Genomics 117 (2025) 111074

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Research Paper

Comparative genomics of *Bacillus anthracis* A and B-clades reveals genetic variation in genes responsible for spore germination

Sankwetea P. Mokgokong [a,b,*], Ayesha Hassim [a], Tendo Mafuna [c], Wendy C. Turner [d], Henriette van Heerden [a], Kgaugelo E. Lekota [a,b]

[a] Department of Veterinary Tropical Diseases, University of Pretoria Veterinary Campus, Onderstepoort, South Africa
[b] Unit of Environmental Sciences and Management, North-West University, Potchefstroom, South Africa
[c] Department of Biochemistry, University of Johannesburg, Auckland Park, South Africa
[d] U.S. Geological Survey, Wisconsin Cooperative Wildlife Research Unit, Department of Forest and Wildlife Ecology, University of Wisconsin–Madison, Madison, WI 53706, USA

ARTICLE INFO

Keywords:
Bacillus anthracis
pan-genomics
Tryptophan operon
Bacillus collagen-like protein of *anthracis*
Single-nucleotide polymorphism (SNP)
Whole genome sequencing

ABSTRACT

Bacillus anthracis, the causative agent of anthrax, is composed of three genetic clades (A, B, and C). Clade-A is the most common and distributed worldwide, B-clade has a narrow geographic distribution, and C-clade is rare. South Africa's Kruger National Park (KNP) has high diversity of *B. anthracis*, with strains from A and B clades described from its northernmost region, Pafuri. We employed whole genome sequencing to investigate the genomic diversity of *B. anthracis* strains isolated from animal carcasses ($n = 34$) during the 2012–2015 outbreaks. Whole-genome single-nucleotide polymorphism (wgSNP) analysis assigned the 2012–2015 *B. anthracis* genomes to the A-clade branch, distributed across the branch's two minor sub-clades A.Br.005/006. Additionally, pan-genomic analysis distinguished the A- and B-clade genomes, identifying unique accessory genes. Notable genetic differences include the biosynthetic spore cell wall genes; long-chain fatty acid CoA ligases (*FaD13*), *Bacillus* collagen-like protein of *anthracis* (BclA) involved in the exosporium germination, as well as a truncated murein DD-endopeptidase (*mepH*) found in the pXO2 plasmid of the B-clade strains. The tryptophan synthase subunit alpha gene (*trpA*), which results in a pseudogene in B-clade genomes separates the A- and B-clade genomes. These differences in biosynthetic cell wall genes suggest variation in adaptability or cell growth of the B-clade strains in the environment, further influenced by the truncation of the *trpA* gene involved in spore germination. The A. Br.005/006-clade strains in KNP exhibit higher genetic diversity, which may enhance their resilience to environmental stressors. In contrast, the KNP B-clade (B.Br.001/002) strains show limited genetic variation, potentially reducing their adaptability. This pattern is evident through whole-genome SNP analysis and pan-genomics investigating the evolution of *B. anthracis*.

1. Introduction

Bacillus anthracis, the causative agent of anthrax, is a Gram-positive endospore-forming soil bacterium that can persist in soils for extended periods. The disease affects herbivores, with regular outbreaks occurring in the endemic regions. Anthrax occurs worldwide; however, there are re-occurring outbreaks in endemic regions where the soil type and environmental factors favour the *B. anthracis* endospore survival. Outbreaks in these endemic regions are associated with alkaline, nutrient (calcium and/or nitrogen) rich soils and changes in seasonal climatic factors such as rainfall and ambient temperatures [1,2].

Bacillus anthracis is considered a monomorphic organism with low genetic diversity [3,4] On a worldwide scale, *B. anthracis* genetic diversity has been well-described [5,6,7,8,9,10,11]. Globally, the genetic structure of *B. anthracis* strains is classified into three primary clades (A, B, and C), with the A clade being the most common and globally prevalent [4]. *B. anthracis* B clade strains have been detected in more limited or specific geographical locations [12,10]. Although lineages are endemic to regions such as the Kruger National Park (KNP) of South Africa, Zimbabwe, Mozambique, and France, there is evidence of multiple variants co-circulating [6,10,4].

Differences between A- and B-clade strains of *B anthracis* have been

* Corresponding author at: Unit of Environmental Sciences and Management, North-West University, Potchefstroom, South Africa.
E-mail address: prudent.mokgokong@nwu.ac.za (S.P. Mokgokong).

https://doi.org/10.1016/j.ygeno.2025.111074
Received 13 June 2024; Received in revised form 9 May 2025; Accepted 3 July 2025
Available online 7 July 2025

linked to the tryptophan biosynthesis pathway [13,14,15]. The inability of the *B. anthracis* B-lineage strains to synthesize tryptophan will have detrimental effects on their survival in the environment [14]. Studies on tryptophan dependence can provide information on the intraspecific evolution of *B. anthracis*. Moreover, the Bacillus collagen-like protein of anthracis (BclA) of the exosporium [16] varies among *B. anthracis* strains. In the BclA protein, repeat motifs vary in length and range between a single to 100 copies [17]. When the variable number of tandem repeat (VNTR) spans across a coding region, the repeat length may influence the species' phenotype. In particular, the Bams13 Multi-Locus Variable-number Tandem Repeat Analysis (MLVA) marker used for genotyping *B. anthracis* strains exhibits the greatest variability in fragment size among VNTRs and influences phenotypic presentation on the endospore surface [16]. In this study, we investigated functional differences in the genomic composition of strains found in the A- and B-clades, to understand why B-clade strains are less common than A-clade strains.

In South Africa, the rare B-clade (B.Br.001) strains were dominant in the northern region of Pafuri, while the A-clade isolates were dominant in the central region of the park [10]. However, since the 1990s, only the A-clade isolates were observed in the northern parts of the Pafuri region of KNP [7,10]. Globally, there are fewer reports of B-clade strains, and some lineages of *B. anthracis* seem to be disappearing [12]. This could be related to genetic variations in genes that are important for phenotypic traits. In the KNP, spore survival in the northern parts (Pafuri) is supported by neutral-to-alkaline soils and the increased calcium levels [18,10], particularly as most outbreaks were attributed to B-clade strains in this park prior to 1975.

This genetic diversity is also evident when comparing the enzootic regions in South Africa. The *B. anthracis* strains from the Northern Cape Province and KNP groups occur in heterogenous genetic lineages that include A.Br.005/006 (Ancient A), A.Br.001/002 (Ames/Sterne), A.Br.0057 (V770), and B.Br.010 (B-branch) clades with considerable geographic and genetic diversity of the A.Br.005/006 clade observed in KNP [7]. KNP consists of heterologous strains that have mostly been genotyped based on MLVA. Fine-scale evolutionary relationships of *B. anthracis* within KNP are not fully resolved because most of the previous work has been performed by MLVA [6]. Laboratory studies investigating genetic diversity are limited; however, a protein expression comparison of seven wild and four laboratory *B. anthracis* A-clade strains grown under identical conditions revealed that wild strains are more highly primed for sporulation, a distinction that may assist in differentiating wild from laboratory strains during anthrax outbreaks [19]. This presents a challenge when assessing the evolution and population dynamics of monomorphic bacteria, such as *B. anthracis*; therefore, describing the genomic diversity of anthrax outbreaks could help to trace the dissemination of *B. anthracis* in the KNP. SNP analysis exploits the complete genomic region and is a powerful tool to genotype monomorphic pathogenic bacteria such as *B. anthracis*.

The genetic differences between the A and B strains suggest that the environment in South Africa may have played a role in the evolution of different *B. anthracis* clades. To assess genetic variations and genomic diversity between A- and B-clade *B. anthracis* strains, we focused on strains found in South Africa's Kruger National Park (KNP), which is known for its high genetic diversity of *B. anthracis* strains, particularly within the A-clade. Additionally, variations in genetic makeup may influence *B. anthracis* phenotypic traits, and these were evaluated in the study. Whole genome sequencing was employed to investigate the genomic diversity of *B. anthracis* strains isolated from animal carcasses in KNP between 2012 and 2015, and pan-genomics was used to explore the evolution of *B. anthracis*.

2. Methods and materials

2.1. Sample collection

A culture collection of *B. anthracis* strains ($n = 34$) isolated from animals in Pafuri, KNP, was used in this study. These isolates were collected from animal carcasses diagnosed with anthrax in 2012–2015. The Pafuri *B. anthracis* strains were isolated from plains zebra (*Equus quagga*, $n = 3$), African elephant (*Loxodonta africana*, $n = 2$), impala (*Aepyceros melampus*, $n = 27$), nyala (*Tragelaphus angasii*, $n = 1$), and white rhinoceros (*Ceratotherium simum*, $n = 1$). A total of 125 *B. anthracis* isolates were used for comparative genomics analysis. This included the genomes sequenced for this study ($n = 34$), and some global strains ($n = 91$) available on GenBank.

2.2. DNA isolation and whole genome sequencing

Bacillus anthracis strains were inoculated in 2 mL of nutrient broth and incubated at 37 °C overnight; the cells were harvested by centrifugation. Genomic DNA of *B. anthracis* strains ($n = 34$) was extracted from the harvested cells using the DNA Blood Mini Kit (Qiagen, Germany) following the manufacturer's protocol for Gram-positive bacteria with 20 mg/mL lysozyme (Sigma Aldrich, United States of America). The DNA was quantified on the qubit fluorometric quantization system using the Invitrogen™ Broad Range assay kit (Thermo Fisher Scientific, United States of America). The quality of the DNA was analyzed by electrophoresis on a 0.8 % agarose gel using ethidium bromide and visualized under UV-light. DNA extracts of the *B. anthracis* strains were processed, and libraries were prepared using the Nextera XT DNA Sample Prep Kit (Illumina, United States of America). Sequencing of paired end libraries was performed on the Illumina MiSeq sequencer using the 200-cycle SBS (sequencing by synthesis) sequencing v4 kit (Illumina, United States of America).

2.3. Genome quality assessment, assembly, and annotation

The raw reads quality was assessed using the FastQC software 0:10.1 [20] and the sequence adapters and ambiguous nucleotides were removed using Trimommatic (v0.32). The paired end trimmed reads of *B. anthracis* strains were *de novo* assembled using Shovill Faster SPAdes v1.1.0 pipeline [21]. The minimum contig length was set to 500 bp and kmer sizes 21, 33, 55, 77, 99, 127 were used for the assembly. CheckM [22] was additionally used to assess the potential contaminants in individual assembled *B. anthracis* genomes. Genomes that did not meet the criteria of 98 % completeness for *B. anthracis* were excluded for downstream analysis. Quast v 2.3 [23] was used to evaluate the draft genome assemblies ($n = 125$). The generated contigs were ordered using the moving contigs function in Mauve version 2.3.1 [24]. The assemblies were aligned to the reference *B. anthracis* Ames ancestor. Contigs not aligned to the reference genome were checked for sequence similarity using the nucleotide Basic Local Alignment Tool (BLASTn) [25] to confirm if they were *B. anthracis*. Draft assemblies were annotated using Rapid Annotation Using Subsystem Technology (RAST) annotation server [26,27,28]. The *B. anthracis* genomes were annotated using Prokka v.1.14.0 [29].

2.4. Read mapping and single nucleotide polymorphism (SNP) variant detection

The trimmed reads of the *B. anthracis* strains ($n = 34$) were aligned to the *B. anthracis* Ames ancestor reference genome (GenBank: AE017334.2) using the Burrows-Wheeler Aligner (BWA) [30]. Sequence reads of *B. anthracis* global genomes were downloaded and retrieved using the SRA-tool kit (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software). These included *B. anthracis* complete and draft genomes from different canonical clades available from NCBI Genbank

(http://www.ncbi.nom.nih.gov). SAMtools v1.10 [30,31] was deployed to convert the aligned mapped reads to bam files and to sort and index the aligned sequenced reads. Picard-tools (http://picard.sourceforge.net/) was used to generate a sequence dictionary of the *B. anthracis* Ames ancestor, to mark duplicate reads and to build bam index of the mapped reads. SNPs were called using the UnifiedGenotyper method in GATK v4.0.12.0 [32,33,34], and ambiguous variants were filtered prior to selecting variants. SNPs positioning sets were deducted from the aligned genomes of *B. anthracis* Ames ancestor using molecular evolutionary genetics analysis software (MEGAX) [35]. The core SNPs were used for the phylogenetic tree construction using MEGAX [35] and/or Bayesian Evolutionary Analysis by Sampling Trees (BEAST) v2 [36].

2.5. Pan-genome analysis

Pafuri *B. anthracis* sequenced and globally available genomes (n = 125) were used to determine the pan-genome structure of *B. anthracis*, which included both A-and B-clade determined using Roary [37,38]. Similarity searches between the coding domain sequences (CDS) of assembled genomes were conducted using pair-wise BLASTp [25] and Markov Cluster Algorithm (MCL) [39]. Clusters were created, paralogs identified, and the isolates were ordered by presence/absence of orthologs [37]. Pan-genome clusters were defined as follows: core-genes present in all isolates; soft core-genes present in at least 95 % of isolates; shell-genes present between 15 and 95 % of isolates; cloud-genes in less than 15 % of isolates [40,41]. BEAST2 [36] was used to construct the phylogenetic tree of the aligned accessory genes using default parameters. The phylogenetic tree was visualized using Figtree v1.16.6 [42] and ITOL [43]. Binary-accessory genes that indicated the presence-absence of unique genes were extracted from A- and B-clade *B. anthracis* genomes using Roary [37]. The unique genes were compared with annotated sequence data from Prokka [29] and further compared and validated with output files annotated by RAST server [26]. Multiple nucleotide sequence alignment for the unique accessory-binary genes were performed using Multiple Alignment using Fast Fourier Transform (MAFFT) [44] and phylogenetic trees constructed using BEAST [36].

2.6. Genetic variations between A- and B-clade B. anthracis strains using tryptophan operon and bclA genes

Multiple sequence alignment of the tryptophan operon (*trp*) genes were performed using MAFFT [44]. The phylogeny of the genes was inferred using the Maximum likelihood method of evolution in MEGAX [45]. Phylogenetic trees were visualized on FigTree v1.16.6 [42] and ITOL [43]. The Bacillus collagen-like protein of anthracis (*bclA)* gene sequences (n = 117) were extracted among the A and B-clade annotated *B. anthracis* genomes and submitted to tandem repeat finder (TRF v 4.09) [46] to determine the copy number of repeats. The *bclA* sequences were translated into amino acid sequences using Molecular Evolutionary Genetics Analysis (MEGAX) [45], aligned using MAFFT [44] and the phylogeny was inferred using the maximum parsimony method.

Accession number: The sequenced *B. anthracis* strains are available under the BioProject ID PRJNA1075343.

3. Results

3.1. Phylogenetic diversity of B. anthracis strains using wgSNP

To infer the phylogeny of the sequenced Pafuri *B. anthracis* isolates (n = 32), paired end sequencing was performed (Table 1). Filter passed, trimmed (high quality) reads were mapped to the *B. anthracis* Ames ancestor as a reference genome. Whole genome single nucleotide

Table 1
Summary statistics for the *de novo* assembly of the sequenced Pafuri *Bacillus anthracis* strains.

Strain ID	Location	Species	Year	Number of contigs	Genome length	N50
Ames ancestor*	United States of America	Unknown	Unknown	3	5,503,926	5,227,419
AX2012291	Pafuri	White Rhinoceros	2012	61	5,459,246	225,720
AX2012435	Pafuri	Elephant	2012	75	5,483,514	365,209
DS201334	Pafuri	Burchell's zebra	2013	320	5,452,352	38,647
DS201333	Pafuri	Impala	2013	67	5,455,546	1,162,137
DS201316	Pafuri	Impala	2013	56	5,454,271	1,161,939
DS201317	Pafuri	Nyala	2013	72	5,455,107	565,551
RL2014115RbM	Pafuri	Impala	2014	38	5,465,476	1,161,938
SVD201453	Pafuri	Impala milk	2014	34	5,460,646	596,642
1298	Pafuri	Impala	2015	111	5,503,314	426,149
1365	Pafuri	Impala	2015	42	5,457,919	1,052,322
1370	Pafuri	Impala	2015	37	5,453,394	404,441
12,460,315	Pafuri	Burchell's zebra	2015	38	5,463,529	596,566
12,610,315	Pafuri	Burchell's zebra	2015	55	5,505,748	596,647
13,030,315	Pafuri	Impala	2015	52	5,484,198	565,838
13,610,315	Pafuri	Impala	2015	29	4,661,115	614,414
13,680,315	Pafuri	Impala	2015	32	5,460,411	596,643
1257032015UP	Pafuri	Impala	2015	33	5,459,929	565,919
1367PETA	Pafuri	Impala	2015	33	5,459,637	1,197,885
DS201505	Pafuri	Impala	2015	40	5,456,388	412,401
DS20150663	Pafuri	Impala	2015	33	5,454,537	455,808
DS20150761	Pafuri	Impala	2015	53	5,482,839	596,643
DS201510	Pafuri	Impala	2015	36	5,453,456	483,213
DS201523	Pafuri	Impala	2015	48	5,465,545	1,162,327
Rbk32	Pafuri	Impala	2015	33	5,460,646	596,642
DS201525	Pafuri	Impala	2015	966	5,508,258	10,487
DS201577	Pafuri	Impala	2015	33	5,460,804	1,197,885
DS201513	Pafuri	Impala	2015	53	5,455,255	267,729
13,670,315	Pafuri	Impala	2015	31	5,450,054	472,345
DS201584	Pafuri	Impala	2015	33	5,455,976	1,162,214
DS201588	Pafuri	Impala	2015	35	5,457,146	703,210
KM20150330	Pafuri	Elephant tusk	2015	60	5,485,514	365,209
RL201531	Pafuri	Impala	2015	794	5,530,942	12,351

* *Bacillus anthracis* Ames ancestor was used as a reference. Plains zebra (*Equus quaggai*), African elephant (*Loxodonta africana*), impala (*Aepyceros melampus*), nyala (*Tragelaphus angasii*), and white rhinoceros (*Ceratotherium simum*).

polymorphisms (wgSNPs) were determined using Pafuri *B. anthracis* strains (n = 32) and global genomes obtained from NCBI (Table 1). The wgSNP phylogeny was defined by 7713 parsimony informative SNPs that clustered the KNP and canonical global *B. anthracis* genomes (Fig. 1). Other compared South African clades noticeable included previously reported A.Br.001/002, A.Br.101, A.Br.102 (Aust94) and A. Br.064 (V770). The Pafuri *B. anthracis* strains in the A.Br.005/006 (Ancient A) clade and available B.Br.001/002 (B-clade) genomes were resolved (Fig. 1B). The dominant A.Br.005/006 clade could further be assigned into three different sub-clades, dispersed across different outbreak years. Previous outbreaks (2012–2014) were attributed to greater kudu, however, the 2015 outbreak was attributed to impalas and constitutes most of the sequenced isolates ($n = 24$). The Pafuri isolates appear to have diverse genotypes within the A.Br.005/006 clade (Fig. 1).

3.2. De novo assembly and annotation

Sequence reads of *B. anthracis* strains were *de novo* assembled and resulted in draft genomes ($n = 32$) of different sizes. The size of the draft genomes was approximately 5.5 Mb with contigs greater than 1000 bp (*de novo* assembled genomes are in bold in Table 1). The generated contigs for the Pafuri KNP isolates varied from 28 to 99. The minimum length of contigs in which half of the bases of the assembly are covered (N50) ranged from 9767 to 5,227,966. Genome assembly statistics showed that the draft genome sizes of the sequenced *B. anthracis* genomes in this study were approximately 5.5 Mb and were comparable to the reference Ames ancestor (Table 1). The GC content of the draft genomes was 35 %, similar to the GC content in other species in the *B. cereus* group. The alignment of the Pafuri draft genomes revealed an exceptionally high level of homogeneity among the studied genomes, with over 99 % of the reads aligning to the reference Ames ancestor chromosome and its plasmids. The RAST annotation server annotated the genes in the draft assemblies and calculated over 5000 coding sequences for the Kruger A- and B-clades draft genomes.

3.3. Pan-genome analysis and gene classification

Pan-genomic analysis was conducted on genomes that included the sequenced Pafuri strains from A clades ($n = 32$) and compared global genomes. The pan-genome of *B. anthracis* was identified by 11,374 clusters of protein-coding sequences (CDS) in this study. The core genes in *B. anthracis* consisted of 3532 CDS, while the soft-core and shell genes were identified to be 1404 and 1038, respectively. About 5497 CDS were identified as cloud genes predicted among these genomes (Table 2). The cloud genes (binary accessory genes) and shell genes were used to investigate the placement of the South African sequenced genomes with global *B. anthracis* strains. Furthermore, this study identified a murein

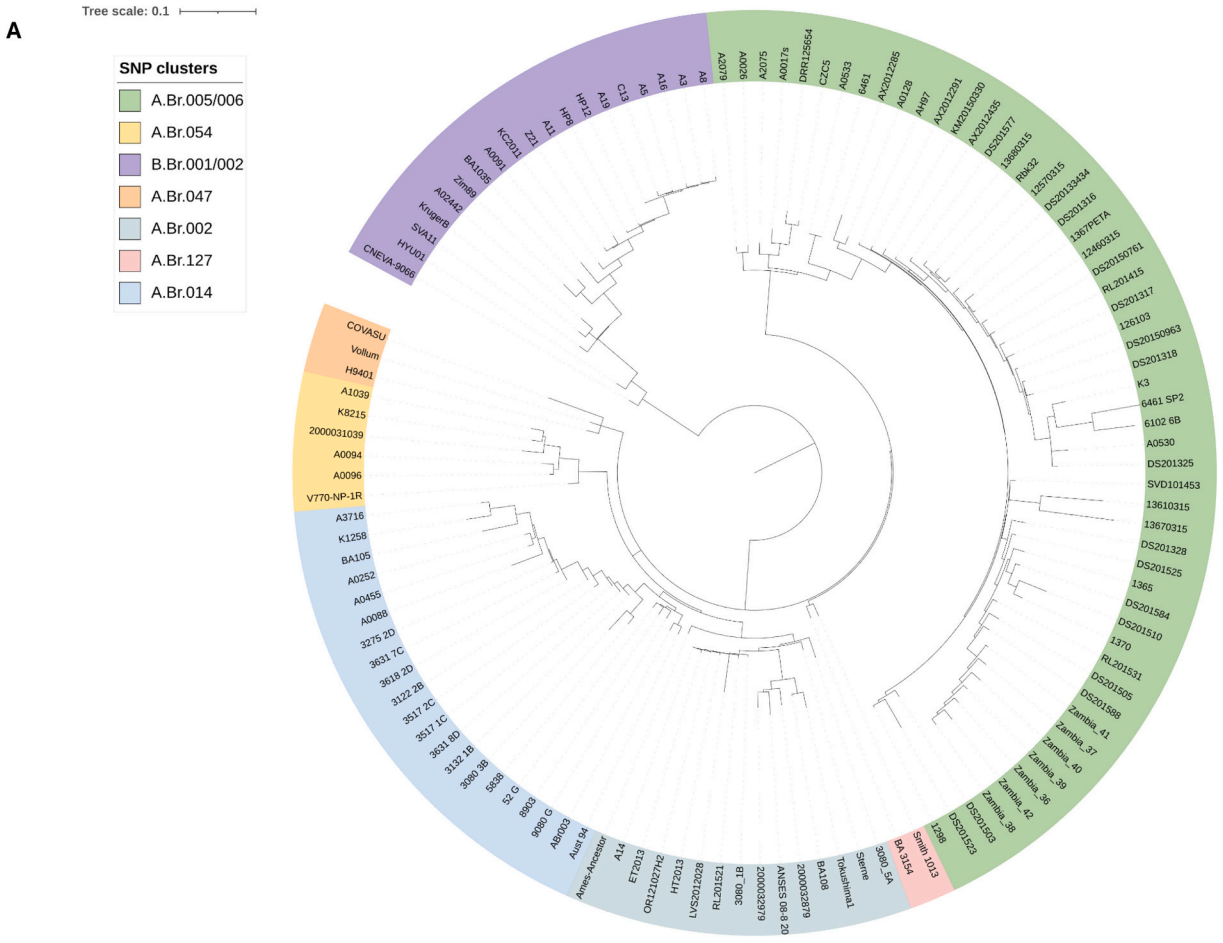

Fig. 1. Phylogenetic relationship of *Bacillus anthracis* strains based on whole-genome single nucleotide polymorphism (SNP) analysis of the chromosome indicating the clustering of Pafuri, South Africa isolates in relation to global genomes. About 7713 parsimony informative SNPs were used for the phylogeny inference using Maximum likelihood method. **A** Colour ranges indicate grouping of the isolates in different major and minor sub-clades. The Pafuri isolates clustered in the A.Br.005/006 clade and A. Br.001/002 SNP clades. **B**: Phylogenetic relationships of *B. anthracis* isolates from Pafuri, Kruger National Park. Colour ranges indicate the clustering of the isolates according to the year and hosts or sources they were isolated from plains zebra (*Equus quagga*), African elephant (*Loxodonta africana*), impala (*Aepyceros melampus*), greater kudu (*Tragelaphus strepsiceros*), nyala (*Tragelaphus angasii*), and white rhinoceros (*Ceratotherium simum*)]. The 1975 B-clade strains from Pafuri were used as an outgroup.

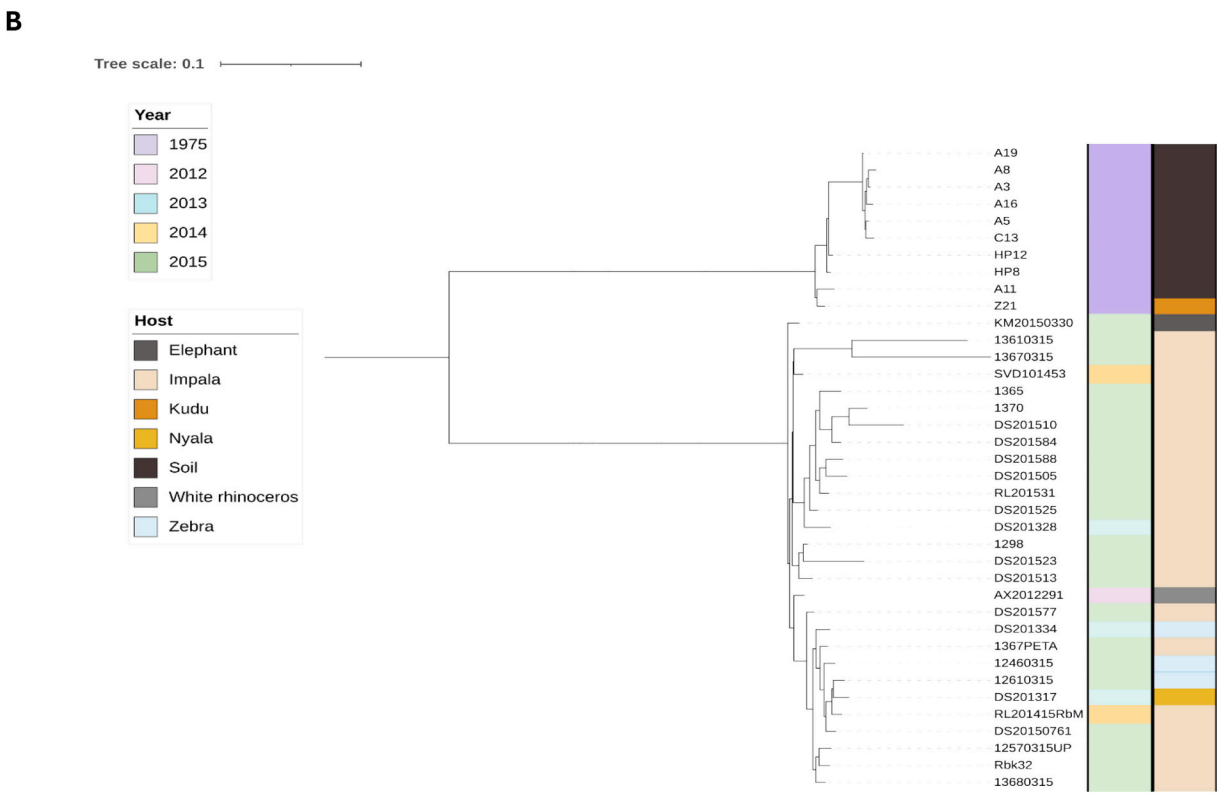

Fig. 1. (*continued*).

Table 2
Pan-genomic analysis of *Bacillus anthracis* strains in B.Br.001/002 and A.Br.005/006. 11,374 clusters of protein coding sequences (CDS) were identified in this study. The core genes in *B. anthracis* consisted of 3532 CDS, while the soft-core and shell genes were identified to be 1404 and 1038, respectively. The cloud genes (binary accessory genes) and shell genes were used to investigate the placement of the South African sequenced genomes with global *B. anthracis* strains.

Gene name	Gene product name in full	Lineage/ strain	Copy number
dprA	DNA processing protein A	B.Br.001/ 002	2
FadD13	long-chain-fatty-acid–CoA ligase	B.Br.001/ 002	4–5
rsmJ	Ribosomal RNA small subunit methyltransferase J	B.Br.001/ 002	2
mdtG	Multidrug resistance protein	B.Br.001/ 002	3
mepH	Murein DD-endopeptidase	B.Br.001/ 002	1
hemAT	Heme-based aerotactic transducer	B.Br.001/ 002	2
	antitoxin of the YeeV-YeeU toxin-antitoxin system	A.Br.005/ 006	1
spxA	Regulatory protein Spx	A.Br.005/ 006	3
yohK	inner membrane protein yohk	A.Br.005/ 006	2
	putative prophage protein	A.Br.005/ 006	2
	membrane-associated lysozyme; Qin prophage	A.Br.005/ 006	1
	Major tail protein V Phage	A.Br.005/ 006	1

endopeptidase (*mepH*) that encodes the lipoprotein (NLP/P60 family) murein endopeptidase in the chromosome and pXO2. The gene arrangement of *mepH* located on pXO2 in B-clade genomes is distinct from the gene in A-clade genomes. In the B-clade, the *mepH* gene is composed of two subunits of 209 and 148 aa residues, respectively, as compared to a complete 381 aa residue protein in A-clade strains (Fig. 2). The sequence alignment of the *mepH* gene 5′ residue in B-clade is similar to the A-clade strains; however, it has a stop codon at position 209 and results in the deletion of 127 amino acid residues (Fig. 2).

3.4. Genetic differences of A- and B-clade genomes based on tryptophan genes

The tryptophan operon structural genes (*trpE, trpD, trpC, trpB,* and *trpA*) were aligned with the global and sequenced *B. anthracis* strains in this study. The multiple sequence alignment of the *trpE*, *trpD*, *trpC*, and *trpB* genes showed no nucleotide variation between A and B clade strains. In contrast, the alignment of the *trpA* gene depicts variation in gene size among A and B-clade *B. anthracis* strains. The size of the *trpA* gene in A-clade strains is 777 bp, compared to 655 bp in B-clade strains (Fig. 3). The 122-bp deletion at the 3′ end of the B-clade strains results in a truncated gene. The presence of a G/T SNP at position 652 bp was observed on the aligned *trpA* gene and signals a stop codon during protein translation (Fig. 3).

3.5. Genetic differences in the A- and B-clades genomes based on the bclA gene

The copy number of the repeat units varied between A and B lineages, as well as within the strains. The Pafuri A-clade (A.Br.005/006) strains had 4 to 6 *bclA* repeat units as compared to 3 units located in the B-clade strains. Sequence comparison showed that the copy number of repeats within the A-clade from Pafuri (A.Br.005/006) group is 4, except for strain 1365 (3 repeats) and strain DS201578 (5 repeats). Low VNTR

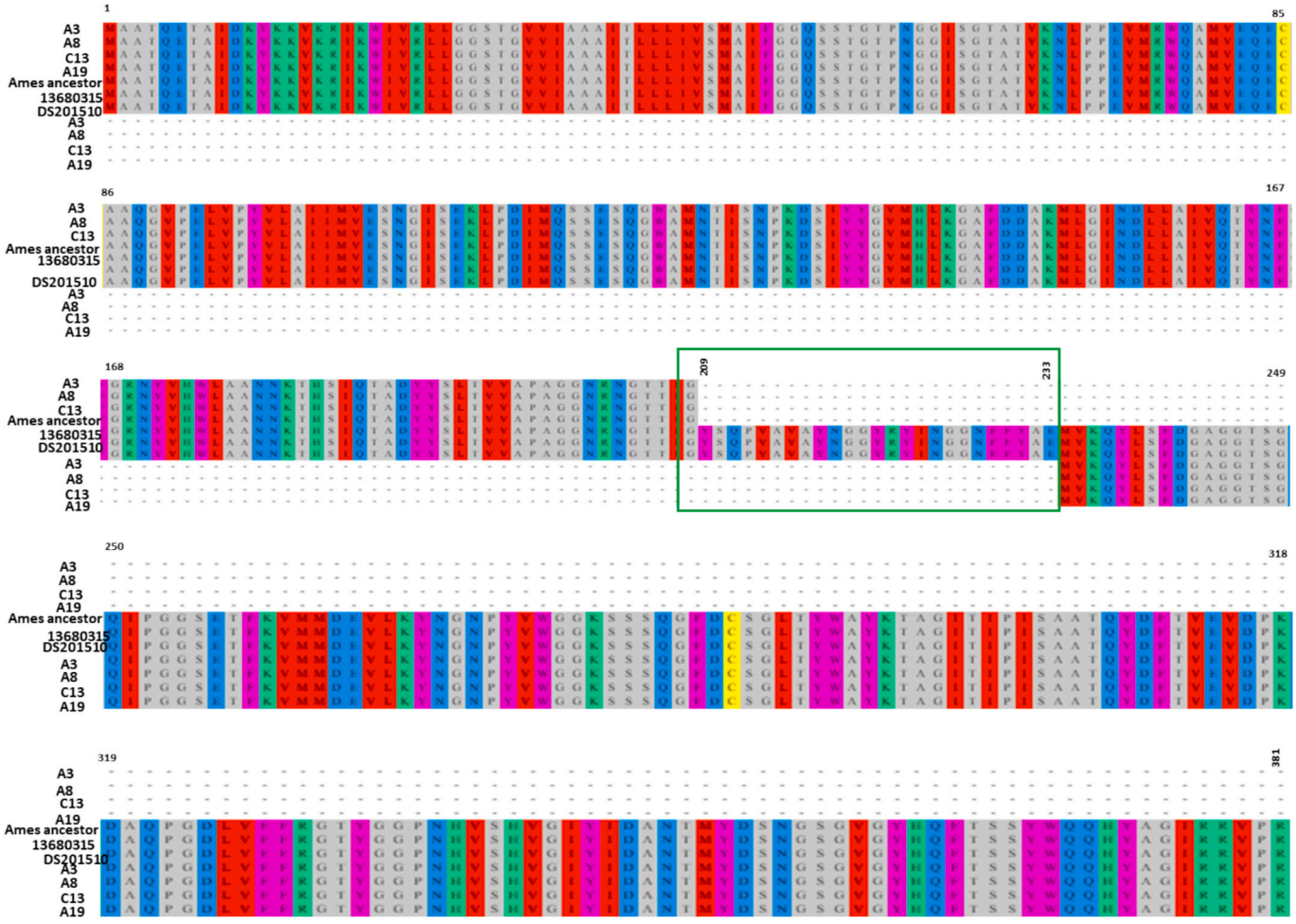

Fig. 2. Amino acid sequence alignment of the *mepH* located on plasmid pXO2 across selected *Bacillus anthracis* genomes from the B-clade (A3, A8, C13, and A19) and A-clade (Ames Ancestor, 13680315, and DS201510). The protein is truncated at amino acid position 209 in B-clade isolates, whereas in A-clade genomes, the full-length protein comprises 381 amino acid residues.

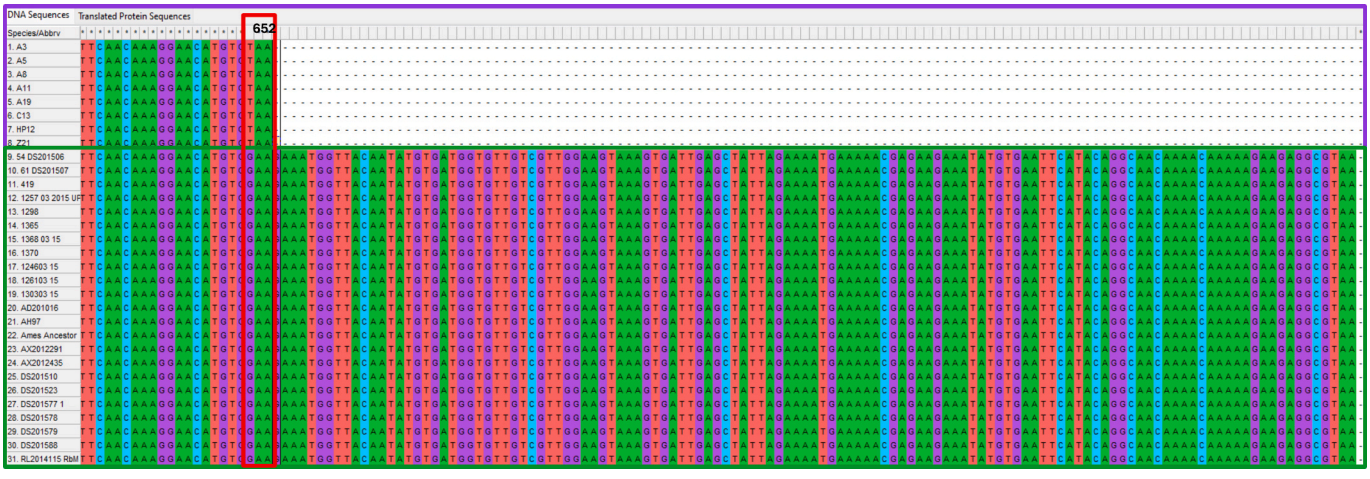

Fig. 3. Multiple sequence alignment of the tryptophan synthase subunit alpha gene (*trpA*, region 643–777 bp) from *Bacillus anthracis* strains highliting the difference between the A and B- clade strains of Pafuri section, Kruger National Park, South Africa.The B-clade strains (indicated in the purple box) exhibit a 122 bp deletion within *trpA*, whereas the A-clade strains, including the reference strain Ames Ancestor (shown in the green box), retain the full-length sequence. A G/T single nucleotide polymorphism (SNP) at position 652 (marked in red) results in a stop codon, signalling termination of protein translation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

copy numbers were observed in the B.Br.001/002 strains, whereby the numbers varied between 2 and 3.

4. Discussion

In this study whole genome single nucleotide polymorphism (wgSNP) and pan-genome analysis were employed to characterize the differences between A- and B-clade *B. anthracis* strains and to determine if there are genetic differences that may influence spatial distribution and environmental associates (as hypothesized by [10]). Pan-genomics unveiled that *B. anthracis* global B-clade strains have noticeable truncations on the biosynthesis spore wall genes. The variability in the genes (*FadD13* and *mepH*) and exosporium surface gene (*bclA*) will affect germination of endospores into vegetative cells and the persistence of A-clade strains over B-clade strains in the environment, a similar observation made by Eremenko et al. [13]. Additionally, the occurrence of SNPs and deletions in the tryptophan synthase subunit alpha gene (*trpA*), separates the A- and B-clade strains. Using genome-wide SNP analysis and pan-genomics, this study found that the A.Br.005/006-clade strains in Kruger National Park (KNP) exhibit higher genetic diversity, which may enhance their resilience to environmental stressors. In contrast, the KNP B-clade (B.Br.001/002) strains display limited genetic variation, potentially reducing their adaptability. This pattern, revealed through the investigation of *B. anthracis* evolution, contributes to the surveillance of *B. anthracis* in the KNP and advances of understanding of anthrax evolution in South Africa.

Whole genome SNP analysis revealed that different genotypes are circulating in the different regions of the KNP. The genetic structure in Pafuri based on the sequenced isolates from the 2012–2015 outbreaks assigned the genomes into the A- clade branch and showed a dispersion into three minor sub-clades that presented a non-stabilized genetic diversity, augmented with novel SNPs of the A.Br.005/006 branch. Previously, endemic anthrax outbreaks in KNP were attributed to dry seasons after heavy rainfalls [47,1], with the main affected host being greater kudu [48]. In contrast, more recent outbreaks occurred in the wet season, infecting predominantly impala [46,49]. The 2015 anthrax outbreak also revealed that impala was mostly affected and contributed to the sequenced genomes in this study.

Accessory genes contribute to an organism's lifestyle and adaptation qualities to its environment [50]. These genes are expressed to aid proliferation in different niches and are thus hypothesized to play a role in selection and contribute to the evolution of bacterial species [51]. Pan-genome analyses of *B. anthracis* strains showed that there are conserved gene clusters between the A- and B-clade strains. The hypothetical proteins were determined to be unique to A-clade strains. Further investigation on the sequence diversity of the phages found in A- and B-clade strains is crucial to understand their influence on *B. anthracis* evolution and persistence of A-clade strains worldwide. These prophages may contain several genes influencing gene expression, potentially driving increased sporulation and observable phenotypic differences [52]. Bacteriophages play a vital role in the ecology and evolution of bacteria by interacting with host and phage genomes in various ways [53,54]. The co-evolution of phages and their hosts affect their respective genomes' survival, persistence, and evolution [55,56,53,52]. *Bacillus anthracis* strains contain unique four phages found across genomes that can assist in bacterial identification [57,7]. However, other additional and questionable prophages have been found across several compared genomes, with seemingly no link to relatedness or geography among isolates [58]. Prophage signatures identified in isolates from some of the A-clade genomes may offer an added survival advantage and warrant further investigation. However, the phages' persistence over B-clade strains in the environment is less transparent [59].

Additionally, coding sequences linked to antitoxin of the YeeV-YeeU toxin-antitoxin (TA) system were identified in genomes from the A.Br.005/006 clade ($n = 45$). Toxin-antitoxins are not necessary for normal cell growth; however, their abundance is evident on bacterial plasmids and chromosomes [60]. It has been proposed that TAs are vital in cell survival in nature for transitioning to a latent, drug-resistant state to enable tolerance to excessive amounts of antibiotic stress [37]. The function of TAs is comparable to that of antibiotics: they inhibit cell growth by targeting different critical cellular processes including DNA replication, transcription, and cell wall construction [61,62]. Toxin-antitoxins may influence virulence evolution (due to enhanced antibiotic resistance) and could aid the persistence of A lineage strains in the environment. The presence of unique accessory genes, prophage signatures, and toxin-antitoxin systems in A-clade strains suggests enhanced adaptability, stress tolerance, and persistence, providing a survival advantage over B-clade strains in the KNP environment. While these genetic elements likely contribute to the ecological success and broader distribution of A-clade strains, understanding their specific functions will necessitate additional research.

Sequence comparison of cell wall biosynthesis proteins identified multiple copies and variants of long-chain-fatty-acid–CoA ligases (*FadD13*) between strains from A and B lineages. These ligases are essential in metabolic and regulatory processes, such as lipid biosynthesis and fatty acid degradation, intermediary metabolism and gene expression [63,64,65]. The result of this study suggests that the survival of A-clade strains in different niches is dependent on the cell wall structure. The protein composition and the interaction of the cell wall proteins aid in maintaining cell wall integrity and inhibit the desiccation of spores [63,64,65]. This study revealed a high copy number of the ligases in the B-clade genomes. In *Mycobacterium tuberculosis*, the gene activates fatty acids for further use in mycobacterial lipid metabolism [66]. The over-expression of *FadD13* inhibits cell membrane and adherence of bacteria and could influence survival in harsh environments [67]. The comparative analysis of cell wall biosynthesis proteins, particularly long-chain-fatty-acid–CoA ligases (*FadD13*), provides important evidence for differential survival mechanisms between A- and B-clade *B. anthracis* strains in KNP. In A-clade strains, the diversity and interaction of cell wall-associated proteins contribute to maintaining cell wall integrity and enhancing spore resistance to desiccation, supporting survival across various environmental niches. Although B-clade strains exhibit a higher copy number of *FadD13*, which in other bacterial systems is linked to fatty acid metabolism and stress response, overexpression may compromise membrane stability and adherence, potentially reducing environmental persistence. These findings highlight that the structural and functional variations in cell wall biosynthesis proteins likely underpin the greater ecological success and resilience of A-clade strains relative to the B-clade in KNP.

Variability in terms of the size and structure of the *mepH* gene on pXO2 was observed in addition to the *mepH* gene found in the chromosome. The *mepH* gene on pXO2 encodes two lipoproteins in B-clade genomes as compared to only one in A-clade genomes. The lipoproteins (*mepH*) are generally found in the chromosome and are responsible for cell-wall biosynthesis [68]. However, this gene was also found in the pXO2 plasmid of *B. anthracis* genomes. Amino acid sequence comparison of the *mepH* depicts a truncated protein in B-clade strains as compared to A-clade strains. It is not clear if the truncated proteins (in the B-clade genomes) observed in this study have any functional significance, and there is currently limited information on the interactions of chromosomal cell wall structure proteins linked with the *mepH* protein of pXO2. This study suggests that there is a cross-link between the *mepH* of pXO2 with chromosomal cell wall structure proteins. Further experimental investigation could test for any links between the presence of the truncated proteins and the stability of the cell wall on the survival of the strains. Although there were some similarities in the cell wall proteins, changes or mutation of proteins (*e.g.*, murein hydrolases) can affect cell wall structure, and thus influence the spore survival in various environments [68]. Any differences may contribute to cell wall biosynthesis and degradation since murein hydrolases are involved in the recycling of peptidoglycan, maintaining cell wall integrity, and programmed cell

death [68]. We hypothesize that differences in spore cell wall structure influence the dominance of A-clade strains over the B-clade strains, however, further investigation to elucidate how the genetic differences correlate with cell wall structure, the protein composition, and their functions/interactions and survival in the environment is necessary.

The amino acid tryptophan requires a lot of energy to synthesize in the cell, and thus its metabolic pathway is expected to be tightly controlled. The tryptophan biosynthetic pathway consists of highly conserved multiple enzymatic processes in various microbial genomes. The tryptophan pathway genes vary in arrangements, operon structure and are regulated differently [15]. Five structural genes (*trpE, trpD, trpC, trpB, and trpA*) encode enzymes needed for tryptophan biosynthesis. These genes form a single transcriptional unit, namely, the trp operon which is tightly regulated [69]. Differences in gene and operon organization reflect evolutionary divergence and adaptability in various environments [13]. Multiple sequence alignment of the *trpA* gene of the KNP *B. anthracis* strains presented a SNP mutation on B-clade strains. The G/T SNP at position 652 bp signals termination of the tryptophan operon in the B-clade strains and thus results in no synthesis of tryptophan. Phylogenetic analysis of this gene separates A and B lineages. This study confirms the findings by Eremenko et al. [13], showing the variability of SNPs and InDels in trp operon genes between these two clades. Studies on tryptophan dependence can provide information on intraspecies evolution of *B. anthracis*. The identified SNP mutation in the *trpA* gene of B-clade strains, which disrupts tryptophan biosynthesis, suggests a metabolic disadvantage that may impair survival under nutrient-limited conditions. In contrast, intact tryptophan operons in A-clade strains likely confer a selective advantage, supporting their persistence in diverse environments within KNP.

The phylogenetic analysis based on the bclA gene is unable to independently cluster isolates in the A- and B-clades since it is unable to split *B. anthracis* strains into distinct significant SNP clades. However, *B. anthracis* lineages can be distinguished based on the number of tandem repeats using MLVA Bams markers. These markers have a strong discriminatory capacity to separate subpopulations [70,8]. Since the exosporium filaments form the spore's outermost structure, variations in BclA protein may influence the spore's characteristics in response to environmental perturbations [16]. The variation in the VNTR within a surface structural gene shows that *B. anthracis*, like other pathogens, may employ the variation to regulate its ability for disease transmission and pathogen fitness, thus altering rates of survival or virulence among genetic variants [71,72]. It is still unclear how the exosporium structure contributes to the survival of A-clade strains compared to the B-clade. However, spore survival in the northern parts of KNP has been hypothesized to be supported by neutral-to-alkaline soils and the increased calcium levels [18,10], particularly as most outbreaks were attributed to the B-clade strain in KNP, especially before 1990 when the A-clade became dominant. Moreover, additional research is necessary to understand the germination dynamics of *B. anthracis* spores under favourable conditions.

5. Conclusions

This study revealed the ubiquity of the ancient A- clade strains present in Pafuri. Whole-genome SNP and pan-genome analyses provided more profound insight into the predominance of the A.Br.005/006 lineage and its genetic diversity within the KNP. These analyses revealed distinct survival advantages between A- and B-clade *B. anthracis* strains, with A-clade strains exhibiting higher genomic plasticity, including novel SNPs, unique accessory genes, and functional diversity in key loci such as *trpA*, *mepH*, and long-chain-fatty-acid–CoA ligases. These features, alongside unique prophage content and toxin-antitoxin systems, likely enhance stress tolerance, niche adaptation, and long-term persistence of A-clade strains in the environment, although their exact roles require further investigation. In contrast, B-clade (B.Br.001/002) strains displayed limited genetic variation, including truncated or mutated genes, which may constrain their adaptability and reduce their prevalence under current environmental conditions.

Pan-genomic analyses also revealed differences in gene clusters associated with spore germination and cell wall biosynthesis that may contribute to the differential persistence of these lineages. The presence of multiple hypothetical proteins and variable gene copy numbers in the accessory genome between A- and B-clade strains underscores the potential benefit of further investigation into their functional significance. Although the study's objectives were met, additional sequencing of *B. anthracis* strains, particularly from the B-clade and deeper analysis of the pan-genome, especially genes linked to cell wall integrity and environmental resilience, would be essential for elucidating the evolutionary trajectories of anthrax and informing its surveillance and control in South Africa and globally.

Ethics declarations

Ethics approval was obtained at the University of Pretoria (REC066–20). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Acknowledgements and disclaimers

This study was funded by AgriSETA for project BC21UP58-18.1, NRF Thuthuka award UID129525 and NSF Grant DEB-1816161/DEB-2106221 through the NSF-NIH-USDA Ecology and Evolution of Infectious Diseases program.

Credit authorship contribution statement

Sankwetea P. Mokgokong: Writing – original draft, Validation, Data curation, Project administration, Formal analysis, Methodology, Writing – review & editing, Visualization, Software, Investigation, Conceptualization. **Ayesha Hassim:** Methodology, Writing – review & editing, Formal analysis, Validation, Investigation, Data curation. **Tendo Mafuna:** Data curation, Writing – review & editing, Validation. **Wendy C. Turner:** Validation, Resources, Funding acquisition, Writing – review & editing, Supervision, Investigation, Conceptualization. **Henriette van Heerden:** Writing – review & editing, Validation, Investigation, Funding acquisition, Resources, Supervision, Project administration, Conceptualization, Methodology. **Kgaugelo E. Lekota:** Visualization, Investigation, Conceptualization, Supervision, Methodology, Funding acquisition, Data curation, Writing – review & editing, Validation, Resources, Formal analysis.

Declaration of competing interest

The authors declare that they have no competing interests. The data (SSR files) are available and can be accessed on https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA1075343. The corresponding author, Sankwetea Prudent Mokgokong (prudent.mokgokong@nwu.ac.za), can be contacted with questions about the data used in this study. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2025.111074.

Data availability

Data will be made available on request.

References

[1] P.J. Steenkamp, H. Van Heerden, L. Van Schalkwyk, Ecological suitability modeling for anthrax in the Kruger National Park South Africa, PLoS One 13 (1) (2018 Jan 29) e0191704, https://doi.org/10.1371/journal.pone.0191704. PMID: 29377918; PMCID: PMC5788353.

[2] WHO, Anthrax in Humans and Animals, 4th ed, World Health Organization, 2008.

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

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

[5] A. Hassim, E.H. Dekker, C. Byaruhanga, T. Reardon, H. van Heerden, 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 (1) (2017 Sep 28) e1–e15, https://doi.org/10.4102/ojvr.v84i1.1414. PMID: 29041790; PMCID: PMC8616768.

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

[7] K.E. Lekota, Genomic Study of *Bacillus anthracis* and Bacillus Species Isolated from anthrax Cases in South Africa, PhD University of Pretoria, 2018.

[8] T. Pearson, J.D. Busch, J. Ravel, T.D. Read, S.D. Rhoton, J.M. U'Ren, T. S. Simonson, S.M. Kachur, R.R. Leadem, M.L. Cardon, M.N. Van Ert, L.Y. Huynh, C. M. Fraser, P. Keim, Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing, Proc. Natl. Acad. Sci. USA 101 (2004) 13536–13541, https://doi.org/10.1073/pnas.0403844101.

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

[10] K.L. Smith, V. DeVos, H. Bryden, L.B. Price, M.E. Hugh-Jones, P. Keim, *Bacillus anthracis* diversity in Kruger National Park, J. Clin. Microbiol. 38 (2000) 3780–3784, https://doi.org/10.1128/jcm.38.10.3780-3784.2000.

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

[12] K.E. Lekota, A. Hassim, E. Madoroba, C.A. Hefer, H. van Heerden, Phylogenomic structure of *Bacillus anthracis* isolates in the northern Cape Province South Africa revealed novel single nucleotide polymorphisms, Infect. Genet. Evol. 80 (2020) 104146, https://doi.org/10.1016/j.meegid.2019.104146.

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

[14] E. Merino, R.A. Jensen, C. Yanofsky, Evolution of bacterial trp operons and their regulation, Curr. Opin. Microbiol. 11 (2) (2008) 78–86, https://doi.org/10.1016/j.mib.2008.02.005. PMID: 18374625; PMCID: PMC2387123.

[15] V.K. Priya, S. Sarkar, S. Sinha, Evolution of tryptophan biosynthetic pathway in microbial genomes: A comparative genetic study, Syst. Synth. Biol. 8 (1) (2014 Mar) 59–72, https://doi.org/10.1007/s11693-013-9127-1. Epub 2013 Oct 19. PMID: 24592292; PMCID: PMC3933628.

[16] P. Sylvestre, E. Couture-Tosi, M. Mock, Polymorphism in the collagen-like region of the *Bacillus anthracis* BclA protein leads to variation in exosporium filament length, J. Bacteriol. 185 (5) (2003 Mar) 1555–1563, https://doi.org/10.1128/JB.185.5.1555-1563.2003. PMID: 12591872; PMCID: PMC148075.

[17] P. Le Flèche, Y. Hauck, L. Onteniente, A. Prieur, F. Denoeud, V. Ramisse, P. Sylvestre, G. Benson, F. Ramisse, G. Vergnaud, A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*, BMC Microbiol. 1 (2001) 1–14, https://doi.org/10.1186/1471-2180-1-2.

[18] D.C. Dragon, R.P. Rennie, The ecology of anthrax spores: tough but not invincible, Can. Vet. J. 36 (5) (1995 May) 295–301 (PMID: 7773917; PMCID: PMC1686874).

[19] O.P. Leiser, J.K. Blackburn, T.L. Hadfield, H.W. Kreuzer, D.S. Wunschel, C. J. Bruckner-Lea, Laboratory strains of *Bacillus anthracis* exhibit pervasive alteration in expression of proteins related to sporulation under laboratory conditions relative to genetically related wild strains, PLoS One (2018), https://doi.org/10.1371/journal.pone.0209120.

[20] S. Andrews, FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc, 2010.

[21] T. Seemann, Shovill: Faster SPAdes Assembly of Illumina Reads, 2017.

[22] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates single cells and metagenomes, Genome Res. 25 (7) (2015) 1043–1055, https://doi.org/10.1101/gr.186072.114.

[23] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics 29 (2013) 1072–1075, https://doi.org/10.1093/bioinformatics/btt086.

[24] A.C.E. Darling, B. Mau, F.R. Blattner, N.T. Perna, Mauve: multiple alignment of conserved genomic sequence with rearrangements, Genome Res. 14 (2004) 1394–1403, https://doi.org/10.1101/gr.2289704.

[25] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410, https://doi.org/10.1016/S0022-2836(05)80360-2.

[26] R.K. Aziz, D. Bartels, A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R. A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: rapid annotations using subsystems technology, BMC Genomics 9 (2008) 1–15, https://doi.org/10.1186/1471-2164-9-75.

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

[28] R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: rapid annotations using subsystems technology, BMC Genomics 9 (2008) 1–15.

[29] T. Seemann, Prokka: rapid prokaryotic genome annotation, Bioinformatics 30 (2014) 2068–2069, https://doi.org/10.1093/bioinformatics/btu153.

[30] H. Li, R. Durbin, Fast and accurate short read alignment with burrows -wheeler transform, Bioinformatics 25 (2009) 1754–1760, https://doi.org/10.1093/bioinformatics/btp324.

[31] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079, https://doi.org/10.1093/bioinformatics/btp352.

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

[33] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo, The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (2010) 1297–1303, https://doi.org/10.1101/gr.107524.110.

[34] G.A. Van der Auwera, M.O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K.V. Garimella, D. Altshuler, S. Gabriel, M.A. DePristo, From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline, Curr. Protoco. Bioinform. (2013), https://doi.org/10.1002/0471250953.bi1110s43.

[35] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (2016) 1870–1874, https://doi.org/10.1093/molbev/msw054.

[36] M.A. Suchard, P. Lemey, G. Baele, D.L. Ayres, A.J. Drummond, A. Rambaut, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10, Virus Evol. 4 (1) (2018 Jun 8) vey016, https://doi.org/10.1093/ve/vey016. PMID: 29942656; PMCID: PMC6007674.

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

[38] F. Sitto, F.U. Battistuzzi, Estimating pangenomes with roary, Mol. Biol. Evol. 37 (2020) 933–939, https://doi.org/10.1093/molbev/msz284.

[39] S. Van Dongen, Graph clustering via a discrete uncoupling process, SIAM J. Matrix Anal. Appl. 30 (1) (2008) 121–141, https://doi.org/10.1137/040608635.

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

[41] H. Tettelin, V. Masignani, M.J. Cieslewicz, C. Donati, D. Medini, N.L. Ward, S. V. Angiuoli, J. Crabtree, A.L. Jones, A.S. Durkin, R.T. Deboy, T.M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J.D. Peterson, C.R. Hauser, J.P. Sundaram, W.C. Nelson, R. Madupu, L.M. Brinkac, R.J. Dodson, M.J. Rosovitz, S.A. Sullivan, S. C. Daugherty, D.H. Haft, J. Selengut, M.L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K.J.B. O'Connor, S. Smith, T.R. Utterback, O. White, C.E. Rubens, G. Grandi, L.C. Madoff, D.L. Kasper, J.L. Telford, M. R. Wessels, R. Rappuoli, C.M. Fraser, Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome", Proc. Natl. Acad. Sci. USA 102 (39) (2005 Sep 27) 13950–13955, https://doi.org/10.1073/pnas.0506758102. Epub 2005 Sep 19. Erratum in: Proc Natl Acad Sci U S A. 2005 Nov 8;102(45):16530. PMID: 16172379; PMCID: PMC1216834.

[42] A. Rambaut, FigTree v1.2.2. http://tree.bio.ed.ac.uk/software/figtree/, 2009.

[43] I. Letunic, P. Bork, Interactive tree of life (iTOL) v4: recent updates and new developments, Nucleic Acids Res. 47 (2019) 256–259, https://doi.org/10.1093/nar/gkz239.

[44] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780, https://doi.org/10.1093/molbev/mst010.
[45] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549, https://doi.org/10.1093/molbev/msy096.
[46] L. Basson, A. Hassim, A. Dekker, A. Gilbert, W. Beyer, J. Rossouw, H. van Heerden, Blowflies as vectors of *Bacillus anthracis* in the Kruger National Park, Koedoe 60 (2018) 1–6, https://doi.org/10.4102/koedoe.v60i1.1468.
[47] V. de Vos, The Epidemiology / Ecology of Anthrax Eco-historical Perspective. https://repository.up.ac.za/bitstream/handle/2263/74462/2020_WAHVM_DeVosValerius.pdf?, 1990.
[48] M.E. Hugh-Jones, V. de Vos, Anthrax and wildlife, Rev. Sci. Tech. OIE 21 (2002) 359–383, https://doi.org/10.20506/rst.21.2.1336.
[49] Y.H. Huang, K. Kausrud, A. Hassim, S.O. Ochai, O.L. van Schalkwyk, E.H. Dekker, A. Buyantuev, C.C. Cloete, J.W. Kilian, J.K.E. Mfune, P.L. Kamath, H. van Heerden, W.C. Turner, Environmental drivers of biseasonal anthrax outbreak dynamics in two multihost savanna systems, Ecol. Monogr. (2022), https://doi.org/10.1002/ecm.1526 n/a e1526.
[50] W. Yichao, N. Zaiden, B. Cao, The core- and pan-genomic analyses of genus comamonas: from environmental adaptation to potential virulence, Front. Microbiol. 9 (2018 Dec 12) 3096, https://doi.org/10.3389/fmicb.2018.03096. PMID: 30619175; PMCID: PMC6299040.
[51] F.J. Whelan, R.J. Hall, J.O. McInerney, Evidence for selection in the abundant accessory gene content of a prokaryote pangenome, Mol. Biol. Evol. 38 (9) (2021) 3697–3708, https://doi.org/10.1093/molbev/msab139. PMID: 33963386; PMCID: PMC8382901.
[52] B.C.M. Ramisetty, P.A. Sudhakari, Bacterial 'grounded' prophages: hotspots for genetic renovation and innovation, Front. Genet. 12 (10) (2019) 65, https://doi.org/10.3389/fgene.2019.00065. PMID: 30809245; PMCID: PMC6379469.
[53] A. Nasir, K.M. Kim, G. Caetano-Anollés, Long-term evolution of viruses: a Janus-faced balance, BioEssays 39 (2017), https://doi.org/10.1002/bies.201700026.
[54] M.J. Roossinck, The good viruses: viral mutualistic symbioses, Nat. Rev. Microbiol. 9 (2011) 99–108, https://doi.org/10.1038/nrmicro2491.
[55] A. Buckling, M. Brockhurst, Bacteria-virus coevolution, Adv. Exp. Med. Biol. 751 (2012) 347–370, https://doi.org/10.1007/978-1-4614-3567-9_16 (PMID: 22821466).
[56] B. Koskella, M.A. Brockhurst, Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities, FEMS Microbiol. Rev. 38 (2014) 916–931, https://doi.org/10.1111/1574-6976.12072.
[57] A. Hassim, Distribution and Molecular Characterization of South African *Bacillus anthracis* Strains and their Associated Bacteriophages, PhD University of Pretoria, 2016.
[58] S.A. Bruce, Y.H. Huang, P.L. Kamath, H. van Heerden, W.C. Turner, The roles of antimicrobial resistance, phage diversity, isolation source and selection in shaping the genomic architecture of *Bacillus anthracis*, Microb. Genom. 7 (8) (2021) 000616, https://doi.org/10.1099/mgen.0.000616.
[59] S.O. Ochai, L. Snyman, A.C. Dolfi, A. Ramoelo, B.K. Reilly, J.M. Botha, E. H. Dekker, O.L. van Schalkwyk, P.L. Kamath, E. Archer, W.C. Turner, H. van Heerden, Roles of host and environment in shift of primary anthrax host species in Kruger National Park, PLoS One 19 (2024) e0314103, https://doi.org/10.1371/journal.pone.0314103.
[60] E.M. Fozo, K.S. Makarova, S.A. Shabalina, N. Yutin, E.V. Koonin, G. Storz, Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families, Nucleic Acids Res. 38 (2010) 3743–3759, https://doi.org/10.1093/nar/gkq054.
[61] J. Davies, D. Davies, Origins and evolution of antibiotic resistance, Microbiol. Mol. Biol. Rev. 74 (2010), https://doi.org/10.1128/mmbr.00016-10.
[62] Q.E. Yang, T.R. Walsh, Toxin-antitoxin systems and their role in disseminating and maintaining antimicrobial resistance, FEMS Microbiol. Rev. 41 (2017) 343–353, https://doi.org/10.1093/femsre/fux006.
[63] P.N. Black, C.C. DiRusso, Transmembrane movement of exogenous long-chain fatty acids: proteins enzymes and vectorial esterification, Microbiol. Mol. Biol. 67 (2003) 454–472, https://doi.org/10.1128/mmbr.67.3.454-472.200327.
[64] Y. Dong, H. Du, C. Gao, T. Ma, L. Feng, 2012. Characterization of two long-chain fatty acid CoA ligases in the gram-positive bacterium *Geobacillus thermodenitrificans* NG80-2, Microbiol. Res. 167 (10) (2012 Dec 20) 602–607, https://doi.org/10.1016/j.micres.2012.05.001. Epub 2012 Jun 12, 22694860.
[65] N.J. Færgeman, C.C. DiRusso, A. Elberger, J. Knudsen, P.N. Black, 1997. Disruption of the *Saccharomyces cerevisiae* homologue to the murine fatty acid transport protein impairs uptake and growth on long-chain fatty acids, J. Biol. Chem. 272 (13) (1997 Mar 28) 8531–8538, https://doi.org/10.1074/jbc.272.13.8531. 9079682.
[66] C.A.K. Lundgren, M. Lerche, C. Norling, M. Högbom, Solution and membrane interaction dynamics of Mycobacterium tuberculosis fatty acyl-CoA synthetase FadD13, Biochemistry 60 (2021) 1520–1532, https://doi.org/10.1021/acs.biochem.0c00987.
[67] S. Wei, D. Wang, H. Li, L. Bi, J. Deng, G. Zhu, J. Zhang, C. Li, M. Li, Y. Fang, G. Zhang, J. Chen, S. Tao, X.E. Zhang, Fatty acylCoA synthetase FadD13 regulates proinflammatory cytokine secretion dependent on the NF-κB signalling pathway by binding to eEF1A1, Cell. Microbiol. 21 (12) (2019) e13090, https://doi.org/10.1111/cmi.13090. PMID: 31364251; PMCID: PMC6899955.
[68] A. Vermassen, S. Leroy, R. Talon, C. Provot, M. Popowska, M. Desvaux, Cell wall hydrolases in bacteria: insight on the diversity of cell wall amidases glycosidases and peptidases toward peptidoglycan, Front. Microbiol. 10 (2019), https://doi.org/10.3389/fmicb.2019.00331.
[69] C. Yanofsky, Tryptophan operon of *Escherichia coli*, in: Brenner's Encyclopedia of Genetics, Second edition 221–223, 2013, https://doi.org/10.1016/B978-0-12-374984-0.01676-4.
[70] F. Lista, G. Faggioni, S. Valjevac, A. Ciammaruconi, J. Vaissaire, C. Le Doujet, O. Gorgé, R. De Santis, A. Carattoli, A. Ciervo, A. Fasanella, F. Orsini, R. D'Amelio, C. Pourcel, A. Cassone, G. Vergnaud, Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis, BMC Microbiol. (2006), https://doi.org/10.1186/1471-2180-6-33.
[71] D.C. Shields, D. McDevitt, T.J. Foster, Evidence against concerted evolution in a tandem array in the clumping factor gene of *Staphylococcus aureus*, Mol. Biol. Evol. 12 (1995) 963–965, https://doi.org/10.1093/oxfordjournals.molbev.a040275.
[72] A. van Belkum, S. Scherer, L. van Alphen, H. Verbrugh, Short-sequence DNA repeats in prokaryotic genomes, Microbiol. Mol. Biol. Rev. 62 (1998) 275–293, https://doi.org/10.1128/MMBR.62.2.275-293.1998.