Ann. Intern. Med. 121, 379-380.

Fasanella, A., Galante, D., Garofolo, G., Jones, M.H., 2010. Anthrax undervalued zoonosis. Vet. Microbiol. 140, 318-331.

Flicek, P., Birney, E., 2009. Sense from sequence reads: methods for alignment and assembly. Nat. Methods 6, S6-S12.

Fouet, A., Mock, M., 2006. Regulatory networks for virulence and persistence of *Bacillus anthracis*. Curr. Opin. Microbiol. 9, 160-166.

Fox, A., Stewart, G.C., Waller, L.N., Fox, K.F., Harley, W.M., Price, R.L., 2003. Carbohydrates and glycoproteins of *Bacillus anthracis* and related bacilli: targets for biodetection. J. Microbiol. Methods 54, 143-152.

Gierczynski, R., Zasada, A., Kałuzewski, S., Jagielski, M., 2006. Molecular characterisation of the haemolytic isolates of *Bacillus anthracis* originated in Poland. Med. Dosw. Mikrobiol. 58, 345.

Green, B., Battisti, L., Koehler, T., Thorne, C., Ivins, B., 1985. Demonstration of a capsule plasmid in *Bacillus anthracis*. Infect. Immun. 49, 291-297.

Gnade, B.T., Moen, S.T., Chopra, A.K., Peterson, J.W., Yeager, L.A., 2010. Emergence of anthrax edema toxin as a master manipulator of macrophage and B cell functions. Toxins 2, 1881-1897.

Hambleton, P., Carman, J.A., Melling, J., 1984. Anthrax: the disease in relation to vaccines. Vaccine 2, 125-132.

Helgason, E., Okstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I., Kolsto, A.B., 2000. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. Appl. Environ. Microbiol. 66, 2627-2630.

Helgason, E., Tourasse, N.J., Meisal, R., Caugant, D.A., Kolsto, A., 2004. Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. Appl. Environ. Microbiol. 70, 191-201.

Henton, M., Briers, G., 1998. Anthrax spores isolated above burning carcases. In: Proceedings of the ARC-Onderstepoort international congress on anthrax, brucellosis, CBPP, clostridial and mycobacterial diseases, Kruger Park, South Africa, 9–15 August 1998. Onderstepoort, Onderstepoort Veterinary Institute 37-40.

Hernandez, D., François, P., Farinelli, L., Osteras, M., Schrenzel, J., 2008. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res. 18, 802-809.

Hoffmaster, A.R., Ravel, J., Rasko, D.A., Chapman, G.D., Chute, M.D., Marston, C.K., De, B.K., Sacchi, C.T., Fitzgerald, C., Mayer, L.W., 2004. Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. Proc. Natl. Acad. Sci. U. S. A. 101, 8449-8454.

Hoffmaster, A.R., Hill, K.K., Gee, J.E., Marston, C.K., De, B.K., Popovic, T., Sue, D., Wilkins, P.P., Avashia, S.B., Drumgoole, R., 2006. Characterization of *Bacillus cereus* isolates associated with fatal pneumonias: strains are closely related to *Bacillus anthracis* and harbor *B. anthracis* virulence genes. J. Clin. Microbiol. 44, 3352-3360.

Hoffmaster, A.R., Meyer, R.F., Bowen, M.P., Marston, C.K., Weyant, R.S., Thurman, K., Messenger, S.L., Minor, E.E., Winchell, J.M., Rasmussen, M.V., 2002. Evaluation and validation of a real-time polymerase chain reaction assay for rapid identification of *Bacillus anthracis*. Emerg. Infect. Dis. 8, 1178-1182.

Hossain, M.S., Azimi, N., Skiena, S., 2009. Crystallizing short-read assemblies around seeds. BMC Bioinformatics 10, S16.

Hu, X., Van der Auwera, G., Timmery, S., Zhu, L., Mahillon, J., 2009. Distribution, diversity, and potential mobility of extrachromosomal elements related to the *Bacillus anthracis* pXO1 and pXO2 virulence plasmids. Appl. Environ. Microbiol. 75, 3016-3028.

Hudson, M.J., Beyer, W., Bohm, R., Fasanella, A., Garofolo, G., Golinski, R., Goossens, P.L., Hahn, U., Hallis, B., King, A., 2008. *Bacillus anthracis*: Balancing innocent

research with dual-use potential. Int. J. Med. Microbiol. 298, 345-364.

Hugh-Jones, M., Blackburn, J., 2009. The ecology of *Bacillus anthracis*. Mol. Aspects Med. 30, 356-367.

Hugh-Jones, M.E., de Vos, V., 2002. Anthrax and wildlife. Rev. Sci. Tech. 21, 359-383.

Hurtle, W., Bode, E., Kulesh, D.A., Kaplan, R.S., Garrison, J., Bridge, D., House, M., Frye, M.S., Loveless, B., Norwood, D., 2004. Detection of the gyrA gene by using a minor groove binder probe. J. Clin. Microbiol. 42, 179-185.

Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Friedlander, A.M., Hauer, J., McDade, J., Osterholm, M.T., O'Toole, T., 1999. Anthrax as a biological weapon. JAMA 281, 1735-1745.

Inglesby, T.V., O'Toole, T., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Friedlander, A.M., Gerberding, J., Hauer, J., Hughes, J., 2002. Anthrax as a biological weapon, 2002. JAMA 287, 2236-2252.

Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., Bhattacharyya, A., Reznik, G., Mikhailova, N., Lapidus, A., 2003. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature 423, 87-91.

Jensen, G., Hansen, B., Eilenberg, J., Mahillon, J., 2003. The hidden lifestyles of *Bacillus cereus* and relatives. Environ. Microbiol. 5, 631-640.

Jernigan, J.A., Stephens, D.S., Ashford, D.A., Omenaca, C., Topiel, M.S., Galbraith, M., Tapper, M., Fisk, T.L., Zaki, S., Popovic, T., 2001. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. Emerg. Infect. Dis. 7, 933-944.

Kamal, S., Rashid, A., Bakar, M., Ahad, M., 2011. Anthrax: an update. Asian Pac. J. Trop. Biomed. 1, 496-501.

Keim, P., Klevytska, A., Price, L., Schupp, J., Zinser, G., Smith, K., Hugh-Jones, M., Okinaka, R., Hill, K., Jackson, P., 1999. Molecular diversity in *Bacillus anthracis*. J. Appl. Microbiol. 87, 215-217.

Keim, P., Price, L., Klevytska, A., Smith, K., Schupp, J., Okinaka, R., Jackson, P., Hugh-Jones, M., 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. 182, 2928-2936.

Keim, P., Gruendike, J.M., Klevytska, A.M., Schupp, J.M., Challacombe, J., Okinaka, R., 2009. The genome and variation of *Bacillus anthracis*. Mol. Aspects Med. 30, 397-405.

Keim, P., Van Ert, M.N., Pearson, T., Vogler, A.J., Huynh, L.Y., Wagner, D.M., 2004. Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. Infect. Genet. Evol. 4, 205-213.

Kim, K., Cheon, E., Wheeler, K.E., Youn, Y., Leighton, T.J., Park, C., Kim, W., Chung, S., 2005. Determination of the most closely related bacillus isolates to *Bacillus anthracis* by multilocus sequence typing. Yale J. Biol. Med. 78, 1-14.

Kisand, V., Lettieri, T., 2013. Genome sequencing of bacteria: sequencing, de novo

assembly and rapid analysis using open source tools. BMC Genomics 14, 211.

Klee, S.R., Brzuszkiewicz, E.B., Nattermann, H., Bruggemann, H., Dupke, S., Wollherr, A., Franz, T., Pauli, G., Appel, B., Liebl, W., 2010. The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids. PLoS One 5, e10986.

Klee, S.R., Ozel, M., Appel, B., Boesch, C., Ellerbrok, H., Jacob, D., Holland, G., Leendertz, F.H., Pauli, G., Grunow, R., 2006. Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. J. Bacteriol. 188, 5333-5344.

Knisely, R.F., 1966. Selective medium for *Bacillus anthracis*. J. Bacteriol. 92, 784.

Koehler, T.M., 2009. *Bacillus anthracis* physiology and genetics. Mol. Aspects Med. 30, 386-396.

Kolsto, A.B., Tourasse, N.J., Okstad, O.A., 2009. What sets *Bacillus anthracis* apart from other *Bacillus* species? Annu. Rev. Microbiol. 63, 451-476.

Koser, C.U., Ellington, M.J., Cartwright, E.J., Gillespie, S.H., Brown, N.M., Farrington, M., Holden, M.T., Dougan, G., Bentley, S.D., Parkhill, J., 2012. routine use of microbial whole genome sequencing in diagnostic and public health Microbiology. PLoS Pathogens 8, e1002824.

Kotiranta, A., Lounatmaa, K., Haapasalo, M., 2000. Epidemiology and pathogenesis of *Bacillus cereus* infections. Microb. Infect. 2, 189-198.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A., Alloni, G., Azevedo, V., Bertero, M., Bessieres, P., Bolotin, A., Borchert, S., 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390, 249-256.

Lane, D., 1991. 16S/23S rRNA sequencing. In: Stackerbrandt, E., Goodfellow M. (eds), Nucleic acid techniques in bacterial systematics, John Wiley and Sons Chichester, New York, 115-175.

Lapidus, A., Goltsman, E., Auger, S., Galleron, N., Segurens, B., Dossat, C., Land, M.L., Broussolle, V., Brillard, J., Guinebretiere, M., 2008. Extending the group genomics to putative food-borne pathogens of different toxicity. Chem. Biol. Interact. 171, 236-249.

Le Fleche, 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 Microbiol 1, 2.

Lee, K., Costerton, J., Ravel, J., Auerbach, R.K., Wagner, D.M., Keim, P., Leid, J.G., 2007. Phenotypic and functional characterization of *Bacillus anthracis* biofilms. Microbiol 153, 1693-1701.

Leendertz, F.H., Ellerbrok, H., Boesch, C., Couacy-Hymann, E., Mätz-Rensing, K., Hakenbeck, R., Bergmann, C., Abaza, P., Junglen, S., Moebius, Y., 2004. Anthrax kills wild chimpanzees in a tropical rainforest. Nature 430, 451-452.

Leendertz, F.H., Lankester, F., Guislain, P., Neel, C., Drori, O., Dupain, J., Speede, S., Reed, P., Wolfe, N., Loul, S., 2006. Anthrax in Western and Central African great apes. Am. J. Primatol. 68, 928-933.

Leppla, S.H., 1982. Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells. PNAS. 79, 3162-3166.

Leppla, S.H., Robbins, J.B., Schneerson, R., Shiloach, J., 2002. Development of an improved vaccine for anthrax. J. Clin. Invest. 110, 141-144.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754-1760.

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.

Li, H., Homer, N., 2010. A survey of sequence alignment algorithms for next-generation sequencing. Brief. Bioinform. 11, 473-483.

Lindeque, P., Turnbull, P., 1994. Ecology and epidemiology of anthrax in the Etosha National Park, Namibia. Onderstepoort J. Vet. Res. 61, 71-83.

Lista, F., Faggioni, G., Valjevac, S., Ciammaruconi, A., Vaissaire, J., Le Doujet, C., Gorge, O., De Santis, R., Carattoli, A., Ciervo, A., 2006. Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. BMC Microbial. 6, 33.

Liu, H., Bergman, N.H., Thomason, B., Shallom, S., Hazen, A., Crossno, J., Rasko, D.A., Ravel, J., Read, T.D., Peterson, S.N., 2004. Formation and composition of the *Bacillus anthracis* endospore. J. Bacteriol. 186, 164-178.

Logan, N., Carman, J., Melling, J., Berkeley, R., 1985. Identification of *Bacillus anthracis* by API tests. J. Med. Microbiol. 20, 75-85.

Makino, S.I., Iinuma-Okada, Y., Maruyama, T., Ezaki, T., Sasakawa, C., and Yoshikawa, M., 1993. Direct Detection of *Bacillus anthracis* DNA in Animals by Polymerase Chain Reaction. J. Clin. Microbiol. 31, 547- 551.

Makino, S., Uchida, I., Terakado, N., Sasakawa, C., Yoshikawa, M., 1989. Molecular characterization and protein analysis of the cap region, which is essential for encapsulation in *Bacillus anthracis*. J. Bacteriol. 171, 722-730.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y., Chen, Z., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376-380.

Marston, C.K., Allen, C.A., Beaudry, J., Price, E.P., Wolken, S.R., Pearson, T., Keim, P., Hoffmaster, A.R., 2011. Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States. PloS One 6, e28274.

Marston, C.K., Gee, J.E., Popovic, T., Hoffmaster, A.R., 2006. Molecular approaches to identify and differentiate *Bacillus anthracis* from phenotypically similar *Bacillus* species isolates. BMC Microbial. 6, 22.

Metzker, M.L., 2009. Sequencing technologies—the next generation. Nature Rev. Genet.

11, 31-46.

Moayeri, M., Haines, D., Young, H.A., Leppla, S.H., 2003. *Bacillus anthracis* lethal toxin induces TNF-α–independent hypoxia-mediated toxicity in mice. J. Clin. Invest. 112, 670-682.

Mock, M., Fouet, A., 2001. Anthrax. Ann. Rev. Microbiol. 55, 647-671.

Mock, M., Mignot, T., 2003. Anthrax toxins and the host: a story of intimacy. Cell. Microbiol. 5, 15-23.

Modrie, P., Beuls, E., Mahillon, J., 2010. Differential transfer dynamics of pAW63 plasmid among members of the *Bacillus cereus* group in food microcosms. J. Appl. Microbiol. 108, 888-897.

Mutz, K., Heilkenbrinker, A., Lonne, M., Walter, J., Stahl, F., 2012. Transcriptome analysis using next-generation sequencing. Curr. Opin. Biotechnol. .

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.

Odendaal, M., Pieterson, P., De Vos, V., Botha, A., 1991. The antibiotic sensitivity patterns of *Bacillus anthracis* isolated from the Kruger National Park. Onderstepoort J. Vet. Res. 58, 17-19.

Oh, S., Budzik, J.M., Garufi, G., Schneewind, O., 2011. Two capsular polysaccharides enable *Bacillus cereus* G9241 to cause anthrax-like disease. Mol. Microbiol. 80, 455-470.

OIE, AHS. Manual of diagnostic tests and vaccines for terrestrial animals. 2008. Chapter 2.1.1. Anthrax, OIE terrestrial manual, Paris, France, 133-144.

Okinaka, R., Pearson, T., Keim, P., 2006. Anthrax, but not *Bacillus anthracis*? PLoS Pathogens 2, e122.

Okinaka, R., Cloud, K., Hampton, O., Hoffmaster, A., Hill, K., Keim, P., Koehler, T., Lamke, G., Kumano, S., Mahillon, J., 1999. Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. J. Bacteriol. 181, 6509-6515.

Olsen, J.S., Skogan, G., Fykse, E.M., Rawlinson, E.L., Tomaso, H., Granum, P.E., Blatny, J.M., 2007. Genetic distribution of 295 *Bacillus cereus* group members based onadk-screening in combination with MLST (Multilocus Sequence Typing) used for validating a primer targeting a chromosomal locus in *B. anthracis*. J. Microbiol. Methods 71, 265-274.

Owen, M., Schauwers, W., Hugh-Jones, M., Kiernan, J., Turnbull, P., Beyer, W., 2013. A simple, reliable M'Fadyean stain for visualizing the *Bacillus anthracis* capsule. J. Microbiol. Methods 92, 264-269.

Pannucci, J., Okinaka, R.T., Sabin, R., Kuske, C.R., 2002. *Bacillus anthracis* pXO1 plasmid sequence conservation among closely related bacterial species. J. Bacteriol. 184, 134-141.

Papaparaskevas, J., Houhoula, D.P., Papadimitriou, M., Saroglou, G., Legakis, N.J., Zerva, L., 2004. Ruling out *Bacillus anthracis*. Emerg. Infect. Dis. 10, 732-735.

Papazisi, L., Rasko, D.A., Ratnayake, S., Bock, G.R., Remortel, B.G., Appalla, L., Liu, J., Dracheva, T., Braisted, J.C., Shallom, S., 2011. Investigating the genome diversity of *B. cereus* and evolutionary aspects of *B. anthracis* emergence. Genomics 98, 26-39.

Patra, G., Sylvestre, P., Ramisse, V., Therasse, J., Guesdon, J., 1996. Isolation of a specific chromosomic DNA sequence of *Bacillus anthracis* and its possible use in diagnosis. FEMS Immunol. Med. Microbiol 15, 223-231.

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., 2004. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. Proc. Natl. Acad. Sci. U. S. A. 101, 13536-13541.

Pearson, T., Okinaka, R.T., Foster, J.T., Keim, P., 2009. Phylogenetic understanding of clonal populations in an era of whole genome sequencing. Infect. Genet. Evol. 9, 1010-1019.

Peng, J., Tsai, W., Chou, C., 2001. Surface characteristics of *Bacillus cereus* and its adhesion to stainless steel. Int. J. Food Microbiol. 65, 105-111.

Pienaar, U.D., 1967. Epidemiology of anthrax in wild animals and the control of anthrax epizootics in the Kruger National Park, South Africa. Fed. Proc. 26, 1496-1502.

Pilo, P., Frey, J., 2011. *Bacillus anthracis*: Molecular taxonomy, population genetics, phylogeny and patho-evolution. Infect. Genet. Evol. 11, 1218-1224.

Pop, M., Salzberg, S.L., 2008. Bioinformatics challenges of new sequencing technology. Trends in Genetics 24, 142-149.

Pribil, P.A., Patton, E., Black, G., Doroshenko, V., Fenselau, C., 2005. Rapid characterization of Bacillus spores targeting species- unique peptides produced with an atmospheric pressure matrix- assisted laser desorption/ionization source. J. Mass Spectrom 40, 464-474.

Priest, F.G., Barker, M., Baillie, L.W., Holmes, E.C., Maiden, M.C., 2004. Population structure and evolution of the *Bacillus cereus* group. J. Bacteriol. 186, 7959-7970.

Prince, A.S., 2003. The host response to anthrax lethal toxin: unexpected observations. J. Clin. Invest. 112, 656-658.

Pushkarev, D., Neff, N.F., Quake, S.R., 2009. Single-molecule sequencing of an individual human genome. Nat. Biotechnol. 27, 847-850.

Qi, Y., Patra, G., Liang, X., Williams, L.E., Rose, S., Redkar, R.J., DelVecchio, V.G., 2001. Utilization of the rpoB gene as a specific chromosomal marker for real-time PCR detection of *Bacillus anthracis*. Appl. Environ. Microbiol. 67, 3720-3727.

Radnedge, L., Agron, P.G., Hill, K.K., Jackson, P.J., Ticknor, L.O., Keim, P., Andersen, G.L., 2003. Genome differences that distinguish *Bacillus anthracis* from *Bacillus cereus* and *Bacillus thuringiensis*. Appl. Environ. Microbiol. 69, 2755-2764.

Rajendhran, J., Gunasekaran, P., 2011. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. Microbiol. Res. 166, 99-110.

Ramisse, V., Patra, G., Vaissaire, J., Mock, M., 1999. The Ba813 chromosomal DNA sequence effectively traces the whole *Bacillus anthracis* community. J. Appl. Microbiol. 87, 224-228.

Ramisse, V., Patra, G., Garrigue, H., Guesdon, J., Mock, M., 1996. Identification and characterization of *Bacillus anthracis* by multiplex PCR analysis of sequences on plasmids pXO1 and pXO2 and chromosomal DNA. FEMS Microbiol. Lett. 145, 9-16.

Rasko, D.A., Altherr, M.R., Han, C.S., Ravel, J., 2005. Genomics of the *Bacillus cereus* group of organisms. FEMS Microbiol. Rev. 29, 303-329.

Rasko, D.A., Worsham, P.L., Abshire, T.G., Stanley, S.T., Bannan, J.D., Wilson, M.R., Langham, R.J., Decker, R.S., Jiang, L., Read, T.D., 2011. *Bacillus anthracis* comparative genome analysis in support of the amerithrax investigation. PNAS. 108, 5027-5032.

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., 2009. The complete genome sequence of *Bacillus anthracis* Ames “Ancestor”. J. Bacteriol. 191, 445-446.

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., 2003. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature 423, 81-86.

Richardson, E.J., Watson, M., 2013. The automatic annotation of bacterial genomes. Briefings in bioinformatics 14, 1-12.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M., Nyren, P., 1996. Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem. 242, 84-89.

Sacchi, C.T., Whitney, A.M., Mayer, L.W., Morey, R., Steigerwalt, A., Boras, A., Weyant, R.S., Popovic, T., 2002. Sequencing of 16S rRNA gene: a rapid tool for identification of *Bacillus anthracis*. Emerg. Infect. Diseases 8, 1117.

Saile, E., Koehler, T.M., 2002. Control of anthrax toxin gene expression by the transition state regulator abrB. J. Bacteriol. 184, 370-380.

Saile, E., Koehler, T.M., 2006. *Bacillus anthracis* multiplication, persistence, and genetic exchange in the rhizosphere of grass plants. Appl. Environ. Microbiol. 72, 3168-3174.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D., Dean, D., 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. 62, 775-806.

Schuch, R., Fischetti, V.A., 2009. The secret life of the anthrax agent *Bacillus anthracis*: bacteriophage-mediated ecological adaptations. PLoS One 4, e6532.

Schwartz, M., 2009. Dr. Jekyll and Mr. Hyde: a short history of anthrax. Mol. Aspects Med. 30, 347-355.

Segerman, B., De Medici, D., Ehling Schulz, M., Fach, P., Fenicia, L., Fricker, M., Wielinga, P., Van Rotterdam, B., Knutsson, R., 2011. Bioinformatic tools for using

whole genome sequencing as a rapid high resolution diagnostic typing tool when tracing bioterror organisms in the food and feed chain. Int. J. Food Microbiol. 145, S167-S176.

Shafazand, S., Doyle, R., Ruoss, S., Weinacker, A., Raffin, T.A., 1999. Inhalational AnthraxEpidemiology, Diagnosis, and Management. CHEST Journal 116, 1369-1376.

Shendure, J., Ji, H., 2008. Next-generation DNA sequencing. Nat. Biotechnol. 26, 1135-1145.

Shoop, W., Xiong, Y., Wiltsie, J., Woods, A., Guo, J., Pivnichny, J., Felcetto, T., Michael, B., Bansal, A., Cummings, R., 2005. Anthrax lethal factor inhibition. PNAS. 102, 7958-7963.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., Birol, İ., 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 19, 1117-1123.

Slamti, L., Lereclus, D., 2002. A cell–cell signaling peptide activates the PlcR virulence regulon in bacteria of the *Bacillus cereus* group. EMBO J. 21, 4550-4559.

Smith, I., 1973. A brief review of anthrax in domestic animals. Postgrad. Med. J. 49, 571-572.

Smith, K., DeVos, V., Bryden, H., Price, L., Hugh-Jones, M., Keim, P., 2000. *Bacillus anthracis* diversity in Kruger National Park. J. Clin. Microbiol. 38, 3780-3784.

Sozhamannan, S., Chute, M., McAfee, F., Fouts, D., Akmal, A., Galloway, D., Mateczun, A., Baillie, L., Read, T., 2006. The *Bacillus anthracis* chromosome contains four conserved, excision-proficient, putative prophages. BMC microbiol 6, 34.

Stenfors Arnesen, L.P., Fagerlund, A., Granum, P.E., 2008. From soil to gut: *Bacillus cereus* and its food poisoning toxins. FEMS Microbiol. Rev. 32, 579-606.

Sterne, M., 1939. The use of anthrax vaccines prepared from avirulent (uncapsulated) variants of *Bacillus anthracis*. Onderstepoort J.Vet.Sci.Anim.Ind 13, 307-312.

Sterne, M., 1946. Avirulent anthrax vaccine. Onderstepoort J. Vet. Sci. Anim. Ind. 21, 41-43.

Sterne, M., 1967. Distribution and economic importance of anthrax. Federation Proceedings. 26, 149-1495.

Stothard, P., Wishart, D.S., 2006. Automated bacterial genome analysis and annotation. Curr. Opin. Microbiol. 9, 505-510.

Stothard, P., Wishart, D.S., 2006. Automated bacterial genome analysis and annotation. Curr. Opin. Microbiol. 9, 505-510.

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.

Sussman, A.S., Halvorson, H.O., 1966. Spores: their dormancy and germination. Spores: their dormancy and germination. Ed. Schwartz, W. New York, London, 349-354.

Tang, W., Guo, Q., 2009. The adenylyl cyclase activity of anthrax edema factor. Mol. Aspects Med. 30, 423-430.

Tomaso, H., Bartling, C., Al Dahouk, S., Hagen, R.M., Scholz, H.C., Beyer, W., Neubauer, H., 2006. Growth characteristics of *Bacillus anthracis* compared to other *Bacillus* spp. on the selective nutrient media anthrax blood agar and cereus ident agar. Syst. Appl. Microbiol. 29, 24-28.

Tournier, J., Ulrich, R.G., Quesnel-Hellmann, A., Mohamadzadeh, M., Stiles, B.G., 2009. Anthrax, toxins and vaccines: a 125-year journey targeting *Bacillus anthracis*. Expert Review of Anti-infective Therapy.7, 219-236

Turnbull, P.C.B., 2002. Introduction: anthrax history, disease and ecology. Anthrax. Springer, pp. 1-19.

Turnbull, P.C.B., 1998. Guidelines for the surveillance and control of anthrax in humans and animals (fourth edition). World Health Organization Department of Communicable Diseases Surveillance and Response report WHO/EMC/ZDI/98.6, 29-30.

Valjevac, S., Hilaire, V., Lisanti, O., Ramisse, F., Hernandez, E., Cavallo, J., Pourcel, C., Vergnaud, G., 2005. Comparison of minisatellite polymorphisms in the *Bacillus cereus* complex: a simple assay for large-scale screening and identification of strains most closely related to *Bacillus anthracis*. Appl. Environ. Microbiol. 71, 6613-6623.

Van der Auwera, Geraldine A, Andrup, L., Mahillon, J., 2005. Conjugative plasmid pAW63 brings new insights into the genesis of the Bacillus anthracis virulence plasmid pXO2 and of the Bacillus thuringiensis plasmid pBT9727. BMC Genomics 6, 103.

Van der Auwera, G.A., Timmery, S., Mahillon, J., 2008. Self-transfer and mobilisation capabilities of the pXO2-like plasmid pBT9727 from *Bacillus thuringiensis* subsp.*konkukian* 97-27. Plasmid 59, 134-138.

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., 2007. Strain-specific single-nucleotide polymorphism assays for the *Bacillus anthracis* Ames strain. J. Clin. Microbiol. 45, 47-53.

Van Ness, G.B., 1971. Ecology of anthrax. Science 172, 1303-1307.

Vilas-Boas, G., Peruca, A., Arantes, O., 2007. Biology and taxonomy of *Bacillus cereus*, *Bacillus anthracis*, and *Bacillus thuringiensis*. Can. J. Microbiol. 53, 673-687.

Voelkerding, K.V., Dames, S.A., Durtschi, J.D., 2009. Next-generation sequencing: from basic research to diagnostics. Clin. Chem. 55, 641-658.

Watson, A., Keir, D., 1994. Information on which to base assessments of risk from environments contaminated with anthrax spores. Epidemiol. Infect. 113, 471-490.

Whiteley, H., Schnepf, H.E., 1986. The molecular biology of parasporal crystal body formation in *Bacillus thuringiensis*. Ann. Rev. Microbiol 40, 549-576.

Woese, C.R., 1987. Bacterial evolution. Microbiol. Rev. 51, 221.

Wooley, J.C., Godzik, A., Friedberg, I., 2010. A primer on metagenomics. PLoS Comput. Biol. 6, e1000667.

World Health Organization (WHO), 2003. Manual for laboratory diagnosis of anthrax. World Health Organization, Geneva, Switzerland. http://w3.whosea.org/bct/ anthrax.

World Health Organization (WHO), 2008. Anthrax in Humans and Animals. World Health Organisation, Geneva Fourth edition. (http://www.who.int/csr/resources/publications/AnthraxGuidelines2008/en/index.html)

Young, J.A.T., Collier, R.J., 2007. Anthrax toxin: receptor binding, internalization, pore formation, and translocation. Annu. Rev. Biochem. 76, 243-265.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821-829.

Zhang, J., Chiodini, R., Badr, A., Zhang, G., 2011. The impact of next-generation sequencing on genomics. J. Genet. Genomics 38, 95-109.

Chapter 3

POLYPHASIC CHARACTERIZATION OF *BACILLUS* SPECIES FROM ANTHRAX OUTBREAKS IN ANIMALS FROM SOUTH AFRICA AND LESOTHO

K.E. Lekota[1,2,3], A. Hassim[2], J. Mafofo[1,],J. Rees[1], F.C. Muchadeyi[1], H. van Heerden[2], E. Madoroba[3].

[1] The Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort 0110, South Africa
[2]University of Pretoria, Department of Veterinary Tropical Diseases, Private bag X4, Onderstepoort, 0110, South Africa
[3]Agricultural Research Council-Onderstepoort Veterinary Institute, Private bag X5, Onderstepoort 0110, South Africa

Abstract

Bacillus anthracis is the causative agent of anthrax, a disease endemic in the Northern Cape Province (NCP) and Kruger National Park (KNP) of South Africa. *Bacillus anthracis* and *Bacillus* isolates have been recovered and identified from animals that died of anthrax symptoms in South Africa and Lesotho. The aim of this study was to characterize *Bacillus* species that were isolated alongside *B. anthracis* during anthrax outbreaks in the Northern Cape, Limpopo, Mpumalanga Provinces and Lesotho, using a polyphasic approach that includes phenotypic and molecular analyses. For this purpose, *B. anthracis* and 10 *Bacillus* isolates were subjected to microbiology tests, Biolog OmniLog identification system (Biolog), 16S ribosomal RNA (rRNA) sequence analysis, PCR detection of protective antigen (*pag)* and capsule (*cap*) regions, real-time PCR using hybridization probes targeting *pag* and *cap*C genes, and multi-loci variable number of tandem repeat analysis (MLVA). The *Bacillus* isolates were non-haemolytic, non-motile, and susceptible to penicillin, which is typical of *B. anthracis*, but resistant to gamma phage, unlike typical *B. anthracis*. The Biolog system assigned the *Bacillus* isolates to primarily *B. endophyticus* (7 out of 10), but also identified *B. pumilus* and *B. amyloliquifaciens*. The16S rRNA gene sequence analysis assigned *Bacillus* isolates to primarily *B. endophyticus* (8 out of 10) but also *Brevibacterium frigoritolerans* and *B. thuringiensis*. Conventional PCR revealed that most of the *B. anthracis*-like isolates contained *pag and cap* gene regions, however PCR produced varying results amongst the *cap*C, *cap*B/C and *cap*A regions. Real-time PCR, MLVA, 16S rRNA sequencing and confirmatory microbiology tests including phage resistance distinguished *B. anthracis*-like isolates from *B. anthracis. Bacillus anthracis* can be differentiated from other similar *Bacillus* species using classical microbiological techniques and molecular technique. However only microbiological and 16S rRNA sequencing enable identification. The existence of *B. anthracis*-like isolates causing

symptoms that are reminiscent of anthrax in animals needs further investigation, especially the *B. anthracis* virulence genes that are present in these isolates.

Key words: *Bacillus* species, anthrax, polyphasic approach, bacteriology, molecular

3.1. Introduction

Bacillus anthracis is the causative agent of anthrax that primarily affects herbivorous animals (WHO, 2008). This pathogenic bacteria form the *B. cereus* group (*B. cereus* sensu lato) together with *B. cereus* and *B. thuringiensis* (Read *et al*., 2003). Homologous recombination and horizontal transfer of genetic material within the *B. cereus* sensu lato group, including phage transmission has been reported (Modrie *et al.,* 2010; Zwick *et al.,* 2012). The genetic backbone of *B. anthracis*, which is made up of plasmids pXO1 and pXO2 (containing the virulent genes) has been found in other members of the *B. cereus* group and is not only restricted to *B. anthracis* as previously thought (Klee *et al*., 2006; 2010).

The principal virulence factors of *B. anthracis* are encoded on two double stranded circular plasmids. The pXO1 plasmid (182 kb) encodes a tripartite protein exotoxin complex and plasmid pXO2 (95 kb) encodes the polypeptide capsule genes. The toxin genes on pXO1 consist of the protective antigen (*pag*), lethal factor (*lef*), and edema factor (*cya*) (Mock and Fouet, 2001; Mock and Mignot, 2003; Young and Collier, 2007). The capsule encoded by pXO2 consists of a five-gene operon (*capBCADE*) that synthesizes the poly-γ-D-glutamic acid capsule of *Bacillus* species, which enable host immune system evasion by protecting the vegetative cells from being phagocytosed by macrophages (Drysdale *et al.,* 2004; Kolsto *et al.,* 2009).

It is paramount to provide rapid and accurate diagnosis of *B. anthracis* to curb the spread of this zoonotic pathogen. For this purpose, *B. anthracis* can be distinguished from closely related *B. cereus* members based on criteria that are recommended by the World Health Organization (WHO) and Centres for Disease Control (CDC). Based on these criteria, *B. anthracis* are non-motile, non-haemolytic, and they are sensitive to penicillin and gamma-phage. Nevertheless, it is imperative to make use of DNA-based methods for consistent accurate diagnosis of anthrax due to the challenge associated with inconsistent attributes of some isolates that resemble *B. anthracis*. The use of 16S ribosomal RNA sequence analysis for identification of *B. anthracis* revealed that the genes are homologous to the *B.*

cereus group, and the group could be considered as a single taxon (Helgason *et al*., 2000; Rasko *et al*., 2005). These pathogenic species share highly conserved chromosomes but differ in pathogenicity, which is mostly plasmid-encoded. *Bacillus cereus* is a food borne pathogen due to the production of an emetic toxin, enterotoxins and degradative enzymes (Kotiranta *et al*., 2000; Stenfors Arnesen *et al*., 2008), and *B. thuringiensis* is widely used in agriculture as an insect pathogen with plasmid-encoding insecticidal crystal proteins (Schnepf *et al*., 1998), whilst *B. anthracis* is a pathogen due to the presence of genes for toxin on plasmids.

One of the molecular based techniques for *B. anthracis* characterization involves the use of polymerase chain reaction (PCR) for amplification of genes that are found on the *B. anthracis* plasmids and other specific chromosomal markers. However, this method may present challenges for the discrimination of *B. anthracis* from closely related bacteria with similar capsule genes and *B. anthracis* virulence gene(s) (Beyer and Pocivalsek, 1999; Hoffmaster *et al.,* 2004; Klee *et al.,* 2010). Multiple-loci variable-number tandem repeat (VNTR) analysis (MLVA) has been proven to reveal the genetic relationships within *B. anthracis* strains (Keim *et al.,* 2000; Le Fleche *et al.,* 2001; Beyer *et al.,* 2012), which is important for differentiating the alleles that are present amongst *B. anthracis*. Although not traditionally used, MLVA has been reported to differentiate closely related organisms in the *B. cereus* group (Valjevac *et al.,* 2005).

The Agricultural Research Council-Onderstepoort Veterinary Institute (ARC-OVI) in South Africa performs classical microbiological tests and PCR targeting pXO1 and pXO2 genes for diagnosis of anthrax cases or outbreaks. Numerous samples from cases of animals showing symptoms that resembled anthrax were received at ARC-OVI for diagnosis in 2001, 2009 and 2011. However, classical microbiological tests showed the presence of both *B. anthracis* and *Bacillus* species (referred as *B. anthracis*-like) isolates causing anthrax in animals from the same geographic region. For instance, diagnostic tests found both *B. anthracis* and *Bacillus* strains from the same farm in the NCP, where the *Bacillus* isolates resembled *B. anthracis* with respect to lack of haemolysis, morphology, sensitivity to penicillin and non-motility except they were resistant to gamma phage. Therefore, the aim of this study was to characterize *B. anthracis* and *Bacillus* isolates using a polyphasic approach that incorporates phenotyping and DNA-based techniques. Such information is important for insight into virulence genes that might be associated with *Bacillus* species bacteria.

3.2. Materials and Methods

3.2.1. *Bacillus* species isolates

Soil and tissue samples were obtained during an anthrax outbreak in 2008-2009 at Klipfontein and Kimberly in the NCP, as well as from cases in Limpopo and Mpumalanga Provinces and Maseru in Lesotho (Figure 1.1, Table 3.1). The animals showed clinical symptoms such as sudden death and bleeding from orifices that resembled anthrax. The samples were processed at the ARC-OVI and identified. All isolates from the suspected anthrax cases were freeze-dried and stored in the Bacteriology culture collection at ARC-OVI, South Africa (Table 3.1). *Bacillus anthracis* Sterne, Ames and Vollum strains, *B. thuringiensis* and *B. cereus* were included as controls in this study (Table 3.1).

3.2.2. Phenotypic identification of *Bacillus* isolates

3.2.2.1. Microbiology tests

Isolates obtained from the ARC-OVI culture collection were grown on 5% sheep blood agar (SBA), followed by incubation at 37 °C for 24 hours to determine of hemolytic activity (OIE, 2008). Capsule visualization of the *Bacillus* isolates was conducted by transferring purified colonies on trypticase soy agar (TSA) containing 0.8% sodium bicarbonate, and incubated at 5% CO_2 at 37 °C for 24 hours. The capsules were stained using India ink and visualized by light microscopy. The confirmatory gamma-phage test was conducted as described by OIE (2008) on SBA plates. The concentration of the liquid culture was compared with MacFarlane (10^{-1}) standard. Penicillin sensitivity/resistance was determined after 6-8 hours incubation at 37 °C on SBA (OIE, 2008). *Bacillus anthracis* Sterne strain was used as a positive control and *B. cereus* and *B. thuringiensis* strains were included as negative controls.

Table 3.1. *Bacillus* isolates from South Africa and Lesotho that where isolated from animals with anthrax clinical symptoms as well as control strains from *B. cereus sensu lato*.

Cases	Strain number	Animal species	Isolate source	Isolation date	Location[#]	*Bacillus* species
1	3617-3C	Kudu 1	Soil	May-09	NCP (Klipfontein)	*Bacillus* sp.
1	3617-2C	Kudu 1	Ear	May-09	NCP (Klipfontein)	*Bacillus* sp.
2	3618-1C [*]	Kudu 2	Ear	May-09	Kimberly, NCP	*Bacillus* sp.
3	3566-3D	Kudu 3	Rib bone	Jan-09	NCP (Klipfontein)	*Bacillus* sp.
3	3566-1B	Kudu 3	Ear	Jan-09	NCP (Klipfontein)	*Bacillus* sp.
4	3631-6C	Kudu 4	Bone	May-09	NCP (Kimberly)	*Bacillus* sp.
4	3631-9D	Sheep 5	Ear	May-09	NCP (Kimberly)	*Bacillus* sp.
4	3631-10C	Sheep 6	Ear	May-09	NCP (Kimberly)	*Bacillus* sp.
5	8334	Giraffe 1	Soil	1995	Maseru (Lesotho)	*Bacillus* sp.
6	7424	Buffalo 1	Lung	Nov-11	Limpopo (Eulalie)	*Bacillus* sp.
2	3618-2D[*]	Kudu 2	Soil	May-09	NCP (Klipfontein)	*B. anthracis*
4	3631-1C	Kudu 5	Ear	May-09	NCP (Klipfontein)	*B. anthracis*
7	20SD	Sheep 1	Ear	2001	MP (Standerton)	*B. anthracis*
	34F2					*B. anthracis* Sterne
	BT		Soil	Jun-75	ARC-OVI	*B. thuringiensis*
	BC	Dog	Ear	2005	OVAH	*B. cereus*
	Ames	Cattle		2001	Sarita, USA	*B. anthracis* Ames
	Vollum	Cow		1935	Oxfordshire, USA	*B. anthracis* Vollum

* Isolate 3618-2D from soil under dead kudu was identified as *Bacillus anthracis*, whereas 3618-1C from ear from the same dead kudu that died of anthrax symptoms on Klipfontein farm in the NCP was identified as *Bacillus* species.

[#] NCP: Northern Cape Province; MP: Mpumalanga Province; ARC-OVI: Agricultural Research council- Onderstepoort Veterinary Institute and; OVAH: Onderstepoort Veterinary Academic Hospital.

3.2.2.2. Biochemical characteristics

Classical biochemical identification was conducted using standard protocols as described by CDC/ASM/APHL (2002). The indole, citrate, oxidase, catalase activity were tested for all isolates. The indole test was performed to determine the ability of the isolates to produce indole from degradation of tryptophan by trytophanase. The indole reacts with the aldehyde in the Kovac's reagent to give the red colour that indicates a positive indole test. *Bacillus* isolates were grown in peptone broth incubated overnight at 37 °C. A few drops of Kovac's reagent were added to the overnight-incubated peptone broth culture, with a colour change observed within 5 seconds on the upper layer of the liquid.

The citrate test was performed to establish whether citrate is utilized as a sole carbon and energy source. The end product (sodium bicarbonate, Na_2CO_3) as well as ammonia from the utilization of sodium citrate and ammonia salt result in alkaline that changes the medium's colour from green to blue. The citrate test was done from overnight grown culture in peptone broth. The culture were inoculated and incubated at 37 °C for 24 hours on a Simmons citrate slant that contained sodium citrate and bromothymol blue as pH indicator as well as inorganic ammonium salts. With the oxidase test the culture was smearedonto filter paper, followed by 1-2 drops of the oxidase reagent, with recording of colour change after 10 second.

The oxidase test was based on the ability of the isolates to produce intracellular oxidase enzyme. The substrate N,N,N,N- tetramethyl-p-henylenediaminedi- hydrochloride acts as an artificial electron acceptor for oxidase which oxidize to give a colour change of purple for positive isolates. The catalase test was performed on a microscope slide, where 1-2 drops of 3% hydrogen peroxide was added to bacterial cells on the slide. The formation of gas bubble was observed when the hydrogen was converted into water and oxygen indicating the present of catalase. The urease test was conducted by streaking culture on entire surface of a Christensen slant surface. After 24 hours incubation at 37 °C colour changes are noted. Urease production is indicated by a bright pink (fuchsia) colour on the slant. Any degree of pink (weakly positive) was considered as a positive result for urease. An un-inoculated slant was included as negative control in addition to other positive and negative controls.

3.2.2.3. API 50CHB assay (Carbohydrates fermentation)

The *Bacillus* isolates were subjected to a battery of carbohydrate fermentation tests based on the API 50 CHB system. The tests were undertaken according to the manufacturer's instructions (BioMerieux, France) and results were interpreted using the Analytical Profile Index (API) database (Apiweb software version 1.2.1; Biomerieux, France).

3.2.2.4. Biolog OmniLog Identification

Suspensions of *Bacillus* isolates were prepared according to the manufacturer's instructions (OmniLog ID System User Guide, Biolog, USA). Briefly, the bacterial suspensions were inoculated into wells of microplates containing 95 different carbon sources. The plates were incubated in the OmniLog incubator at 30 °C for 4 to 24 hours depending on the growth requirements of the organisms. The microplates were read using the Biolog's microbial identification system software (OminiLog® Data Collection), which contains biochemical fingerprints of different species (Biolog, USA).

3.2.3. Molecular characterization

3.2.3.1. Bacterial DNA extraction

All isolates (Table 3.1) were inoculated in 2 ml nutrient broth, followed by overnight incubation at 37 °C. The bacterial cells were then harvested by centrifugation at 5000 ×g for 10 minutes. Genomic DNA was extracted from the harvested cells using the DNAeasy Tissue kit (Qiagen, Germany) according to the manufacturer's instructions. The DNA was suspended in 100 µl of Buffer AE instead of 200 µl. The DNA was quantified with the Qubit® fluorometric quantization method (Life Technologies, USA) using the broad range assay kit according to the manufacturer's instruction. The DNA elute was further diluted to 5:100 (~5 ng/µl) with AE Buffer and stored at -20 °C. The quality of DNA extracted was visualized on 0.8% agarose gel with 0.4 µl of ethidium bromide under a UV light.

3.2.3.2. Conventional PCR

Primers R1 and R2 (Patra *et al.,* 1996, 2005; OIE, 2008); PA-5 and PA-8 (Beyer *et al.,* 1995; OIE, 2008) and Cap 1234 and 1301 (OIE, 2008) were used to amplify the chromosomal gene Ba813 as well as the virulence genes on pXO1 and pXO2 (Table 3.2). The 25 µl PCR reaction contained 2x Dream *Taq*^{TM} Green PCR Master mix (Fermentas, Lithuania), 0.2 µM of each primer and 2 ng/µl target DNA. The PCR conditions consisted of an initial denaturation at 95 °C for 5 minutes, followed by 30 cycles of denaturation at 95 °C for 5 seconds, annealing for 30 seconds and extension at 72 °C for 30 seconds with a final extension for 5 minutes at 72 °C. The PCR annealing temperature was 55 °C for the chromosomal region (*Ba*813), *pag*A, and *cap*B/C regions, 58 °C for *cap*C and 65 °C for *cap*A.

Agarose Melt-MAMA assay of the PlcR marker was amplified as described by Birdsell *et al*. (2012). The synthetic positive control (PC) templates of derived and ancestral were generated by conventional PCR. The 10 µl PCR reaction contained 1x MyTaq^{TM} PCR Master mix (Bioline, USA), 2 mM $MgCl_2$, 0.2 µM of each primer and 1 ng/µl target DNA. The PCR conditions consisted of denaturation at 94 °C for 30 seconds, annealing at 58 °C for 30 seconds, extension at 72 °C for 30 seconds for 35 cycles, and a final extension at 72 °C for 5 minutes. The PCR amplicons of the PC-derived and PC-ancestral were used as control templates for their specific assay at 1/100 000 dilution. The primer ratio of 1:1:1 of derived, ancestral and common reverse was used respectively for Agarose Melt-MAMA (Table 3.2). The same PCR reaction and conditions except for the annealing temperature of 55 °C was used as described above. The PCR amplicons were electrophoresed at 100 V for 90 minutes on a 3% agarose gel.

Table 3.2. Summary of the primer and probe sequences, gene targets, and expected PCR products sizes.*

Primer or probe[#]	Sequence (5'-3')	Target gene	PCR fragment size (bp)
PCR			
1234	CTGAGCCATTAATCGATATG	*Cap*B/C	846
1301	TCCCACTTACGTAATCTGAG		
R1	ACTCGTTTTTAATCAGCCCG	*Ba*813	
R2	GGTAACCCTTGTCTTTGAAT	chromosome	152
PA8	AGGTAGAAGGATATACGGT	*pag*A	
PA5	TCCTAACACTAACGAAGTCG		596
57	ACTCGTTTTTAATCAGCCCG	*Cap*C	244
58	GGTAACCCTTGTCTTTGAAT		
MO1	GCTGATCTTGACTATGTGGGTG	*Cap*A	287
MO2	GGCTTCCTGTCTAGGACTCGG		
Melt-MAMA PCR			
PlcR derived	CGGGGCGGGGCGGGGCGGGCTTATTTGCATGACAAAGCGCATA	*PlcR*	70-90bp**
PlcR ancestral	TTTGCATGACAAAGCGCCTC		
Reverse	AAAGCATTATACTTGGACAATCAATACG		
Real-time PCR			
BAPA-S	CGGATCAAGTATATGGGAATATAGCAA	*Pag*	204
BAPA-R	CCGGTTTAGTCGTTTCTAATGGAT		
BAPA-FL[#]	TGCGGTAACACTTCACTCCAGTTCGA-X		
BAPA-LC[#]	CCTGTATCCACCCTCACTCTTCCATTTTC-P		
CapS	ACGTATGGTGTTTCAAGATTCATG	*Cap*C	291
CapA	GATTGCAAATGTTGCACCACTTA		
CapC-FL[#]	TATTGTTATCCTGTTATGCCATTTGAGATTTTT-X		
CapC-LC[#]	AATTCCGTGGTATTGGAGTTATTGTTCC-P		
ANT-F+	GCTAGTTATGGTACAGAGTTTGCGAC	*Sasp*	102
ANT-Amt+	CCATAACTGACATTTGTGCTTTGAAT		
ANT-FL[#]	CAAGCAAACGCACAATCAGAAGCTAAG-X		
ANT-LC[#]	GCGCAAGCTTCTGGTGCTAGC-P		

*The information in Table 2 is based on Makino *et al*. (1993); Beyer *et al*. (1995); Patra *et al*. (1996); Ramisse *et al*. (1996); and OIE (2008) where *cap* refer to capsule genes, *pag* refer to protective antigen, SASP refer to small acid soluble protein and *Ba*813 refer to region in *B. anthracis* chromosome. Melt-MAMA refer to melt analysis of mismatch amplification mutation assays

[#] Indicate probe sequences.

** Estimated between 70-90 bp, size depends on derived or ancestral *PlcR* region

3.2.3.3. Real-time PCR

Real-time PCR was conducted for the detection of protective antigen (*pag*) and capsule (*capC)* on both virulence plasmids and a specific *B. anthracis* chromosomal target (small acid soluble protein, SASP) using the LightCycler Nano instrument (Roche Applied Science, Germany). The 20 µl PCR mixture consisted of 4 mM $MgCl_2$, 5 ng of DNA, 1/10 volume of FastStart master mix (Roche, Germany), 0.5 µM of each primer and 0.2 µM of each probe (Table 3.2). The LightCycler experimental protocol was as indicated by OIE experimental protocol (OIE, 2008), but the cutoff was after 35 cycles. The LightCycler Nano software was used to analyze the amplification of the genes. The PCR products were run on 2% agarose gels in order to confirm the size amplification of the genes and hybridization of the probes.

3.2.3.4. 16S rRNA PCR

The 16S rRNA genes were amplified using universal primers 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-AGAGTTTGATCMTGGCTCAG-3') (Lane, 1991). The 67F (5'TGAAAACTGAACGAAACAAAC 3') and 1671R (5'CTCTCA AAACTGAACAAAACGAAA 3') (Sacchi *et al*., 2002) primers were used to amplify the 16S rRNA gene of *B. anthracis*, *B. thuringiensis*, *and B. cereus*. The 25 µl reaction contained 1x Dream *Taq*^{TM} Green PCR Master mix (Fermentas, Lithuania), 0.2 mM of each primer and 50 ng of template DNA. The PCR conditions for both the universal and 16S rRNA primers consisted of an initial denaturation at 95 °C for 5 minutes followed by 30 cycles of denaturation at 94 °C for 30 seconds, annealing at 52 °C for 30 seconds, and extension at 72 °C for 1 minutes with a final elongation at 72 °C for 7 minutes. The PCR products were electrophoresed in 1.5% ethidium bromide-stained agarose gels, followed by visualization under UV light.

3.2.3.5. 16S rRNA and virulence genes sequencing

The 16S rRNA amplicons of isolates 3617-3C, 3617-2C, 3618-1C, 3566-3D, 3566-1B, 3631-6C, 3631-10C, 8334, 7424, 3618-2D, 3631-1C and 20SD, and as well as *cap*B/C and *pag*A amplicons for isolates 3566-1B, 3618-2D, *B. cereus*, and *B. anthracis* Ames were purified using a High Pure PCR product Purification Kit (Roche, Germany) and sequenced in both directions at Inqaba Biotechnologies (Pretoria, South Africa). All the

sequences were processed using the CLC Genomic Workbench 6.0. In order to obtain preliminary identifications of the *Bacillus* spp. isolates, the basic local alignment search tool for nucleotide sequences (BLASTN) was used to compare the sequences to those in the database of nucleotides in the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov). The preliminary identifications were refined through phylogenetic analyses. For this purpose, the 16S rRNA sequence alignments were generated with the MEGA 6.0 software using ClustalW multiple sequence alignment. The alignments also incorporated nucleotide sequences of other relevant *Bacillus* species that were obtained from Genbank. The phylogenetic relationships based on maximum likelihood were inferred with the MEGA 6.0 software. Non-parametric bootstrap analysis was used to estimate branch support and it was based on 1000 replicates.

3.2.3.6. Multiloci variable number of tandem repeat analysis (MLVA)

The 31 VNTR markers were initially used to amplify the *B. anthracis* and *Bacillus* isolates as described by by Keim *et al*. (2000), Le Fleche *et al*. (2001) and Beyer *et al*. (2012). Twenty six VNTRs were used for cluster analysis since the copy number of the small VNTRs (6 bp and less) could not be determined using agarose electrophoresis. The 26 VNTR markers for *B. anthracis* constisted of vrrA, vrrB1, vrrB2, vrrC1, vrrC2, CG3 (Keim *et al*., 2000); Bams: 1, 3, 5, 15, 21, 22, 23, 24, 25, 28, 30, 31, 34, 44, 51, 53 (Le Fleche *et al*., 2001, Lista *et al*., 2006); VNTR: 16, 17, 23, (Van Ert *et al*., 2007). The 26 VNTR markers were amplified using 4 ng DNA in a final volume of 15 µl PCR reaction containing 2 mM $MgC1_2$, 20 pM primers, 1x MyTaq^{TM} mix (Bioline, USA). PCR condition 1 (for VNTR loci vrrA, vrrB1, vrrB2, vrrC1, vrrC2, CG3, Bams: 1, 3, 5, 15, 21, 22, 23, 24, 25, 28, 30, 31, 34, 44, 51, 53, and VNTR 17, 23) consisted of 95 °C initial enzyme activation for 5 minutes followed by 35 cycles of 95 °C denaturation for 20 seconds, 60 °C annealing for 30 seconds and 72 °C elongation for 5 minutes for primers amplifying products smaller than 600 bp and 7 minutes for primers amplifying products larger than 600 bp. PCR condition 2 consisted of 95 °C initial enzyme activation for 5 minutes, followed by 35 cycles of 95 °C denaturation for 20 seconds, 56 °C annealing for 30 seconds and 72 °C elongation for 5 minutes for VNTR 16. The PCR products were separated on 3% agarose gel and visualized under UV light. A database was created using BioNumerics version 6.0 software (Applied Maths, Belgium) for analysis of copy numbers of the VNTR units calculated from PCR products sizes obtained from agarose gel images. Allele sizes for the MLVA-26 panel were converted into copy numbers as reported by Le Fleche *et al.* (2001) and Lista *et al.*

(2006). The cluster analysis was done using UPGMA (unweighted pair group method using arithmetic averages) with BioNumerics 6.0. software. Additional MLVA data from confirmed *B. anthracis* isolates from the Northern Cape Province (Hassim, personal communication) were included with MLVA results (Supplementary Figure 3S1 and Figure 3S2).

3.3. Results

3.3.1. Phenotypic Identification of the *Bacillus* species

All *Bacillus* isolates formed white-grey non-haemolytic colonies on blood agar (Figure 3.1) with morphological appearances (Supplementary Table 3S1). These colonies appeared circular, rough and dry, with ground glass appearance. The colonies were similar to *B. anthracis* that are characterized by a "Medusa head", which appears as curl-like projections. The colonies were typically Gram-positive rods (Figure 3.2, Supplementary Table 3S1) showing slightly variable thickness, non-motile, sensitive to penicillin like classic *B. anthracis*, but in contrast to the classic anthrax bacteria, the *Bacillus* bacteria were not lysed by gamma phage. All isolates including *B. cereus* and *B. thuringiensis* controls were catalase positive, oxidase negative and indole negative (Table 3.3). Most *Bacillus* isolates (8334, 3631-6C, 3631-10C, 3566-1B, 3566-3D, 3617-2C, 3618-1C, and 7424) utilized citrate as a sole source of carbon except for isolates 3631-9D, 3617-3C, 3618-2D and positive control *B. anthracis* (Sterne and 20SD), *B. cereus and B. thuringiensis*. The isolates hydrolyzed urea into ammonia and carbon dioxide except for 8334, 20SD, 3618-2D, 3631-1C and *B. anthracis* Sterne isolates.

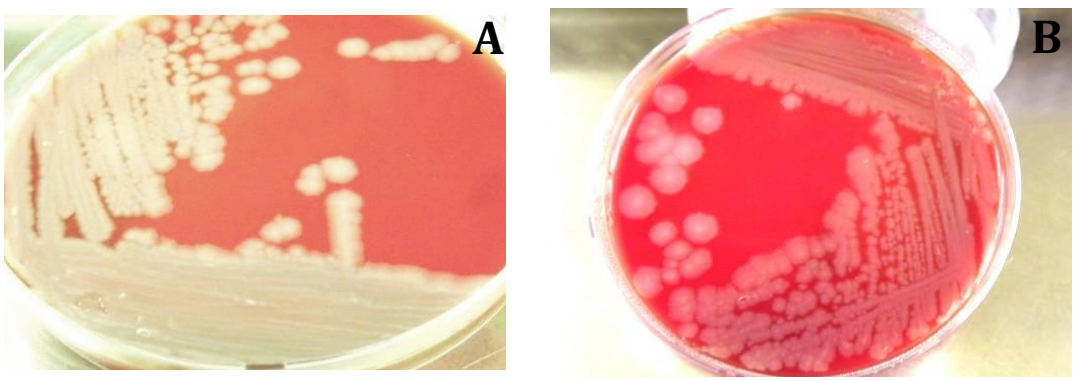


Figure 3.1. Colony morphology of *Bacillus anthracis* Sterne (A) and *Bacillus* species (3618-1C) (B) on sheep blood agar with central located head around the colony.

The microscopic appearances of *Bacillus* isolates from this study are summarized in Figure 3.2. Isolate 3617-3C (Figure 3.2 A) had square ended rods in long chains, which is similar to the typical *B. anthracis* Sterne (Figure 3.2 B). Isolates 3617-2C (Figure 3.2 C) and 3631-9D had small broad rods in short chains that clustered together. Only *Bacillus* species isolates 8334 and 3617-3C (Figure 3.2 D) could produced a capsule when induced on trypticase soy agar (TSA) stained with india ink compared to the $pXO1^{+}$, $pXO2^{-}$ *B. anthracis* 3631-1C and Sterne strains.

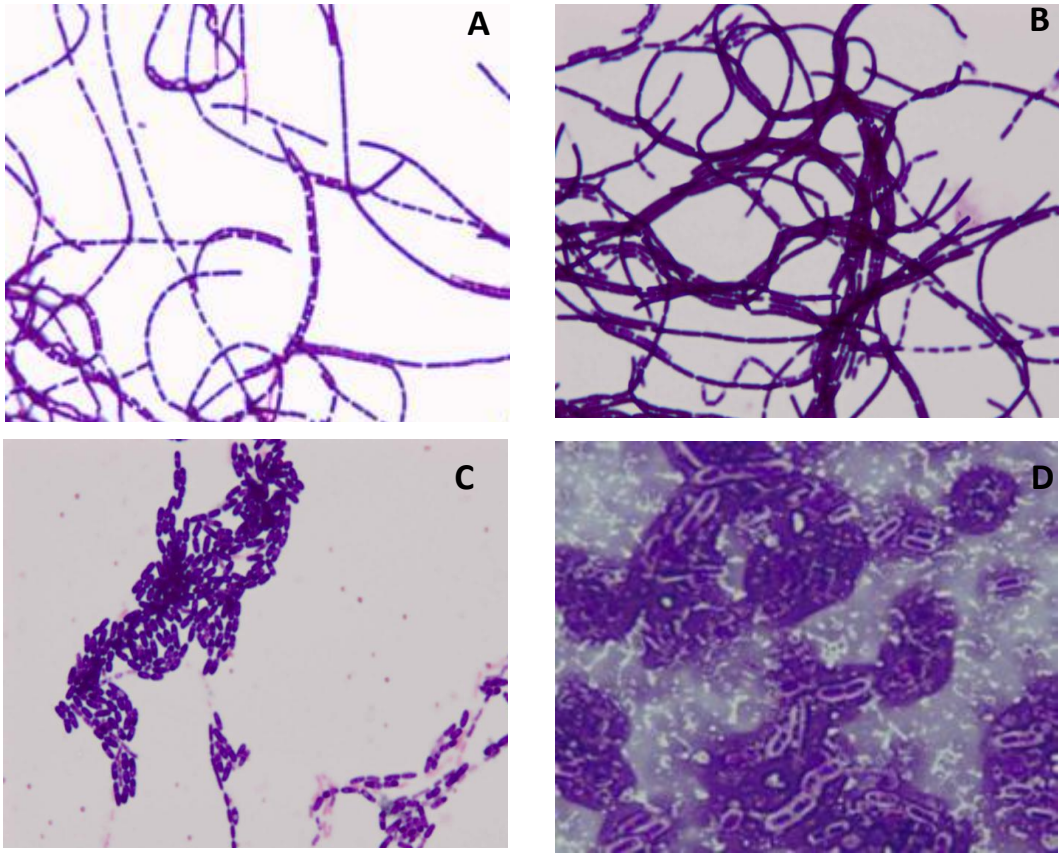


Figure 3.2. Micrographs of Gram stains and capsule stain showing morphological characteristics of *Bacillus anhtracis* and *Bacillus* isolates. Morphology of *Bacillus* isolate 3617-3C (B) compared to the *Bacillus anthracis* Sterne control strain (A) and *Bacillus* isolate 3617-2C (C). *Bacillus* isolate 3617-3C capsule (D) in Indian ink stain.

Table 3.3. Microbiological and biochemical tests of *Bacillus* isolates after incubation at 37 °C for 48 hours**.**

Isolate	Motility (M)[!]	Haemolysis (H)[$]	Penicillin *	Gamma (γ)-phage*	Biochemical tests[#]				
					Citrate	Urease	Indole	Catalase	Oxidase
***Bacillus anthracis* Sterne**	N-M[±]	N-H	S	S	-	-	-	+	-
B. cereus	M	H	R	R	-	W+	-	+	-
B. thuringiensis	M	H	R	R	-	+	-	+	-
3618-2D	N-M[±]	N-H	S	R	-	-	-	+	-
3631-1C	N-M[±]	N-H	S	S	-	-	-	+	-
20SD	N-M[±]	N-H	S	S	-	-	-	+	-
3631-9D	N-M	N-H	S	R	-	W +	-	+	-
8334	N-M[##]	N-H	S	R	+	-	-	+	-
3631-6C	N-M	N-H	S	R	+	W +	-	+	-
3631-10C	N-M	N-H	S	R	+	W +	-	+	-
3566-1B	N-M	N-H	S	R	+	W +	-	+	-
3566-3D	N-M	N-H	S	R	+	+	-	+	-
3617-3C	N-M	N-H	S	R	-	W +	-	+	-
3617-2C	N-M	N-H	S	R	+	W +	-	+	-
3618-1C	N-M	N-H	S	R	+	W +	-	+	-
7424	N-M	N-H	S	R	+	+	-	+	-

! M indicates motile and N-M indicates non-motile, [##] growth along the step line, [±]Inverted fir- tree.

$ N-H indicates Non-haemolysis.

* S = sensitive and R = resistance

W = Weakly, + positive, - negative

3.3.2. The API 50CHB and Biolog OmniLog systems

Bacillus species can be identified using the API 50CHB standardized system that is based on 49 carbohydrate metabolism tests. The API 50CHB was rendered irrelevant since it assigned majority of the *Bacillus* isolates with low probability of ID (identification). The results of API 50CH and Biolog OmniLog identification system are summarized in Table 3.4.

The Biolog OmniLog identification system is a powerful carbon source utilization technology that identifies microorganism by producing a characteristic pattern or metabolic finger printing. It test a microorganism's ability to utilize or oxide a panel of 95 tests (Supplementary Table 3S2). The Biolog OmniLog system assigned most of the *Bacillus* isolates to *B. endophyticus*, except isolates 3631-9D and 7424, which were identified as *B. pumilus* and *B. amyloliquefaciens* respectively (Table 3.4). The isolate 8334 from Lesotho could not be identified using the Biolog OmniLog system.

The growth curves of the *Bacillus* isolates represented in Figure 3.3 mostly followed the same carbon utilization patterns to be identified as *B. endophyticus* except for the organism identified as *B. amyloliquefaciens* (7424). This isolate utilize nalidixic acid, lithium chloride and potassium tellurite (Figure 3.3, see G10, G11 and G12). The isolate 8334 from Lesotho had a unique growth curve profile since it could utilize D-Serine (Figure 3.3 see C12) as a substrate and the growth kinetics on guanidine HCl (Figure 3.3. see E11) was poor compared to all the other *Bacillus* isolates in the study. In most instances, 8334 had high peak growth curves in comparison to the other *Bacillus* species, but could not be identified using this system.

Table 3.4. Identification of the *Bacillus* isolates by means of API 50 CH biochemical test, Biolog OmniLog system and 16S ribosomal RNA sequencing.

	API 50 CHB		Biolog OmniLog	16S ribosomal RNA
Isolate ID	**Taxon assigned to**	**Similarity* %**		
3631-9D	*Bacillus subtilis/ amyloliquefaciens*	98.3	*Bacillus pumilus*	*Bacillus endophyticus*
8334	*Bacillus circulans*	69.0	No ID	*Bacillus thurigiensis*
3631-10C	*Bacillus subtilis/amyloliquefaciens*	84.9	*Bacillus endophyticus*	*Bacillus endophyticus*
3566-3D	*Geobacillus thermoglucosidasius*	98.5	*Bacillus endophyticus*	*Brevibacterium frigoritolerans*
3566-1B	*Brevibacillus* non-reactive	96.7	*Bacillus endophyticus*	*Bacillus endophyticus*
3617-3C	*Bacillus subtilis/ amyloliquefaciens*	92.8	*Bacillus endophyticus*	*Bacillus endophyticus*
3617-2C	*Bacillus pumilus*	95.8	*Bacillus endophyticus*	*Bacillus endophyticus*
3618-1C	*Bacillus megaterium*	58.3	*Bacillus endophyticus*	*Bacillus endophyticus*
7424	*Bacillus megaterium*	58.4	*Bacillus amyloliquefaciens*	*Bacillus endophyticus*
3631-6C	NO ID		*Bacillus endophyticus*	*Bacillus endophyticus*

*Identification at the species level is assigned as an acceptable profile with a percentage greater than 75%.

ID = Identification

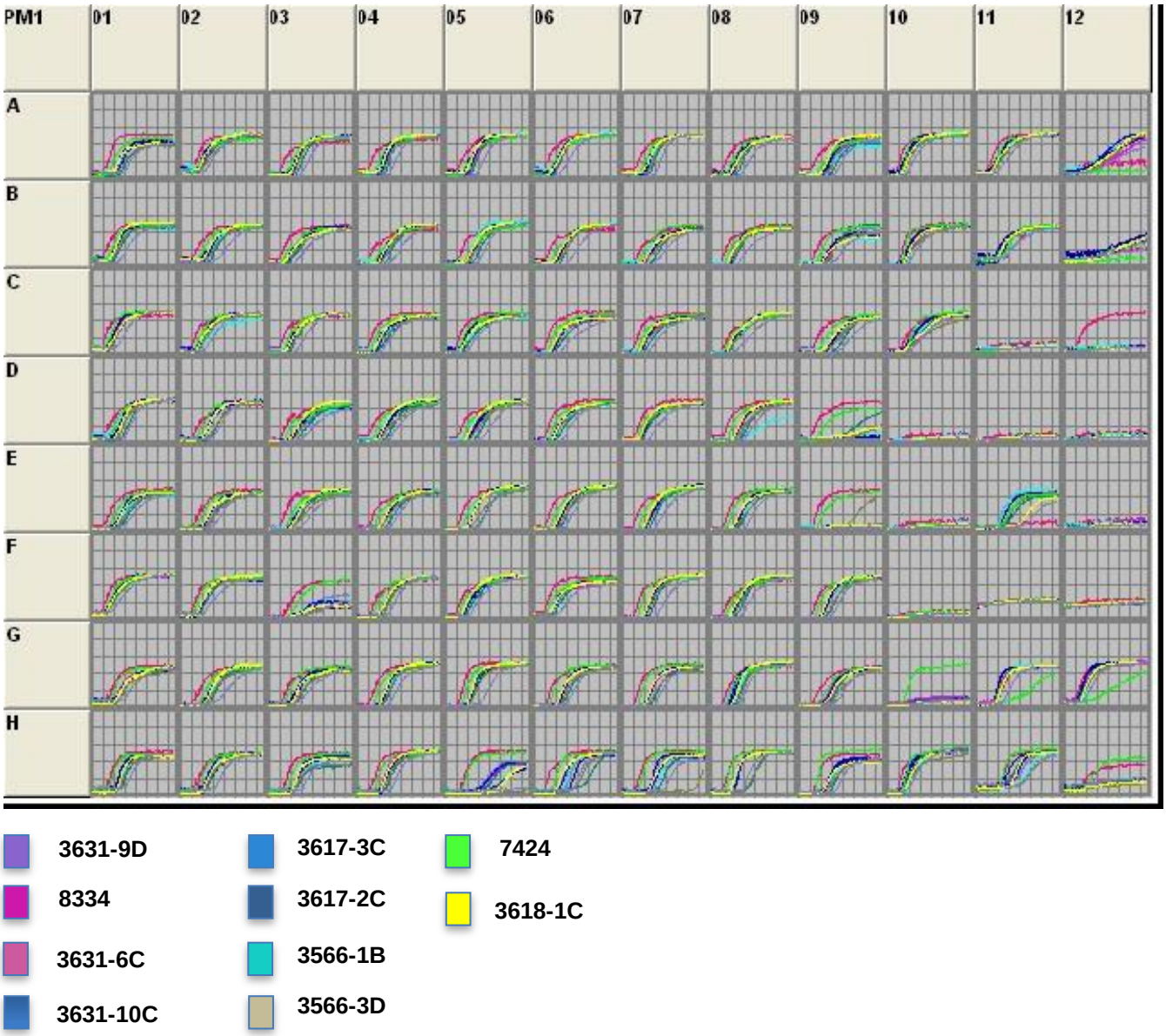


Figure 3.3. Growth curves of the *Bacillus* isolates on the Biolog OmniLog identification system. The cellular responses for kinetic assays for the different *Bacillus* species (colour coded) are indicated.

3.3.3. 16S rRNA gene amplification

The 27F and 1671R primers used for 16S rRNA gene amplification of *B. anthracis*, *B. thuringiensis*, *and B. cereus* amplified 1686 bp products for *B. anthracis* (positive controls, 3618-2D, 3631-1C), *B. thuringiensis* and 8334 (Figure 3.4). The *Bacillus* species isolates had non-specific binding amplicons (Figure 3.4). The 67F and 1671R universal primers amplified all the *B. anthracis* and *Bacillus* species in the study with an amplification of 1.5 kb (Figure 3.5). The 1.5 kb PCR products of the *B. anthracis* and *Bacillus* isolates were sequenced to construct phylogenetic tree (Figure 3.6).

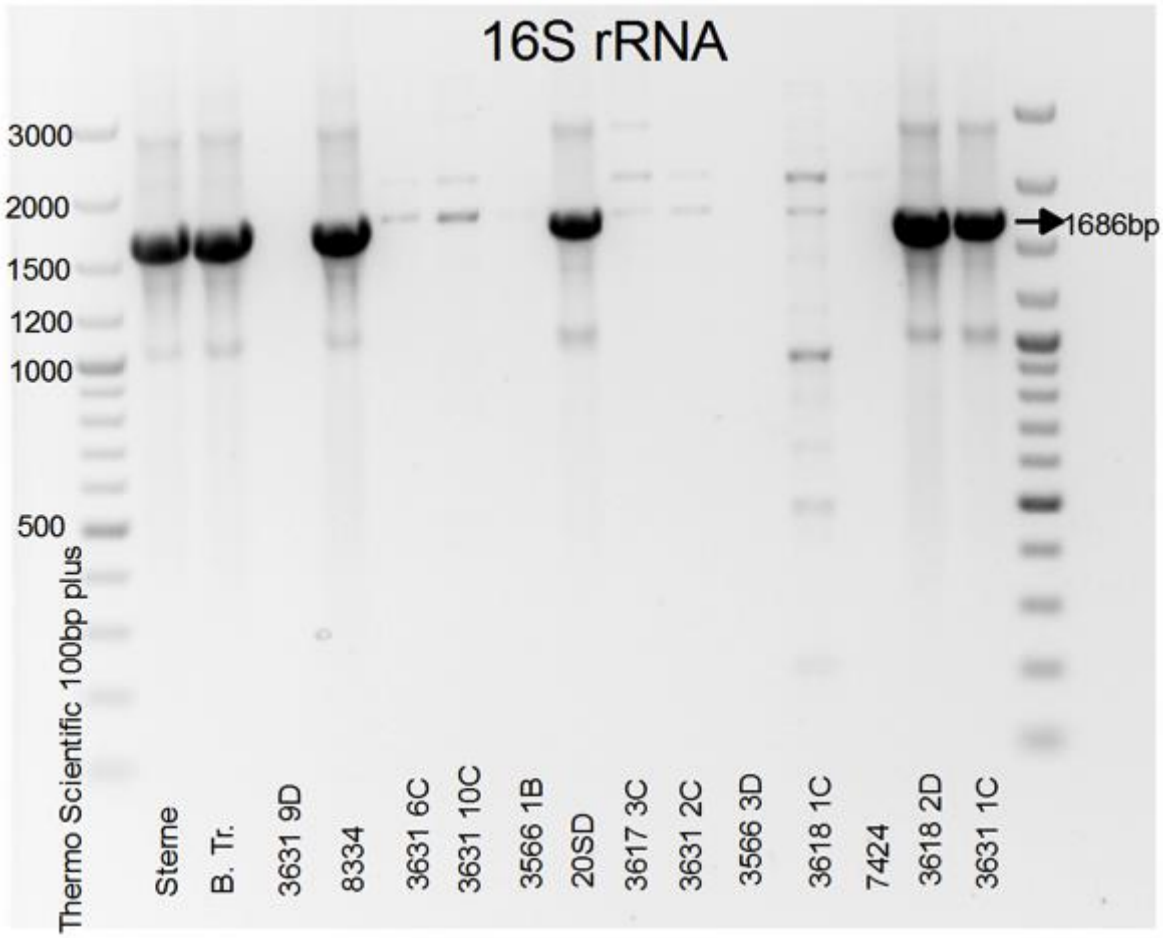


Figure 3.4. The 16S ribosomal RNA gene region amplified with *Bacillus cereus* group specific primers (67F and 1671R) of 1686 bp PCR product. Generuler 100 bp plus (Thermo Scientific) marker was used.

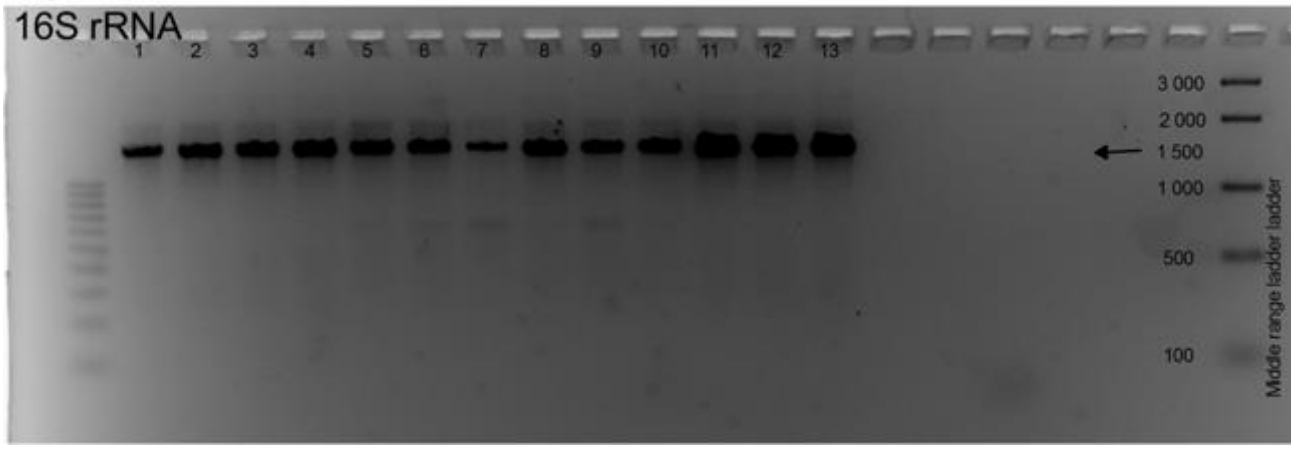


Figure 3.5. The 16S ribosomal RNA gene PCR generated with universal (67F and 1671R) primers. Lanes 1-13 includes *B. anthracis* and *Bacillus species* isolates consisting of 3631-9D, 3631-6C, 3631-10C, 8334, 3617-2C, 3617-3C, 3566-1B, 3566-3D, 3618-1C, 7424, 3631-1C, 3618-2D and 20SD, respectively. FastRuler Middle range ladder (Thermo Scientific) was used.

3.3.4. 16S rRNA sequencing and phylogenetic analysis

The *B. anthracis* Sterne and confirmed *B. anthracis* isolates (20SD, 3631-1C, 3618-1C) with conventional microbiology tests had identical 16S rRNA sequences and clustered with the *B. anthracis* Sterne and *B. anthracis* Ames strains using maximum likelihood phylogeny (Figure 3.6). The 16S rRNA sequence of isolate 8334 from Lesotho clustered with the *B. cereus* group and was closely related to *B. thuringiensis* IAM 12077 in this cluster. The 16S rRNA sequence of isolate 3566-3D was closely related to *Brevibacterium frigoritolerans* and all the other *Bacillus* isolates clustered with *B. endophyticus* (n = 8) (Figure 3.6).

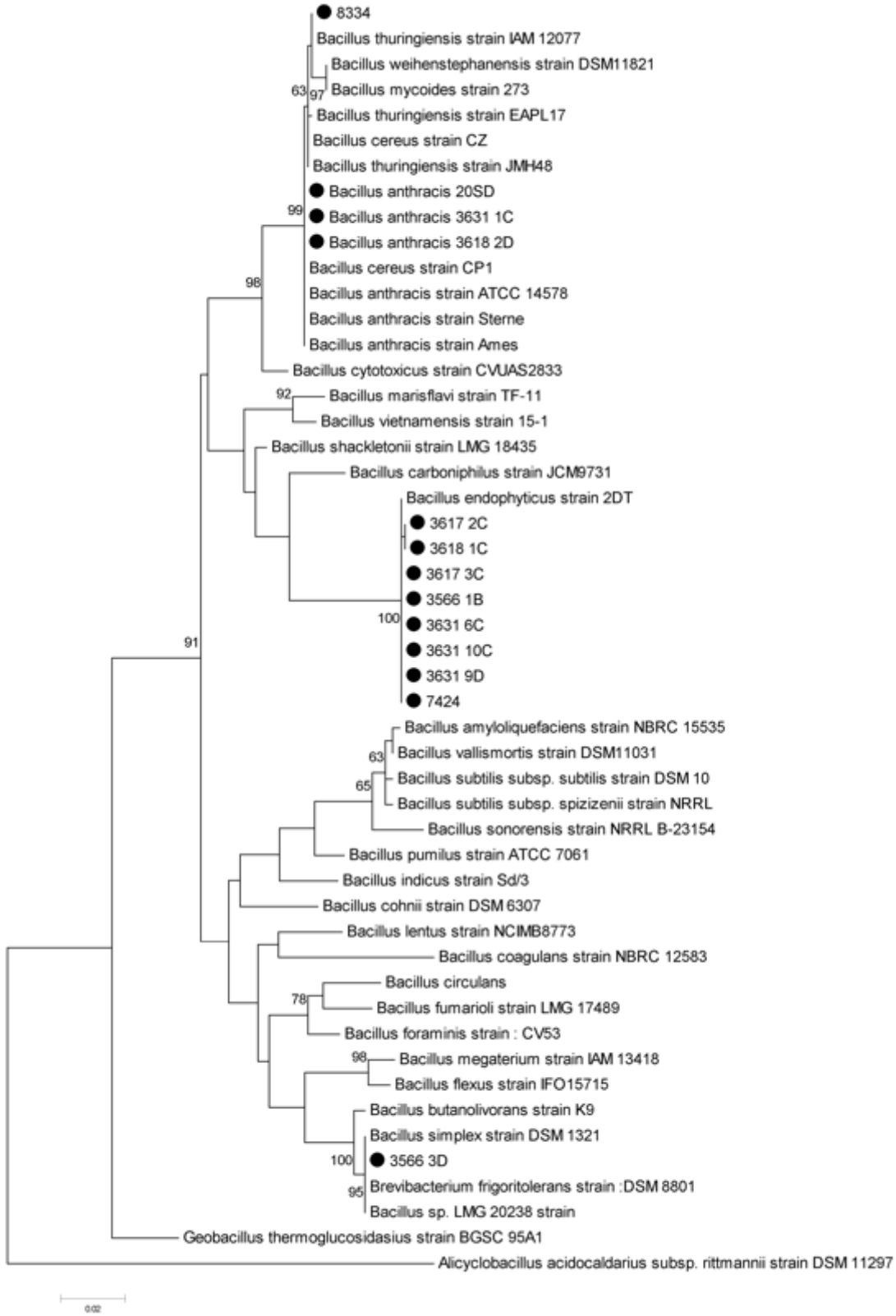


Figure 3.6. Maximum Likelihood (ML) phylogeny for the southern African *Bacillus anthracis* and *Bacillus* species isolates (•) based on 16S rRNA sequences with other Bacillus group obtained from Genbank. Bootstrap values > 60% are indicated at the internodes. The *Geobacillus thermoglucosidasius* and *Alicyclobacillus acidocaldarius* were used as outgroup to root the tree. Scale bar indicates nucleotide substitutions per site.

3.3.5. Conventional PCR assay and sequence analysis

The PCR assays confirmed the presence or absence of the pXO1 and pXO2 virulence genes (Table 3.5). The *Bacillus* species isolates yielded the expected 596 bp amplicons for *pag* region on pXO1 with the exception of isolate 7424 (Supplementary Figure 3S3). The *Ba*813 chromosomal gene, which is associated with *B. anthracis* was detected in *Bacillus* isolates 3631-9D and 3566-3D, which yielded the expected 152 bp PCR product (Figure 3.7). The *cap*C gene (pXO2) was present in most *Bacillus* isolates (3618-2D, 36319D, 8334, 3631-6C, 3631-10C, 3566-1B, 3566-3D, 3617-3C, 3617-2C, 3618-1C) except isolates 7424, 3631-1C and Sterne (Table 3.5, Supplementary Figure 3S3). The *Bacillus* isolates produced varying results with the different regions of the capsule gene, namely *cap*A, *cap*B/C, and *cap*C (Table 3.5, Figure 3.7). The VNTR pXO1 and pXO2 marker as described by Keim *et al*. (2000) were used to amplify target region on plasmids. The pXO1 region was present in all the isolates and pXO2 region was only present in *B. anthracis* isolates 20SD and 3618-2D, and absent from *B. anthracis* Sterne and 3631-1C isolates. The results are summarized in Table 3.5.

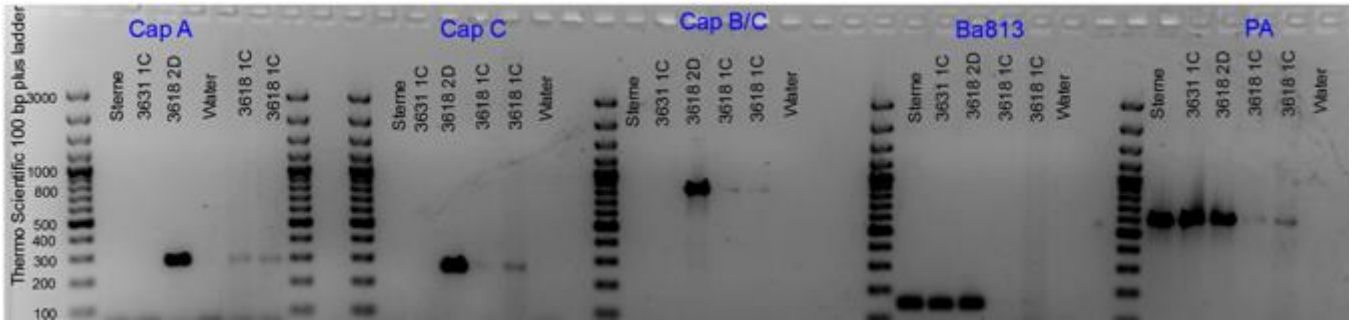


Figure 3.7. Conventional PCR products of *cap*A, *cap*C, *cap*B/C, *Ba813* and *pag* (PA) gene regions for *Bacillus anthracis* and *Bacillus* isolate. The 100 bp plus ladder (Thermo Scientific) and positive control *B. anthracis* Sterne strain was used together with *B. anthracis* isolates (3631-1C and 3618-2D) and *Bacillus* isolate 3618-1C (in duplicate).

Selected PCR products namely *pag*A region using primers PA5 and PA8 and *cap*B/C region using primers 1301 and 1234 of *B. cereus, B. anthracis* Ames, 3618-2D and 3566-1B isolates were sequenced. There was no variation amongst sequence of each target region of sequenced isolates (Supplementary Figure 3S4 and 3S5).

Table 3.5. Summary of the PCR results of *Bacillus anthracis* and *Bacillus* isolates for pXO1, pXO2 plasmids and/or chromosomal genes.

	Conventional PCR[$]							**Real-time PCR[$]**		
Isolates	***pag***	***cap*B/C**	***cap*A**	***cap*C**	***Ba813***	**pXO1 VNTR[&]**	**pXO2 VNTR[&]**	***pag***	***cap*C**	***SASP*[*]**
Sterne[#]	+	-	-	-	+	+	-	+	-	+
3618-2D[#]	+	+	+	+	+	+	+	+	+	+
20SD	+	+	+	+	+	+	+	+	+	+
3631-1C[#]	+	-	-	-	+	+	-	+	-	+
3618-1C	+	+	+	+	-	+	-	-	-	-
3631-9D	+	-	+	+	+	+	-	-	-	-
8334	+	-	+	+	-	+	-	-	-	-
3631-6C	+	+	+	+	-	+	-	-	-	-
3631-10C	+	-	-	+	-	+	-	-	-	-
3566-1B	+	+	-	+	-	+	-	-	-	-
3566-3D	+	-	+	+	+	+	-	-	-	-
3617-3C	+	-	-	+	-	+	-	-	-	-
3617-2C	+	-	-	+	-	+	-	-	-	-
7424	-	-	-	-	-	+	-	-	-	-

+ Indicates amplification of the gene.
- Indicates no amplification.
[#] Identified as *Bacillus anthracis* based on conventional microbiology methods.
* SASP: Small acid soluble proteins in *B. anthracis* chromosome.
[$] See Table 3.2 for primers used to amplify the target regions and probes for hybridization.
[&] Variable number of tandem repeat (VNTR) marker using primers for pXO1 and pXO2 described by Keim *et al*. (2000).

The pleiotropic transcriptional regulator melt analysis of mismatch amplification mutation assays (PlcR-Melt-MAMA) marker showed no amplification on the *Bacillus* isolates (7424, 3618-1C, 8334, 3631-9D, 3566-3D, 3566-1B, 3631-6C, 2617 2C, 3617-3C) including *B. cereus* and *B. thuringiensis* (Supplementary Figure 3S6). The PlcR positive controls *B. anthracis* (Vollum, Ames, and Sterne) strains amplified the same PCR size product as the positive (PC) derived control as indicated by Birdsell *et al*. (2012).

3.3.6. Real-time PCR assay

The real-time PCR assay was used for the detection of regions specific to *B. anthracis* in *pag* (pXO1), *cap*C (pXO2) and chromosomal SASP markers. None of the *B. anthracis* specific probes hybridized with the *Bacillus* isolates. The probes hybridized with the *B. anthracis* strains (Sterne, 3618-2D, 3631-1C and 20SD), except *cap*C probe did not hybridize with Sterne and 3631-1C (Supplementary Figure 3S7).

3.3.7. Multiloci variable number of tandem repeat analysis assay (MLVA)

MLVA-26 of the *B. anthracis* and *Bacillus* isolates reported in this study amplified most of the VNTR markers. The *B. anthracis* isolates reported in the study produced different copy numbers of the VNTR markers relative to the *B. anthracis* Sterne, whereas the non-anthrax (*Bacillus* species) isolates produced different amplicon sizes (copy numbers) and non-specific binding bands for the other VNTR markers (Supplementary Figure 3S1 and 3S2). The NCP *B. anthracis* strains reported in the study clustered with *B. anthracis* Sterne and additional A-lineages *B. anthracis* strains (Figure 3.8). The *Bacillus* isolates clustered seperately from *B. anthracis* with *B. thuringiensis* (Figure 3.8).

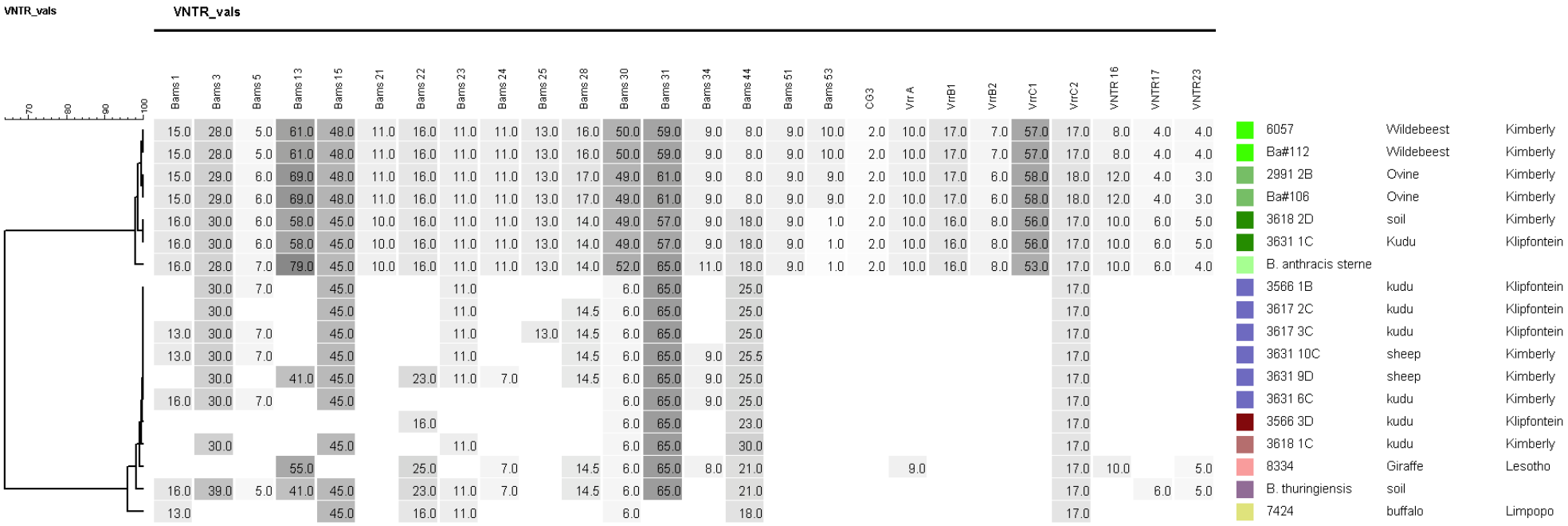


Figure 3.8. MLVA-26 cluster analysis using unweighted pair group method arithmetric averages (UPGMA) algorithm of *Bacillus anthracis* Sterne vaccine strain and other *Bacillus* group reported in the study. Additional confirmed *Bacillus anthracis* isolates from Northern Cape Province (6057, 2991 2B, Ba#112, and Ba#112) A lineage were included with *B. anthracis* Sterne control strain. *Bacillus thuringiensis* was included as a negative control.

3.4. Discussion

In South Africa and Lesotho, *B. anthracis* and *Bacillus* isolates were obtained either from animals that showed symptoms reminiscent of anthrax or from soil that was beneath animals that died of anthrax symptoms. We therefore investigated the isolates using a combination of classical microbiological techniques, API CH50B, Biolog OmniLog, conventional and real-time PCR targeting various virulence genes, 16S rRNA gene sequencing and MLVA. In this study we determined that *B. anthracis* can be differentiated from other similar *Bacillus* species using a combination of classical microbiological techniques, real-time PCR, sequencing and MLVA. With the polyphasic approach we could not identify the other *Bacillus* species (*B. anthracis*-like), but some methods suggested that most of the *Bacillus* isolates were close relatives of *B. endophyticus* and that these isolates might contain virulence genes of *B. anthracis.* Conventional PCR revealed that all the isolates harbored the pXO1 gene and some contained pXO2 capsule genes (Table 3.5). This study highlights the need to determine the significance of *B. anthracis*-like isolates as potential causative agents of anthrax.

The classical microbiological tests indicated that *B. anthracis*-like isolates from this study differed from the classical *B. anthracis* as they were resistant to gamma phage (OIE, 2008; Koehler, 2009). Currently, classical *B. anthracis* diagnosis is based on the observation of capsule, non-motile, non-hemolytic isolates that are sensitive to penicillin and gamma phage (Koehler, 2009). However, the resistance to gamma phage was observed in 15% of *B. anthracis* in previous study (Buck *et al*., 1963) and in South Africa, a survey indicated that up to 16% of *B. anthracis* isolates from soil and carcasses were resistant to gamma phage (Odendaal *et al*., 1991). Although it is a known fact that a single diagnostic attribute may not necessarily compromise the correct identification of classical *B. anthracis* (Klee *et al*., 2006), the combination of various tools used in study has shown that the non-conformance of one diagnostic trait may exclude *B. anthracis.* The other identifying methods (API CH50B, Biolog OmniLog and 16S rRNA sequencing) identified isolates as *B. endophyticus*, *B. thuringiensis* and *Brevibacterium frigotoleransis* despite the fact that

resistance to gamma phage was the only differential microbiological criteria that excluded these isolates from being *B. anthracis*.

The API 50CHB identification system for *Bacillus* species results lacked correlation with 16S rRNA sequences and Biolog OmniLog. The latter technique identified isolates 3617-2C, 3618-1C, 2617 3C, 3566-1B, 3631-10C and 3631-6C (n=6) as *B. endophyticus* (Table 3.4). Indeed, the API 50 CHB system was rendered irrelevant for diagnosis of *B. anthracis*-like isolates in this study. This was further evidenced by the low probabilities of identities obtained (Table 3.4) and the inability for identification of *Bacillus* isolates in the study by Klee *et al.* (2006).

The *B. endophyticus* identified by Biolog OmniLog and 16S rRNA sequencing reported in this study are associated with plants and their role in anthrax-associated animal deaths could not be established in the study. However, the role of virulence genes that were identified in these isolates may not be ruled out. For this reason, further investigation of the virulence genes found in the *B. endophyticus* is paramount. This could improve the understanding of the role played by these isolates when they are diagnosed during anthrax outbreaks.

Most *B. anthracis*-like isolates in this study contained the protective antigen gene, which indicates that they were *B. anthracis*-like (Klee *et al*., 2006). The sequence alignment of the *pag*A PCR product showed that the *Bacillus* strain 3566-1B, *B. cereus*, *B. anthracis* Ames and *B. anthracis* 3618-2D has identical sequences (Supplementary Figure 3S4). The *pag* gene is a known carrier of the two toxin proteins in the cell that pass through the cell membrane, hence it is tempting to infer that the *B. anthracis*-like isolates in this study contains anthrax virulence genes that caused symptoms that were reminiscent of anthrax. However the sequence results indicate that the *pag*A region amplified using PA5 and PA8 primers is not specific to *B. anthracis*. *Bacillus anthracis*-like strains were reported from chimpanzees that died of anthrax symptoms in Tai National Park, Cote d'Ivoire (CI) and a gorilla that died in Cameroon (CA) as a result of *B. cereus* biovar *anthracis* (Klee *et al*., 2006; Leendertz *et al*., 2006; Klee *et al.,* 2010). The *B. anthracis* virulence genes, pXO1

and pXO2, were present in the *B. cereus* biovar *anthracis* CI and CA strains (Klee *et al*., 2006; 2010). Nevertheless, thorough investigation is required of *B. anthracis*-like isolates and anthrax virulence genes as the molecular mechanism contributing towards pathogenicity of the *B. anthracis*-like isolates is unknown. The plasmids, pXO1 and/or pXO2, have also been reported in other *B. cereus* strains (Hoffmaster *et al*., 2004; 2006). Hoffmaster *et al*. (2004) also identified anthrax toxin genes in *B. cereus* G9241 that is capable of causing a severe inhalation anthrax-like disease. The genome sequence of *B. cereus* G9241 revealed the plasmid pBXO1 that is 99.6% similar to pXO1. Weak PCR products of pXO1 target suggest that the primer sequences are not 100% identical to the sequence of the genes in *B. endophyticus* strains (Figure 3.7). The plasmid pXO1 genes of *B. anthracis* have been found significantly similar to the chromosomal encoded genes from the other members of the *B. cereus* group, *B. subtilis*, *B. hakodurans*, *Listeria*, and *Staphylococcus* (Okinaka *et al.*, 1999; Rasko *et al*., 2007).

The *B. anthracis*-like isolates from this study also contained the *cap*C genes (Table 3.5) except for isolate 7424. The capsular genes *Cap*BCADE play a role in the synthesis of the poly-glutamic acid or polyglutamate (PGA) capsule of *B. anthracis* and these are located on the pXO2 plasmid. PGA has the ability to evade the host immune system by protecting the vegetative cells from phagocytotic killing by macrophages (Candela and Fouet, 2005; 2006). However, poly-glutamic acid is synthesized by many *Bacillus* species (Beesley *et al*., 2010) and other organisms (Candela and Fouet, 2005). Three *B. anthracis* genes, *cap*B, *cap*C, and *cap*A, were shown to be necessary to drive the production of PGA weakly in *Escherichia coli* but could not promote PGA synthesis in *B. subtilis* (Makino *et al.,* 1989; Candela and Fouet, 2005). Candela and Fouet (2005) investigated the five *cap* genes and found *capA*, *cap*B, *cap*C, and *cap*E genes all necessary for sufficient PGA synthesis by *B. anthracis*. The *cap*D gene encodes a Ɣ-glutamyl transpeptidase or PGA depolymerase, which is important for the covalent anchoring of the PGA to the peptidoglycan in *B. anthracis* (Candela and Fouet, 2005; Bleesley *et al*., 2010). Therefore, isolates that lack *cap*D capsular gene produce a loose slime layer of the PGA instead of a covalently linked capsule (Candela and Fouet, 2005; Bleesley *et al*., 2010). Similar capsule genes in *Bacillus* species therefore explain varying results obtained with PCR using different capsule

gene regions in this study. In this study, some of the isolates tested positive for the three genes (*cap*A, *cap*B/C and *cap*C) using conventional PCR. The PCR amplification of *cap*B/C amplified with primers 1301 and 1234 is not specific to *B. anthracis* as it generated the PCR products for *B. anthracis*, *B. cereus*, and *Bacillus* species (Table 3.5). The sequence alignment of the capsular region (*cap*B/C) shows that the *Bacillus* strain 3566-1B, *B. cereus*, *B. anthracis* Ames and *B. anthracis* 3618-2D are 100% similar (Supplementary Figure 3S5). *Bacillus thuringiensis*, *B. subtilis* 168 IFO3336, *B. licheniformis* ATCC 14580 and *Staphyloccus epidermidis* ATCC 12228 also contain the *cap*B, *cap*C, *cap*A and *cap*E genes similar to *B. anthracis*. Annotations of these genes are named differently amongst the species but they synthesise the PGA (Kunst *et al*., 1997, Rey *et al*., 2004; Candela *et al.*, 2005). In the *B. subtilis and B. licheniformis* they are named *pgs* (polyglutamate synthase) when PGA is released, which might help the bacterium to survive in high salt concentrations or confers resistance to adverse environment (Candela and Fouet, 2006; Bajaj and Singhal, 2011).

The amplified capsule gene regions of the *Bacillus* species isolates in this study suggests that the genes could be responsible for the production of the PGA or they may be involved in virulence since the capsule genes are present. In other studies where *B. megaterium* and other *Bacillus* species were characterized, the *cap*B, *cap*C, *cap*A and *cap*D primers specific to *B. anthracis* could not amplify regions from isolates using PCR (Beesley *et al*., 2010). However, phenotypically the isolates produced PGA suggesting that they are too divergent to be amplified using PCR primers specific for *B. anthracis* (Beesley *et al.,* 2010). Phenotypically the members of the *Bacillus* species reported in the study produce microscopically visible capsules when encapsulated. Therefore, the *B. anthracis*-like isolates from South Africa and Lesotho contained different capsular genes (Table 3.5), which might have contributed either to virulence or environmental survival.

The chromosomal region *Ba*813 was regarded as a useful gene with respect to discriminating *B. anthracis* from other *Bacillus* isolates (Ramisse *et al*., 1996), but members of the *B. cereus* group including *B. thuringiensis* and *B. cereus* have been reported to contain the *Ba*813 DNA sequence (Ramisse *et al.,* 1999; Klee *et al.,* 2006). In this study,

the *Ba*813 gene was detected in close relatives of *B. endophyticus* (3631-9D) and *Brevibacterium frigoritolerans* (3566-3D). This highlights the need to search for this gene in other bacterial species that are not necessarily members of the *B. cereus* group.

The MLVA technique (Keim *et al*., 2000; Lista *et al*., 2006; Van Ert *et al*., 2007; Beyer *et al*., 2012) was designed to type *B. anthracis* strains but has been reported to differentiate closely related organisms in the *B. cereus* group (Valjevac *et al*., 2005). In this study, the MLVA-26 differentiated the *B. anthracis*-like isolates from the typical *B. anthracis* as the *B. anthracis* isolates clustered separately from the other *Bacillus* species (Figure 3.8). The VNTR markers were not only specific to the *B. anthracis* but could amplify other *Bacillus* species, however in most cases the amplicon sizes differered clearly from those found with *B. anthracis*. For the *Bacillus* species isolates, non-specific bindings were observed for several of the VNTR markers (Supplementary Figure 3S1 and 3S2). This is an indication of differences in the genomes as the sequences are more specifically designed for the *B. anthracis*.

The members of the *B. cereus* group are under the control of pleiotropic transcriptional regulator (PlcR) and its regulatory peptide, PapR that transcribe genes encoding collagenases, haemolysins, phospholipases and enterotoxins (Aggaisse *et al*., 1999). The PlcR gene product of *B. cereus* and *B. thuringiensis* is known to upregulate the production of numerous extracellular enzymes, while *B. anthracis* has a nonsense mutation that is responsible for non-motile and non-heamolytic (Agaisse *et al*., 1999; Gohar *et a*l., 2008). The PlcR- melt-MAMA marker can be used as a specific marker for the *B. anthracis*, since it differentiated *B. anthracis* strains from *Bacillus* isolates (Supplementary Figure 3S6).

The real-time PCR assay is a more sensitive PCR assay compared to conventional PCR method for diagnosis of anthrax. The SASPs (small acid soluble proteins) found in *B. anthracis* spores were used as *B. anthracis* specific protein marker (Pribil *et al*., 2005). However, this marker has been found to be insufficient in discriminating amongst closely related organisms (Callahan *et al.,* 2008; 2009). However, the real-time PCR consists of confirmation of three targets namely the SASP, and plasmids targets that improves the

specificity of the test. In this study real-time PCR, MLVA, 16S rRNA sequencing and conventional microbiology tests accurately discriminated *B. anthracis* from closely related *Bacillus* species. However, closely related *Bacillus* species with virulence genes that may cause anthrax in animals in South Africa and Lesotho need to be characterized using whole genome sequencing, which will verify the accuracy of identification methods used in this study.

3.5. Conclusion

The use of Biolog OmniLog system and 16S rRNA gene sequencing could identify most of the isolates as *B. endophyticus.* These techniques can differentiate *B. anthracis* from *B. anthracis*-like isolates. However, the Biolog OmniLog system may not be sufficient for diagnosis of *B. anthracis*-like isolates on its own and neither of these methods will detect the virulence genes. The presence of virulence genes can be detected using PCR. However, the reliability of PCR analysis provides limited information about the genetic basis or virulence as it relies on previous provided genetic sequences. Therefore, whole genome sequencing is needed to resolve the variability within species and sub-species groups that are closely related amongst the *B. cereus*/*subtilis* group. This is important for comparative analysis of the virulence genes that might be associated with anthrax symptoms.

3.6. References

Agaisse, H., Gominet, M., Okstad, O.A., Kolsto, A., Lereclus, D., 1999. PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. Mol. Microbiol. 32, 1043-1053.

Ash, C., Farrow, J., Wallbanks, S., Collins, M., 1991. Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small- subunit- ribosomal RNA sequences. Lett. Appl. Microbiol. 13, 202-206.

Bajaj, I., Singhal, R., 2011. Poly (Glutamic Acid)–An emerging biopolymer of commercial interest. Bioresour. Technol. 102, 5551-5561.

Battisti, L., Green, B.D., Thorne, C.B., 1985. Mating system for transfer of plasmids among *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*. J. Bacteriol. 162, 543-550.

Beesley, C.A., Vanner, C.L., Helsel, L.O., Gee, J.E., Hoffmaster, A.R., 2010. Identification and Characterization of clinical *Bacillus* spp. isolates phenotypically similar to *Bacillus anthracis*. FEMS Microbiol. Lett. 313, 47-53.

Beyer, W., Pocivalsek, S., 1999. Polymerase chain reaction- ELISA to detect *Bacillus anthracis* from soil samples-limitations of present published primers. J. Appl. Microbiol. 87, 229-236.

Beyer, W., Bellan, S., Eberle, G., Ganz, H.H., Getz, W.M., Haumacher, R., Hilss, K.A., Kilian, W., Lazak, J., Turner, W.C., 2012. Distribution and molecular evolution of *Bacillus anthracis* genotypes in Namibia. PLoS Neglect. Trop. Dis. 6, e1534.

Beyer, W., Glockner, P., Otto, J., Bohm, R., 1995. A nested PCR method for the detection of Bacillus anthracis in environmental samples collected from former Tannery sites. Microbiol. Res. 150, 179-186.

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., 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, e32866.

Buck, C., Anacker, R., Newman, F., Eisenstark, A., 1963. Phage Isolated from lysogenic *Bacillus anthracis*. J. Bacteriol. 85, 1423-1430.

Callahan, C., Castanha, E.R., Fox, K.F., Fox, A., 2008. The *Bacillus cereus* containing sub-branch most closely related to *Bacillus anthracis*, have single amino acid substitutions in small acid-soluble proteins, while remaining sub-branches are more variable. Mol. Cell. Probes 22, 207-211.

Callahan, C., Fox, K., Fox, A., 2009. The small acid soluble proteins (SASP α and SASP β)

of *Bacillus weihenstephanensis* and *Bacillus mycoides* group 2 are the most distinct among the *Bacillus cereus* group. Mol. Cell. Probes 23, 291-297.

Candela, T., Fouet, A., 2006. Poly-gamma-glutamate in bacteria. Mol. Microbiol. 60, 1091-1098.

Candela, T., Fouet, A., 2005. *Bacillus anthracis* CapD, belonging to the γ-glutamyltranspeptidase family, is required for the covalent anchoring of capsule to peptidoglycan. Mol. Microbiol. 57, 717-726.

Centres for disease control and prevention/American society for microbiology/Association of public health laboratories (CDC/ASM/APHL). 2002. Protocols for level a laboratories: For the presumptive identification of *Bacillus anthracis*. ban.asm.la.cp.031802. Page 1-22.

Drysdale, M., Heninger, S., Hutt, J., Chen, Y., Lyons, C.R., Koehler, T.M., 2004. Capsule synthesis by *Bacillus anthracis* is required for dissemination in murine Inhalation anthrax. EMBO J. 24, 221-227.

Gohar, M., Faegri, K., Perchat, S., Ravnum, S., Okstad, O.A., Gominet, M., Kolsto, A., Lereclus, D., 2008. The PlcR virulence regulon of *Bacillus cereus*. PLoS One 3, e2793.

Helgason, E., Okstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I., Kolsto, A.B., 2000. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. Appl. Environ. Microbiol. 66, 2627-2630.

Helgason, E., Tourasse, N.J., Meisal, R., Caugant, D.A., Kolsto, A., 2004. Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. Appl. Environ. Microbiol. 70, 191-201.

Hoffmaster, A.R., Ravel, J., Rasko, D.A., Chapman, G.D., Chute, M.D., Marston, C.K., De, B.K., Sacchi, C.T., Fitzgerald, C., Mayer, L.W., 2004. Identification of anthrax toxin genes in *a Bacillus cereus* associated with an illness resembling inhalation anthrax. Proc. Natl. Acad. Sci. U. S. A. 101, 8449-8454.

Hoffmaster, A.R., Hill, K.K., Gee, J.E., Marston, C.K., De, B.K., Popovic, T., Sue, D., Wilkins, P.P., Avashia, S.B., Drumgoole, R., 2006. Characterization of *Bacillus cereus* isolates associated with fatal pneumonias: strains are closely related to *Bacillus anthracis* and harbor *B. anthracis* virulence genes. J. Clin. Microbiol. 44, 3352-3360.

Keim, P., Price, L., Klevytska, A., Smith, K., Schupp, J., Okinaka, R., Jackson, P., Hugh-Jones, M., 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. 182, 2928-2936.

Klee, S.R., Brzuszkiewicz, E.B., Nattermann, H., Brüggemann, H., Dupke, S., Wollherr, A., Franz, T., Pauli, G., Appel, B., Liebl, W., 2010. The genome of a *Bacillus* isolate

causing anthrax in chimpanzees combines chromosomal properties of B. cereus with *B. anthracis* virulence plasmids. PLoS One 5, e10986.

Klee, S.R., Ozel, M., Appel, B., Boesch, C., Ellerbrok, H., Jacob, D., Holland, G., Leendertz, F.H., Pauli, G., Grunow, R., 2006. Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. J. Bacteriol. 188, 5333-5344.

Koehler, T.M., 2009. *Bacillus anthracis* physiology and genetics. Mol. Aspects Med. 30, 386-396.

Kolsto, A.B., Tourasse, N.J., Okstad, O.A., 2009. What Sets *Bacillus anthracis* apart from other *Bacillus* species? Annu. Rev. Microbiol. 63, 451-476.

Kotiranta, A., Lounatmaa, K., Haapasalo, M., 2000. Epidemiology and pathogenesis of *Bacillus cereus* infections. Microb. Infect. 2, 189-198.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A., Alloni, G., Azevedo, V., Bertero, M., Bessieres, P., Bolotin, A., Borchert, S., 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390, 249-256.

Lane, D., 1991. 16S/23S rRNA sequencing. In: Stackerbrandt, E., Goodfellow, M. (eds), Nucleic acid techniques in bacterial systematics-John Wiley and Sons Chichester, New York, 115-175.

Leendertz, F.H., Lankester, F., Guislain, P., Neel, C., Drori, O., Dupain, J., Speede, S., Reed, P., Wolfe, N., Loul, S., 2006. Anthrax in Western and Central African great apes. Am. J. Primatol. 68, 928-933.

Le Fleche, 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 Microbiol. 1, 2.

Lista, F., Faggioni, G., Valjevac, S., Ciammaruconi, A., Vaissaire, J., Le Doujet, C., Gorge, O., De Santis, R., Carattoli, A., Ciervo, A., 2006. Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. BMC Microbiol. 6, 33.

Makino, S.I., Iinuma-Okada, Y., Maruyama, T., Ezaki, T., Sasakawa, C., and Yoshikawa, M., 1993. Direct Detection of *Bacillus anthracis* DNA in Animals by Polymerase Chain Reaction. J. Clin. Microbiol. 31, 547- 551.

Makino, S., Uchida, I., Terakado, N., Sasakawa, C., Yoshikawa, M., 1989. Molecular characterization and protein analysis of the cap region, which is essential for encapsulation in *Bacillus anthracis*. J. Bacteriol. 171, 722-730.

Mock, M., Fouet, A., 2001. Anthrax. Annu. Rev. Microbiol. 55, 647-671.

Mock, M., Mignot, T., 2003. Anthrax toxins and the host: a story of intimacy. Cell. Microbiol. 5, 15-23.

Modrie, P., Beuls, E., Mahillon, J., 2010. Differential transfer dynamics of pAW63 plasmid among members of the *Bacillus cereus* group in food microcosms. J. Appl. Microbiol. 108, 888-897.

Odendaal, M., Pieterson, P., De Vos, V., Botha, A., 1991. The antibiotic sensitivity patterns of *Bacillus anthracis* isolated from the Kruger National Park. Onderstepoort J. Vet. Res. 58, 17.

OIE, AHS. Manual of diagnostic tests and vaccines for terrestrial animals. 2008. Chapter 2.1.1. Anthrax, OIE terrestrial manual, Paris, France, 133-144.

Okinaka, R., Cloud, K., Hampton, O., Hoffmaster, A., Hill, K., Keim, P., Koehler, T., Lamke, G., Kumano, S., Mahillon, J., 1999. Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. J. Bacteriol. 181, 6509-6515.

Okinaka, R., Pearson, T., Keim, P., 2006. Anthrax, but not *Bacillus anthracis*? PLoS pathogens 2, e122.

Pannucci, J., Okinaka, R.T., Sabin, R., Kuske, C.R., 2002. *Bacillus anthracis* pXO1 plasmid sequence conservation among closely related bacterial species. J. Bacteriol. 184, 134-141.

Papazisi, L., Rasko, D.A., Ratnayake, S., Bock, G.R., Remortel, B.G., Appalla, L., Liu, J., Dracheva, T., Braisted, J.C., Shallom, S., 2011. Investigating the genome diversity ofb. cereusand evolutionary aspects of *Bacillus anthracis* emergence. Genomics 98, 26-39.

Patra, G., Sylvestre, P., Ramisse, V., Therasse, J., Guesdon, J., 1996. Isolation of a specific chromosomic dna sequence of *Bacillus anthracis* and its possible use in diagnosis. FEMS Immunol. Med. Microbiol. 15, 223-231.

Pribil, P.A., Patton, E., Black, G., Doroshenko, V., Fenselau, C., 2005. Rapid characterization of *Bacillus* spores targeting species- unique peptides produced with an atmospheric pressure matrix- assisted laser desorption/ionization source. J. Mass Spectrom. 40, 464-474.

Ramisse, V., Patra, G., Vaissaire, J., Mock, M., 1999. The *Ba813* chromosomal DNA sequence effectively traces the whole *Bacillus anthracis* community. J. Appl. Microbiol. 87, 224-228.

Ramisse, V., Patra, G., Garrigue, H., Guesdon, J., Mock, M., 1996. Identification and characterization of *Bacillus anthracis* by multiplex PCR analysis of sequences on

plasmids pXO1 and pXO2 and chromosomal DNA. FEMS Microbiol. Lett. 145, 9-16.

Rasko, D.A., Altherr, M.R., Han, C.S., Ravel, J., 2005. Genomics of the *Bacillus cereus* group of organisms. FEMS Microbiol. Rev. 29, 303-329.

Rasko, D.A., Rosovitz, M., Okstad, O.A., Fouts, D.E., Jiang, L., Cer, R.Z., Kolsto, A., Gill, S.R., Ravel, J., 2007. Complete sequence analysis of novel plasmids from emetic and periodontal *Bacillus cereus* isolates reveals a common evolutionary history among the *B. cereus* group plasmids, including *Bacillus anthracis* pXO1. J. Bacteriol. 189, 52-64.

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., 2003. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature 423, 81-86.

Rey, M., Ramaiya, P., Nelson, B., Brody-Karpin, S., Zaretsky, E., Tang, M., De Leon, A., Xiang, H., Gusti, V., Clausen, I.G., 2004. Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. Genome Biol. 5, r77.

Sacchi, C.T., Whitney, A.M., Mayer, L.W., Morey, R., Steigerwalt, A., Boras, A., Weyant, R.S., Popovic, T., 2002. Sequencing of 16S rRNA gene: a rapid tool for identification of *Bacillus anthracis*. Emerg. Infect. Dis. 8, 1117.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D., Dean, D., 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. 62, 775-806.

Stenfors Arnesen, L.P., Fagerlund, A., Granum, P.E., 2008. From soil to gut: *Bacillus cereus* and its food poisoning toxins. FEMS Microbiol. Rev. 32, 579-606.

Valjevac, S., Hilaire, V., Lisanti, O., Ramisse, F., Hernandez, E., Cavallo, J., Pourcel, C., Vergnaud, G., 2005. Comparison of minisatellite polymorphisms in the *Bacillus cereus* complex: a simple assay for large-scale screening and identification of strains most closely related to *Bacillus anthracis*. Appl. Environ. Microbiol. 71, 6613-6623.

Van der Auwera, G.A., Timmery, S., Mahillon, J., 2008. Self-transfer and mobilisation capabilities of the pXO2-like plasmid pBT9727 from *Bacillus thuringiensis* subsp.konkukian97-27. Plasmid 59, 134-138.

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., 2007. Strain-specific single-nucleotide polymorphism assays for the *Bacillus anthracis* Ames strain. J. Clin. Microbiol. 45, 47-53.

World Health Organization (WHO). Anthrax in humans and animals. 4th edition. Edited by Turnbull PCB. World Health Organization, Geneva, http://www.who.int/csr/resources/publications/AnthraxGuidelines2008/en/index.html.

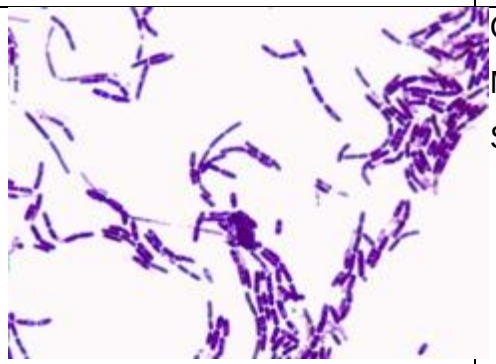
World Health Organization (WHO), 2003. Manual for Laboratory Diagnosis of Anthrax. World Health Organization, Geneva, Switzerland, http://w3.whosea.org/bct/anthrax.

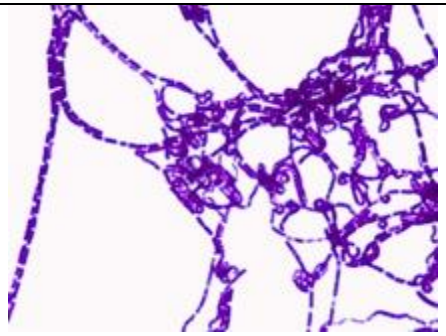
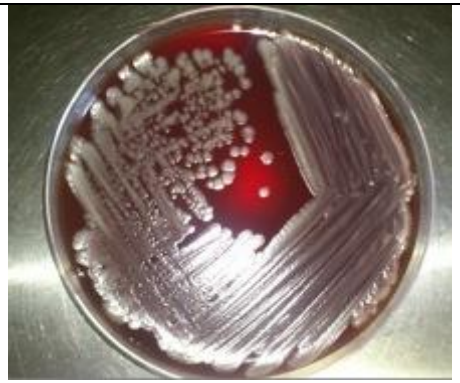
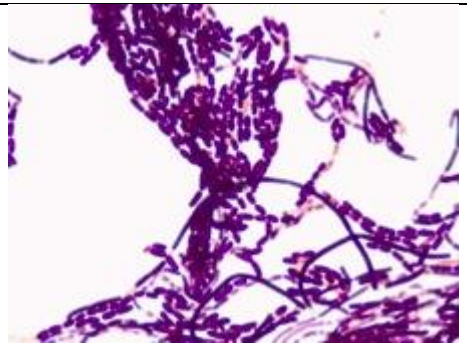
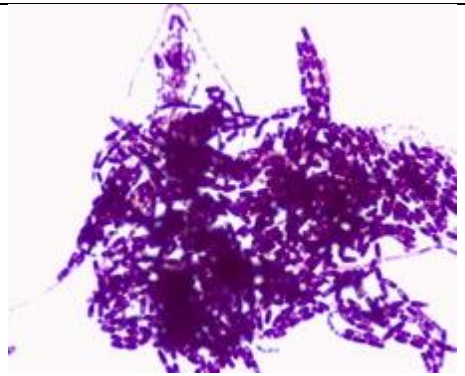
Young, J.A.T., Collier, R.J., 2007. Anthrax toxin: receptor binding, internalization, pore formation, and translocation. Annu. Rev. Biochem. 76, 243-265.

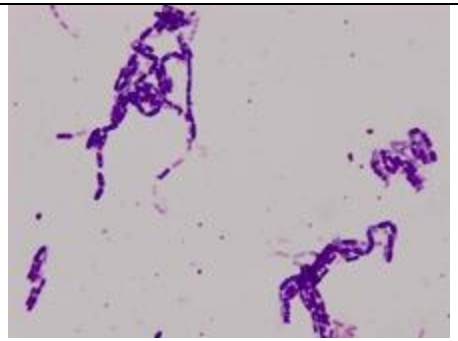
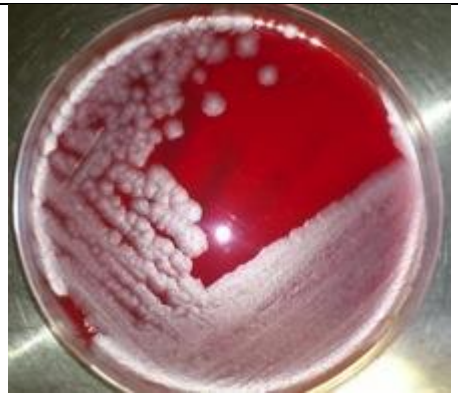
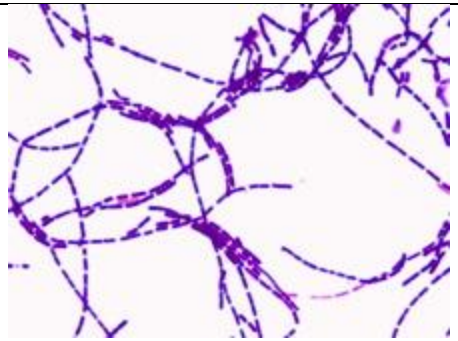
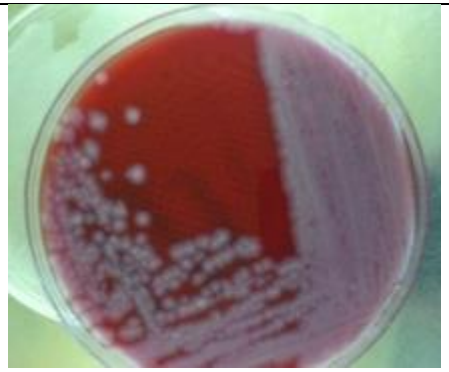
Zwick, M.E., Joseph, S.J., Didelot, X., Chen, P.E., Bishop-Lilly, K.A., Stewart, A.C., Willner, K., Nolan, N., Lentz, S., Thomason, M.K., 2012. Genomic characterization of the *Bacillus cereus* sensu lato species: backdrop to the evolution of *Bacillus anthracis*. Genome Res. 22, 1512-1524.

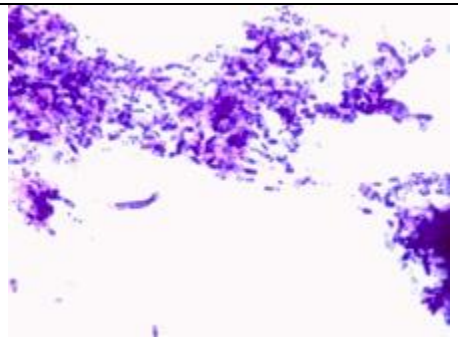
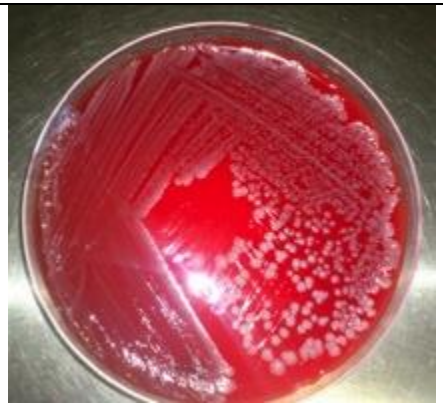
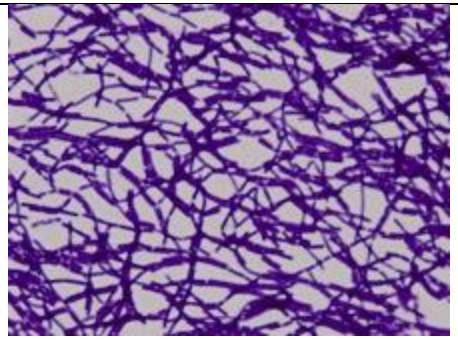
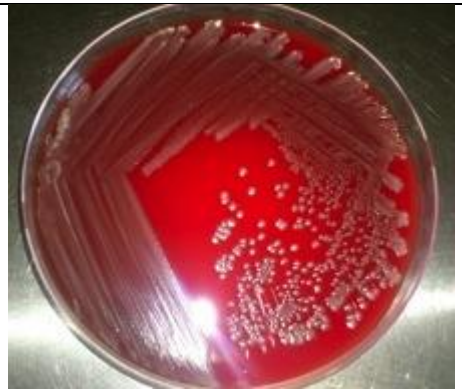
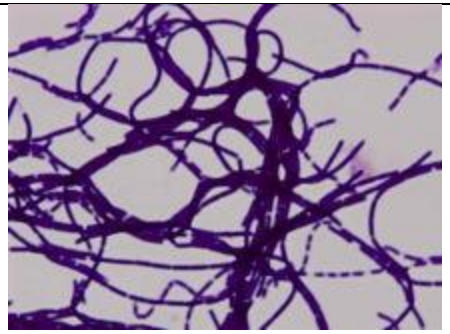
3.7. Supplementary tables and figures

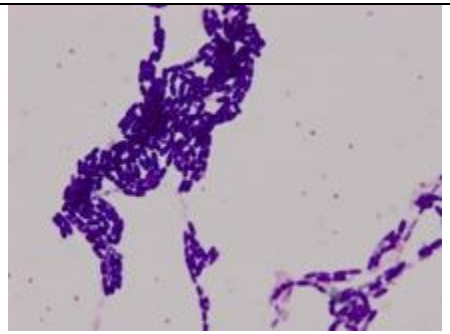
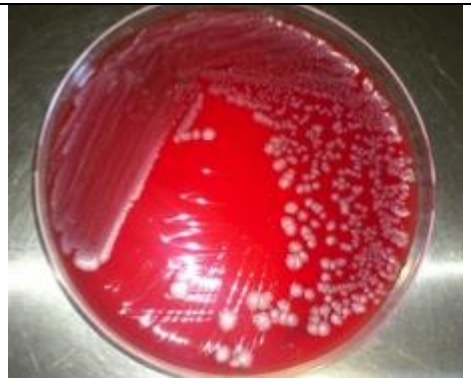
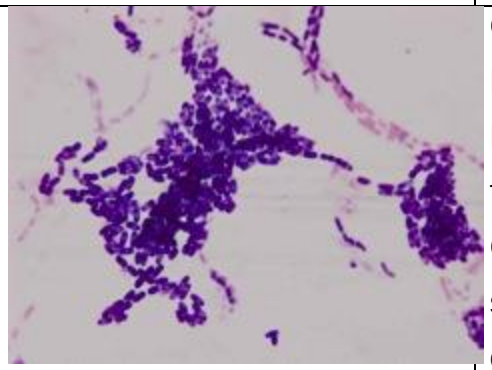
Supplementary Table 3S1. Morphology characterization of *Bacillus* species isolated from animals in the South Africa and Lesotho.

Stain ID	Gram stain		Morphology	
	Reaction	**Observation**	**Blood Agar**	**Observation**
Sterne		Gram + square ended rods, single and pair long chains		White glass appearance, irregular edges, With medusa heads
3631-9D		Gram + Medium rods Small chain		White circular, smooth appearance, with medusa heads

Strain ID	Gram stain	Observation	Blood agar	Observation
8334		Gram + Large folding rods Long chains		White-grey, irregular edges
3631-6C		Gram + Medium rods Short and long chains		White circular colonies, smooth
3631-10C		Gram + Medium rods Clustered and Long chains		White circular colonies, smooth

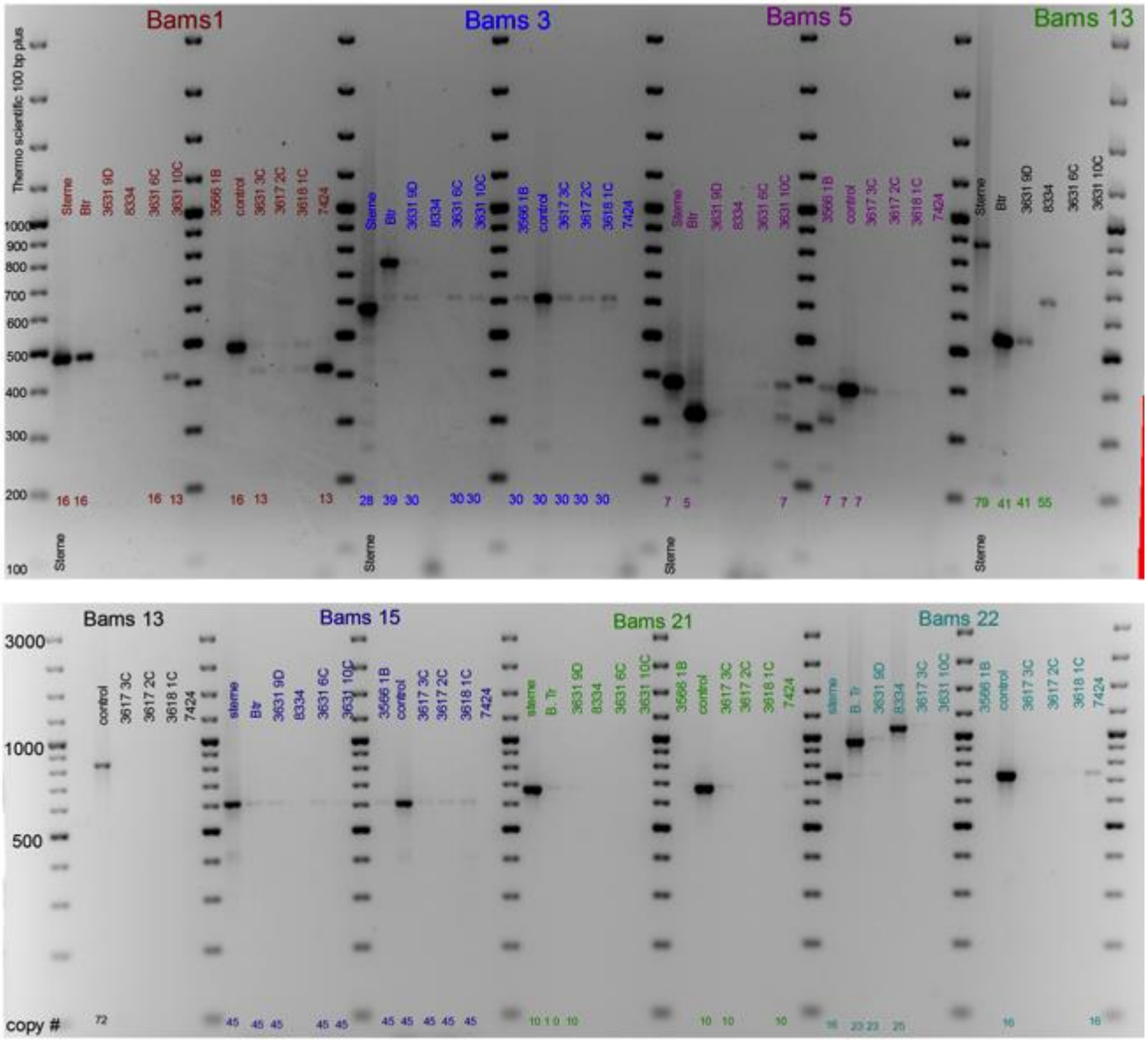
Strain ID	Gram stain	Observation	Blood agar	Observation
7424		Gram + small rods, short chains		White circular, smooth and small colonies
3631-1C		Gram + square ended rods, single and pair long chains		White glass appearance, irregular edges, with medusa heads
3618-2D		Gram + square ended rods, single and pair long chains		White glass appearance, irregular edges, with medusa heads

Strain ID	Gram stain	Observation	Blood agar	Observation
3566-1B		Gram + small rods		White circuar, rough or dry colonies
3566-3D		Gram + Small rods Clustered Short chains		Light yellow white, smooth colonies
3617-3C		Gram +, square end rods, Single and long chains		White circular, smooth colonies

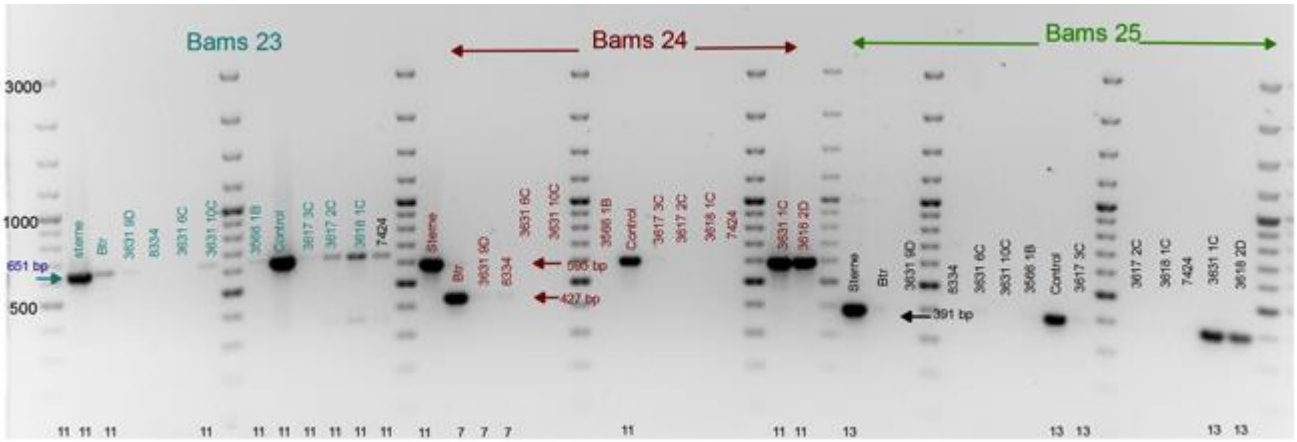
Strain ID	Gram stain	Observation	Blood agar	Observation
3617-2C		Gram + Medium rods Clustered and small in chains		White circular, ground glass appearance
3618-1C		Gram + Medium rods Endospore-formation Clustered and small in chains		White circular, ground glass appearance

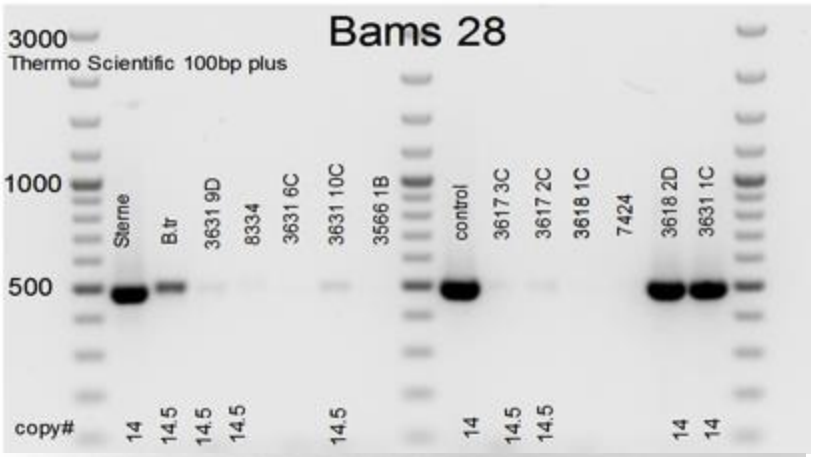
Supplementary Table 3S2. The information of the GEN III MicroPlate™ for the Biolog system indicating the carbon sources and sensitivity assays.

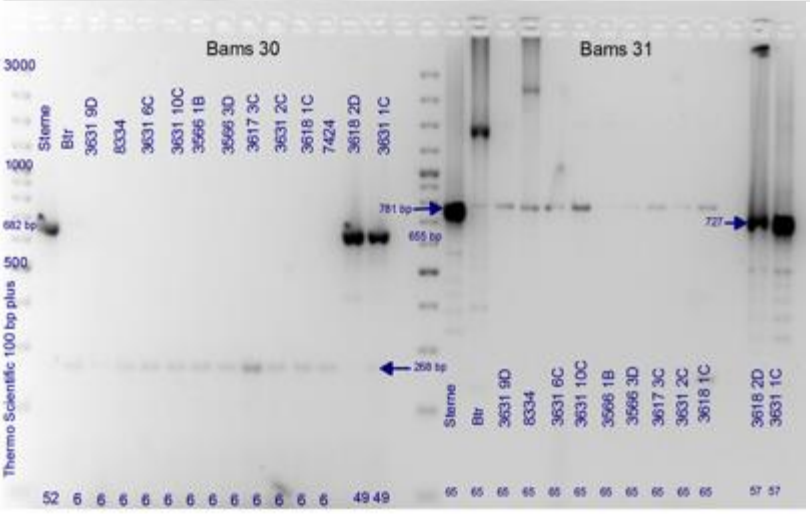
A1 Negative Control	**A2** Dextrin	**A3** D-Maltose	**A4** D-T rehalose	**A5** D-Cellobiose	**A6** Gentiobiose	**A7** Sucrose	**A8** D-T uranose	**A9** Stachyose	**A10** Positive Control	**A11** pH 6	**A12** pH5	**A1** Negative Control
B1 D-Raffinose	B2 α-D-Lactose	B3 D-Melibiose	B4 β-Methyl-D-Glucoside	B5 D-Salicin	B6 N-Acetyl-D-Glucosamine	B7 N-Acetyl-β-D-Mannosamine	B8 N-Acetyl Neuraminic Acid	B9 N-Acetyl Neuraminic Acid	B10 1% NaCl	B11 4% NaCl	B12 8% NaCl	B1 D-Raffinose
C1 α-D-Glucose	C2 D-Mannose	C3 D-Fructose	C4 D-Galactose	C5 3-Methyl Glucose	C6 D-Fucose	C7 L-Fucose	C8 L-Rhamnose	C9 Inosine	C10 1% Sodium Lactate	C11 Fusidic Acid	C12 D-Serine	C1 α-D-Glucose
D1 D-Sorbitol	D2 D-Mannitol	D3 D-Arabitol	D4 myo-Inositol	D5 Glycerol	D6 D-Glucose- 6-PO4	D7 D-Fructose- 6-PO4	D8 D-Aspartic Acid	D9 D-Serine	D10 Troleandomycin	D11 Rifamycin SV	D12 Minocycline	D1 D-Sorbitol
E1 Gelatin	E2 Glycyl-L-Proline	E3 L-Alanine	E4 L-Arginine	E5 L-Aspartic Acid	E6 L-Glutamic Acid	E7 L-Histidine	E8 L-Pyroglutamic Acid	E9 L-Serine	E10 Lincomycin	E11 Guanidine HCl	E12 Niaproof 4	E1 Gelatin
F1 Pectin	F2 D-Galacturonic Acid	F3 L-Galactonic Acid Lactone	F4 4 D-Gluconic Acid	F5 D-Glucuronic Acid	F6 Glucuronamide	F7 Mucic Acid	F8 Quinic Acid	F9 D-Saccharic Acid	F10 Vancomycin	F11 Tetrazolium Violet	F12 Tetrazolium Blue	F1 Pectin
G1 p -Hydroxy-Phenylacetic Acid	G2 Methyl Pyruvate	G3 D-Lactic Acid Methyl Ester	G4 L-Lactic Acid	G5 Citric Acid	G6 α-Keto-Glutaric Acid	G7 D-Malic Acid	G8 L-Malic Acid	G9 Bromo-Succinic Acid	G10 Nalidixic Acid	G11 Lithium Chloride	G12 Potassium Tellurite	G1 p -Hydroxy-Phenylacetic Acid
H1 Tween 40	H2 γ-Amino-Butyric Acid	H3 α Hydroxy-Butyric Acid	H4 β-Hydroxy-D,L- Butyric Acid	H5 α-keto-butyric Acid	H6 Acetoacetic Acid	H7 Propionic Acid	H8 Acetic Acid	H9 Formic Acid	H10 Aztreonam	H11 Sodium Butyrate	H12 Sodium Bromate	H1 Tween 40



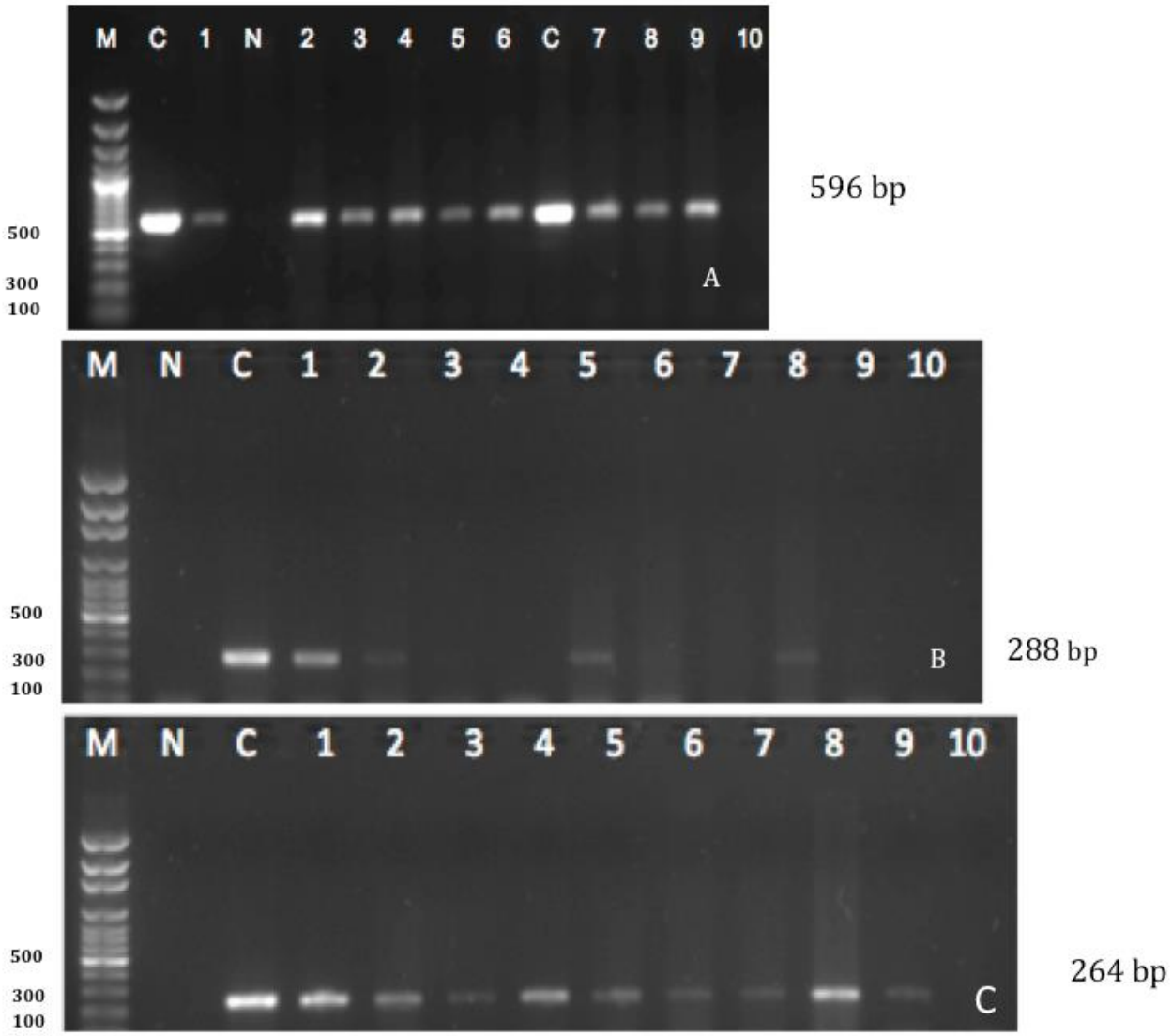
Supplementary Figure 3S1. The variable number of tandem repeats (VNTRs) PCR products (of markers Bams 1-22) of *Bacillus* species on 3 % agarose. The 100 bp plus ladder (Thermo Scientific) and positive control *B. anthracis* Sterne strain was used together with *B. anthracis* isolates.



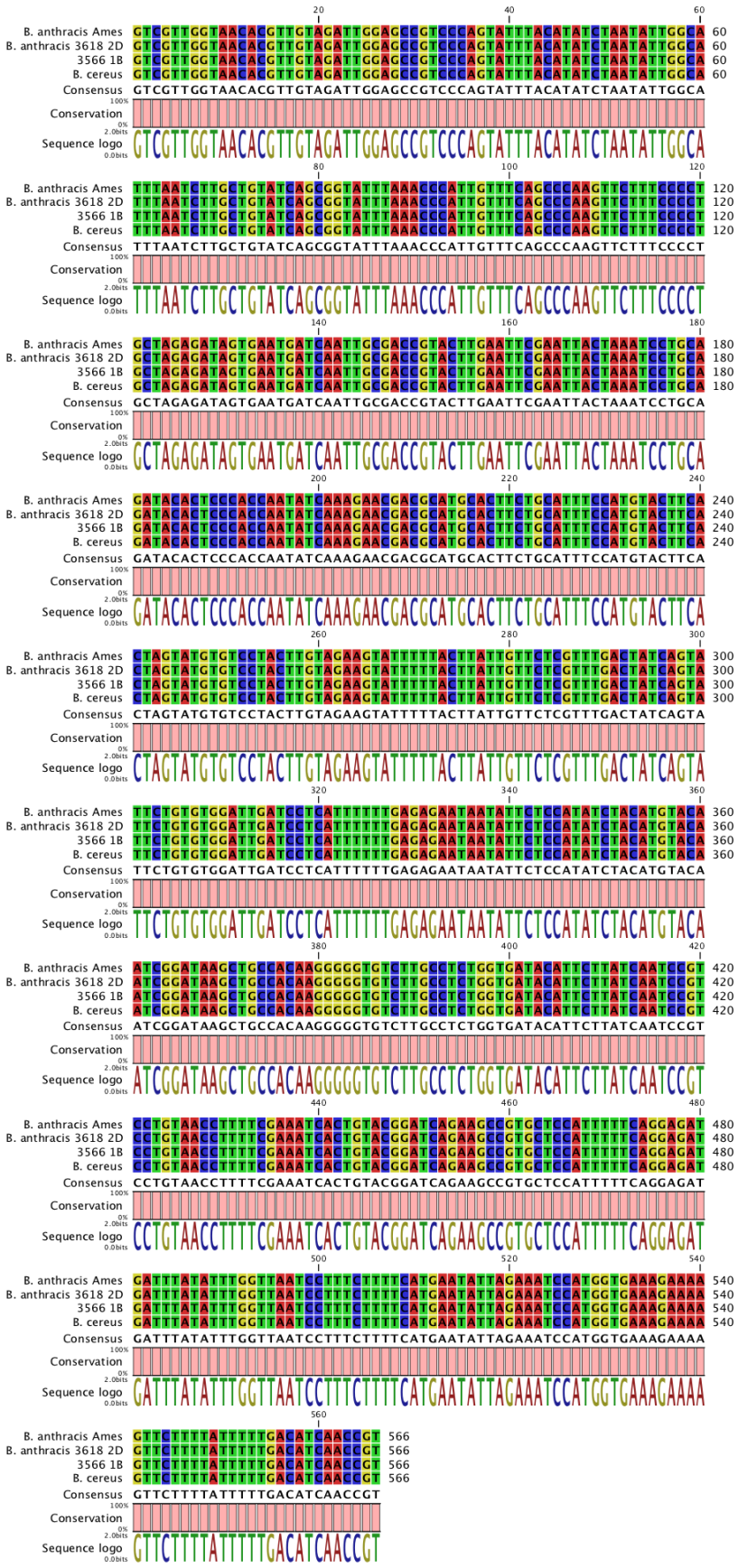




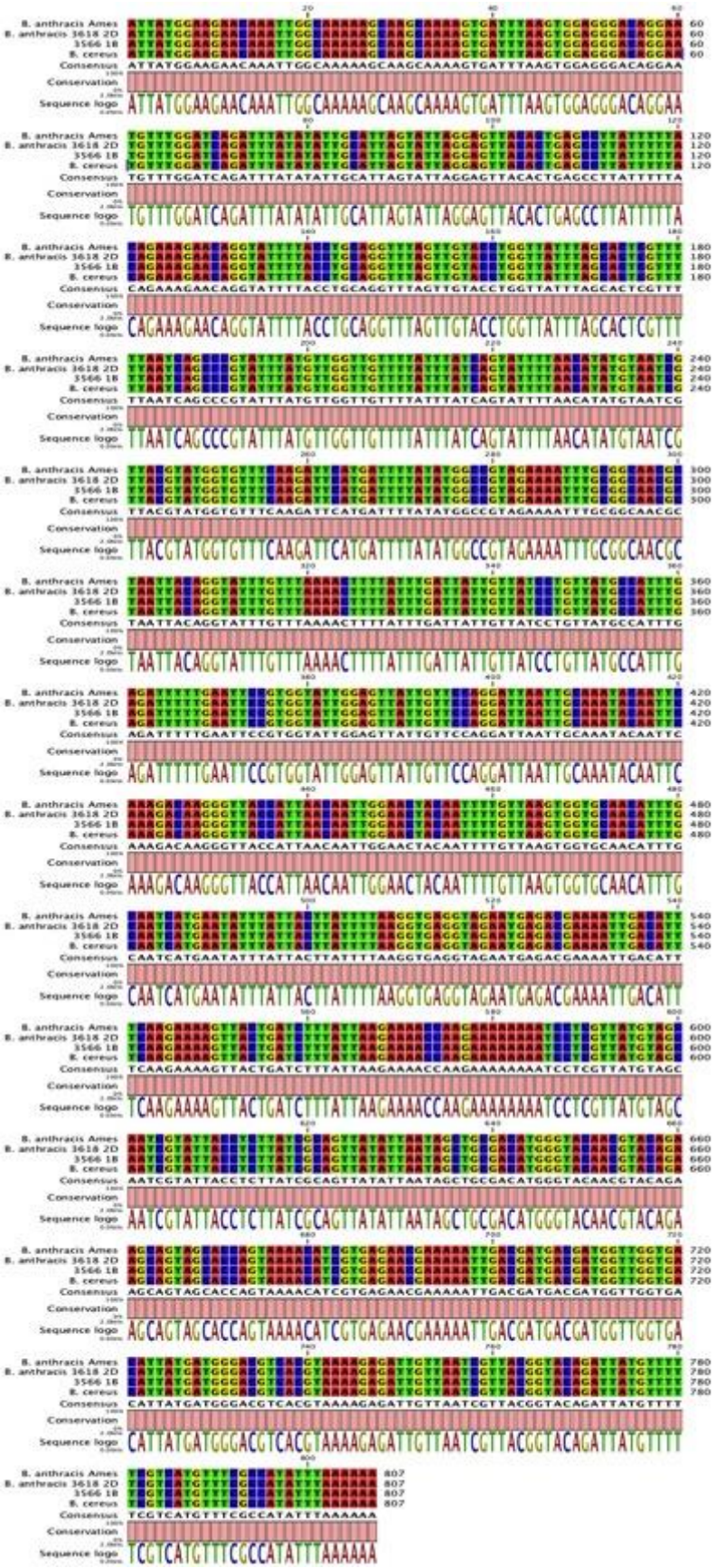
Supplementary Figure 3S2. The variable number of tandem repeats (VNTR) markers (Bams 23-31) for *B. anthracis* and *Bacillus* isolates on 3% agarose gel. The 100 bp plus ladder (Thermo Scientific) and positive control *B. anthracis* Sterne strain was used together with *B. anthracis* isolates as well as *Bacillus* isolates.



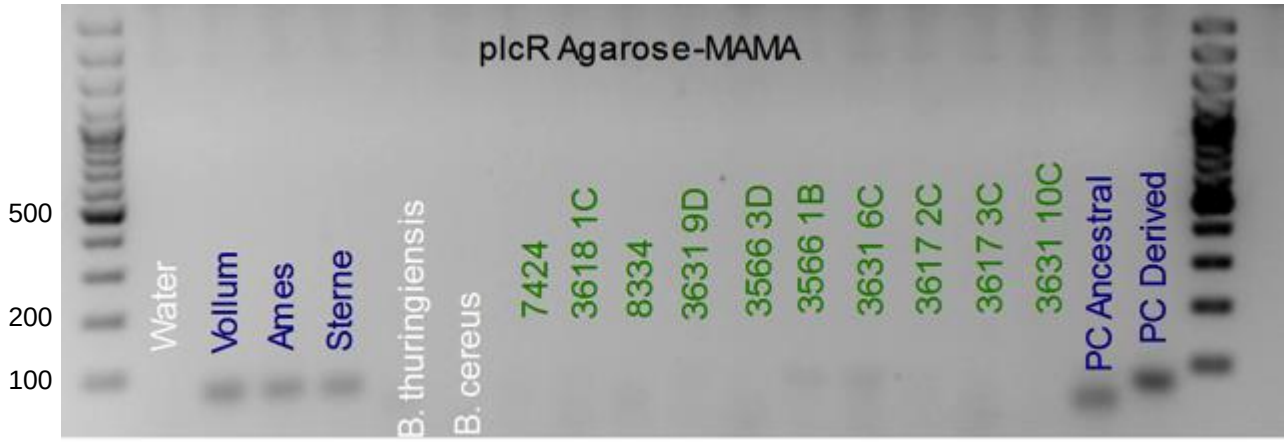
Supplementary Figure 3S3. Conventional PCR products of pag (A), capA (B), and capC (C) gene regions for *Bacillus anthracis* isolates. The 100 bp plus ladder (Fermentas), positive control *B. anthracis* Ames (Indicated as C) and negative control *B. anthracis* Sterne strain (lacks pXO2; indicated as N) were used together with *Bacillus* isolates (lane 1-10). Lane 1-10 consist of isolates 3631-9D, 3631-6C, 3631-10C, 8334, 3617-2C, 3617-3C, 3566-3D, 3618-1C, 3566-1B, and 7424.



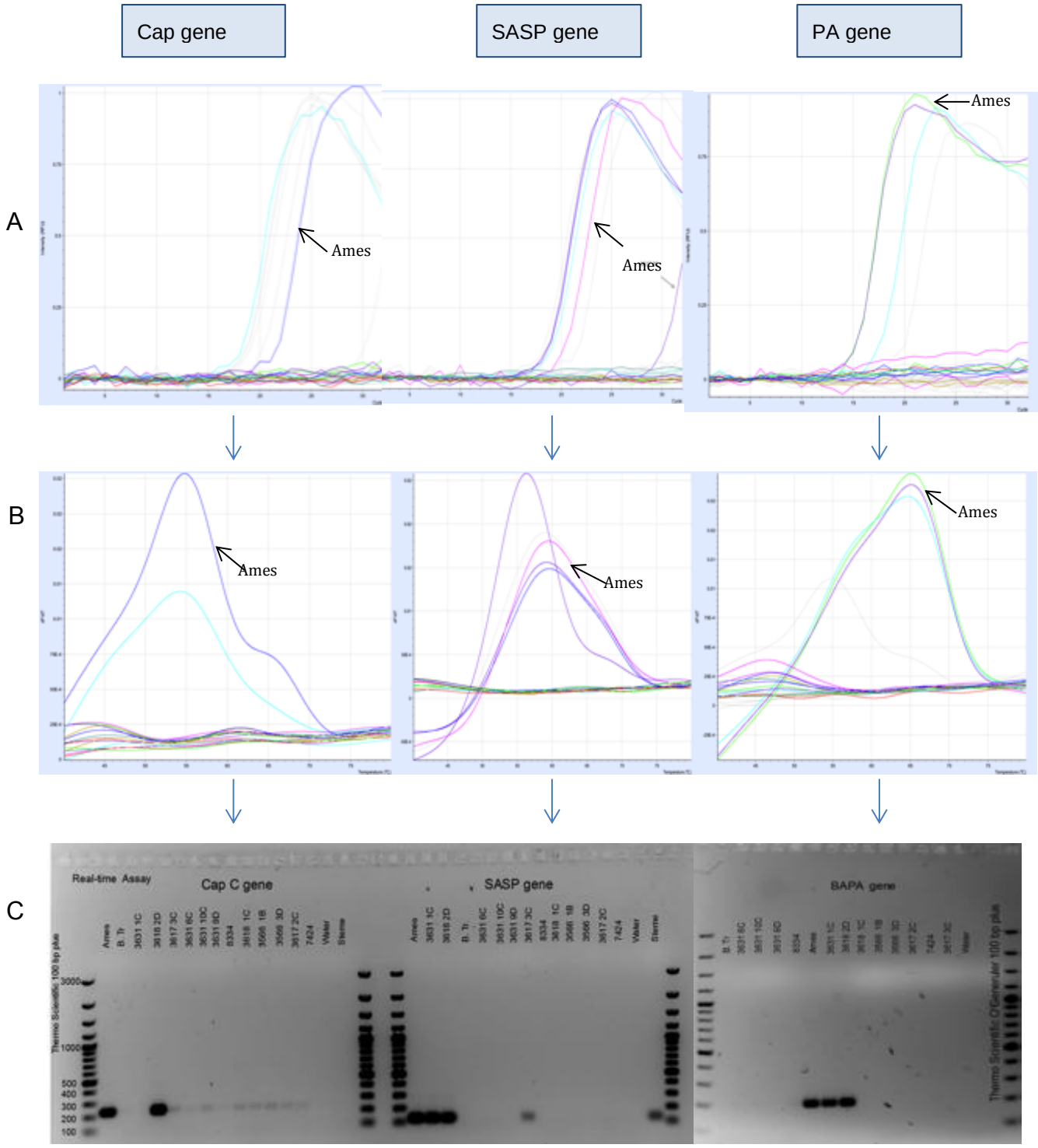
Supplementary Figure 3S4. *Bacillus anthracis* Ames, 3618-2D, *B. cereus*, *Bacillus* 3566-1B protective antigen (*pag*A) gene alignment using CLC Genomic workbench. *Bacillus cereus* (OVAH) was included for sequencing and alignment. No difference in the available sequence of selected strains are observed.



Supplementary Figure 3S5. *Bacillus anthracis* Ames, 3618-2D, *B. cereus*, *Bacillus* 3566-1B protective antigen (*cap*B/C) gene alignment using CLC Genomic workbench. No difference in the gene between the selected strains.



Supplementary Figure 3S6. Conventional Melt-MAMA PCR products using the pleiotropic transcriptional regulator (PlcR) marker for the *Bacillus* species on agarose. The 100 bp plus generuler ladder (Thermo Scientific) was used. Positive controls *Bacillus anthracis* Vollum, Ames, and Sterne amplified the same PCR size product as the positive (PC) derived control as indicated by Birdsell *et al*. (2012) with negative controls *B. thuriengiensis* and *B. cereus*. All *Bacillus* species did not amplify the ancestral or derived PC PlcR region.



Supplementary Figure 3S7. Real-time PCR of the chromosomal (SASP), pXO1 (BAPA) and pXO2 (*cap*C). (A) Cycle threshold, (B) melting temperature points and, (C) PCR products of *Bacillus anthracis* and *Bacillus* isolates on 1.5 % agarose gel. *B. anthracis* Ames (indicated by black arrows) was used as a positive control, and *B. thuringiensis* as negative control.

Chapter 4

WHOLE GENOME SEQUENCING AND GENETIC VARIANT ANALYSIS OF TWO SOUTH AFRICAN *BACILLUS ANTHRACIS* STRAINS

K.E. Lekota[1,2,3], J. Mafofo[1] , E. Madoroba[3], H. van Heerden[2], J. Rees[1], F.C. Muchadeyi[1]

[1] The Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort 0110, South Africa
[2] University of Pretoria, Department of Veterinary Tropical Diseases, Private Bag X4, Onderstepoort, 0110, South Africa
[3] Agricultural Research Council-Onderstepoort Veterinary Institute, Private Bag X5, Onderstepoort 0110, South Africa

Abstract

Bacillus anthracis is the causative agent of anthrax. It is found in the soil and can infect most mammals including humans. In South Africa, anthrax occurs sporadic and is endemic in the Northern Cape region (NCP) and Kruger National Park (KNP). Whole genome characterization through sequencing is becoming more relevant in diagnostics monitoring and control of anthrax outbreaks since it facilitates the identification of the infecting strains and virulent factors. In this study, a strain from the endemic Northern Cape Province (3631-1C) and another (20SD) from a sporadic outbreak in the Mpumalanga Province in South Africa were sequenced. These two South Africa strain were compared with each other and with the *B. anthracis* Ames ancestor using single nucleotide polymorphism (SNPs), insertion and deletions (INDELs), virulent genes/ regions and functional comparison of metabolites. The genome sequencing was performed using the Illumina HIScan SQ platform that generated approximately 2.0 and 3.0 million reads for *B. anthracis* 20SD and 3631-1C strains respectively. Approximately 99% of the reads generated for the 20SD strain mapped to the Ames ancestor reference chromosome and virulence plasmids pXO1 and pXO2. The 3631-1C strain lacks the plasmid pXO2. A total of 458 SNPs and 149 INDELs were identified in strain 3631-1C, while 403 SNPS and 126 INDELS were detected in strain 20SD. High numbers of prophages regions were represented on strains *B. anthracis* A0442 (46), KrugerB (42), 3631 1C (7), 20SD (6) and Ames (5) respectively. The study showed that the *B. anthracis* strains have highly conserved similar regions but differed in the numbers of SNPs and INDELs and

prophage regions. The four-prophage lambdas (01-04) are shared amongst the *B. anthracis* strains.

4.1. Introduction

Anthrax is a notoriously acute disease that affects mostly herbivorous animals but also domesticated and wild animals including humans (Turnbull, 1998; Beyer and Turnbull, 2009). South Africa has experienced many anthrax epidemics mostly in the Northern Cape Province (NCP) and Kruger National Park (KNP) regions. Outbreaks occurs in the two endemic regions, like the large anthrax outbreak in 2008 and 2009 on the escarpment of the Ghaap Plataea, NCP involving goats, sheep, horses, cattle and thousands of game including kudu, springbok and roan antelope.

The Agricultural Research Council-Onderstepoort Veterinary Institute (ARC-OVI) is a reference laboratory in South Africa responsible for receiving suspected anthrax samples from animal for diagnosis purposes. ARC-OVI has received various samples from anthrax outbreaks with special reference to samples from outbreaks at Klipfontein farm in the Northern Cape province (NCP) during an anthrax outbreak in 2009, and a virulent strain used in a challenge animal study (Ndumnego, 2012) originally isolated from a sporadic case in Standerton, Mpumalanga Province (MP). Currently there are methods of diagnosing anthrax based on microbiological and molecular techniques. Microbiological test includes confirmatory assay such as capsule visualization, motility, hemolysis, gamma-phage and penicillin susceptibility (Turnbull, 1998; Fasenella *et al*., 2010). Traditional methods such as classical microbiology tests give rise to equivocal results for anthrax genotyping because they are unable to detect small genomic differences (Fasenella *et al*., 2010). Molecular routine methods involve the use of PCR assay targeting chromosomal regions and virulence plasmid genes (Rammise *et al*., 1996). Little diversity is reported worldwide among *B. anthracis* isolates. Therefore numerous techniques have recently been developed for genetic characterization of *B. anthracis*.

Multi-locus sequence typing (MLST), multi-locus variable number of tandem repeat (VNTR) analysis (MLVA) and canonical single nucleotide polymorphism (canSNP)

have been developed to analyze genetic variation and relatedness within *B. anthracis* (Helgason *et al*., 2004; Keim *et al*., 2004; Pearson *et al*., 2004). MLST does not sufficiently discriminate amongst *B. anthracis* isolates (Helgason *et al*., 2004). The development of MLVA and canSNPs enhanced the identification of genetic markers that established phylogenetic relationship amongst *B. anthracis* isolates (Van Ert *et al*., 2007; Marson *et al*., 2011). Cost-effective mismatch amplification mutation assay (MELT-MAMA) SNP genotyping can be used to replace expensive canSNP for *B. anthracis* (Birdsell *et al.,* 2012).

MLVA and canSNP (or MELT-MAMA) can distinguish the three major lineages of *B. anthracis* (Clade A, B and C) (Van Ert et al., 2007; Marston *et al*., 2011; Birdsell *et al*., 2012). Clade A is considered the most important one, distributed around the world, and has been associated with over 90 % of reported anthrax cases (Van Ert *et al*., 2007; Keim *et al*., 2009). Clade A is a complex lineage that radiates into multiple closely related and widely dispersed subgroups (Keim *et al*., 2009; Van Ert *et al*., 2007), however the relatedness of the sub-lineages has not been completely resolved. The B clade includes sub-clade B (B. Br KrugerB) strains that are restricted mostly to the Kruger National Park region (Smith *et al*., 2000). Clade C is considered to be the difficult lineage due to its source and rare occurrence (Keim *et al*., 2009; Van Ert *et al*., 2007). South Africa is considered to harbor more *B. anthracis* diversity due to the occurrence of both clade A and B isolates (Smith *et al*., 2000; Van Ert *et al*., 2007). MLVA and canSNP have made it possible to separate closely related lineages within species groups. However the use of a single locus or multiple loci in these methods is still limiting as loci reflect small portions of the genome, and therefore give rise to limited resolution (Koser *et al*., 2012; Segerman *et al*., 2011).

Molecular typing of *B. anthracis* strains is imperative for epidemiological investigations of natural outbreaks. Genetic characterization helps with the identification of particular variants, to trace the animals' transmission events of the bacterium and the classification of virulence and antibiotic resistance genes that might be harbored amongst the bacterial strains. Sequencing bacterial genomes by next generation sequencing (NGS) technologies has recently emerged as a cost effective and convenient approach (Didelot *et al*., 2012). Whole genome sequencing (WGS) provides more promise in diagnostic microbiological practice, by allowing

whole genome variability analysis within and between bacterial species and subspecies groups (Koser *et al*., 2012; Segerman *et al*., 2011). The genome sequence of *B. anthracis* Ames was compared with closely related bacteria and the pXO1 and pXO2 gene sequences were found to be more variable between the strains, suggesting plasmid mobility amongst the strains studied (Read *et al.,* 2003).

More recent WGS indicated that isolate CDC 684 from Centres of Disease Control (CDC) initially identified as *B. megaterium* that was reidentified as *B. anthracis* contains inversions and is a close relative of Vollum strain (Okinika *et al.,* 2011). WGS of *B. anthracis* H9401 from a Korean patient with anthrax shows high pathogenicity and genome sequences similar to Ames ancestor (Chun *et al.,* 2012). Vollum and Ames ancestor group into different subgroups in the A clade of *B. anthracis*. Whole genome sequencing studies indicated approximately 3500 SNPs and 12 canSNPs that can be used as binary markers to define subgroups within the *B. anthracis* species (Read *et al*., 2002; Pearson *et al*., 2004).

Results from the analysis of isolates from the South African outbreaks in the Northern Cape and MP regions (Chapter 3) called for further investigations of the genomic structure of the *B. anthracis* to determine the unique genetic elements in these isolates, analyze diversity within this group and possibly build an understanding and review of the current diagnostic protocols. The genetic features of an endemic NCP strain 3631-1C and a 20SD isolate from a sporadic outbreak in Standerton area in MP South Africa, were therefore characterized in relation to *B. anthracis* Ames ancestor and other South African *B. anthracis* genome sequences (KrugerB and A0442) that were from KNP. Hence, the aims of this study were to (i) complete genome sequences of South African *B. anthracis* strains from the A clade sampled from a recent outbreak in the NCP and MP regions of South Africa; (ii) determine the genetic variations amongst the South African *B. anthracis* strains; and (iii) compare the two *B. anthracis* strains to a *B. anthracis* Ames ancestor, which was considered a reference genome in this study and two other South African isolates (KrugerB and A0442) that had been sequenced in previous studies

4.2. Materials and methods

4.2.1 *Bacillus anthracis* strains

The *B. anthracis* isolates (3631-1C and 20SD) from NCP and MP were characterized by whole genome sequencing. Isolate 3631-1C was isolated from a kudu that died on the Klipfontein farm (Ghaap area) in the NCP in 2009, and the 20SD isolate was isolated from a sheep that died of anthrax in 2001 in the Standerton area, MP. The reference laboratory at the Agricultural Research Council, Onderstepoort Veterinary Institute (ARC-OVI), Onderstepoort, South Africa diagnosed these isolates as *B. anthracis*. They were shown to be Gram positive, non-motile, non-hemolytic, capsulated, sensitive to penicillin and γ-phages (Chapter 3, session 3.3.1). The Department of Veterinary Tropical Diseases confirmed the identification as *B. anthracis* using TaqMan real-time PCR that targets the plasmids and the small acid soluble protein (SASP) (Hoffmaster *et al*., 2002). The real-time PCR indicated that isolate 3631-1C lacked plasmid pXO2.

4.2.2. DNA extraction

The *B. anthracis* strains (20SD and 3631-1C) were inoculated in 2 ml nutrient broth, followed by overnight incubation at 37 °C. The bacterial cells were then harvested by centrifugation at 5000 ×g for 10 min. Genomic DNA was extracted using the DNAeasy Tissue kit (Qiagen, Germany) according to the manufacturer's instructions except that the DNA was suspended in 100 µl of Buffer AE instead of 200 µl. The DNA was quantified with the Qubit® fluorometric quantization method (Life Technologies, USA) using the broad range assay kit according to the manufacturer's instructions.

4.2.3. Library preparations for sequencing

Preparation of the shotgun paired-end sequencing libraries was performed using the Nextera DNA Sample Prep Kit (Illumina-compatible, Epicentre Biotechnologies, USA) protocol with alterations. Genomic DNA was fragmented to generate fragments with a size range of 100-300 bp using the tagmentation approach. Approximately

27.5 ng of total genomic DNA was incubated for 5 minutes at 55 °C in the tagmentation reaction. The fragmented genomic DNA was purified using the QIAGEN MiniElute PCR purification kit (Qiagen, Germany). Limited-cycle PCR (9 cycles) in a 25 µl reaction was prepared using 5 µl of purified tagmentation reaction, 1X Nextera PCR buffer, 2X Nextera primer cocktail, 2X Nextera adaptor 2 and Nextera PCR enzyme. The library was then purified using QIAGEN MiniElute PCR purification kit (Qiagen, Germany) according to the manufacturer's instructions. The sequencing library assessed for size distribution through electrophoresis on a 2% agarose gel and quantified using the Qubit ® fluorometric method (Life Technologies, USA).

4.2.4. Whole genome sequencing (WGS)

The sequencing library was diluted to 2.5 pM and hybridized to a flow cell. Clusters were generated on the flow cell using the Illumina TruSeq™ paired end cluster kit v2-cBot-HS (Illumina, USA). Sequencing of the paired-end library was performed on the Illumina HiScan SQ sequencer using the 200-cycle SBS (sequencing by synthesis) sequencing v3 kit (Illumina, USA). The 5.38 kbp coliphage phiX-174 was used as a sequencing positive control. The whole genome sequencing and genetic variation pipeline is shown in Figure 4.1.

4.2.5. Quality and Trimming

Sequencing adapters were then trimmed using the Nextera adapter index in the CLC-bio Workbench 6.0 (Aarhus, Denmark) using the following parameters: quality scores limit 0.05, and maximum number of ambiguities set to 2 per 100 nucleotide sequence read. The sequence data was analyzed using the FastQC software version 0:10.1 (Andrews, Brabraham Bioinformatics, 2011) to assess the quality metrics.

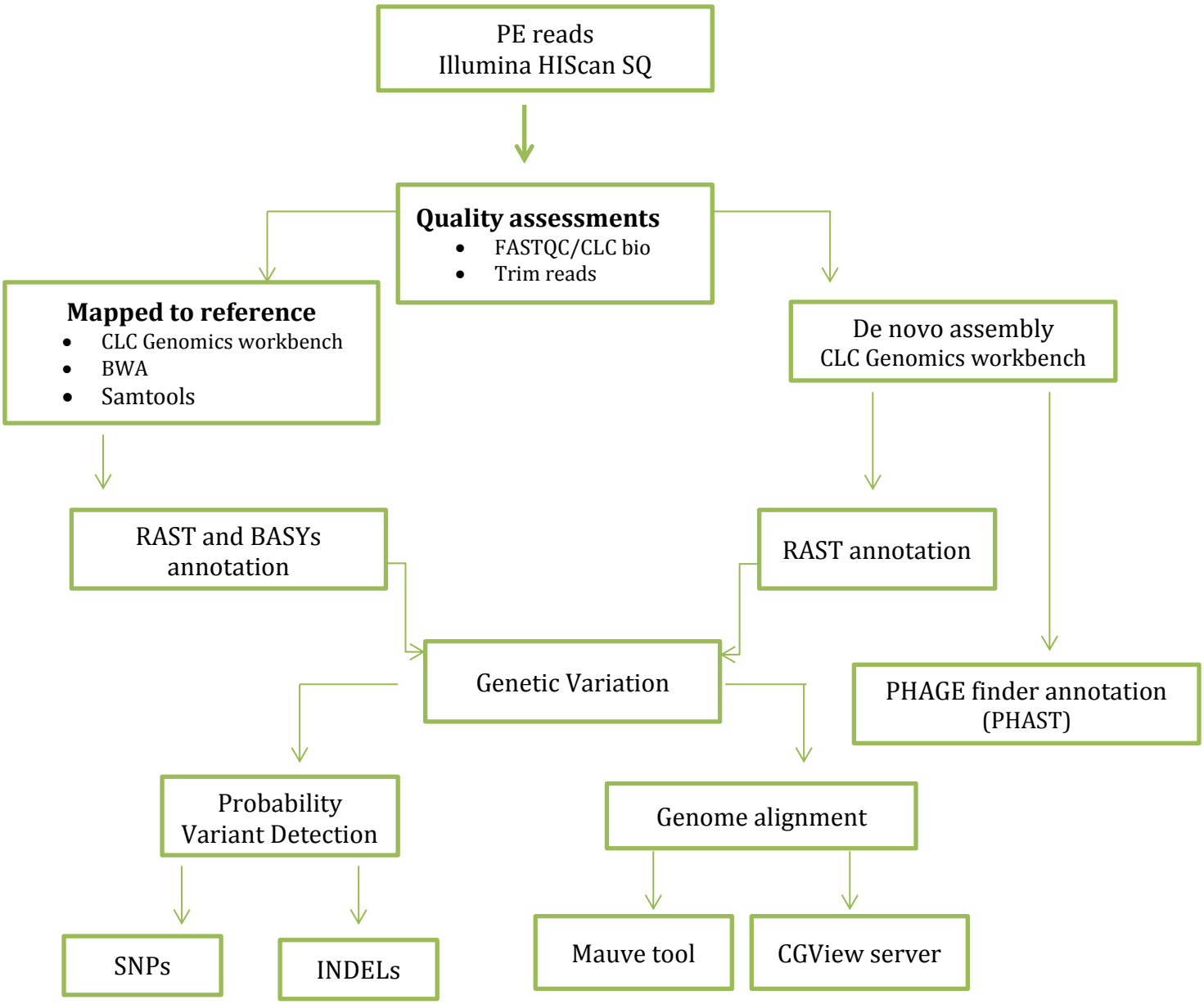


Figure 4.1. An overview of whole genome sequencing (WGS) and genetic variation analysis pipeline using next generation sequencing (NGS) platform. PE paired end; RAST rapid annotation using subsystem technology; SNPs: Single nucleotide polymorphism, INDELs: Insertion and deletions.

4.2.6. Mapping to the reference

The sequenced reads were mapped to *B. anthracis* Ames ancestor (reference chromosome; accession AE017334) plasmid pXO1 (accession AE017336) and pXO2 (accession AE017335) using CLC-bio Genomic Workbench 6.0, BWA (Li and Durbin, 2009) and Samtools (Li *et al*., 2009). The sequence reads were then aligned to the reference genome with mismatch cost, insertion cost, deletion cost, length fraction, and similarity fraction parameters set to 2, 3, 3, 0 and 0.8 respectively.

4.2.7. Analysis of unique regions from the un-mapped reads

De novo assembly was performed on all of the reads that did not map to either the chromosome or the two plasmids of the reference genome. This was performed using the CLC Genomics Workbench 6.0 (Aarhus, Denmark) with the mismatch cost, insertion cost, deletion cost, minimum paired distance, maximum paired distance parameters set to 3, 3, 3, 50 bp and 800 bp respectively. Contigs generated from assembly were further analyzed by using BLASTN on CLC Genomic Workbench 6.0.

4.2.8. *De novo* assembly of the *B. anthracis* strains

The *de novo* assemblies of sequence data for the two *B. anthracis* strains were performed using the CLC Genomic Workbench 6.0 with the mismatch cost, insertion cost, deletion cost, length fraction, similarity fraction parameters set to 2, 3, 3, 0.5, 0.8 respectively.

4.2.9. Annotation of strains

The assembled contigs from the reference mapping and *de novo* assembly were subjected to online tools RAST (Aziz *et al*., 2008) and BASys (Van Domselaar *et al*., 2005). The assembled contigs of the *B. anthracis* strains were submitted to the RAST annotation server for subsystems classification and functional annotation.

4.2.10. Genetic variant analysis

The CLC Genomic Workbench 6.0 was used to detect SNPs and INDELS using the probability variant detection program with the following parameters: minimum coverage =10; variant probability 90 and; maximum expected variants=2. Mauve tool (Darling *et al*., 2004) was used to align the genome assemblies of the South African strains (*B. anthracis* 20SD, 3631-1C, KrugerB (accession AAEQ00000000)) and A0442 (accession NZ_ABKG00000000) against *B. anthracis* Ames ancestor. CGView (Grant and Stothard, 2008) was used for BLASTN comparison of the *B. anthracis* Ames ancestor against the two South African strains. PHAST (Zhou *et al*., 2011) was used to predict the prophage regions on the *B. anthracis* strains, which uses the BLAST for homology search of the regions using Glimmer 3.02 amongst available sequences on the database.

4.3. Results and Discussion

4.3.1. Sequencing and mapping to the reference

The genome of Northern Cape *B. anthracis* 3631-1C strain was sequenced from paired-end shotgun libraries. About 3 050 254 sequence reads were generated at an average length of 90 bp after trimming. The sequence data generated for strain 3631-1C was about 98% in relative to *B. anthracis* Ames ancestor (Table 4.1). The mapping assembly gave a chromosome of approximately 5.22 Mb, relative to the 5.22 Mb of *B. anthracis* Ames ancestor (Ravel *et al.,* 2009), with a 53-fold sequence coverage. The pXO1 was of 181 623 bp long with a 127-fold coverage. The second plasmid pXO2 (Table 4.1) was absent, thus confirming the results of real-time PCR (Chapter 3, session 3.3.6). With MLVA analyses, the *B. anthracis* 3631-1C group in the A clade, subclade A. Br 002 which is the same clade for *B. anthracis* Ames (A. Hassim personal communication; Figure 2.8; Chapter 3, session 3.3.7).

About 2 042 788 sequence reads were obtained for *B. anthracis* 20SD, these had an average read length of 90 bp after trimming. About 99% of the sequence data mapped to the *B. anthracis* Ames ancestor (Table 4.1). The mapping report indicated

the presence of a chromosome with a length of 5.22 Mb, a plasmid pXO1 with length 181 677 bp, and plasmid pXO2 that is 94 830 bp long (Table 4.1). The sequence data gave a 30.5-fold coverage of the chromosome, 98-fold coverage of pXO1, and 48.37-fold coverage of pXO2. Representation of the fold coverage level for this strain suggests a pXO1-pXO2-chromosome molecular ratio of 3:2:1 which is identical to *B. anthracis* Ames Florida, *B. anthracis* Ames ancestor and *B. anthracis* H9401 (Read *et al*., 2002; Chun *et al*., 2012).

Table 4.1. Whole genome mapping of the South African *Bacillus anthracis* strains to the *B. anthracis* Ames ancestral reference genome.

Strain	Name	*B. anthracis* Ames (bp)	Consensus length (bp)	Total read count	Average coverage
B. anthracis 20 SD	Chromosome	5 227 419	5 219 377	1 781 282	30.54
	pXO1	181 677	181 362	195 695	97.99
	pXO2	94 830	94 764	50 580	48.37
B. anthracis 3631-1C	Chromosome	5 227 419	5 226 054	2 984 646	53.09
	pXO1	181 677	181 623	244 957	126.59
	pXO2	94 830	1 322	250	0.06

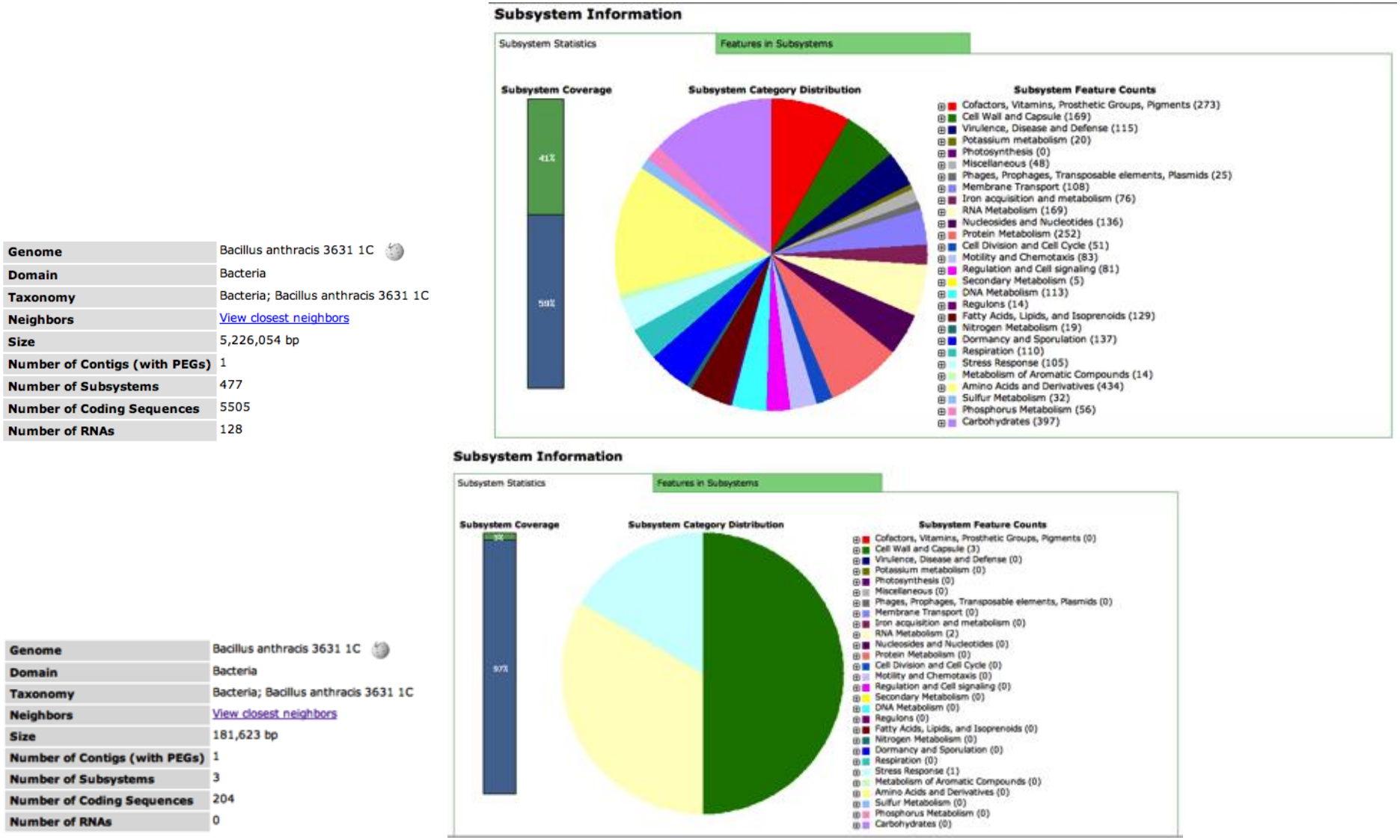
4.3.2. Analysis of unique regions from the un-mapped reads

The *de novo* assembly of unmapped reads revealed the represents of the phage minor proteins that are common to the *B. anthracis* CDC 684 and *B. anthracis* H9401 in both *B. anthracis* strains 3631-1C and 20SD. The occurrence of the phage minor proteins is common to the *B. anthracis* strains (session 4.4.12) since they are sensitive to specific bacteriophages. The genetic mobile elements such as protein phages have been reported in the *B. anthracis* strains representing evidence of horizontal gene transfer (Sozhamannan *et al*., 2006). These regions that came out of the unmapped reads are unique regions that are found in the South African 20SD and 3631-1C strains and were absent in *B. anthracis* Ames ancestor.

4.3.3. Annotation of *de novo* assemblies

The RAST server predicted 204 coding sequences for pXO1 of strain 3631-1C and 5 505 for the chromosome. About 201 coding sequences were predicted for pXO1 of strain 20SD, 117 for pXO2 and 5 494 for the chromosome. Strain 3631-1C had 59% of the chromosome proteins as hypothetical proteins that were not covered in the subsystem. Ninety-seven percent of the pXO1 proteins were also hypothetical (Figure 4.2). Only 43% of the chromosomal proteins of strain 20SD were covered in the RAST subsystem. Of the plasmids proteins, 2% of pXO1, and none of pXO2 were in the subsystems and hence pX02 was not incorporated in the subsystem features (Figure 4.3). The percentages of proteins or coding sequences that were not covered in the subsystems were regarded as hypothetical proteins. Although these proteins were annotated in the subsystems they still remain functionally unidentified. It has been reported that on *B. anthracis* pXO1, 148 out of the 204 (72.5 %) potential coding sequences remain functionally unidentified (Okinaka *et al*., 1999; Rasko *et al*., 2005).

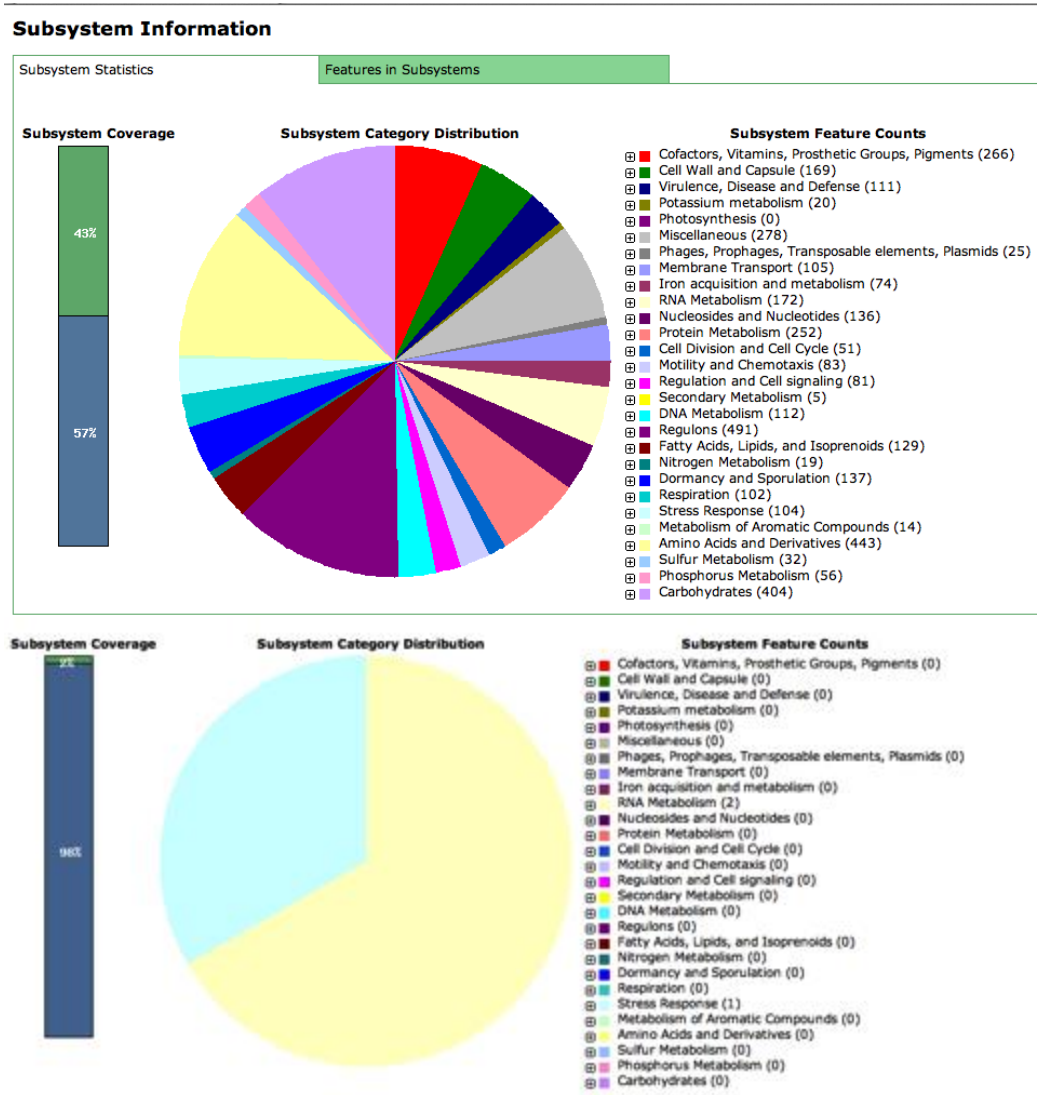
Graphical representation of the *B. anthracis* strains using BASys annotation server (Supplementary Figure 4S1 and 4S2) shows their genome sizes and predicted annotated genes on cluster of orthologous groups of proteins (COG) functional classification. BASys predicted about 199 coding sequences on the plasmid XO1 and 5 561 on the chromosome of *B. anthracis* 3631-1C whereas *B. anthracis* 20SD encoded 5 570, 196 and 104 predicted coding sequences on the replicons chromosome, pXO1 and pXO2 respectively.



Genome	Bacillus anthracis 3631 1C
Domain	Bacteria
Taxonomy	Bacteria; Bacillus anthracis 3631 1C
Neighbors	View closest neighbors
Size	5,226,054 bp
Number of Contigs (with PEGs)	1
Number of Subsystems	477
Number of Coding Sequences	5505
Number of RNAs	128

Genome	Bacillus anthracis 3631 1C
Domain	Bacteria
Taxonomy	Bacteria; Bacillus anthracis 3631 1C
Neighbors	View closest neighbors
Size	181,623 bp
Number of Contigs (with PEGs)	1
Number of Subsystems	3
Number of Coding Sequences	204
Number of RNAs	0

Figure 4.2. RAST annotation of *Bacillus anthracis* 3631-1C strain chromosome and plasmid pXO1 showing the subsystem coverage of the coding sequences and the features covered.



Genome	Bacillus anthracs 20SD
Domain	Bacteria
Taxonomy	Bacteria; Bacillus anthracs 20SD
Neighbors	View closest neighbors
Size	5,219,377 bp
Number of Contigs (with PEGs)	1
Number of Subsystems	470
Number of Coding Sequences	5492
Number of RNAs	128

Genome	Bacillus anthracis
Domain	Bacteria
Taxonomy	Bacteria; Bacillus anthracis
Neighbors	View closest neighbors
Size	181,362 bp
Number of Contigs (with PEGs)	1
Number of Subsystems	3
Number of Coding Sequences	201
Number of RNAs	0

Figure 4.3. RAST annotation of *Bacillus anthracis* 20SD strain chromosome and plasmid pXO1 showing the subsystem coverage of the coding sequences and the features covered.

4.3.4. *De novo* assembly and genetic sequence features

The genomic features of the two South African strains are indicated in Table 4.2. The nucleotide composition of the genomes for the two strains was highly biased towards AT (adenine and thymine), with low GC (guanine and cytosine) of approximately 35% which is a general feature for the *B. cereus* group (Rasko *et al*., 2005). The number of contigs generated for *B. anthracis* 3631-1C and *B. anthracis* 20SD was 214 and 209 respectively. The N50 representing the smallest contig length in 50% of the total contigs in the genome was high on both strains indicating a good assembly of the genome. The annotation system of RAST predicted the total number of coding sequences of *B. anthracis* 3631-1C to be 5 656 compared to be 5 714 of *B. anthracis* 20SD, which was expected since 3631-1C lacked pXO2.

Table 4.2. Genome features of South African *Bacillus anthracis* strains using CLC Genomics workbench and RAST software (Aziz *et al*., 2008).

Genomic features	***B. anthracis* 3631-1C**	***B. anthracis* 20SD**
N75	48 240	31 173
N50	100 222	59 587
N25	207 952	85 743
Minimum contig	191	911
Maximum contig	551 809	183 951
Average contig	25 134	25 983
Number of contigs	214	209
GC content	35%	35%
Number of coding sequences	5 656	5714
Number of RNAs	70	49
Sequence coverage	54	33
Genome coverage	98%	99%
Total base pairs	5 378 709	5 430 511

4.3.5. Probability variant detection

Bacillus anthracis is one of the genetically monomorphic organisms and genetic variation can be analyzed at the genomic level by using the SNPs (Read *et al*., 2002; Achtman *et al*., 2008). About 458 SNPs and 149 INDELs were detected in strain 3631-1C and for 20SD 403 SNPS and 126 INDELs were detected (Table 4.3). The distribution of both INDELS and SNPs from both strains was almost similar (Figure 4.3). Amongst the coding SNPs only about 204 (51%) and 229 (50%) were non-synonymous SNPs in strain 20SD and 3631-1C respectively (Table 4.3). In strain 3631-1C, approximately 20 % of SNPs were located in the intergenic regions and with the remaining mixture of coding SNPs. About 41% of the non-synonymous SNPs were transition and 9% were transversion. In the 20SD strain about 19 % of SNPs were located in the intergenic regions. Isolate 20SD had 19% of SNPs located in the intergenic regions. About 42 % of the non-synonymous SNPs were transition and 9% were transversion in strain 20SD.

Table 4.3. Probability variant detection of the South African *Bacillus anthracis* strains showing single nucleotide polymorphism (SNPs) and insertions and deletions (INDELs) using CLC Genomic workbench software.

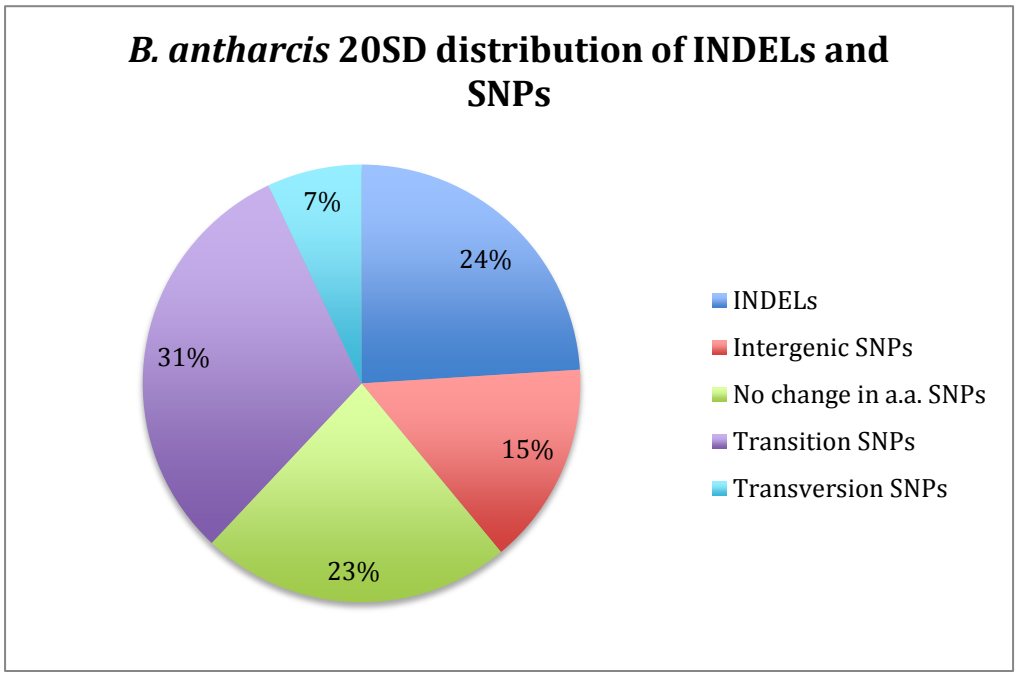
***B. anthracis* strains**	**INDELs**	**Total SNPs**	**Intergenic SNPs**	**Coding SNPs**	**a.a. change SNPs**	**Transition SNPs**	**Transversion SNPs**
20SD	126	403	78	325	204	167	37
3631-1C	149	458	92	366	229	165	54

INDELs indicates Insertion and Deletions

SNPs =Single Nucleotide polymorphism

a.a. = Amino acid

A



B

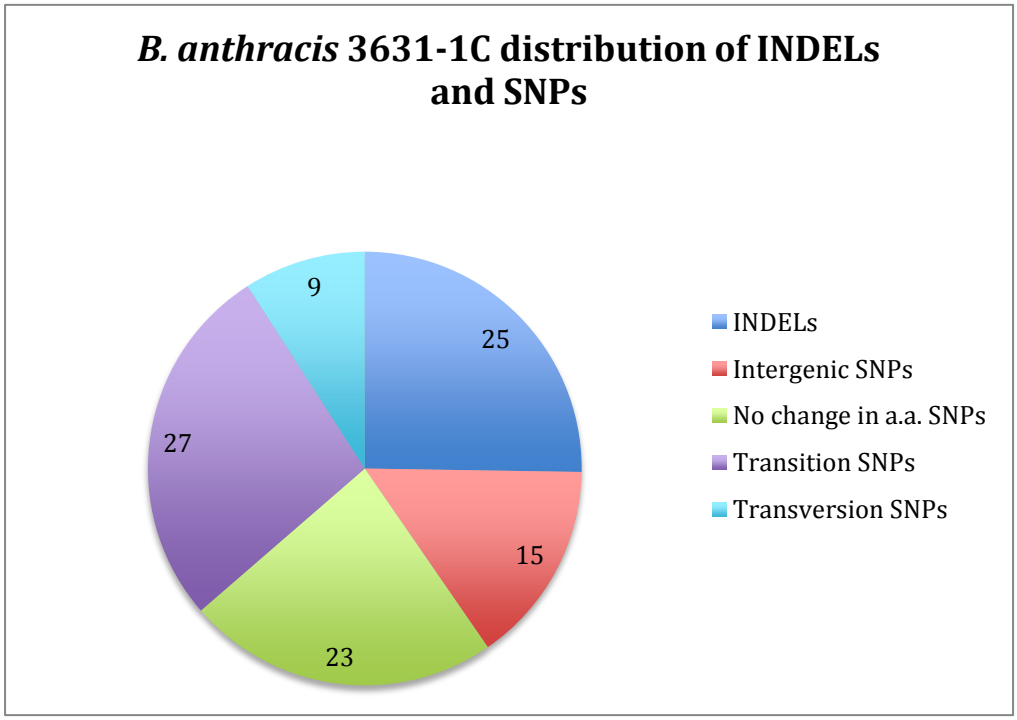


Figure 4.4. Distribution of SNPs and INDELs on *B. anthracis* 20SD (A) and *B. anthracis* 3631-1C (B) relative to *B. anthracis* Ames ancestor using the probability variant detection.

4.3.6. *B. anthracis* plasmids features

In order for *B. anthracis* strains to be virulent, both plasmids pXO1 and pXO2 are required (Okinaka *et al*., 1999; Pannucci *et al*., 2002). Plasmid pXO1 is about 182-kb and pXO1 is 86-kb that consists of genes that encodes for production and regulation of the toxin genes. The pXO2 encodes for the synthesis and degradation of the polyglutamate capsule. Previous studies have shown that many strains of *B. anthracis* especially those isolated from the soil and/or after long term culturing usually lack one or both plasmids (Marston *et al*., 2005; Birdsell *et al*., 2013). The loss of either plasmid pXO1/pXO2 has also been reported in half of the Bulgarian strains that have been collected between 1960 and 1980 (Antwerpen *et al*., 2011). The pathogenicity islands (PAI) on pXO2 include *cap*BCADE that is required for the synthesis of the polyglutamate (PGA) on cell wall (Okinaka *et al*., 1999). The absence of plasmid pXO2 on *B. anthracis* 3631-1C was proven on mapping to the reference genome *B anthracis* Ames ancestor (Table 4.1) and was observed phenotypically when the strain could not produce the capsule layer under induction of sodium bicarbonate on tripticase soy agar (Chapter 3, session 3.3.1). The absence of pXO2 in isolates sampled from an outbreak cannot be fully explained. It could be that the isolate lost its plasmid during culturing or a different genotype without this plasmid, which was picked from the culture leaving behind the representative genotypes. The *B. anthracis* 20SD contained both the pXO1 and pXO2 plasmid replicons. Beyer and Turnbull (2013) indicated that more than one genotype of *B. anthracis* can be found within a culture taken from a single colony from anthrax cases. It is therefore essential for a good microbiological practice in a laboratory to isolate several colonies from the same animal sample.

4.3.7. Plasmids alignment

Comparative genomics of the plasmid pXO1 and pXO2 amongst five *B. anthracis* strains (Ames ancestor, 20SD, 3631-1C, A0442, and KrugerB) are represented in Figure 4.5 and 4.6. The pXO1 and pXO2 of the *B. anthracis* strains presented high levels of similarity with few genomic rearrangements. A few SNPs differences were observed on

strain 20SD and 3631-1C relative to *B. anthracis* Ames ancestor. Previous studies also reports similarities among studied pXO1 of *B. anthracis* strains with only a few small insertions or deletions observed except for the *B. anthracis* Florida isolate that was reported to have two inversions in relation to the *B. anthracis* Sterne genome sequence (Read *et al*., 2002; Rasko *et al*., 2005). The *B. anthracis* KrugerB from the B clade, shows an insertion in the region 161250-161930 bp of pXO1 when analyzed against the *B. anthracis* Ames ancestor (Figure 4.5). Small variations were also noted at non-coding regions of pXO2 20SD and the KrugerB strains relative to *B. anthracis* Ames ancestor (Figure 4.6).

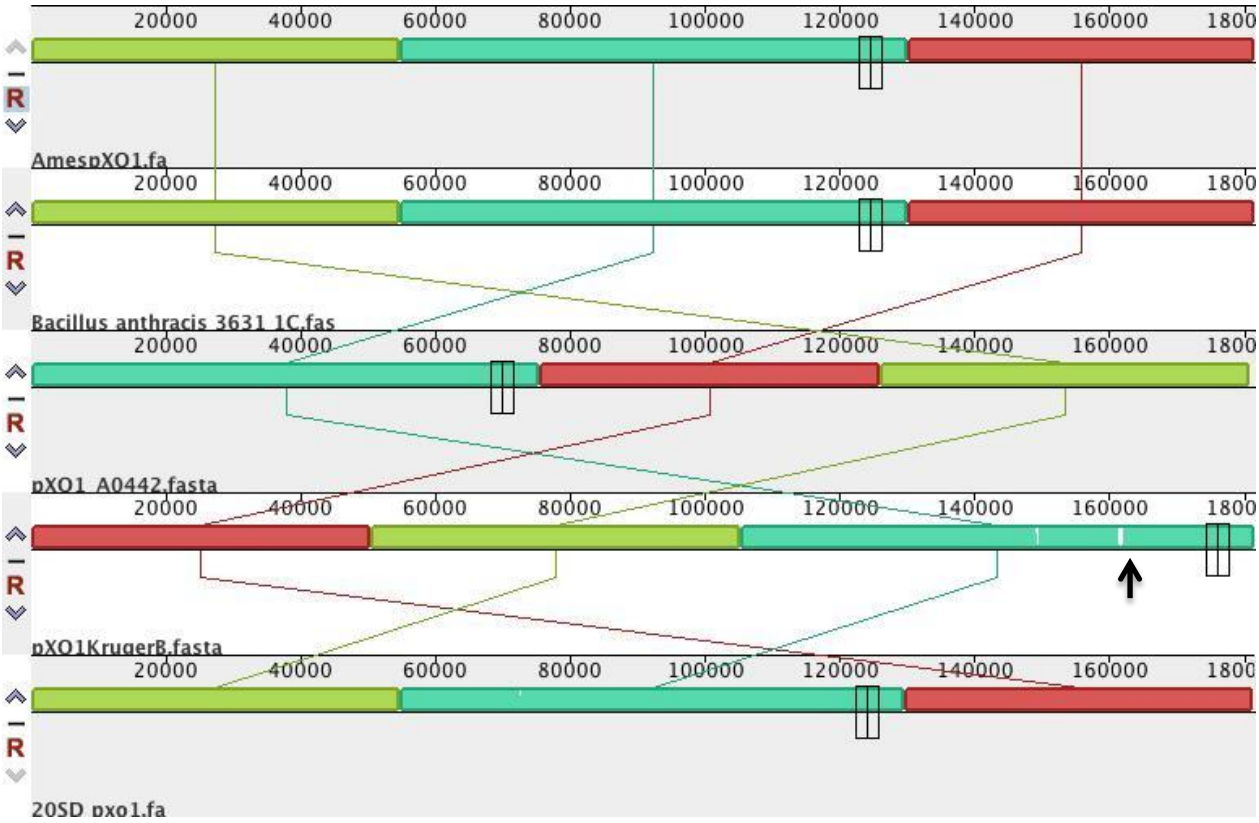


Figure 4.5. Alignment of *Bacillus anthracis* Ames ancestor pXO1 with *B. anthracis* KrugerB, *B. anthracis* A0442, *B. anthracis* 3631-1C and *B. anthracis* 20SD. Each colour block indicates homologous regions of the genome sequences. White areas in colour blocks indicate possible nucleotide sequence absent from the 3631-1C and 20SD strains.

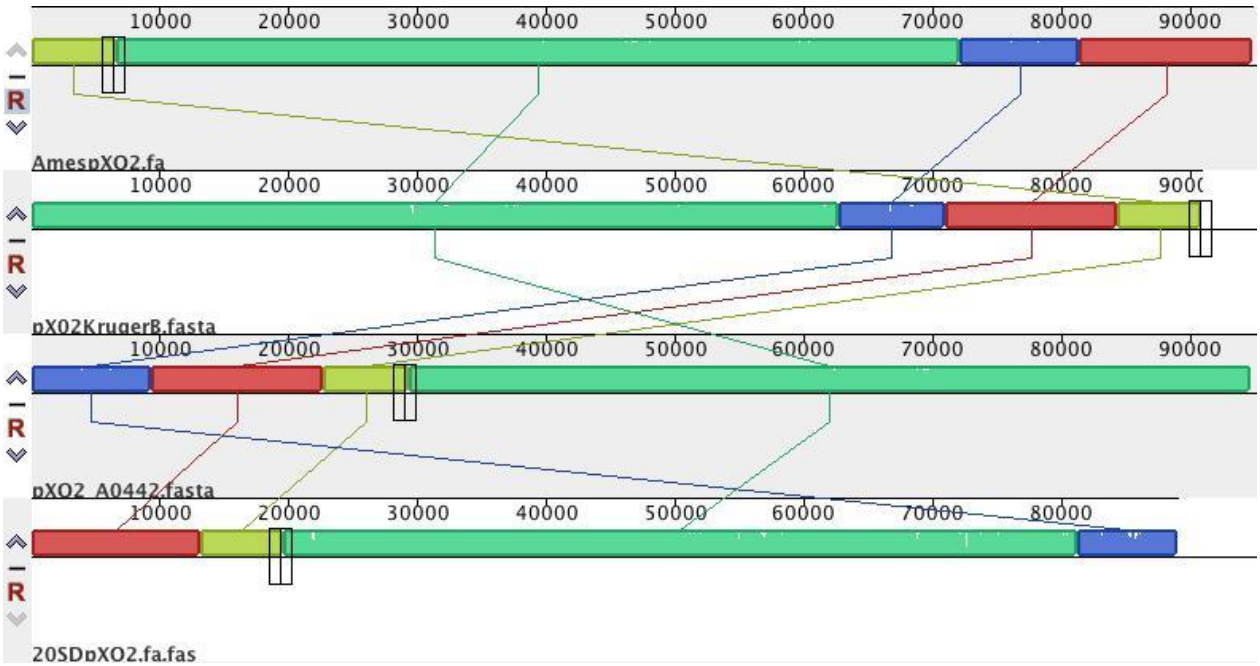


Figure 4.6. Alignment of *Bacillus anthracis* 20SD, A0442, KrugerB pXO2 with *B. anthracis* Ames ancestor. Blocks indicate homologous regions of the genome sequences. White areas in colour blocks indicate possible missing coding or non-coding regions absent in South African sequences *B. anthracis* 20SD. The plasmid pXO2 of the *B. anthracis* strains are highly similar in relation to the reference sequence with few genomic rearrangements.

4.3.8. Single nucleotide polymorphism (SNPs) on plasmids

Few SNPs were found on pXO1 in *B. anthracis* 20SD and 3631-1C strains relative to *B. anthracis* Ames ancestor i.e. 10 and 6 SNPs on strain 3631-1C and 20SD respectively. The coding sequences (CDS) with single nucleotide variations were found in the genome of *Bacillus anthracis* annotation (GBAA)_pXO1 gene in both strains. Strain 3631-1C had SNPs at GBAA_pXO1_0217, 0115, 0104, 0089, 0047, 0011, 0010, and other additional SNPs were located at CDS *top*X and two at CDS *cya* gene. The two genes exhibiting SNPs play an important role on the virulence of plasmid pXO1. The *cya* is one of the toxin genes that encodes for the edema factor (Okinaka *et al*., 1999) while *top*X (topoisomerase) gene plays a role in either the maintenance of or in the regulation of expression of genes carried by the virulence plasmid (Fouet *et al*., 1994). No SNPs were detected in the *cya* and *top*X genes of the 20SD strain. SNPs in this isolate were

only found at the hypothetical proteins GBAA_pXO1_0047, 0089, 0099, 0115, 0137, and 0217. The GBAA_pXO1_009 and 0137 are the two unique SNPs found on pXO1 of strain 20SD relative to 3631-1C. SNPs within GBAA_pXO1_009 and 0137 were unique to the plasmid pXO1 of strain 20SD and not found in 3631-1C. Only two SNPs were found on plasmid pXO2 regions GBAA_pXO2_0009 and GBAA_pXO2_0051 hypothetical proteins.

4.3.9. Genome alignment of the chromosome

Comparative genome alignment was performed for *B. anthracis* strains 3631-1C and 20SD chromosome using Mauve tool against the genome of *B. anthracis* Ames ancestor sequence. The results show that the chromosome structure of *B. anthracis* is relatively conserved with few large rearrangements. In this study, a number of RNA genes including ribosomal RNA, transfer RNA and ribosomal protein genes were represented on each strain, and but were distributed at different positions relative to *B. anthracis* Ames ancestor (Figure 4.7). Few translocated regions were observed in *B. anthracis* strains including *B. anthracis* A0442 and KrugerB strains relative to *B. anthracis* Ames ancestor.

The genomic sequence of *B. anthracis* 3631-1C was highly similar to the Ames ancestor with few genetic rearrangements. A significant structural variation of inversion of a 21,145 bp region was detected between the base nucleotides 460758 and 481903 relative to the *B. anthracis* Ames ancestor genome sequence. This region consists of prophage lamdaBa04, phage minor capsids and hypothetical proteins of the prophage, which are found amongst *B. anthracis* strains.

Three large inversions regions are observed in strain 20SD. The strain exhibited inversions of about 6 kb between the base nucleotide 492 0000 and 498 0000 relative to the reference genome. This region carries the mobile element protein region of 1230 bp at base pair coordinates 4945770-4947000 of *B. anthracis* Ames ancestor. Another inversion of about 16 838 nucleotides in length was detected between nucleotides 955210 and 972048 of *B. anthracis* Ames ancestor. This inverted region includes coding

sequences of FNM independent NADH azoreductase, dehydroxyacetone kinase family protein, ATP dependent, NAD dependent epimerase, and other hypothetical proteins. Keim *et al*. (2009) has reported genomic arrangements amongst three *B. anthracis* strains (Australia 94, Ames, and CDC). The *B. anthracis* CDC strain (Okinika *et al*., 2011) has many genomic rearrangements that include inversions. The functional effect of these rearrangements and whether they affect virulence is unknown.

This study showed that regions that have capsid or mobile phage elements were likely to change the structural complex of the strain resulting in inversions. Previous studies also showed that strains that were exposed to bacteriophages were likely to modify *B. anthracis* phenotypically in terms of virulence, sporulation, metabolic enzymes, and transcriptional receptors among other effects (Schuch and Fischetti, 2009). This explains why prophages such as conjugative transposons, insertion sequences, introns and other elements that make up the mobile portion of bacterial genomes often account for large-scale genomic rearrangements, insertions and deletions in bacterial chromosomes (Brussow *et al*., 2004; Sozhamannan *et al*., 2006).

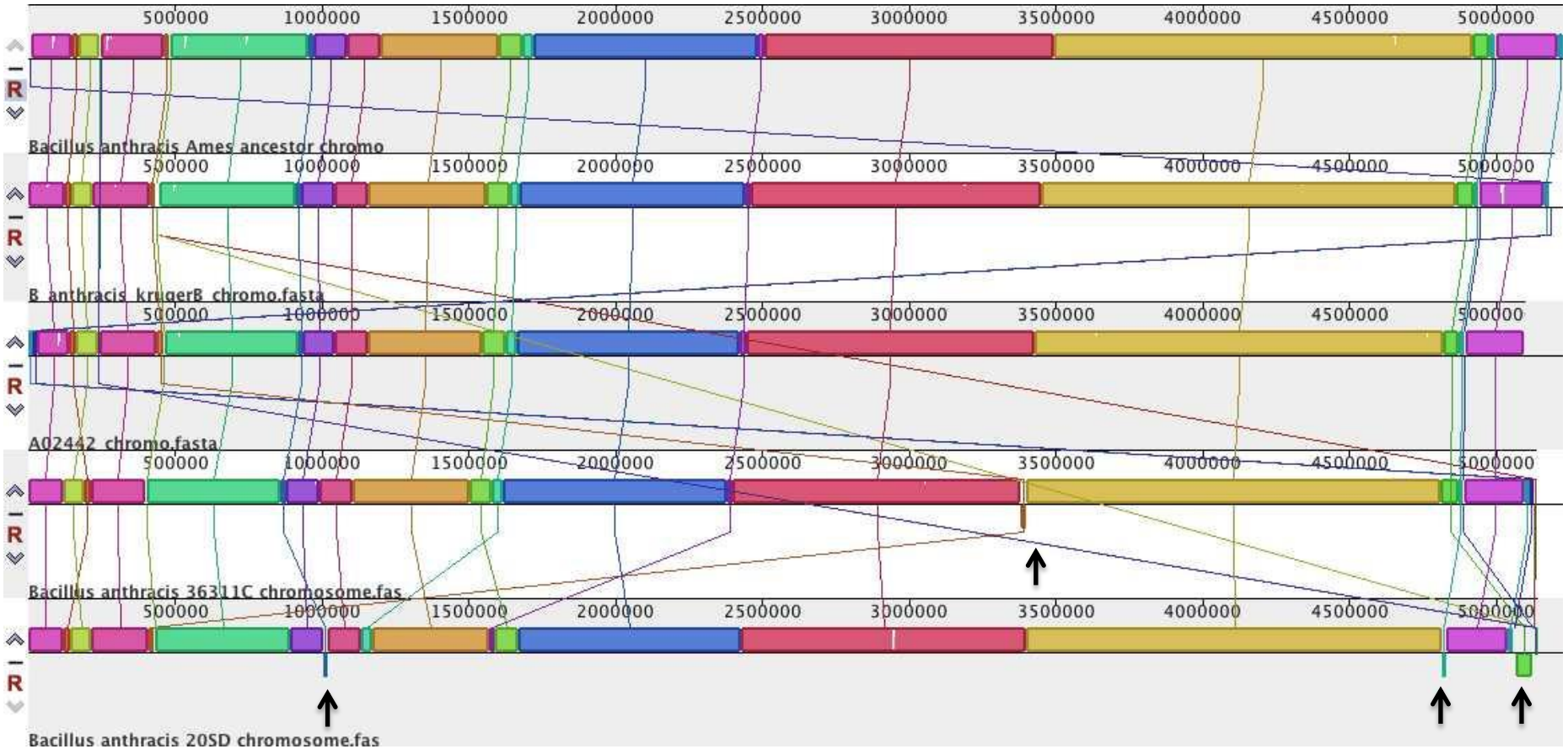


Figure 4.7. Alignment of chromosomes of the *Bacillus anthracis* Ames ancestor with *B. anthracis* KrugerB, *B. anthracis* A0442, *B. anthracis* 3631-1C and *B. anthracis* 20SD. Each colour blocks indicates homologous regions in the genome sequences being compared. The inversions are indicated by arrows. White areas in colour blocks indicate insertion or deletions at intergenic or coding regions, small and large subunits of ribosomal RNA absent or present in *B. anthracis* strains.

4.3.10. Single nucleotide polymorphism (SNPs) on chromosome

Both strains in this study exhibited a high number of SNPs in the chromosomes compared to the plasmids. , About 397 SNPs were found in strain 20SD chromosome, and 448 SNPs in strain 3631-1C *relative to B. anthracis* Ames ancestor. Majority of the coding region SNPs were found on hypothetical proteins, phage tail fiber proteins, acetyltransferase GNAT family, and *VrrA* coding sequences of the two strains. The proportion of SNPs between the two strains was almost equal (session 4.4.5, Figure 4.4). However *B. anthracis* 3631-1C had 51 unique SNPs compared to *B. anthracis* 20SD. The two *B. anthracis* strains had different genotypes that could be explained by differences in geographic regions of origin. Comparative analysis of the chromosome in *B. anthracis* 20SD and 3631-1C relative to *B. anthracis* KrugerB revealed 1 801 SNPs 20SD and 2 027 for 3631-1C respectively. This high proportion of SNPs is expected since *B. anthracis* KrugerB is found in the B clade whilst 20SD and 3631-1C belong to the A clade. Both *B. anthracis* A0442 and KrugerB were isolated from KNP with A0442 being isolated from a kudu. High proportions of SNPs were also observed from *B. anthracis* 20SD (2 030) and 3631-1C (1 802) relative to *B. anthracis* A0442.

4.3.11. Circular genome map

The genome map was constructed with the genome sequences assembled *de novo* using CGView comparative tool (CCT) (Grant and Stothard, 2008). *B. anthracis* Ames ancestor sequence was compared to the *B. anthracis* 3631-1C and 20SD strains (Figure 4.8). Genome variations were found at non-coding regions and SNPs not shown in the Figure 4.8. The circular genomes indicated that the chromosome of the *B. anthracis* strains have highly conserved regions using BLASTn.

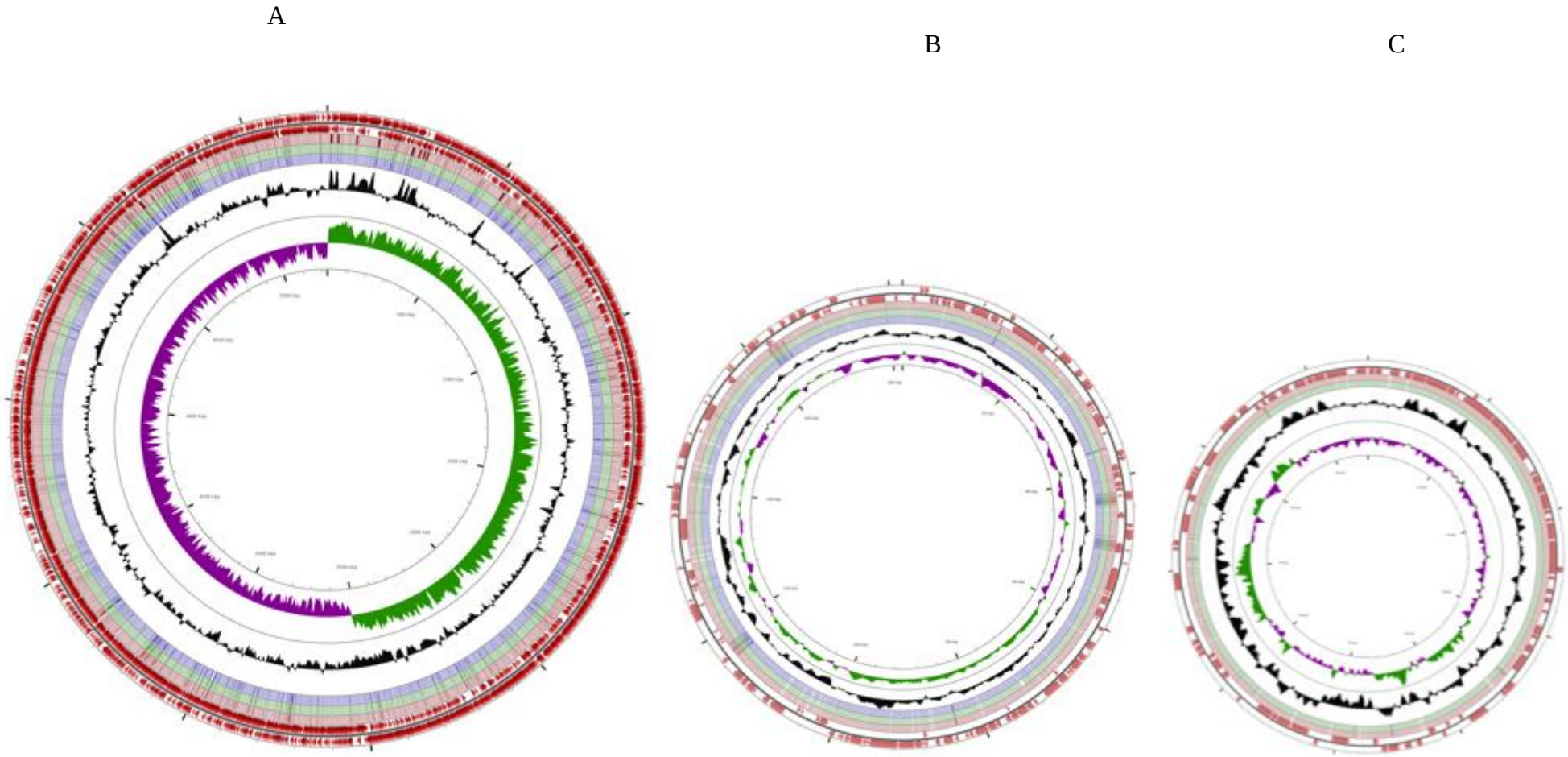


Figure 4.8. The CGView Comparative tool (CCT) maps comparing the *B. anthracis* Ames ancestor genome sequence to the *B. anthracis* 3631-1C and 20SD strains. A= chromosome alignment, B =plasmid pXO1 alignment, C = plasmid pXO2 alignment. Circles display from inside to outside: genomic position in kilobases, GC skew (purple and green), GC content (black). BLASTX results with *B. anthracis* Ames (light blue), *B. anthracis* 3631-1C (light green), and *B. anthracis* 20SD (light red) and the open reading frames are shown in red.

4.4.12. Identification of phage sequences

The *B. anthracis* chromosome is distinguished by having four integrated, putative prophage regions (Sozhamannan *et al*., 2006) constituting about 3% of the 5.2 Mb chromosome (Read *et al*., 2003). These regions make a higher percentage of unique genes in comparison to *B. cereus* group strains (Read *et al*., 2003; Kolsto *et al*., 2009). It was shown using PHAST that *B. anthracis* Ames ancestor has 5 prophages which included the 4 lambda regions (lambda01-lambda04) and the PHAGE_Sphing_PAU (Supplementary Figure 4S3**).** *Bacillus anthracis* 3631-1C consists of 7 phages found on the genome, while 6 were detected on the *B. anthracis* 20SD genome (Table 4.6). The strain 3631-1C had one questionable prophage (28.7 Kb) region identified as PHAGE_Bacill_1 (lambda01), similarly to *B. anthracis* Ames ancestor, while strain 20SD had two questionable prophage regions identified as PHAGE_Bacill_1 (lambda01) and PHAGE_Lactob_c5 (lambda3). Some phages were intact in the regions while some are incomplete (Supplementary Figure 4S3, 4S4, 4S5, 4S6, 4S7). Incomplete phages indicated that some other regions on the phages could have been missing.

Bacillus anthracis 20SD and 3631-1C had different sizes and numbers of phages relative to *B. anthracis* Ames ancestor. The PHAGE_Entero_lato prophages were intact and present only in the two *B. anthracis* strains 20SD and 3631-1C and absent to *B. anthracis* Ames ancestor. About 43 PHAGE_Entero EF62 phages were observed on the *B. anthracis* A0442, and 2 in *B. anthracis* str. KrugerB (PHAGE_Entero RB51 and RB16). However the prophages are not incomplete in *B, anthracis* KrugerB and A0442, and are unique to the *B. anthracis* 20SD, 3631-1C and *B. anthracis* Ames ancestor.

The 4 prophage regions lambda01 (PHAGE_Bacill_1), lambda02 (PHAGE_Gamma), lambda03 (PHAGE_Lactob_c5), and lambda04 (PHAGE_Geobac) sequences are present in the *B. anthracis* strains 3631-1C, 20SD, Ames ancestor, as well as *B. anthracis* KrugerB and *B. anthracis* A0442. These unique prophage sequences set *B. anthracis* apart from its nearest neighbors (Sozhamannan *et al*., 2006; Kolsto *et al*., 2009). The *B. anthracis* strains had prophage regions for gamma (ϒ) phage that is used as a diagnostic tool in *B. anthracis* identification (Schuch *et al*., 2006), although some non-susceptible *B. anthracis* strains have been reported (Abshire *et al*., 2005; Davison *et al*., 2005; Arabi *et al*., 2013). The phages identified in this study are unique and useful unique regions of *B. anthracis*

that can be used to distinguish *B. anthracis* amongst *B. cereus* group (Sozhamannan *et al*., 2006).

Table 4.4. Prophage regions of *Bacilllus anthracis* 3631-1C (7) and 20SD (6) strains identified using the phage search tool*.

Region	**Phage Region Length (Kbp)**	**Region Position (*B. anthracis* 3631-1C)**	**Possible phage**
1	49.5	1932083-1981678	PHAGE_Bacill_1 (Lambda01), PHAGE Bacilli BtCSS33
2	16.5	3078000-3094501	PHAGE_Sphing_PAU, PHAGE_Escher_phAPEC8
3	21.3	3617692-3639026	PHAGE_Geobac_GBSV1 (Lambda04), PHAGE_Bacill_1,
4	16.6	4098972-4115642	PHAGE_Lactob_c5 (Lambda03), PHAGE_Bacill_BtCS33
5	49.6	4205266-4254942	PHAGE_Bacill_Gamma (Lambda02), PHAGE_Bacill_phIS3501
6	101.1	4991999-5093149	PHAGE_Entero_lato, PHAGE_Bacill_PBC1
7	8.1	5082157-5090300	PHAGE_Entero_lato, PHAGE_Plankt_PaV_
Region	**Phage Region Length (Kbp)**	**Region Position (*B. anthracis* 20SD)**	**Possible phage**
1	5.3	685181-690527	PHAGE_Entero_lato, PHAGE_Entero_alpha3
2	28.2	1497088-1525299	PHAGE_Bacill_1 (Lambda01), PHAGE_Geobac_GBSV1
3	32.3	1896564-1928942	PHAGE_Bacill_Gamma (Lambda02), PHAGE_Bacill_WBeta
4	21.3	3430802-3452136	PHAGE_Bacill_1, PHAGE_Geobac_GBSV
5	35.6	4255179-4290840	PHAGE_Lactob_c5 (Lambda03), PHAGE_Bacill_BtCS33
6	20.6	4556873-4577491	PHAGE_Bacill_1 (Lambda04), PHAGE_Geobac_GBSV1

***** Indicate different colours that display the prophage types: pink= intact, grey= incomplete, and green= questionable.

4.4. Conclusion

The two *B. anthracis* strains in the study were analyzed using genome wide characterization. The *B. anthracis* 20SD had the chromosome, plasmids pXO1 and pXO2, while 3631-1C lacked pXO2. The two strains carried unique phage minor structural proteins that are common to *B. anthracis* H9401 and CDC 684 strains. The regions consisting of genetic mobile elements were most likely responsible for the genomic rearrangements in the form of inversions. *Bacillus anthracis* strains are defined by having four lambda prophages that were also observed in the *B. anthracis* strains used in this study. The number and size of the prophages differed between strains. The *B. anthracis* 20SD and 3631-1C had different number of coding SNPs. The SNPs observed from these *B. anthracis* strains will be used in further studies to construct whole genome SNP phylogeny, which will provide more information on the clade structure of the *B. anthracis* strains.

4.5. Reference

Abshire, T., Brown, J., Ezzell, J., 2005. Production and validation of the use of gamma phage for identification of *Bacillus anthracis*. J. Clin. Microbiol. 43, 4780-4788.

Achtman, M., 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu. Rev. Microbiol. 62, 53-70.

Andrews, S., 2010. FASTQC.A quality control tool for high throughput sequence data. http://www. bioinformatics. babraham. ac. uk/projects/fastqc.

Antwerpen, M., Ilin, D., Georgieva, E., Meyer, H., Savov, E., Frangoulidis, D., 2011. MLVA and SNP analysis identified a unique genetic cluster in Bulgarian *Bacillus anthracis* strains. Eur. J. Clin. Microbiol. Infect. Dis. 30, 923-930.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75.

Beyer, W., Turnbull, P., 2009. Anthrax in animals. Mol. Aspects Med. 30, 481-489.

Beyer, W., Turnbull, P., 2013. Co- infection of an animal with more than one genotype can occur in anthrax. Lett. Appl. Microbiol. 57, 380-384.

Birdsell, D.N., Antwerpen, M., Keim, P., Hanczaruk, M., Foster, J.T., Sahl, J.W., Wagner, D.M., Grass, G., 2013. Draft genome sequences of two Bulgarian *Bacillus anthracis* strains. Genome announc. 1.

Brussow, H., Canchaya, C., Hardt, W., 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. 68, 560-602.

Chun, J., Hong, K., Cha, S.H., Cho, M., Lee, K.J., Jeong, D.H., Yoo, C., Rhie, G., 2012. Complete Genome Sequence of *Bacillus anthracis* H9401, an Isolate from a Korean Patient with Anthrax. J. Bacteriol. 194, 4116-4117.

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

Davison, S., Couture-Tosi, E., Candela, T., Mock, M., Fouet, A., 2005. Identification of the *Bacillus anthracis* γ phage receptor. J. Bacteriol. 187, 6742-6749.

Didelot, X., Bowden, R., Wilson, D.J., Peto, T.E., Crook, D.W., 2012. Transforming clinical microbiology with bacterial genome sequencing. Nature Rev. Genet. 13, 601-612.

Edwards, D.J., Holt, K.E., 2013. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. Microb. Inform. Exp. 3, 2.

El-Arabi, T.F., Griffiths, M.W., She, Y., Villegas, A., Lingohr, E.J., Kropinski, A.M., 2013. Genome sequence and analysis of a broad-host range lytic bacteriophage that infects the *Bacillus cereus* group. Virol. J.10, 48.

Fasanella, A., Galante, D., Garofolo, G., Jones, M.H., 2010. Anthrax undervalued zoonosis. Vet. Microbiol. 140, 318-331.

Fouet, A., Mesnage, S., 2002. *Bacillus anthracis* cell envelope components. Curr. Top. Microbiol. Immunol. 271, 87.

Fouet, A., Sirard, J., Mock, M., 1994. *Bacillus anthracis* pXO1 virulence plasmid encodes a type 1 DNA topoisomerase. Mol. Microbiol. 11, 471-479.

Grant, J.R., Stothard, P., 2008. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 36, W181-W184.

Hoffmaster, A.R., Meyer, R.F., Bowen, M.P., Marston, C.K., Weyant, R.S., Thurman, K., Messenger, S.L., Minor, E.E., Winchell, J.M., Rasmussen, M.V., 2002. Evaluation and validation of a real-time polymerase chain reaction assay for rapid identification of *Bacillus anthracis*. Emerg. Infect. Diseases 8, 1178.

Keim, P., Gruendike, J.M., Klevytska, A.M., Schupp, J.M., Challacombe, J., Okinaka, R., 2009. The genome and variation of *Bacillus anthracis.* Mol. Aspects Med. 30, 397-405.

Keim, P., Price, L., Klevytska, A., Smith, K., Schupp, J., Okinaka, R., Jackson, P., Hugh-Jones, M., 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. 182, 2928-2936.

Keim, P., Van Ert, M.N., Pearson, T., Vogler, A.J., Huynh, L.Y., Wagner, D.M., 2004. Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. Infect. Genet. Evol. 4, 205-213.

Koser, C.U., Ellington, M.J., Cartwright, E.J., Gillespie, S.H., Brown, N.M., Farrington, M., Holden, M.T., Dougan, G., Bentley, S.D., Parkhill, J., 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathogens 8, e1002824.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754-1760.

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.

Lista, F., Faggioni, G., Valjevac, S., Ciammaruconi, A., Vaissaire, J., Le Doujet, C., Gorge, O., De Santis, R., Carattoli, A., Ciervo, A., 2006. Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. BMC Microbiol. 6, 33.

Marston, C.K., Allen, C.A., Beaudry, J., Price, E.P., Wolken, S.R., Pearson, T., Keim, P., Hoffmaster, A.R., 2011. Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States. PloS one 6, e28274.

Marston, C.K., Hoffmaster, A.R., Wilson, K.E., Bragg, S.L., Plikaytis, B., Brachman, P., Johnson, S., Kaufmann, A.F., Popovic, T., 2005. Effects of long-term storage on plasmid stability in *Bacillus anthracis*. Appl. Environ. Microbiol. 71, 7778-7780.

Mock, M., Fouet, A., 2001. Anthrax. Ann. Rev. Microbiol. 55, 647-671.

Mock, M., Mignot, T., 2003. Anthrax toxins and the host: a story of intimacy. Cell. Microbiol. 5, 15-23.

Ndumnego, O.C., 2012. Immunogenicity and protectivity of a live spore *Bacillus anthracis* vaccine in goats. M.Sc thesis, University of Pretoria, South Africa.

Okinaka, R., Price, E., Wolken, S., Gruendike, J., Chung, W., Pearson, T., Xie, G., Munk, C., Hill, K., Challacombe, J., 2011. An attenuated strain of *Bacillus anthracis* (CDC 684) has a large chromosomal inversion and altered growth kinetics. BMC Genomics 12, 477.

Okinaka, R., Cloud, K., Hampton, O., Hoffmaster, A., Hill, K., Keim, P., Koehler, T., Lamke, G., Kumano, S., Mahillon, J., 1999. Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. J. Bacteriol. 181, 6509-6515.

Pannucci, J., Okinaka, R.T., Sabin, R., Kuske, C.R., 2002. *Bacillus anthracis* pXO1 plasmid sequence conservation among closely related bacterial species. J. Bacteriol. 184, 134-141.

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., 2004. Phylogenetic discovery bias in Bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing. Proc. Natl. Acad. Sci. U. S. A. 101, 13536-13541.

Pearson, T., Okinaka, R.T., Foster, J.T., Keim, P., 2009. Phylogenetic understanding of clonal populations in an era of whole genome sequencing. Infect. Genet. Evol. 9, 1010-1019.

Price, E.P., Seymour, M.L., Sarovich, D.S., Latham, J., Wolken, S.R., Mason, J., Vincent, G., Drees, K.P., Beckstrom-Sternberg, S.M., Phillippy, A.M., 2012. Molecular Epidemiologic investigation of an anthrax outbreak among heroin users, Europe. Emerg. Infect. Dis. 18, 1307.

Ramisse, V., Patra, G., Garrigue, H., Guesdon, J., Mock, M., 1996. Identification and characterization of *Bacillus anthracis* by multiplex PCR analysis of sequences on plasmids pXO1 and pXO2 and chromosomal DNA. FEMS Microbiol. Lett. 145, 9-16.

Rasko, D.A., Altherr, M.R., Han, C.S., Ravel, J., 2005. Genomics of the *Bacillus cereus* group of organisms. FEMS Microbiol. Rev. 29, 303-329.

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., 2009. The complete genome sequence of *Bacillus anthracis* Ames "Ancestor". J. Bacteriol. 191, 445-446.

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., 2003. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature 423, 81-86.

Read, T.D., Salzberg, S.L., Pop, M., Shumway, M., Umayam, L., Jiang, L., Holtzapple, E., Busch, J.D., Smith, K.L., Schupp, J.M., 2002. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. Science 296, 2028-2033.

Schuch, R., Fischetti, V.A., 2009. The secret life of the anthrax agent *Bacillus anthracis*: bacteriophage-mediated ecological adaptations. PLoS One 4, e6532.

Schuch, R., Fischetti, V.A., 2006. Detailed genomic analysis of the Wβ and γ phages infecting *Bacillus anthracis*: implications for evolution of environmental fitness and antibiotic resistance. J. Bacteriol. 188, 3037-3051.

Segerman, B., De Medici, D., Ehling Schulz, M., Fach, P., Fenicia, L., Fricker, M., Wielinga, P., Van Rotterdam, B., Knutsson, R., 2011. Bioinformatic tools for using whole genome sequencing as a rapid high resolution diagnostic typing tool when tracing bioterror organisms in the food and feed chain. Int. J. Food Microbiol. 145, S167-S176.

Smith, K., DeVos, V., Bryden, H., Price, L., Hugh-Jones, M., Keim, P., 2000. *Bacillus anthracis* diversity in Kruger National Park. J. Clin. Microbiol. 38, 3780-3784.

Sozhamannan, S., Chute, M., McAfee, F., Fouts, D., Akmal, A., Galloway, D., Mateczun, A., Baillie, L., Read, T., 2006. The *Bacillus anthracis* chromosome contains four conserved, excision-proficient, putative prophages. BMC microbiology 6, 34.

Sozhamannan, S., McKinstry, M., Lentz, S.M., Jalasvuori, M., McAfee, F., Smith, A., Dabbs, J., Ackermann, H., Bamford, J.K., Mateczun, A., 2008. Molecular characterization of a variant of *Bacillus anthracis*-specific phage AP50 with improved bacteriolytic activity. Appl. Environ. Microbiol. 74, 6792-6796.

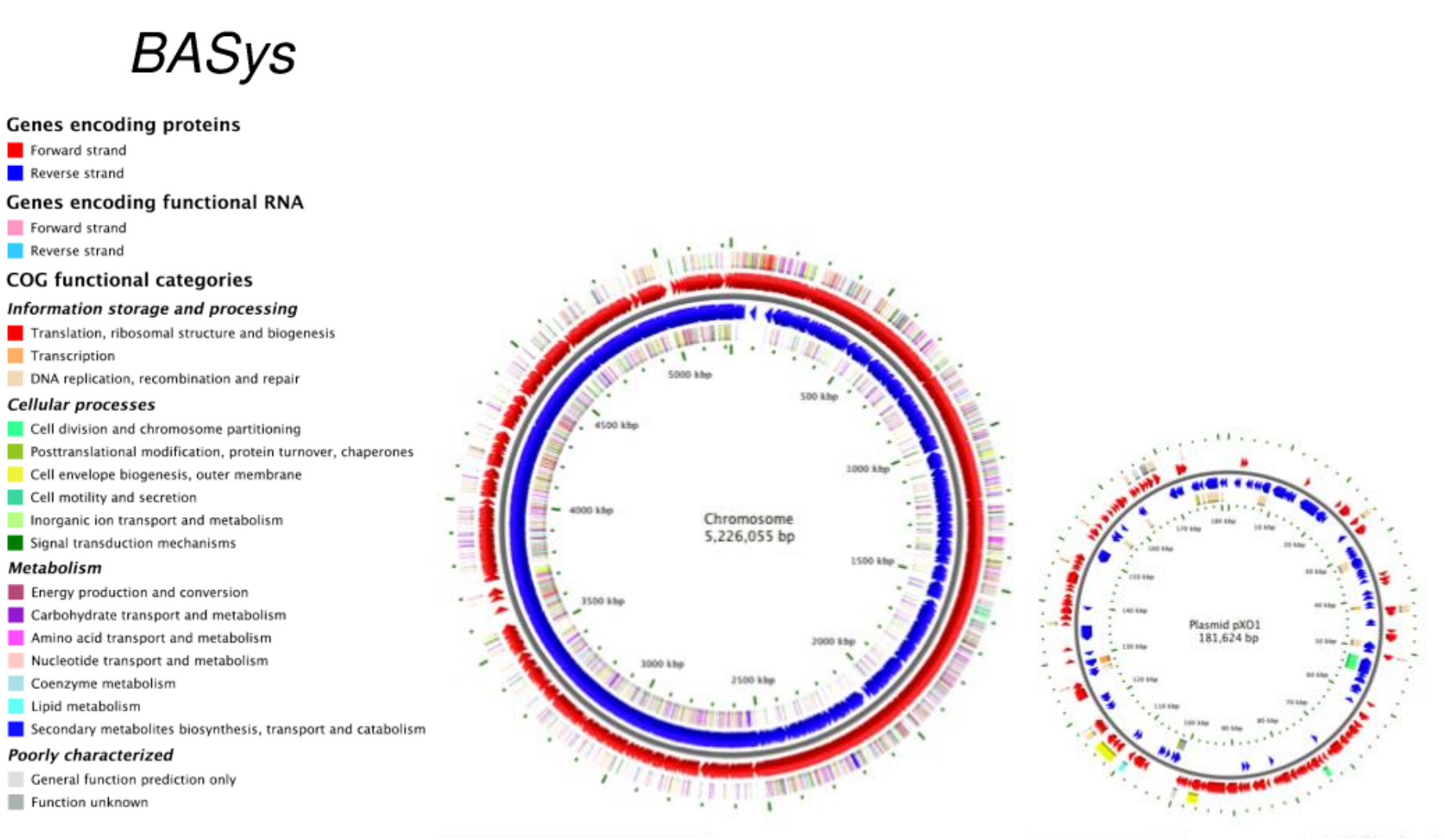
Turnbull, P.C.B., 1998. WHO, EMC/ZDI/98 Guidelines for the surveillance and control of anthrax in humans and animals (fourth edition). World Health Organization Department of Communicable Diseases Surveillance and Response report WHO/EMC/ZDI/98.6, 29-30.

Van Domselaar, G.H., Stothard, P., Shrivastava, S., Cruz, J.A., Guo, A., Dong, X., Lu, P., Szafron, D., Greiner, R., Wishart, D.S., 2005. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res. 33, W455-W459.

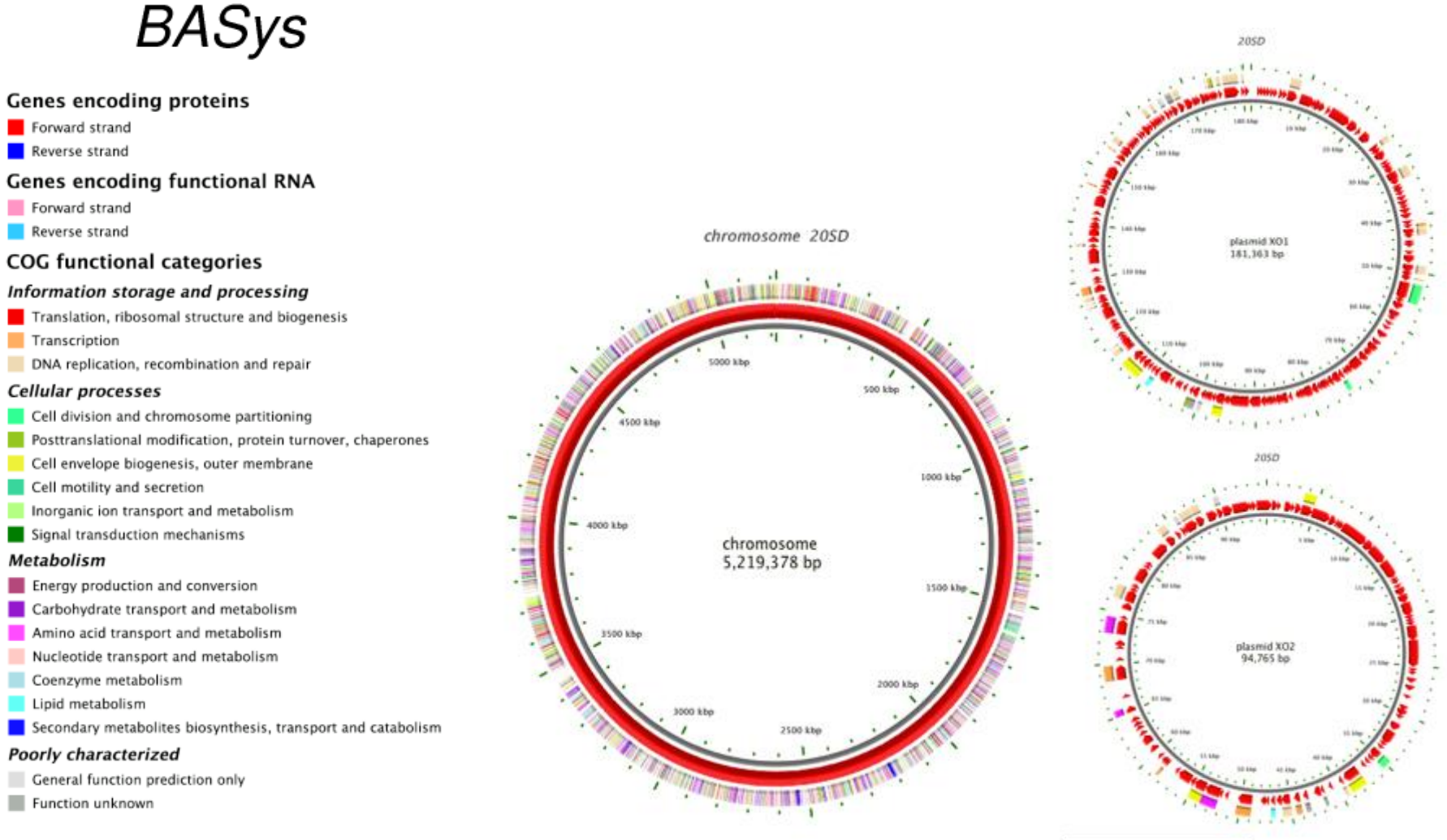
Van Ert, M.N., Easterday, W.R., Huynh, L.Y., Okinaka, R.T., Hugh-Jones, M.E., Ravel, J., Zanecki, S.R., Pearson, T., Simonson, T.S., U'Ren, J.M., 2007. Global genetic population structure of *Bacillus anthracis*. PLoS One 2, e461.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., Wishart, D.S., 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39, W347-W352.

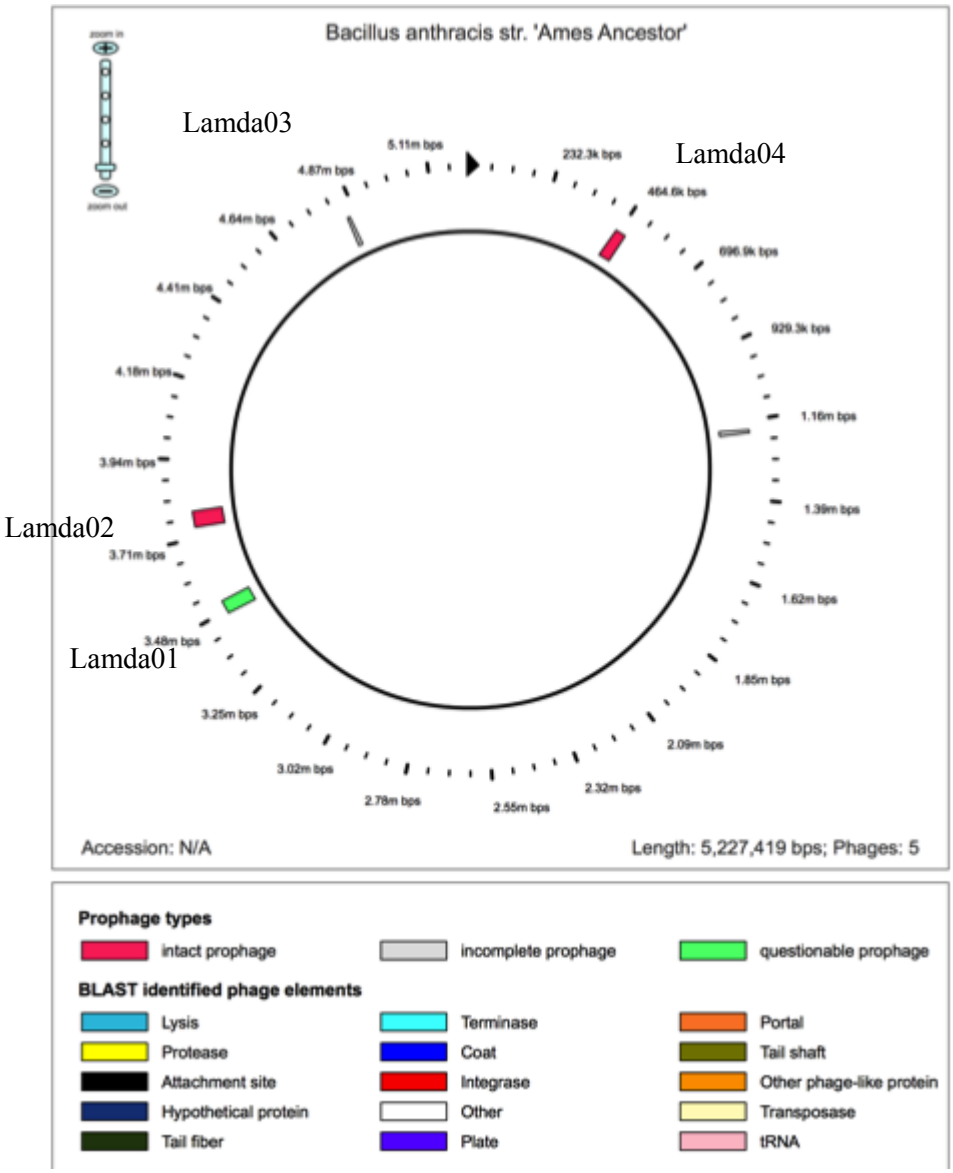
4.6. Supplemenatary tables and figures



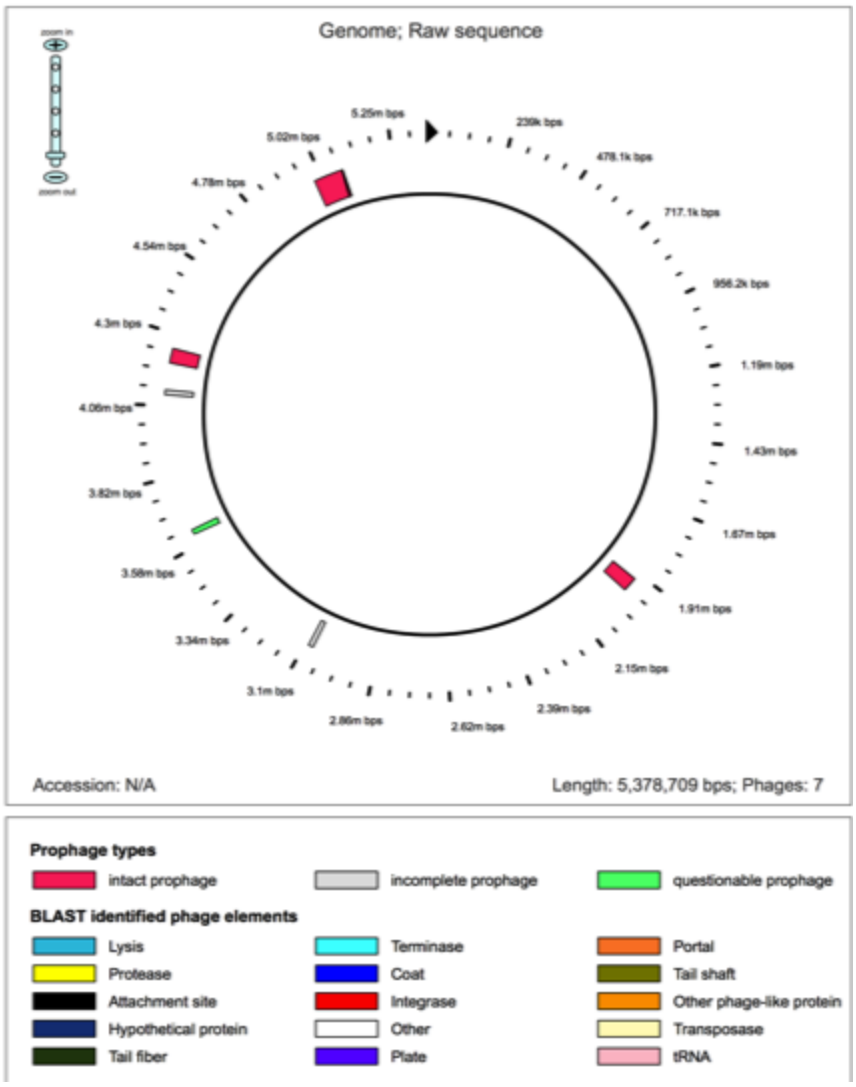
Supplementary Figure 4S1. A circular genome map of *Bacillus anthracis* 3631-1C strain chromosome and plasmid pXO1 isolated from kudu in Northen Cape Province on the farm Klipfontein, generated using the BASys annotation.



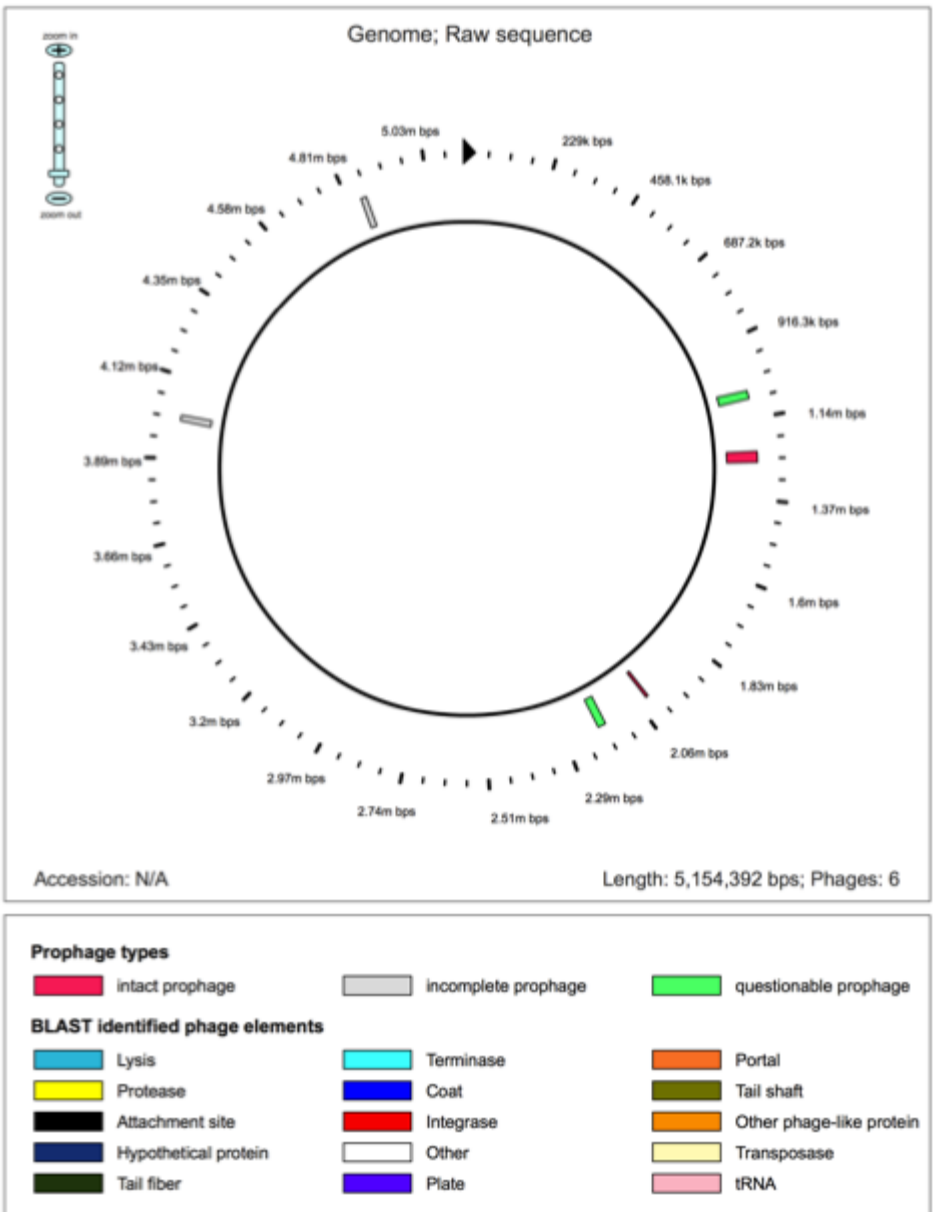
Supplementary Figure 4S2. A circular genome map of *Bacillus anthracis* 20SD strain chromosome and plasmids isolated from ovine in Standerton area, Mpumalanga Province, generated using the BASys annotation.



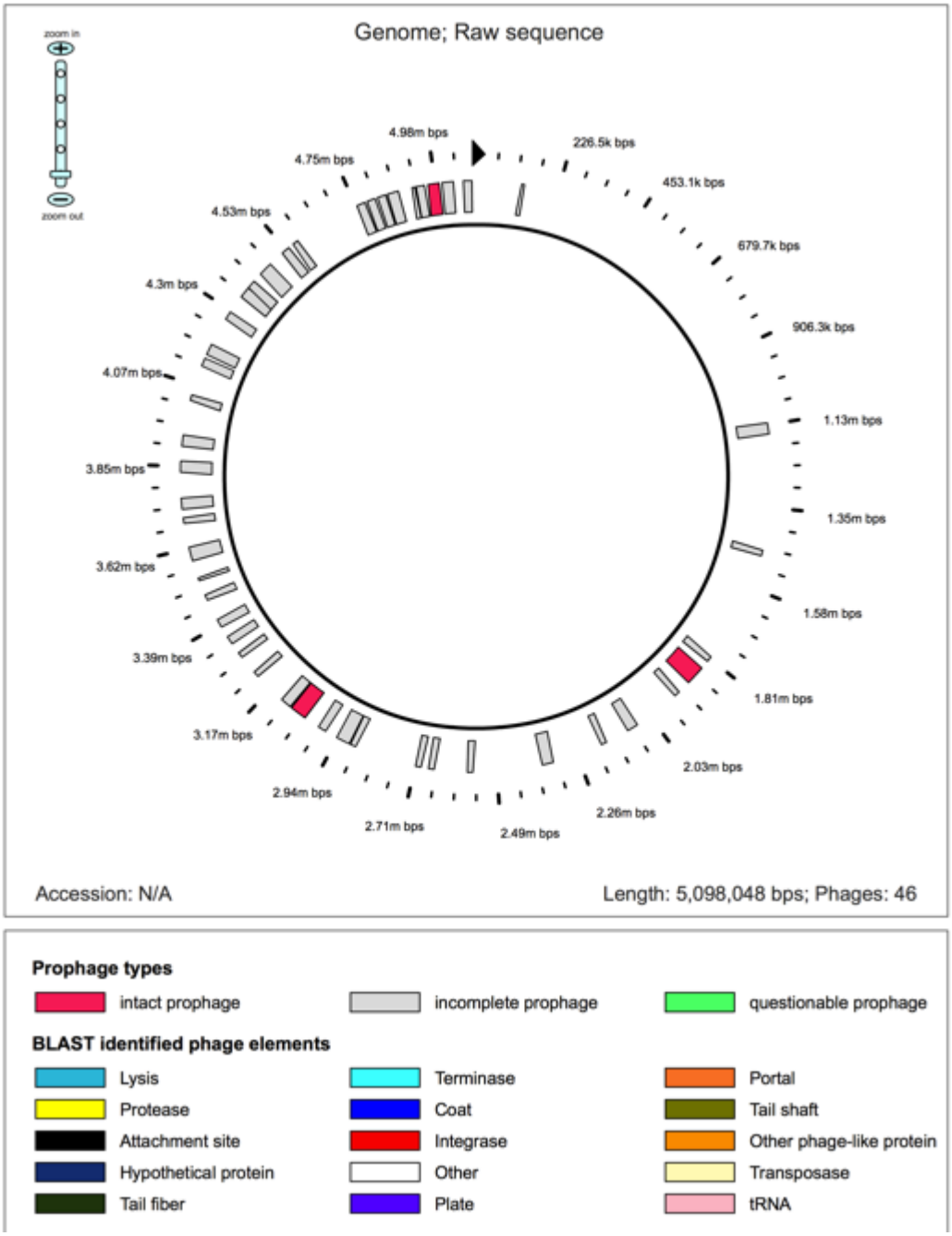
Supplementary Figure 4S3. A graphical representation of *Bacillus anthracis* Ames ancestor circular chromosome with annotated prophages using PHAST server.



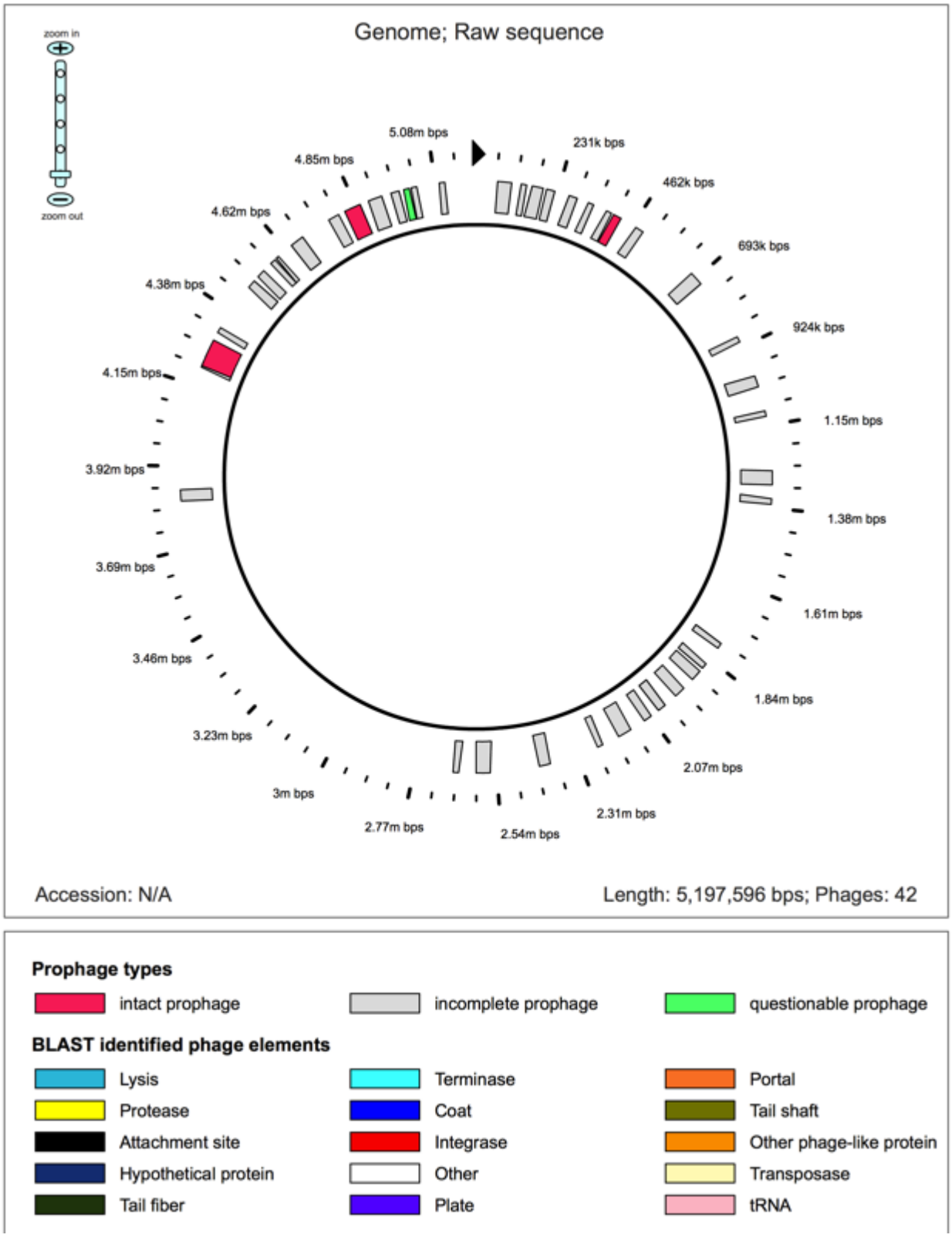
Supplementary Figure 4S4. A graphical representation of *Bacillus anthracis* 3631-1C circular chromosome with annotated prophages using PHAST server.



Supplementary Figure 4S5. A graphical representation of *Bacillus anthracis* 20SD circular chromosome with annotated prophages using PHAST server.



Supplementary Figure 4S6. A graphical representation of *Bacillus anthracis* A0442 circular chromosome with annotated prophages using PHAST server.



Supplementary Figure 4S7. A graphical representation of *Bacillus anthracis* KrugerB circular chromosome with annotated prophages using PHAST server.