Phylogenomic structure of *Bacillus anthracis* isolates in the Northern Cape Province, South Africa revealed novel single nucleotide polymorphisms

Kgaugelo Edward Lekota[a,b,c,d],*, Ayesha Hassim[a], Evelyn Madoroba[b,1], Charles A.Hefer[c,e,2] and Henriette van Heerden[a]

[a]University of Pretoria, Faculty of Veterinary Science, Department of Veterinary Tropical diseases, Onderstepoort 0110, South Africa
[b]Agricultural Research Council-Onderstepoort Veterinary Institute, Bacteriology section, Onderstepoort 0110, South Africa
[c]Agricultural Research Council, Biotechnology Platform, Onderstepoort 0110, South Africa
[d]College of Agriculture and Environmental Sciences, University of South Africa, Florida campus, Florida 1710, South Africa
[e]University of Pretoria, Faculty of Natural and Agricultural Sciences, Department of Biochemistry, Genetics and Microbiology, Centre for Bioinformatics and Computational Biology, Pretoria 0001, South Africa

*Corresponding author at: University of Pretoria, Faculty of Veterinary Science, Department of Veterinary Tropical diseases, Onderstepoort 0110, South Africa. lekotae@gmail.com

Highlights

- wgSNP analysis of the Northern Cape Province, South African *B. anthracis* strains cluster in to new A-sub-clades.
- Strains were defined by A.Br.002 (Sterne), A.Br.101, A.Br.005/006, and A.Br.V770 clades featuring unique regional SNPs.
- The prevailing A.Br.101 branch is dissimilar from the A.Br.013 (Australia94 and Georgia strains) sub-clade.
- Mozambique and Zimbabwe strains (A.Br.102 branch) are derived from the established A.Br.172 branch that is shared with NCP strains.

Abstract

Bacillus anthracis, the aetiological agent of anthrax, is regarded as a highly monomorphic pathogen that presents a low genetic diversity using standard molecular techniques. Whole genome sequencing and single nucleotide polymorphisms (SNPs) are definitive signatures for subtyping of *B. anthracis*. Here we employed whole genome single nucleotide polymorphism (wgSNP) analysis to investigate the genetic diversity of *B. anthracis* in the historically endemic region of Northern Cape Province (NCP), South Africa. Twenty-six isolates from anthrax outbreaks that occurred between 1998 and 2008/9 in NCP as well as from Namibia-South Africa Transfontier Conservation area and Botswana were compared to global *B. anthracis* genomes. Most NCP *B. anthracis* strains (n = 22) clustered in the A.Br.003/004 (A.Br.101) branch and are closely related to the Zimbabwe and Mozambique strains (A.Br.102 branch). A total of 4923 parsimony informative-SNPs accurately established the A.Br.003/004 phylogenetic relationships of the NCP isolates into two distinct sub-clades and SNP markers designated as A.Br.172 and A.Br.173 were developed. Other NCP strains (n = 2) grouped in the A.Br.001/002 (Sterne) branch while strains (n = 2) from the Namibia-South Africa Transfontier Conservation area and Botswana clustered in A.Br.005/006 (Ancient A) branch. The sequenced *B. anthracis* strains (A0094, A0096 and A0097) that clustered in the A.Br.064 (V770) clade were isolated from Vaalbos National Park and similar strains have not been isolated. The *B. anthracis* A0088 strain cluster with the NCP strains in

the A.Br.003/004 (A.Br.172) SNP branch which has been isolated in NCP, South Africa. This study highlights the phylogenetic structure of NCP *B. anthracis* strains with distinctive SNP branches important for forensic tracing and novel SNP discovery purposes. The sequenced strains will serve as a means to further trace the dissemination of *B. anthracis* outbreaks in NCP, South Africa, and on the continent, as well as for forensic tracking on a global scale.

Keywords: *Bacillus anthracis;* Whole genome sequencing (WGS); Single nucleotide polymorphisms (SNPs); Whole genome single nucleotide polymorphism (wgSNP)

1. Introduction

Bacillus anthracis, the causative agent of anthrax, is a spore forming zoonotic bacterium (Sterne, 1967) that affects both wild and domestic animals (Turnbull, 2008). Due to protracted dormancy periods between spore germination, the bacterium is monomorphic and evolves slowly (Keim et al., 2000). Ascertaining the source and spread of the disease is essential to determine the evolutionary relationships amongst isolates and to trace the distribution and paths of transmission. Sequencing bacterial genomes with next generation sequencing (NGS) technologies has recently emerged as a convenient approach for the comprehensive genetic profiling of microbes (Didelot et al., 2012). A whole genome sequencing approach provides more diagnostic resolution power to differentiate lineages by identifying thousands of SNPs retrieved from aligned *B. anthracis* genomes (Keim et al., 2004; Pearson et al., 2009). This provides more robust information in forensic microbiology thus allowing trace back of anthrax outbreaks from where a particular variant arises and determining evolutionary relationships.

Major lineages and sub-lineage of the *B. anthracis* strains have been determined using canonical single nucleotide polymorphisms (canSNPs) markers which were identified using hybridization real-time PCR (qPCR) assay (Van Ert et al., 2007b) and the more cost effective melt analysis of mismatch amplification mutation assay (melt-MAMA) using SYBR green real-time PCR or conventional PCR (Birdsell et al., 2012). Major lineages of *B. anthracis* include A, B, and C clades which have been identified using canSNPs (Van Ert et al., 2007a). The A-clade is commonly distributed worldwide, while B and C are the rare clades and more regionally restricted (Keim et al., 2004). CanSNPs and melt-MAMA differentiate the common and divergent SNPs of each *B. anthracis* strain (Keim et al., 2004; Van Ert et al., 2007a). However, these markers have been developed using available genomes mostly from the USA and Europe (Birdsell et al., 2012; Girault et al., 2014a; Khmaladze et al., 2014) and are therefore limited in the identification of strains from other regions.

Before wgSNP analysis, the lower resolution multiple loci variable number tandem repeats analysis (MLVA) enabled *B. anthracis* genotyping (Keim et al., 2004). However, this approach tends to be laborious and has a reputation for high homoplasticity (Achtman, 2008; Le Fleche et al., 2001), which is a drawback in the evaluation of the population structure and evolution of monomorphic bacteria like *B. anthracis* (Achtman and Wagner, 2008). Therefore, wgSNP analysis more accurately enables tracing the lineages and sub-clades of *B. anthracis* as well as to elucidate deep phylogenetic relationships amongst worldwide strains, due to their evolutionary stability (Girault et al., 2014b; Keim et al., 2004).

Anthrax is enzootic in South Africa in the northern part of Kruger National park and regions in the Northern Cape Province (NCP). Anthrax is endemic in the Northern Cape Province (NCP) in South Africa (De Vos, 1990; Hugh-Jones and de Vos, 2002) due to the predominance of wildlife conservancies and game farms (Turnbull, 2008). NCP has experienced anthrax epidemics along the Ghaap escarpment area and Vaalbos National Park (the latter has been decommissioned) (Hassim, 2016; Hassim et al., 2017) with two major outbreaks occurring in 1998–1999 and 2008–2009 (Henton and Briers, 1998). The NCP *B. anthracis* strains group in a sub-clade in the A-clade separately from Ames (A.Br.001; Ames) and Sterne $34F_2$ (A.Br.002; Sterne) when using MLVA genotyping and canSNP (Hassim, 2016). Since these PCR based genotyping methods have proved limiting in the establishment of region-specific clades and sub-clades for southern African strains, in depth SNP mining was used to determine the diversity. There is no *B. anthracis* genomic information available for the SNP lineages and diversity in the NCP, South Africa, to give insight into genetic population structure. In this study, we determined the phylogenomic diversity of *B. anthracis* isolates in NCP using wgSNP analysis from different anthrax outbreaks.

2. Materials and methods

2.1. *Bacillus anthracis* sample collection and microbiology

In order to examine the genomic diversity of *B. anthracis* in NCP (n = 26) isolated from different years' outbreaks, bacterial isolates were collected based on the availability in culture collections from the Agricultural Research Council-Onderstepoort Veterinary Institute (ARC-OVI) Bacteriology and NCP state veterinary services through the Department of Veterinary Tropical Disease at the University of Pretoria, South Africa. *B. anthracis* isolates were isolated from different farms in the Ghaap region (1998 and 2009), Vaalbos National Park (1998) as well as Sendelingsdrift (1998) located within the Richtersveld TCA (Namibia-South Africa Transfrontier Conservation Area). Kimberley provincial veterinary services sent samples to the ARC-OVI reference laboratory for diagnostics (Table 1). A *B. anthracis* 6102_6B isolate from Botswana (that borders South Africa) as well as South African *B. anthracis* 6461_SP2 from Sendelingsdrift (on the border with South Africa and Namibia) were sequenced to enable a better understanding of the relationship of strains in South Africa, Namibia and Botswana (southern Africa bordering regions) (Table 1). Pure colonies of *B. anthracis* were further sub-cultured onto sheep blood agar. Microbiological tests performed included colony morphology of *B. anthracis* on sheep blood agar (SBA), sensitivity of penicillin and gamma-phage, haemolysis and motility (WHO, 2012).

2.2. Genomic DNA extraction and PCR confirmation

Genomic extraction of the *B. anthracis* was carried out using DNA Mini Kit (Qiagen) according to manufacturer's instructions for Gram-positive bacteria. The DNA was quantified using qubit fluorometric quantization (Life Technologies, USA) using the broad range assay according to the manufacture's instruction. Quality of the extracted DNA was visualized on 0.8% agarose gel electrophoresis visualized under UV. Quantitative fluorescence resonance energy transfer PCR (FRET qPCR) targeting the virulence plasmids BAPA (*Bacillus anthracis* protective antigen) and *capC* (capsule) genes along with the genome target SASP (small acid soluble proteins) gene was used to confirm virulence factors of *B. anthracis* (Organization, W.H., Epizootics, I.O.o, 2008). The qPCR assay was performed essentially as described in OIE (Ellerbrok et al., 2002).

Table 1. *Bacillus anthracis* strains subjected to whole genome sequencing with canonical single nucleotide polymorphisms (canSNP) typing and newly discovered South African SNP branches.

Strain[a]	Location and Source	Year	Major canSNP	SNP group	New SNP Branch/Isolate	pXO1	pXO2
6102_6B	Botswana, *Loxodonta*	1998	A.Br.005/006 (Ancient A)	A.Br.034	6102_6B	+	–
6461_SP2	Sendelingsdrift, *Capra aegagrus*	1998	A.Br.005/006 (Ancient A)	A.Br.034	6461_SP2	+	+
3080_5A	Ghaap, Bovine	2009	A.Br.001/002 (Ames/Sterne)	A.Br.002 (Sterne)	3080_5A	+	–
3080_1B	Ghaap, Bovine	2009	A.Br.001/002 (Ames/Sterne)	A.Br.002 (Sterne)	3080_1B	+	–
2110	Klipfontein, Ovine	1998	A.Br.014	A.Br.101	A.Br.173	+	+
5838	Klipfontein, *Alcelaphus buselaphus*	1998	A.Br.014	A.Br.101	A.Br.173	+	+
3631_1C	Klipfontein, *Tragelaphus strepsiceros*	2009	A.Br.014	A.Br.101	A.Br.173	+	–
3080_3B	Ghaap, Bovine	2009	A.Br.014	A.Br.101	A.Br.173	+	+
3079_1C	Vlakfontein, *Oryx gazella*	2009	A.Br.014	A.Br.101	A.Br.173	+	+
3090_1B	Ghaap, unknown	2009	A.Br.014	A.Br.101	A.Br.173	+	+
JB10/NC_14	Vlakfontein, Ghaap, *Equus quagga*	2008	A.Br.014	A.Br.101	A.Br.173	+	+
JB25/NC_29	Dikbosch, *Tragelaphus strepsiceros*	2008	A.Br.014	A.Br.101	A.Br.173	+	+
2991_2B	Ghaap, Ovine	2009	A.Br.014	A.Br.101	A.Br.173	+	–
3618_2D	Ghaap, Soil	2009	A.Br.014	A.Br.101	A.Br.172	+	+
3517_1C	Klipfotein, *Tragelaphus strepsiceros*	2009	A.Br.014	A.Br.101	A.Br.172	+	+
3631_4C	Klipfotein, *Tragelaphus strepsiceros*	2009	A.Br.014	A.Br.101	A.Br.172	+	+
3631_7C	Ghaap, Soil	2009	A.Br.014	A.Br.101	A.Br.172	+	+
3275_2D	Ghaap, Soil	2009	A.Br.014	A.Br.101	A.Br.172	+	+
3122_2B	Klipfotein, Gemsbok	2009	A.Br.014	A.Br.101	A.Br.172	+	+
3008_1B	Klipfotein, Bovine	2009	A.Br.014	A.Br.101	A.Br.172	+	+
2949_1D	Klipfotein, Ovine	2009	A.Br.014	A.Br.101	A.Br.172	+	+
2991_1B	Ghaap, Ovine	2009	A.Br.014	A.Br.101	A.Br.172	+	+
3517_2C	Ghaap, *Tragelaphus strepsiceros*	2009	A.Br.014	A.Br.101	A.Br.172	+	+
3132_1B	Ghaap, *Tragelaphus strepsiceros*	2009	A.Br.014	A.Br.101	A.Br.172	+	+
3631_3D	Ghaap, *Tragelaphus strepsiceros*	2009	A.Br.014	A.Br.101	A.Br.172	+	+
3631_8D	Ghaap, *Tragelaphus strepsiceros*	2009	A.Br.014	A.Br.101	A.Br.172	+	+

[a]All *B. anthracis* were susceptible to penicillin, resistant to gamma-phage and amplified both SASP and BAPA genes.

2.3. Whole genome sequencing and mapping of reads to reference

B. anthracis genomes from NCP (n = 25) and Botswana (6102_6B) were sequenced in this study (Table 1). Sequence libraries of the DNA isolated was generated using Nextera XT Sample Prep Kit (Illumina, USA) protocol. Sequence reads of the paired end library was performed on a HiSeq 2500 and MiSeq sequencers (Illumina, USA). The quality of the genome sequenced reads were assessed using FastQC software 0:10.1 (Andrews, 2010). Trimmomatic version 0.33 (Bolger et al., 2014) was used to remove the sequenced adapters and ambiguous nucleotide reads. The genomic sequence of *B. anthracis* Ames ancestor (GenBank: NC_007322.2 and NC_007323,2) was used as a reference to determine the presence and/or absence of pXO1 and pXO2 respectively on the sequenced strains. Trimmed reads were mapped to the reference using Burrows-Wheeler Aligner (BWA) version 0.7.12 (Li and Durbin, 2009).

2.4. Whole genome SNP analysis

For wgSNP analysis, whole genome sequences and draft genomes (n = 32) of *B. anthracis* available in GenBank were used (Table 2) and also included already sequenced NCP 3631-1C (Lekota et al., 2015). The inclusion of the selected genome sequences were based on the different major clades of *B. anthracis* found globally, as well as South African and those of neighboring countries available at GenBank. Complete and draft genome assemblies retrieved from GenBank were simulated using SAMTools version 1.3 to generate the overlapping 200 bp reads. The trimmed reads of the NCP *B. anthracis* isolates and retrieved genomes were aligned to *B. anthracis* Ames ancestor reference chromosome (GenBank accession: NC_007530.2) using the Burrows-Wheeler Aligner (BWA) version 0.7.12 (Li and Durbin, 2009). SAMtools version 1.3 (Li et al., 2009) was used to sort and index the aligned sequenced reads. Sequence dictionary of the Ames ancestor sequence file was created using Picard-tools version 2.2.1 (http://picard.sourceforge.net/). Picard-tools version 2.2.1 was also used to mark duplicate reads and to build a binary index of the samples. The unified genotyper method in GATK version 3.7 (McKenna et al., 2010) was used to call for SNPs. GATK version 3.7 was used to filter the ambiguous variants prior to select variants. SNPs were called across all isolates with a standard base quality confirmation ≥30 and depth of coverage ≥5 on the chromosome. SNPs positioning sets of all the sequenced and compared genomes (total of 58 genomes) were deducted from the aligned genomes of *B. anthracis* Ames ancestor. SNP identified in genomic regions missing at one or more strains were excluded for selection of aligned informative SNP sites. SNPs with parsimony informative sites that were present in all genome sequences were used for the phylogenetic analysis. A phylogenetic tree of the whole genome filtered SNPs was constructed using molecular evolutionary genetics analysis (MEGA) tool version 7 (Kumar et al., 2016). The trees were generated using maximum likelihood method with a bootstrap replication value of 500.

Table 2. Whole genomes of *Bacillus anthracis* retrieved from public database and used in this study.

Strain	Country	canSNP	Accession number
Ames ancestor	USA	A.Br.001 (Ames)	NC_007530.2
Sterne	Unknown	A.Br.002 (Sterne)	NC_005945
52_G	Georgia	A.Br.013	AZUF00000000.1
8903_G	Georgia	A.Br.013	AZUD00000000.1
9080_G	Georgia	A.Br.013	AZUE00000000.1
A.Br003	Scotland	A.Br.013	JMPV00000000.1
Aust94	Australia	A.Br.013 (Aust94)	GCA_000167335.1
A0088	South Africa	A.Br.014 (A.Br.101)	SRR2968133, PRJNA302749, SAMN04283839
3631_1C	South Africa	A.Br.014 (A.Br.101)	LGCC00000000
A0252	Zimbabwe	A.Br.014 (A.Br.101)	SRR2968160, PRJNA302749, SAMN04283840
A0455	Mozambique	A.Br.014 (A.Br.101)	SRR2968165, PRJNA302749, SAMN04283841
V770 -NP-1R	USA	A.Br.064 (V770)	AZQO00000000
A0094	South Africa	A.Br.064 (V770)	SRR2968135, PRJNA302749, SAMN04283813
A0096	South Africa	A.Br.064 (V770)	SRR2968191, PRJNA302749, SAMN04283814
A0097	South Africa	A.Br.064 (V770)	SRR2968192, PRJNA302749, SAMN04283815
2,000,031,039	USA	A.Br.064 (V770)	JSZR01000000.1
A1039	Bolivia	A.Br.064 (V770)	LAKZ01000000.1
K8215	Argentina	A.Br.064 (V770)	LGIG01000000.1
ATCC14185	Israel	A.Br.064 (V770)	AZQO00000000.1
BA_3154	Bulgaria	A.Br.008/011 (TEA)	ANFF00000000.1
Sen2Co12	Senegal	A.Br.008/011 (TEA)	CAVC010000000
Smith_1013	Unknown	A.Br.008/011 (TEA)	JNOD00000000.1
CZC5	Zambia	A.Br.005/006 (AncientA)	BAVT00000000.1
A2075	Tanzania	A.Br.005/006 (AncientA)	SRR2968187, PRJNA3–2749, SAMN04283799
H9401	Korea	A.Br.H9401 (Vollum)	NC_017729.1
CDC684	USA	A.Br.007 (Vollum)	NC_012581.1
Vollum	UK	A.Br.007 (Vollum)	AAEP00000000.1
CNEVA-9066	France	B.Br.CNEVA (B-branch)	NZ_AAEN00000000.1
KrugerB	South Africa	B.Br.Kruger (B-branch)	AAEQ00000000.1
HYU01	South Korea	B.Br.001/002 (B-branch)	CP008846
SVA11	Sweden	B.Br.001/002 (B-branch)	CP006742.1
A0442	South Africa	B.Br.001/002 (B-branch)	ABKG00000000.1

2.5. SNP discriminating assay using melt-MAMA

The established SNPs identified as A.Br.101, A.Br.172 and A.Br.173 were designed using Primer 3^{+} software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/). The SNP positioning of the designed SNPs in the Ames ancestor genome and primer sequences are listed in Table 3. Validation was performed on LightCycler™ 96 (Roche®) Real-Time PCR system. The 20 µL reaction included 10 µL DNA diluted in 1× FastStart DNA Green Master (Roche®) with an ancestral forward and a derived forward SNP target primer (GC-clamp: no-GC-clamp) and a common reverse primer (Inqaba Biotec™) with a starting concentration of 0.2 µM depending on the ratio indicated that allowed for separation of melt peaks by at least 5 °C. Multiple replicates of primer ratios of the ancestral, derived and reverse included 1:1:1, 4;1:1, and 1:4:1. Thermocycling parameters were 95 °C for 10 min,

Table 3. Melt-MAMA primers targeting canonical SNPs for the newly discovered phylogenetic SNP branches from this study.

Assay name	Reference genome position[a]	Derived MAMA 5′-3′	Ancestral MAMA 5′-3′	Common reverse 5′-3’	Annealing temperature (°C)
A.Br.101	1,876,363	cggggcggggcggggcgggcTAGTTCGAAAATGGGAGGtC	TAGTTCGAAAATGGGAGGcA	ATTCAAATGCACATGGCAAA	58
A.Br.172	1,876,377	cggggcggggcggggcgggcGGGAGGACTTCCTGATTTGGaT	GGAGGACTTCCTGATTTGGgA	TCCTTTGTGCATTGGGTATTC	58
A.Br.173	2,280,989	cggggcggggcggggcgggcGGAAACACAGGTCTGGCAGaT	GGAAACACAGGTCTGGCAGcA	GCTCCCGTTATTCCCGTCGG	58

[a]SNP position are in relation to *Bacillus anthracis* Ames ancestor chromosome (NC_007530.2).

thereafter at 95 °C for 15 s and 55 °C–60 °C (oligonucleotide dependent on Table 3) for 1 min for 35 cycles. End-point PCR amplicons were subjected to melt analysis using a dissociation protocol comprising of 95 °C for 15 s, followed by incremental temperature ramping (0.1 °C) from 60 °C to 95 °C. SYBR Green fluorescence intensity was measured at 530 nm at each ramp interval and plotted against temperature and observed as the separate melt peaks for each SNP marker.

3. Results

3.1. *B. anthracis* characterization

The *B. anthracis* isolates (Table 1) were identified and confirmed using microbiology and FRET qPCR. All *B. anthracis* were susceptible to penicillin and to gamma-phage, non-haemolytic and non-motile (Table 1). Real-time PCR assay detected the presence of SASP and BAPA genes in all the *B. anthracis* isolates used in this study. Table 3

3.2. Whole genome sequencing SNP phylogeny

B.anthracis strains from NCP (n = 25) and Botswana (n = 1) were sequenced and grouped in the A-clade. Mapping of the NCP *B. anthracis* sequence reads to *B. anthracis* Ames ancestor reference genome identified that a few *B. anthracis* NCP strains lacked plasmid pXO2 (Table 1). The wgSNP phylogeny was constructed using the aligned sequences of *B. anthracis* genomes. Whole genome alignment of filtered parsimony informative SNPs was r used to construct the phylogeny. A total of 4923 SNPs determined in the global *B. anthracis* genomes (n = 58) were used to construct the phylogenetic tree (Fig. 1, Supplementary Fig. S1). The wgSNP analysis showed that NCP *B. anthracis* strains group in the A-clade at different phylogeographic regions (Fig. 2). None of the sequenced genomes in this study grouped in the *B. anthracis* B-clade. The wgSNP analysis grouped the A-clade sequenced genomes into minor sub-clades of A.Br.001/002 (Sterne; n = 2), A.Br.003/004 (n = 22) and A.Br.005/006 (Ancient A; n = 2) (Table 1, Fig. 1; Supplementary Fig. S1). The NCP 2008/9 and 1998 outbreaks strains mostly grouped in the A.Br.101 branch of the A.Br.003/004clade separate from A.Br.013 (Aust94). In this study, the established A.Br.101 branch by Sahl et al. (2016) was further divided into two newly discovered sub-clades namely A.Br.172 and A.Br.173 represent of NCP *B. anthracis* strains isolated in the Ghaap Plateau (Supplementary Fig. S2). These two canSNPs were named after the end-numbering point of Sahl et al. (2016). New melt-MAMA primers were designed from whole genome sequences (Table 3) to discriminate and validated the NCP *B. anthracis* strains in newly identified A.Br.172 and A.Br.173 branches within the established A.Br.101 sub-clade of the A.Br.003/004 clade. The border region *B. anthracis* isolates from Sendelingsdrift and Botswana (SA-Namibia; Botswana-SA) clustered in A.Br.005/006 (Ancient A) (Fig. 2). Furthermore, wgSNP discovery revealed additional SNP positions of the established branches (A.Br.101, A.Br.172, and A.Br.173) (Fig. 1). This study followed the canSNP nomenclature from Sahl et al. (2016) derived from older nomenclature of Van Ert et al. (2007a).

3.2.1. A.Br.001/002 (Sterne) phylogenetic analysis

Sequence analysis of the NCP *B. anthracis* 3080_5A and 3080_1B strains isolated from bovine showed that they cluster with the A.Br.002 (Sterne) sub-clade. These isolates were from a 2009 sporadic anthrax outbreak that affected domesticated bovine in the NCP Ghaap plateau region. Comparative genome analysis showed that *B. anthracis* Sterne, NCP

3080_5A, and NCP 3080_1B lack the pXO2 plasmid (Table 1) and are all defined by A.Br.072 SNP branch. Both sequenced genomes presented -SNPs that are distinctive from *B. anthracis* Sterne genome. *B. anthracis* 3080_1B strain resembles *B. anthracis* Sterne based on 4564 unique SNPs, compared to 3080_5A with 4713 SNPs which is defined by a protracted branch length (Fig. 1, Supplementary Fig. S1).

3.2.2. A.Br.003/004 (A.Br.014) phylogenetic analysis

Most NCP *B. anthracis* strains sequenced (n = 22) in this study clustered into A.Br.003/004 (A.Br.014) SNP branch (Fig. 1, Supplementary Fig. S1), which also consists of the Australia94 and Georgia *B. anthracis* strains. The NCP strains group in the A.Br.101 branch that separates them from Australia94 and Georgia strains (A.Br.013). The A.Br.101 phylogenetic analysis was defined by 1286 unique SNPs. The study showed that the A.Br.101 sub-clade is sub-divided into the A.Br.172 and A.Br.173 branches. Branch A.Br.172 is represented by 13 sequenced NCP genomes namely 3132_1B, 3631_8D, 3517_1C, 2991_1B, 3122_2B, 3517_2C, 3631_4C, 3631_3D, 3618_2D, 3275_2D, 3631_7C, 3008_1B, 2949_1D, and 3 previously sequenced strains from South Africa, Zimbabwe and Mozambique (Sahl et al., 2016). Comparative wgSNP analysis of the A.Br.172 genomes presented a total of 426 parsimony informative-SNPs. According to the ARC-OVI sample registry, the SA A0088 (V20) strain was isolated from 1939 in the Vaalbos National Park. This strain clustered within the A.Br.172 branch but separately from the 1998 and 2008/9 NCP sequenced strains in this study.

The A.Br.173 is represented by 9 genomes namely, 5838, 2991_2B, 3631_1C, 2991_2B, 2110, 3080_3B, JB10, JB25, 3079_1C with a high number of unique SNPs that are distinct from A.Br.172 branch. This branch was defined by 745 parsimony informative-SNPs with high number of unique and non-informative SNPs and other sub-branches with lower bootstrap values. The phylogeny based on wgSNP indicate that the 2009 anthrax outbreak strains in A.Br.173 evolved from the 1998 *B. anthracis* 5838 strain isolated from a hartebeest in the Ghaap plateau, NCP, which also highlights distinctive SNPs. In this study, the 2009 isolates grouped separately from the 1998 isolates (5838 and 2110) within the A.Br.173 branch (Fig. 1). The genome SNP profile of *B. anthracis* JB10 was similar to NCP *B. anthracis* 3079_1C displaying the re-occurrence of a homogenous *B. anthracis*. There were no SNP variations identified in these two genomes despite being isolated from different animals and farms within the 2009 anthrax outbreaks (Table 1). The 2009 anthrax isolates in the A.Br.173 reflect a high diversity of *B. anthracis* strains present in NCP farmlands. The differences in genetic diversity of these isolates is due to that they were isolated from different farms with fenced borders in NCP. *B. anthracis* JB25 was isolated from a sub-adult male *Tragelaphus strepsiceros* on Dikbosch farm in February of 2008, while JB10 was isolated from an adult *Equus quagga* carcass from Vlakfontein farm a few months later. The A.Br.101, A.Br.172 and A.Br.173 differentiating SNP markers were used to validate the sequenced *B. anthracis* strains using melt-MAMA assay.

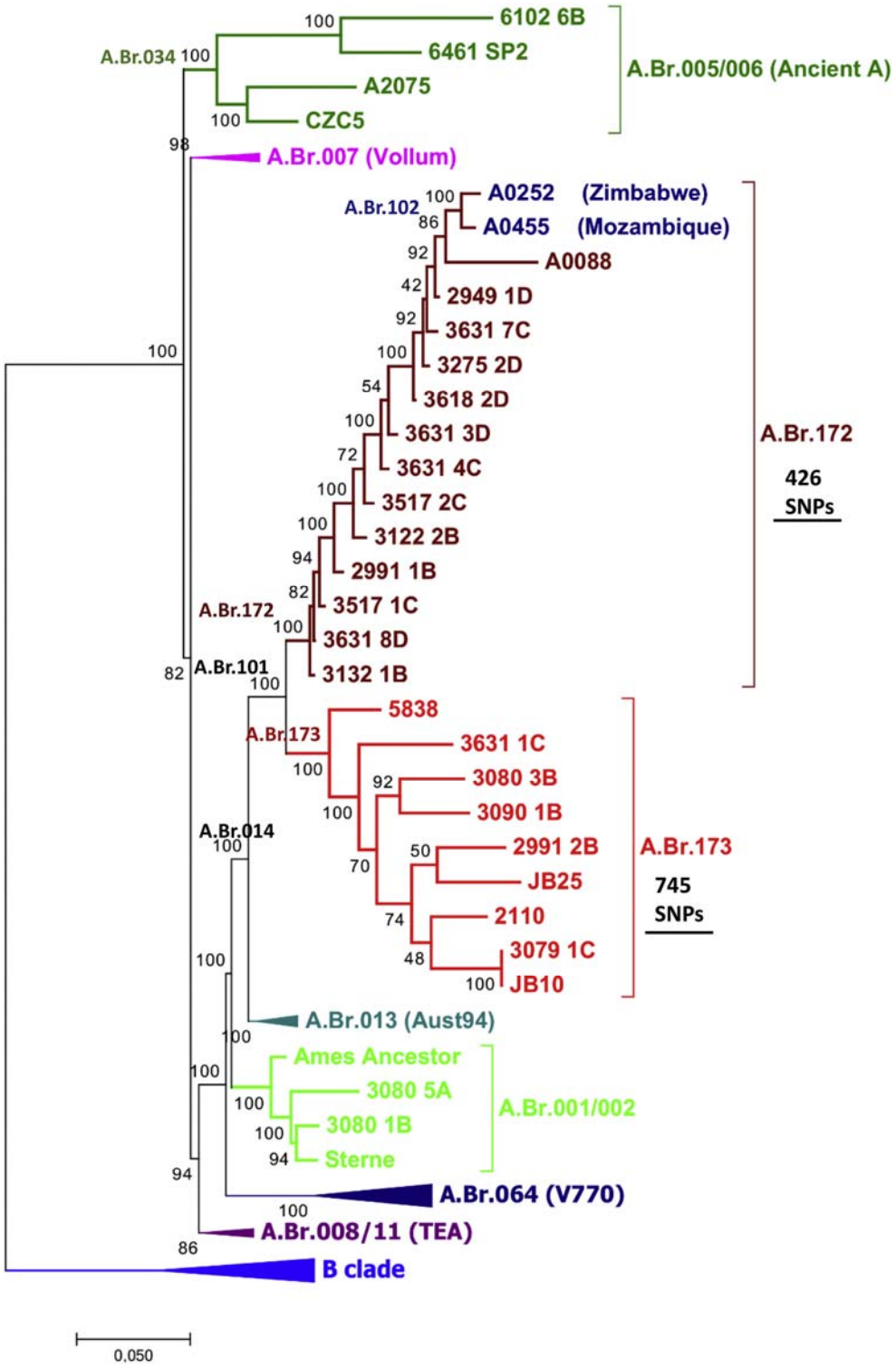

Fig. 1. Global phylogeny of the *Bacillus anthracis* strains indicating the clustering of Northern Cape Province isolates as A.Br.101 (Aust94) in to two unique sub-clades namely A.Br.172 and A.Br.173. A total of 4923 parsimony informative-SNPs from whole genome sequences were analyzed using maximum likelihood to generate the collapsed phylogenetic tree. The sequenced genomes grouped in the A.Br.005/006 (A.B.r.034) (n = 2), dark green, A.Br.101 (n = 22), maroon and red, and A.Br.002 (Sterne) (n = 2), lime green. The major clades are collapsed in this figure, the complete tree is available in the supplementary material (see Supplementary Fig. S1).

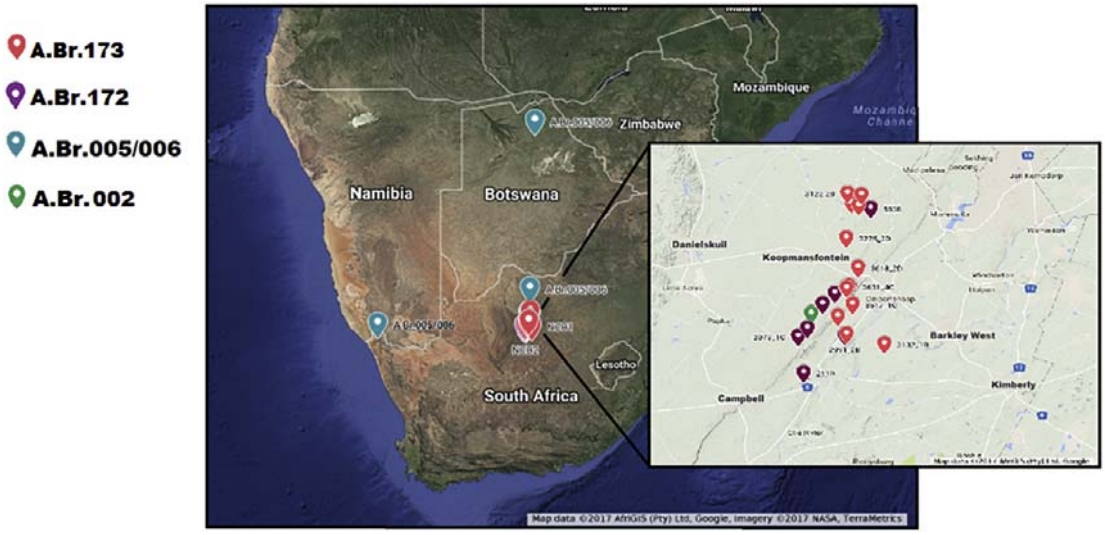

Fig. 2. Phylogeography of the 26 sequenced *B. anthracis* strains that are mapped across southern Africa. The map indicates the two new two sub-clades of the Northern Cape Province with other lineages or clades.

3.2.3. A.Br.005/006 (A.Br.034) phylogenetic analysis

In this study, two sequenced *B. anthracis* genomes namely 6462 SP1 and 6102_6B grouped in the A.Br.005/006 clade. The NCP *B. anthracis* strain 6462_SP1 was isolated from Sendelingsdrift (on the border of South Africa and Namibia). The Botswana 6102_6B strain was included as it borders both Namibia and South Africa and grouped separately from the NCP *B. anthracis* 6462_SP2 genome in the A.Br.005/006 sub-clade. The Botswana strain was differentiated by 503 SNPs from the NCP *B. anthracis* 6462_SP2 strains. The two sequenced *B. anthracis* strains in this study also showed unique SNPs that are distinctive from the sequenced Zambian *B. anthracis* CZC5 and Tanzanian A2075 isolate.

4. Discussion

The phylogenomic structure of *B. anthracis* in this study was consistent with the previously reported clades or SNP lineages (Girault et al., 2014b; Khmaladze et al., 2014; Pearson et al., 2009; Van Ert et al., 2007a; Vergnaud et al., 2016). In this study, all the anthrax report cases in NCP clustered in the A-clade. Phylogenetic analysis defined the predominating clade A into different SNP sub-clades present amongst NCP *B. anthracis* strains. Novel SNPs were discovered and identified amongst the NCP strains that distinguished the prevailing A.Br.101 majority group from A.Br.013 and closely related A.Br.102 branches. The A.Br.014 subclade consists of the A.Br.101 and A.Br.013 SNP branch with Australia94 and Georgia/Turkey strains. The A.Br.101 branch with a previously sequenced South African strain (A0088), and closely related Mozambique and Zimbabwean strains, both group in the A.Br.102 SNP branch. The *B. anthracis* isolates of the 2008/9 anthrax outbreaks in Ghaap plateau, NCP clustered in the A.Br.101 SNP branch. This branch is succeeded by two newly established A.Br.172 and A.Br.173 branches. The A.Br.172 branch consists of sequenced NCP *B. anthracis* strains and the previously sequenced Mozambique and Zimbabwe (A.Br.102) strains that share the A.Br.101 SNP branch (Fig. 1). However, the 2008/9 anthrax outbreaks in NCP have heterogeneity in *B. anthracis* strains with unique SNPs. The 1998 anthrax outbreaks consisted of *B. anthracis* strains that clustered in the A.Br.005/006 (Sendelinglsdrift) (ancient A; n = 1) and A.Br.173 (n = 1) branches with the recognition of

fewer isolates and hence less genomic data being available from these outbreaks. The study suggests that the diversity of the A.Br.173 branch *B. anthracis* strains reported from the 2008/9 anthrax outbreaks in Ghaap plateau likely evolved from 1998 *B. anthracis* 5838 strain. This study couldn't identify common *B. anthracis* SNP branches between the 1998 and 2008/9 anthrax outbreaks in NCP. Nevertheless, there were less strains and data available from the 1998 anthrax outbreaks. A case report of the 2008/9 anthrax outbreaks in the Ghaap plateau showed the presence of homogeneous *B. anthracis* strains that affected different animals across the Ghaap plateau area i.e. clonal outbreaks stemming from a common historical source.

Anthrax is regarded as endemic in KNP and NCP South Africa (Hassim et al., 2017). NCP has a number of state, provincial and privately owned wildlife conservancy areas that service the small communal farms and National Parks such as Kgalagadi Transfrontier Park bordering Botswana and the Richtersveldt Tranfrontier Park bordering Namibia (Hassim et al., 2017). NCP experienced widespread epidemics of animal anthrax outbreaks in 1998 and 2008/9. The 1998 outbreaks mainly occurred in the Vaalbos National Park after heavy rains in 1997 (Henton and Briers, 1998). Similarly at the end of 2007, heavy rainfall is hypothesized to have precipitated anthrax outbreaks during 2008–2009 in the NCP (Nduli, 2009) that affected unvaccinated livestock and wildlife in the region. The NCP veterinary services used vaccination of livestock, treatment of the carcass sites with either chlorine or formalin and burning of the carcasses as control measures during that period. This study used *B. anthracis* isolates that were confirmed positive for anthrax from the samples submitted to the NCP state veterinary services and ARC-OVI at that time.

Protocols and tools for phylogenomic diversity analysis of *B. anthracis* strains from anthrax outbreaks are essential to be developed for control and monitoring purposes (Wielinga et al., 2011). Sequencing technology offers rapid and advanced tools that enable traceability of disease outbreaks and transmission. WGS of more *B. anthracis* isolates enables identification of the discriminative novel SNPs in a given genome wide population (Aarestrup et al., 2012). A genetic database of *B. anthracis* was created in order to delineate the phylogenomic structure of the NCP, South African strains using SNP analysis. In this study, the initially employed melt-MAMA assay could broadly assign the NCP *B. anthracis* strains into major clades such as A.Br.003/004 (A.Br.014), A.Br.001/002 (Sterne) and A.Br.005/006 (Ancient A). However, this assay could not significantly define a high-resolution phylogenetic structure of the NCP *B. anthracis* strains. The SNP/melt-MAMA assay has some limitations with regards to placement of South African *B. anthracis* strains into minor SNP clades due to a paucity of wgSNP data for the region. The melt-MAMA assay was designed from known global *B. anthracis* SNP profiles (Birdsell et al., 2012). This was also found previously with the analysis of French and Georgian *B. anthracis* isolates (Derzelle et al., 2015b; Girault et al., 2014b; Khmaladze et al., 2014; Vergnaud et al., 2016). The phylogenetic structure observed in the NCP *B. anthracis* strains using wgSNPs is the same with previously used MLVA technique or canSNP assays (Hassim, 2016). *B. anthracis* strains from the NCP have unique genotypes that grouped in a separate cluster in the A-clade using MLVA-31 assay (Hassim, 2016). However, in this study there are well-defined branch points with precise lengths that accurately identify the strains for canonical evolutionary relationships. The phylogenetic tree and the identified SNPs laid a baseline for future studies in order to trace anthrax transmission in NCP and southern Africa.

The canSNP group A.Br.014 is a broader lineage that included the established A.Br.101 NCP *B. anthracis* strains, A.Br.013 (Australia94, ABr.003, and the well-established Georgia

strains (Khmaladze et al., 2014) (52-G, 8903-G and 9080-G). Members of this A.Br.014 have also been identified in other countries such as China, Germany, USA, Turkey, England, Great Britain, India, Namibia and Zimbabwe. Some other strains were also reported from the Netherlands (CVI-188678-1 and CVI-56430) (Derzelle et al., 2015a) and grouped separately from the Georgia and Australia94 strains. The 1993 French *B. anthracis* CVI-188678-1 strain isolated from a dairy farm was speculated to have occurred due to importation of contaminated animal products from Asia (Derzelle et al., 2015a) because the lineage is not well established in France. This clade suggests an extensive dispersal since it is also present in South Africa (Khmaladze et al., 2014; Van Ert et al., 2007a). A previous study also indicated that some of the South African *B. anthracis* strains grouped in the A.Br.014 using the melt-MAMA assay (Khmaladze et al., 2014), which was also found in this study with the NCP *B. anthracis* strains grouping in A.Br.101 branch (Fig. 1).

The 2008/9 anthrax outbreaks consisted of *B. anthracis* strains clustering in different branches (A.Br.101 and A.Br.001/002(Sterne)). The NCP phylogenomic structure indicated that different *B. anthracis* sub-clades (A.B.172 and A.Br.173) resulted in the anthrax outbreaks in the NCP along the Ghaap plateau escarpment area (Supplementary Fig. S2). *B. anthracis* strains from these two established SNP branches were found in the same plateau region at the top of the escarpment. The distribution of these diverse *B. anthracis* strains in NCP can be attributed to wildlife movement on livestock farms as well as controlled (fenced) game reserves. The majority of the A.Br.172 branch strains were found at the top and base of the Ghaap Plateau escarpment, which indicates dispersal via topographical factors such as water drainage. wgSNP analysis placed the previously sequenced A0088 (V20) strain (Sahl et al., 2016) with an unknown geographical location into this monophyletic A.Br.172 branch, suggesting that it could have come from NCP. As indicated the ARC-OVI sample registry contains a record of the SA A0088 (V20) strain that was isolated from 1939 in the Vaalbos National Park. However, A0088 exhibited a vastly different SNP profile that is discernable within the A.Br.172 grouping by its' extended branch length (Fig. 1, Supplementary Fig. S1).

The A.Br.173 branch was affected by high number of informative SNPs and long branches (Fig. 1, Supplementary Table S1) defined by few numbers of genomes. Additional genome wide profiling of the A.Br.173 is needed to further define this branch. This branch included strains (2110 and 5838) from the 1998 and 2008/9 outbreaks in the NCP. The A.Br.173 branch is polymorphic and is represented by lower bootstrap values (Fig. 1; Supplementary Fig. S1). The 1998 *B. anthracis* two strains are distinct and cluster in different sub-branches within the A.Br.173 branch. The *B. anthracis* 3079_1C and JB10/NC-14 strains are an example of the same genotype occurring in the 2008 and 2009 anthrax outbreaks. These strains were recovered from the same farm (Vlakfontein), at different times and from different carcasses. These isolates again demonstrate clonal outbreaks.

The Sendelingsdrift, NCP consisted of *B. anthracis* 6461 SP2 strain that grouped in A.Br.005/006, also referred to as the Ancient A clade. The canSNP A.Br.005/006 clade mostly consist of strains from central Africa (Beyer et al., 2012) and southern Africa (Botswana (6102_6B strain in this study) and KNP, South Africa). However, other Ancient A *B. anthracis* strains were reported in Australia, and South Korea (Sahl et al., 2016) showing that it has a global distribution. In this study, the Sendelingsdrift *B. anthracis* 6461_SP2 strain showed unique SNPs and grouped separately from the closest globally sequenced Botswana 6102_6B and Zambian CZC5 strains highlighting regional differences.

The A.Br.064 (V770) is another clade that was present in the Vaalbos National Park. This sub-clade was previously genotyped using MLVA-31 in South Africa (Hassim, 2016), and some of the *B. anthracis* strains were subjected for wgSNP comparison (Sahl et al., 2016). In this study, none of the sequenced *B. anthracis* strains grouped in the A.Br.0064 (V770) clade. This clade also consists of the U.S vaccine V770-NP-1R strain, Argentinian A1030 and SA A0094 (V22) strain (Sahl et al., 2016). Other sequenced South African *B. anthracis* strains that group in the V770 sub-clade include A0096 and A0097 isolated from a 1992 anthrax outbreak in the Vaalbos National Park. The South African genomes grouped separately from the USA and Argentina V770 sub-clades (Sahl et al., 2016) (Supplementary Fig. S1).

With the exception of the established A.Br.101 branch in the Ghaap plateau area, the 2009 anthrax outbreaks also consisted of other distinct *B. anthracis* strains that grouped in the A.Br.002 (Sterne). Worldwide, A.Br.002 (Sterne) groups together with A.Br.001 (Ames) in the A.Br.001/002 (Ames/Sterne) clade (Van Ert et al., 2007b). The A.Br.001 (Ames) consist of *B. anthracis* Ames ancestor, Ames, U.S.A 0248, Florida A2012 and other strains from China, Japan and Indonesia, while A.Br.002 (Sterne) includes the South African *B. anthracis* Sterne vaccine strain and a single clonal cluster of French strains (Derzelle et al., 2016; Girault et al., 2014b; Vergnaud et al., 2016). The A.Br.002 (Sterne) clade was resolute in the bovine anthrax case in this study. The Ghaap Plateau is a wildlife-livestock farming area where *Tragelaphus strepsiceros* (kudu) is free roaming across the fenced livestock and game farms. Therefore, both wildlife-livestock were affected during the 2009 anthrax outbreaks. The two sequenced *B. anthracis* genomes frombovine grouped in the A.Br.002 (Sterne) as different sub-clade with novel SNPs. These strains share common SNPs with *B. anthracis* Sterne in the A.Br.075 branch, designated by Sahl et al., 2016. The A.Br.002 (Sterne) isolates were not re-isolates of the Sterne vaccine strains since the isolates were recovered from non-vaccinated animals. The two genomes in this branch lacked the pXO2 (Table 1). *Bacillus anthracis* 3631_1C strain that lacks a pXO2 has already been reported in the Ghaap Plateau, NCP (Lekota et al., 2015; Lekota et al., 2018). *B. anthracis* strains that loose pXO2 have been recovered from soil and/or after long term storage (Marston et al., 2005). However, other *B. anthracis* strains (3631_4C, 3631_3D and 3631_8D) with both plasmids were also isolated from the same carcass. There is no clear understanding when exactly the strains lose their plasmids, and this needs further investigation in order to obtain insight into pXO2 plasmid stability.

5. Conclusion

The melt-MAMA assay with the existing primers broadly placed the NCP *B. anthracis* strains into major A-sub-clades. The assay can substantively be used to define the *B. anthracis* population structure within the geographical region; provided more SNP markers are assigned based on the availability of genomic data. A high-resolution phylogenetic structure of NCP *B. anthracis* strains was established using wgSNP analysis tree that positioned the genetic diversity of the NCP *B. anthracis* strains in comparison with worldwide strains. Major and novel SNP branches from the A-clade were determined in this study. The wgSNP analysis of the NCP *B. anthracis* strains showed the prevailing A.Br.101 (Aust94) branch is dissimilar from the A.Br.013 sub-clade (Australia94 and Georgia strains). The A.Br.101 branch was well defined with distinctive *B. anthracis* strains that constituted 9 branches. This study shows that the sparsely represented Mozambique and Zimbabwean (A.Br.102) *B. anthracis* strains belong to this branch. The monophyletic A.Br.172 branch also includes these two *B. anthracis* strains. The availability of high-quality genomes in well-populated database will set the stage for epidemiological identification of the *B. anthracis* strains to

trace anthrax outbreaks/cases in South Africa. Furthermore, the number of genomes that become available can later influence the topology of the *B. anthracis* strains in NCP, South Africa.

Author's contribution

Kgaugelo Edward Lekota participated in running the experiments, bioinformatics analysis and writing the manuscript. Ayesha Hassim participated in running the experiments and writing the manuscript. Charles Hefer assisted in running of the bioinformatics scripts and drafting the manuscript. Evelyn Madoroba and Henriette van Heerden participated in the design of the study, drafting the manuscript, revising it critically and provided funding.

Funding

This study was funded by the National Research Foundation (NRF), Institute of Tropical Medicine (ITM), Belgium, and Agriculture Sector Education Training Authority (AgriSETA). The funding organizations only supplied funds and did not review or supply any other input.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

The authors would like to thank the farmers of the Northern Cape for their wholehearted participation in the study of anthrax in the region. We also wish to extend our appreciation to the Northern Cape Provincial government and State Veterinary Services for allowing us to conduct this study and improve our understanding of this zoonotic pathogen.

References

Aarestrup, F.M., Brown, E.W., Detter, C., Gerner-Smidt, P., Gilmour, M.W., Harmsen, D., Hendriksen, R.S., Hewson, R., Heymann, D.L., Johansson, K., Ijaz, K., Keim, P.S.,Koopmans, M., Kroneman, A., Lo Fo Wong, D., Lund, O., Palm, D., Sawanpanyalert,P., Sobel, J., Schlundt, J., 2012. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. *Emerg. Infect. Dis*. 18, e1.

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

Achtman, M., Wagner, M., 2008. Microbial diversity and the genetic nature of microbial species. *Nat. Rev. Microbiol*. 6, 431–440.

Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data.

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

Birdsell, D.N., Pearson, T., Price, E.P., Hornstra, H.M., Nera, R.D., Stone, N., Gruendike, J.,

Kaufman, E.L., Pettus, A.H., Hurbon, A.N., Buchhagen, J.L., Harms, N.J., Chanturia, G., Gyuranecz, M., Wagner, D.M., Keim, P.S., 2012. Melt analysis of mismatch amplification mutation assays (melt-MAMA): a functional study of a cost-effective SNP genotyping assay in bacterial models. *PLoS One* 7, e32866.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.

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

Derzelle, S., Girault, G., Kokotovic, B., Angen, O., 2015a. Whole genome-sequencing and phylogenetic analysis of a historical collection of *Bacillus anthracis* strains from Danish cattle. *PLoS One* 10, e0134699.

Derzelle, S., Girault, G., Roest, H.I., Koene, M., 2015b. Molecular diversity of *Bacillus anthracis* in the Netherlands: investigating the relationship to the worldwide population using whole-genome SNP discovery. *Infect. Genet. Evol.* 32, 370–376.

Derzelle, S., Aguilar-Bultet, L., Frey, J., 2016. Comparative genomics of *Bacillus anthracis* from the wool industry highlights polymorphisms of lineage A.Br.Vollum. *Infect Genet. Evol*. 46, 50–58.

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

Ellerbrok, H., Nattermann, H., Ozel, M., Beutin, L., Appel, B., Pauli, G., 2002. Rapid and sensitive identification of pathogenic and apathogenic *Bacillus anthracis* by real-time PCR. *FEMS Microbiol. Lett*. 214, 51–59.

Girault, G., Blouin, Y., Vergnaud, G., Derzelle, S., 2014a. High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. *BMC Genomics* 15, 288
.
Girault, G., Blouin, Y., Vergnaud, G., Derzelle, S., 2014b. High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. *BMC Genomics* 15, 288.

Hassim, A., 2016. Distribution and Characterisation of South African *Bacillus anthracis* Strains and their Associated Bacteriophages. Department of Veterinary Tropical Diseases, University of Pretoria, Onderstepoort.

Hassim, A., Dekker, E.H., Byaruhanga, C., Reardon, T., Van Heerden, H., 2017. A retrospective study of anthrax on the Ghaap plateau, northern cape province of South Africa, with special reference to the 2007-2008 outbreaks. *Onderstepoort J. Vet. Res.* 84, e1–e15.

Henton, M., Briers, G., 1998. Anthrax Spores Isolated above Burning Carcases, OIE International Congress With WHO-Cosponsorship on Anthrax, Brucellosis, CBPP, Clostrial and Mycobacterial Diseases, Berg-En-dal (South Africa), 9–15 Aug 1998. Onderstepoort Veterinary Inst.

Hugh-Jones, M.E., de Vos, V., 2002. Anthrax and wildlife. *Rev. Sci. Tech.* 21, 359–383.
Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Schupp, J.M., Okinaka, R., Jackson, P.J., Hugh-Jones, M.E., 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis. 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.

Khmaladze, E., Birdsell, D.N., Naumann, A.A., Hochhalter, C.B., Seymour, M.L., Nottingham, R., Beckstrom-Sternberg, S.M., Beckstrom-Sternberg, J., Nikolich, M.P., Chanturia, G., Zhgenti, E., Zakalashvili, M., Malania, L., Babuadze, G., Tsertsvadze, N., Abazashvili, N., Kekelidze, M., Tsanava, S., Imnadze, P., Ganz, H.H., Getz, W.M., Pearson, O., Gajer, P., Eppinger, M., Ravel, J., Wagner, D.M., Okinaka, R.T., Schupp, J.M., Keim, P., Pearson, T., 2014. Phylogeography of *Bacillus anthracis* in the country of Georgia shows evidence of population structuring and is dissimilar to other regional genotypes. *PLoS One* 9, e102651.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol*. 33, 1870–1874.

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.

Lekota, K.E., Mafofo, J., Madoroba, E., Rees, J., van Heerden, H., Muchadeyi, F.C., 2015. Draft genome sequences of two south African Bacillus anthracis strains. *Genome Announc*. 3.

Lekota, K.E., Bezuidt, O.K.I., Mafofo, J., Rees, J., Muchadeyi, F.C., Madoroba, E., van Heerden, H., 2018. Whole genome sequencing and identification of *Bacillus endophyticus* and *B. anthracis* isolated from anthrax outbreaks in South Africa. *BMC Microbiol*. 18, 67.

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

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
.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20, 1297–1303.

Nduli, N., 2009. Department of Agriculture 2008/09 Annual Report. Directorate Agricultural Information Services, South Africa.

Organization, W.H., Epizootics, I.O.o, 2008. Anthrax in Humans and Animals. World Health Organization.

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.

Sahl, J.W., Pearson, T., Okinaka, R., Schupp, J.M., Gillece, J.D., Heaton, H., Birdsell, D., Hepp, C., Fofanov, V., Noseda, R., Fasanella, A., Hoffmaster, A., Wagner, D.M., Keim, P., 2016. A *Bacillus anthracis* genome sequence from the Sverdlovsk 1979 autopsy specimens. *MBio* 7.

Sterne, M., 1967. Distribution and economic importance of anthrax. *Fed. Proc*. 26, 1493–1495.

Turnbull, P.C.B., 2008. *Anthrax in Humans and Animals*, 4th ed. World Health Organization.

Van Ert, M.N., Easterday, W.R., Huynh, L.Y., Okinaka, R.T., Hugh-Jones, M.E., Ravel, J., Zanecki, S.R., Pearson, T., Simonson, T.S., U'Ren, J.M., Kachur, S.M., Leadem-Dougherty, R.R., Rhoton, S.D., Zinser, G., Farlow, J., Coker, P.R., Smith, K.L., Wang, B., Kenefic, L.J., Fraser-Liggett, C.M., Wagner, D.M., Keim, P., 2007a. Global genetic population structure of *Bacillus anthracis. PLoS One* 2, e461.

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

Vergnaud, G., Girault, G., Thierry, S., Pourcel, C., Madani, N., Blouin, Y., 2016. Comparison of French and worldwide Bacillus anthracis strains Favors a recent, post-Columbian origin of the predominant north-American clade. *PLoS One* 11, e0146216.

Wielinga, P.R., Hamidjaja, R.A., Agren, J., Knutsson, R., Segerman, B., Fricker, M., Ehling-Schulz, M., de Groot, A., Burton, J., Brooks, T., Janse, I., van Rotterdam, B., 2011. A multiplex real-time PCR for identifying and differentiating *B. anthracis* virulent types. *Int. J. Food Microbiol.* 145 (Suppl. 1), S137–S144.