


RESEARCH ARTICLE

Open Access



Whole genome sequencing and identification of *Bacillus endophyticus* and *B. anthracis* isolated from anthrax outbreaks in South Africa

Kgaugelo Edward Lekota^{1,2,3,4} , Oliver Keoagile Ignatius Bezuidt¹, Joseph Mafofo¹, Jasper Rees¹, Farai Catherine Muchadeyi¹, Evelyn Madoroba^{3,4,5} and Henriette van Heerden^{2*}

Abstract

Background: *Bacillus endophyticus* is a soil plant-endophytic bacterium, while *B. anthracis* is the causative agent of anthrax. The virulence factors of *B. anthracis* are the plasmid encoded tripartite toxins (pXO1) and poly- γ -glutamic acid (PGA) capsule (pXO2). *B. endophyticus* isolated alongside *B. anthracis* from animals that died of anthrax in Northern Cape Province (NCP), South Africa, harbored polyglutamate genes. The study compared the characteristics of *B. anthracis* and *B. endophyticus* with other *Bacillus* species with a focus on the presence of the PGA capsule or/ and unbound PGA. The morphology and whole genome sequence analysis of *B. endophyticus* strains and *B. anthracis* were compared.

Results: In conventional microbiology, *B. endophyticus* showed gram-positive round-shaped rods in single/short chains, which were endospore-forming, non-motile, non-haemolytic with white and dry colonies, and γ -phage resistant. *B. anthracis* was differentiated from *B. endophyticus* based on the latter's box-shaped rods in pairs/long chains, white-grey and slimy colonies, encapsulated and γ -phage susceptible. The study identified a PGA polyglutamate synthase operon that consisted of *pgsBCA*, γ -glutamyltranspeptidase (*ggt*) and *pgsE* in *B. endophyticus* genomes.

Conclusions: PGA regions of *B. anthracis* contain *capBCADE* genes located in the pXO2 required for capsulation formation, while *B. endophyticus* contain the *pgsBCAE* genes in the chromosome. Whole genome and microbiology analysis identified *B. endophyticus*, as a non-capsuled endospore-forming bacterium that consists of PGA required for biosynthesis. *B. endophyticus* strains do not synthesize surface associated PGA, therefore capsule visualization of *B. anthracis* is a key diagnostic characteristic. The study highlights the significance of using whole genome shotgun sequencing to identify virulence and other important genes that might be present amongst unknown samples from natural outbreaks. None of the *B. anthracis* related plasmids or virulence genes were found in the *B. endophyticus* genomes.

Keywords: *Bacillus endophyticus*, *Bacillus anthracis*, Poly- γ -glutamic acid (PGA), Whole genome sequencing (WGS)

* Correspondence: Henriette.VanHeerden@up.ac.za

²Department of Veterinary Tropical Diseases, University of Pretoria, Private bag X4, Onderstepoort 0110, South Africa

Full list of author information is available at the end of the article



Background

Bacillus endophyticus is regarded as a plant-endophytic bacterium that is found in the inner tissues of plants, specifically cotton [1]. It is present either as gram-positive single rod-shaped cells or as chains that can be short or long, non-haemolytic and non-motile. Biochemical characteristics that differentiate *B. endophyticus* from other *Bacillus* species include the inability to reduce nitrate (NO_3^-) to nitrite, casein and starch, as well as, ampicillin and NaCl resistance [1].

B. anthracis is the causative agent of anthrax, and primarily affects herbivorous animals, although all mammals can also be affected. The vegetative cells of *B. anthracis* appear 'box-shaped' either in pairs or chains. It is phenotypically characterized as gram-positive aerobic rods (3–5 $\mu\text{m} \times 1 \mu\text{m}$), that are non-haemolytic, non-motile, penicillin and γ -phage resistant [2]. However, it is distinguishable from its close relatives by its ability to synthesize virulence factors encoded on plasmids, pXO1 and pXO2. The pXO1 (182 kb) contains genes encoding for the tripartite anthrax toxins (protective antigen, lethal factor and edema factor) and pXO2 (96 kb) encodes a five-gene operon *capBCADE* (capsule biosynthesis genes), which synthesizes a poly- γ -glutamic acid (PGA) capsule [3, 4]. Capsule biosynthesis genes are transcribed as a single operon predicted to encode proteins for the biosynthesis, transport and attachment of D-glutamic acid residue on the bacterial surface [5]. The anthrax capsule activators (*acpA* and *acpB*) located on pXO2 are controlled by anthrax toxin activator (*atxA*) located on pXO1 [5]. The PGA capsule enables host immune system evasion by protecting the vegetative cells from phagocytosis by macrophages [5]. The vegetative cells of *B. anthracis* have also been shown to secrete the PGA capsules under anaerobic conditions and in the presence of bicarbonate [3, 5].

Many pathogenic bacteria require a cell-associated capsule for virulence [6]. Capsule composition of bacteria can be in a form of polypeptide (poly-glutamate) or polysaccharide. Poly- γ -glutamic acid (PGA) is a poly-anionic polymer that may be composed of only D, only L or both glutamate enantiomers [4, 7]. Most strains producing PGA are members of the gram-positive *Bacillus* group. The function of PGA depends on whether it is bound to peptidoglycan or unbound/released. In the bound state it forms the capsule, whereas in the secreted/unbound state it is released into the environment [4, 8]. The uncommon bound PGA capsule only includes *B. anthracis* and *Staphylococcus epidermidis* that synthesize the anchored (surface-associated) PGA, which enables them to act as a virulence factor [4]. The *B. anthracis* PGA synthesis genes are encoded on the pXO2 consisting of *capB*, *capC*, *capA* and *capE*, while *capD* acts as the peptidoglycan binding/anchoring site [4, 7, 9]. The corresponding polyglutamate

biosynthesis pathway orthologs in *B. subtilis* includes *pgsB*, *pgsC* and *pgsAA* [10] and *pgsS* has been suggested to induce the release of PGA [4, 7]. The *capBCADE* genes of *B. anthracis* encoded on pXO2 have functional orthologs encoded on the chromosomes of *B. subtilis/licheniformis* and other *Bacillus* species [4, 11]. Few species such as *B. anthracis* and *S. epidermidis* have been reported to produce the PGA capsule [4]. The unbound PGA have been reported to *Bacillus* species such as *B. cereus* strains ATCC 10987, 14,579 and *B. thuringiensis* 97–27, AI Hakam [4]. *B. cereus* biovar *anthracis* strains isolated from great apes that died of anthrax symptoms in west and central Africa were shown to harbor the *B. cereus* chromosome and pXO2-like plasmid [12] that contained the PGA capsule genes identical to those of *B. anthracis*.

Gene sequences that encode for the formation of PGA and capsules on the pathogenic and non-pathogenic species need to be compared and distinguished from their close relatives [11]. This is essential especially when some of the virulence gene sequences and morphological characteristics are used for identification and diagnosis of anthrax. In this study, *B. endophyticus* strains were isolated alongside *B. anthracis* strains from animals that died of anthrax in Northern Cape Province (NCP), South Africa in an outbreak that occurred in 2009. *B. endophyticus* is regarded a plant-endophyte and it is uncommon to be isolated from blood or animals. The *B. endophyticus* strains that were isolated alongside *B. anthracis* strains had some of the similar morphological, biochemical and some genetic characteristics compared to the anthrax causing bacteria. In our previous study, conventional PCR detected PGA gene regions in both *B. anthracis* and *B. endophyticus* isolates and attempts were made to distinguish and identify these strains using routine and non-routine diagnostic methods [13]. The *B. endophyticus* strains were identified using non-routine diagnostic Omnilog (Biolog) and 16S rRNA sequencing methods and differentiated based on routine diagnostic microbiological tests and real time PCR. Therefore, in order to enhance and contribute to the unequivocal diagnosis of *B. anthracis*, the goal of this study was to perform comparative analysis of the *B. endophyticus* and *B. anthracis* strains from the afore-mentioned outbreak as well as contribute towards the scant genome information of *B. endophyticus*. Thus the virulence genes of *B. anthracis* occurring on the plasmids were investigated, as well as the capsule and phenotypic characteristic of related *Bacillus* species were summarized using results from this study and published literature to enhance and contribute towards anthrax diagnosis.

Results

Phenotypic characterization

B. endophyticus strains reported in the study were isolated from the environment and/or animals that died of

B. anthracis during the 2009 anthrax outbreak in Northern Cape Province (NCP) (Table 1). On sheep blood tryptose agar (SBTA) at 5% CO₂, colonies of *B. anthracis* appeared whitish-grey, smooth, dry and shiny (medusa head), while *B. endophyticus* colonies were circular white, slimy or rough (Additional file 1: Figure S1 (2)). The *B. endophyticus* colonies on nutrient agar supplemented with 0.8% sodium bicarbonate at 5% CO₂ were smaller and circular, non-mucoid and wet (Additional file 1: Figure S1A), whereas *B. anthracis* colonies appeared circular, mucoid and shiny (Additional file 1: Figure S1B). Colony morphology of the *B. endophyticus* strains was observed after 24 h in culture compared to *B. anthracis*, which was observed earlier (12–24 h) on sodium bicarbonate supplemented nutrient agar.

Gram-positive *B. anthracis* cells occurred in box-shaped rods in pairs and/or long chain rods (Fig. 1a) that are encapsulated (cap⁺) after incubation at 5% CO₂ in blood (Fig. 1b), while the gram-positive *B. endophyticus* appeared as round-edged rods either as single and/ or short chains (Fig. 1c, Table 2). No capsules were observed in *B. endophyticus* strains after incubating at 5% CO₂ (Fig. 1d). *B. anthracis* 3631_1C [14] and *B. anthracis* Sterne strains were non-capsulated (cap⁻) since they lack pXO2 while *B. anthracis* 20SD was capsulated (Fig. 1e). The terminal ellipsoidal spores were also observed in *B. endophyticus* 3631_9D strain using the copper sulphate stain after 24 h incubation on nutrient agar containing 0.8% sodium bicarbonate (Fig. 1 and Additional file 2: Figure S2A-D).

The comparison of phenotypic properties of *B. endophyticus*, *B. anthracis*, *B. cereus*, *B. megaterium* and *B. smithii* strains is shown in Table 2. *B. anthracis* and *B. cereus* were compared in Table 2 as they belong to *B. cereus sensu lato* group, while *B. megaterium* is closely related to *B. endophyticus* based on whole genome sequence and some of the microbiological features are similar to *B. anthracis*. *B. smithii* is a closely related species of *B. endophyticus* based on 16S rRNA sequence gene. However, Table 2 shows that *B. cereus* and *B.*

smithii are both motile and can easily be excluded from *B. anthracis*. *B. endophyticus* is a gram-positive, non-capsulated, non-motile, round-edged rod that is endospore-forming, non-hemolytic, penicillin sensitive but γ -phage resistant bacterium. *B. anthracis* is a gram-positive capsulated, non-motile, box-shaped rod that is endospore-forming, non-hemolytic, penicillin and γ -phage sensitive (Table 2). In this study biochemical characterization showed some common results between *B. anthracis* and *B. endophyticus* including the positive reaction for catalase and oxidase and negative reaction for indole (Table 2). Biochemical properties of *B. endophyticus* that differentiated it from other *Bacillus* species included inability to reduce nitrate to nitrite, hydrolyze casein, gelatin, and starch, as well as resistance to NaCl. The absence of lecithinase and Voges Proskauer (VP) can be used to distinguish *B. endophyticus* from *B. anthracis* (Table 2).

16S rRNA gene phylogenetic analysis

The 16S rRNA gene sequences of *B. endophyticus* strains 3631_9D, 3617_2C, 3631_10C and 3618_1C strains were used to mine for other 16S rRNA gene sequences through BLAST homology searches. The sequenced *B. endophyticus* strains 3631_9D, 3617_2C and 3631_10C showed a 100% similarity with the 16S rRNA gene sequences of *B. endophyticus* strains (A6, S160(2), 2DT and uncultured bacterium 12TR2ACLN347) (Additional file 3: Figure S3). Strain 3618_1C grouped with majority of the uncultured bacterium (12TRACLN435 and 12TRACLN431) obtained from NCBI. The *B. cereus sensu lato* group grouped separately from the *B. endophyticus* based on 16S rRNA gene region (Additional file 3: Figure S3).

Average nucleotide identities, pan-genome analyses functional classification of orthologous genes

South African *B. endophyticus* sequences (3617_2C, 3618_1C, 3631_9D, 3631_10C) had a total of approximately 5.1 to 45.3 million reads with an average length

Table 1 *Bacillus endophyticus* and *B. anthracis* strains isolated from animal anthrax cases in Northern Cape province (NCP) in South Africa

Strain number	Animal source ^c	Specimen	Isolation date	Provinces and location (farm or town) in South Africa	<i>Bacillus</i> species
3617_2C ^a	<i>Tragelaphus strepsiceros</i> (Kudu 1)	Ear blood	May-09	NCP (Klipfontein)	<i>B. endophyticus</i>
3618_1C ^a	<i>Tragelaphus strepsiceros</i> (Kudu 2)	Ear blood	May-09	NCP (Kimberly)	<i>B. endophyticus</i>
3631_9D ^a	<i>Ovis aries</i> (Sheep 1)	Ear blood	May-09	NCP (Kimberly)	<i>B. endophyticus</i>
3631_10C ^a	<i>Ovis aries</i> (Sheep 2)	Ear blood	May-09	NCP (Kimberly)	<i>B. endophyticus</i>
3618_2D	<i>Tragelaphus strepsiceros</i> (Kudu 2)	Soil	May-09	NCP (Klipfontein)	<i>B. anthracis</i>
3631_1C ^b	<i>Tragelaphus strepsiceros</i> (Kudu 3)	Ear blood	May-09	NCP (Klipfontein)	<i>B. anthracis</i>
3617_1C	<i>Tragelaphus strepsiceros</i> (Kudu 1)	Ear blood	May-09	NCP (Klipfontein)	<i>B. anthracis</i>

^aThe strains that were subjected for sequencing

^bDraft genome from previous study [14]

^cKudu 1, kudu 2, sheep 1 etc. refer to different animal carcasses with *B. endophyticus* (3617_2C) and *B. anthracis* (3617_1C) isolated from the same kudu carcass (kudu 1)

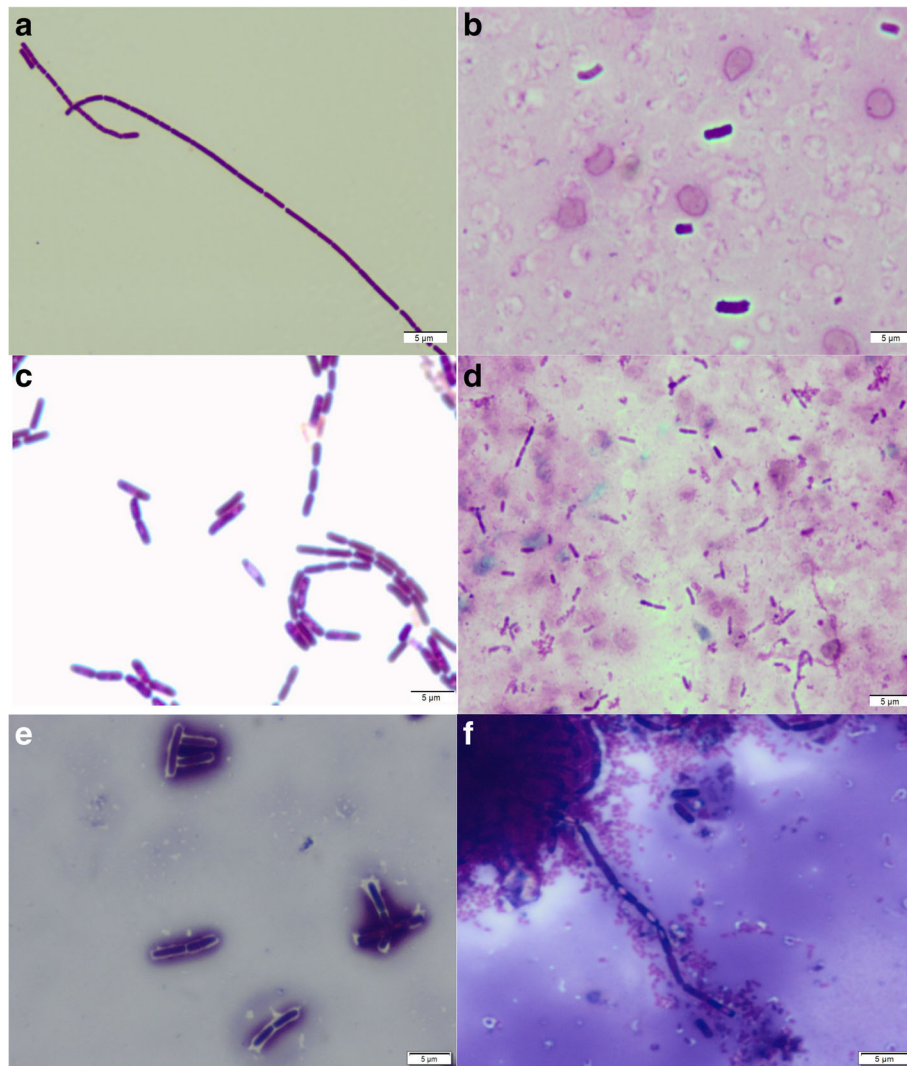


Fig. 1 Phenotypic electron microscopic examination of the morphology of *Bacillus anthracis* and *B. endophyticus* strains. **(a)** Gram-positive vegetative cells of *B. anthracis* Sterne long, bacilli chains, **(b)** *B. anthracis* 3618_2D capsulated in blood serum, **(c)** Gram-positive vegetative cells of *B. endophyticus* short, bacilli chains, and **(d)** *B. endophyticus* 3631_9D non-capsulated in blood, **(e)** capsulated *B. anthracis* 3618_2D and **(f)** *B. endophyticus* 3631_9D non-capsulated with spores after incubation on nutrient agar containing 0.8% sodium bicarbonate in the presence of 5% CO₂

of 94 nucleotides after trimming. Sequenced reads were de novo assembled (Table 3) and annotated using PGAAP for further classification of the *B. endophyticus* strains. The heat map (Fig. 4) indicated the average nucleotide identities of the *B. endophyticus* CDSs of South African sequenced strains and available whole genome sequences (2102, Hbe603, A6, S160(2), 2DT, KCC 13922, DSM13796 and uncultured bacterium 12TR2ACLN347). The sequenced *B. endophyticus* strains in this study as well as *B. endophyticus* DSM 13976 and KCTC 13922 had the same profile (with an ANI score of > 98%); *B. endophyticus* 3617_2C is highly related with this two genomes forming their own sub-clade, but clustered separate from *B. endophyticus* 2102 and Hbe603 strains (Fig. 2). *B. endophyticus*

3618_1C grouped separately amongst the sequenced *B. endophyticus* strains.

The pan-genome homology analysis of 4 South African *B. endophyticus* and Hbe603 strains identified 7154 clusters of protein coding genes with 3711, 3954, 997 and 2203 clusters represented the core, softcore, shell and cloud genomes, respectively (Fig. 3). In this study the *B. endophyticus* has more genes assigned to the core than the accessory genes (shell and cloud clusters), but the latter might increase when more genomes are sequenced and become available (Fig. 3). In the COG category assignments, the core and the accessory genomes have a slightly different number of genes assigned to the defense mechanisms category (Fig. 3

Table 2 Comparison of phenotypic and biochemical characteristics of *Bacillus endophyticus*, *B. anthracis*, *B. cereus*, *B. megaterium* and *B. smithii*

Microbiological characteristics	<i>B. endophyticus</i>	<i>B. anthracis</i>	<i>B. cereus</i>	<i>B. megaterium</i>	<i>B. smithii</i>
Gram reaction	+	+	+	+	+
Rods (µm)	2.5–3.5 × 0.5–1.5	3–6 × 1–1.25	3–5 × 1–1.2	2–5 × 1.2–1.5	5–6 × 0.8–1.0
Spores	Ellipsoidal/cylindrical	Ellipsoidal/cylindrical	Ellipsoidal/cylindrical	Ellipsoidal/Spherical	Ellipsoidal/cylindrical
Haemolysis	NH	NH	H	H ^{96%}	ND
Motility	Non-motile	Non-motile	Motile	Non-motile	motile
Capsule	–	+	–	+	–
Penicillin	S	S	R	ND	ND
Gelatin hydrolysis	–	+ *	+	+	–
Nitrate reduction	–	+	+	(–)	–
Starch hydrolysis	–	+	+	+	W+
Voges Proskauer (VP)	–	+	+	–	–
γ-phage	R	S ⁵	R	S [®]	ND
Casein	–	+	+	+	–
Egg yolk/Lecithinase	–	+	+	–	–
Urea hydrolysis	–	–	–	_%%	–
Citrate	W+	–	+	+	(–)
NaCl	2–10%	5%	2–7%	7%	2–3%
Catalase	+	+	+	+	+
Indole	–	–	–	–	–
Oxidase	+	+	–	(+)	+

Abbreviations: +, positive, W+, weakly positive, –, negative, (), variable, S, susceptible, R, resistance, ND, not determined, NH, non-haemolytic, *, inverted fir tree, ⁵, *B. anthracis* strains resistance to γ-phage have been reported [22], [®] resistant γ-phage *B. megaterium* strain has been reported at Kansas State University [44]. All data of *B. endophyticus* were obtained in the study and supplemented with information of Reva et al. [1]. *B. anthracis* and *B. cereus* information was compiled from WHO [42]; ^{96%} Beesley et al. [27] reported *B. megaterium* non-haemolytic and urea hydrolysis strains

Table 3 Genome comparison features of the *Bacillus endophyticus* strains used in the study

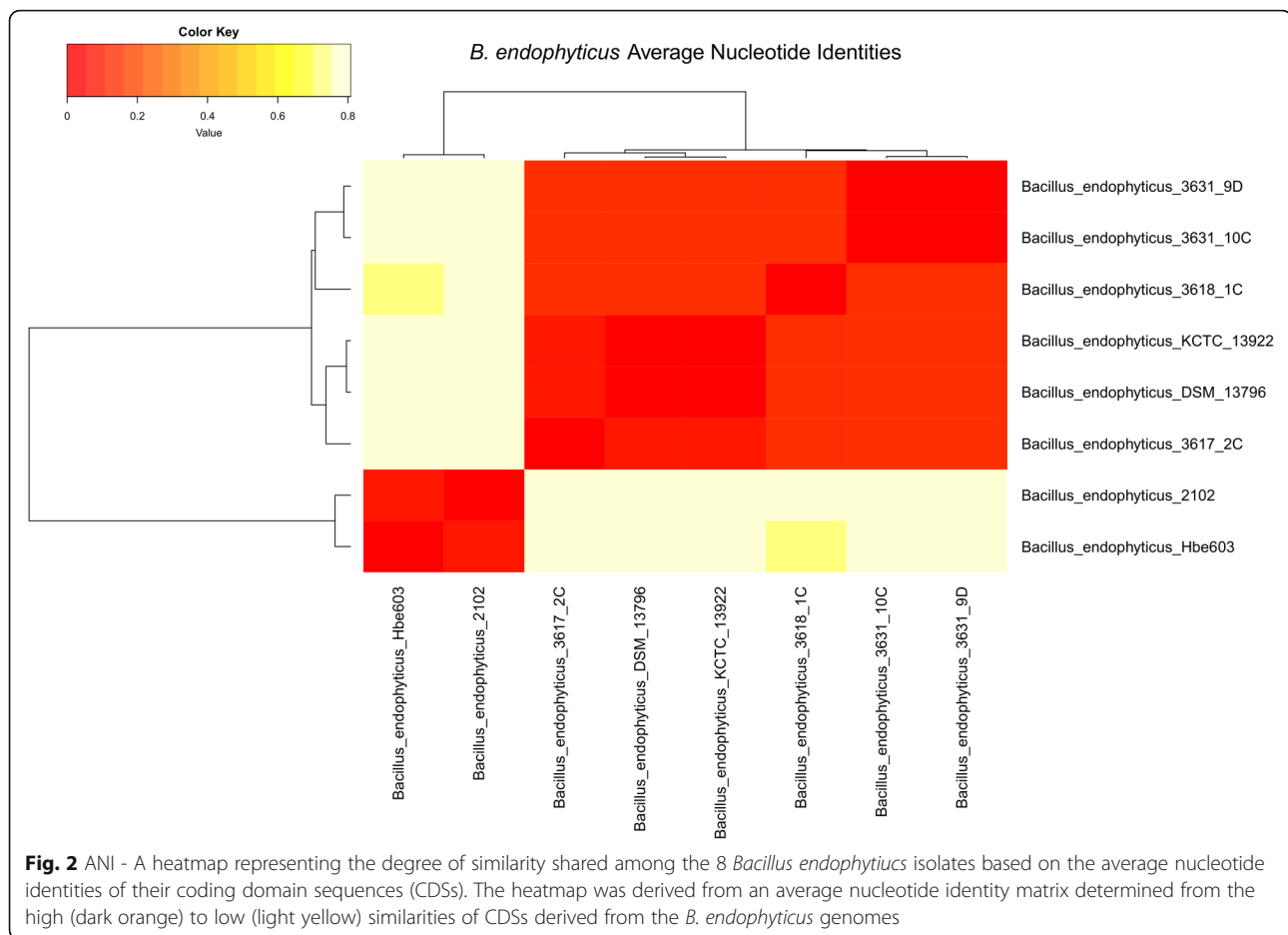
	<i>B. endophyticus</i> 3631_9D	<i>B. endophyticus</i> 3618_1C	<i>B. endophyticus</i> 3631_10C	<i>B. endophyticus</i> 3617_2C	<i>B. endophyticus</i> Hbe603 ^a
Genome size (bp)	5,311,808	5,379,838	5,243,706	5,319,031	5,313,985
GC content (%)	36	36	36	36	36
Total contigs	57	68	60	99	9
N50	304,287	152,600	277,642	99,724	4,865,574
Maximum contig	820,841	657,546	874,718	224,909	NA
Minimum contig	307	141	1164	227	NA
Coding sequences (CDS) ^b	5310	5470	5358	5408	5455
RNAs	68	54	50	47	114
Prophage regions	1	7	2	2	4
Chromosome (size)	5,056,260	5,011,594	5,305,974	5,231,075	4,865,574
Number of plasmids ^c	7	4	6	6	8

^a*Bacillus endophyticus* Hbe603 was used as a reference strain in this study

^bCoding sequences predicted using RAST

^cBLASTn with an e-value of 1e-10 and > 90% identity was used against the *B. endophyticus* Hbe603 plasmids

NA not available



category V) as in most cases this category is mainly abundant in the accessory genome [15]. The core cluster dominates all the other categories, including categories for function unknown (S) and general prediction only (R) in Fig. 3.

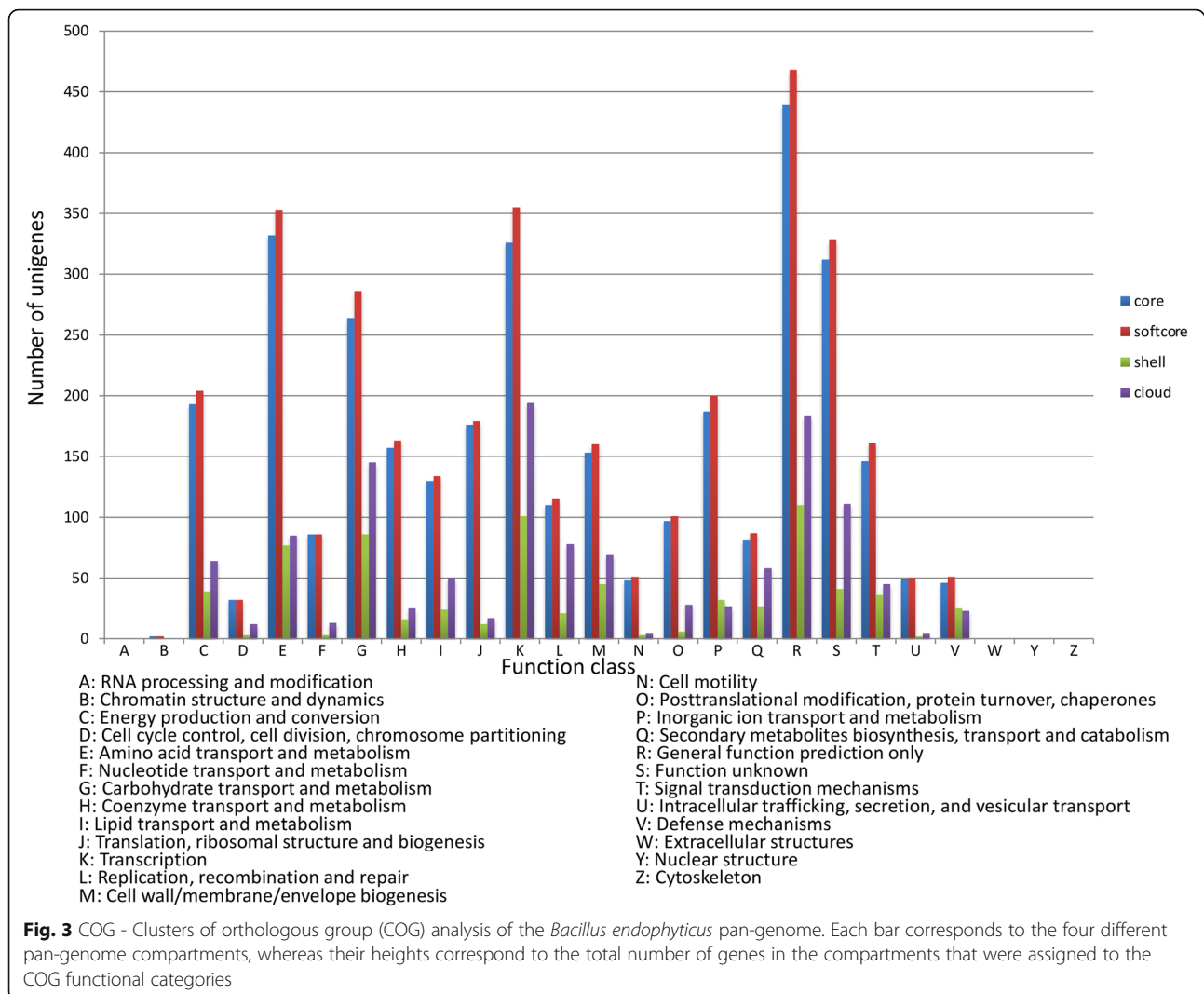
Genomic features of *B. endophyticus* strains

Comparative genomics of the draft sequenced *B. endophyticus* strains in the study and complete genome *B. endophyticus* Hbe603 showed almost equivalent genome sizes with the complete genome of *B. endophyticus* Hbe603 (Table 3). The GC content of the sequenced *B. endophyticus* genomes are approximately 36%, and similar to the *B. endophyticus* Hbe603 and other *B. endophyticus* strains used in pan-genome analysis. The complete genome of Hbe603 is 5.31 Mb and consists of a chromosome and 8 plasmids [16]. Annotation using RAST [17], predicted the number of coding sequences of *B. endophyticus* Hbe603 to be 5455 that is slight higher than the sequenced genomes in this study except for 3618_1C. High numbers of accessory genes of *B. endophyticus* 3618_1C are represented in the unknown function or as hypothetical proteins. A total of 5310,

5431, 5358, and 5408 predicted coding sequences in strains 3631_9D, 3618_1C, 3631_10C and 3617_2C respectively (Table 3). RAST analyses showed *B. megaterium* DSM 319 to be the closest neighbor to the *B. endophyticus* strains with the comparative analysis using sequence similarity option.

Plasmids of *B. endophyticus*

B. endophyticus Hbe603 complete genomes consist of 8 plasmids. The roles of the plasmids have never been reported in *B. endophyticus* Hbe603 strain. The draft genomes of *B. endophyticus* strains sequenced in this study each presented 4–7 plasmids (Table 3, Additional file 4: Table S1). Comparative analysis of the sequenced *B. endophyticus* strains with *B. endophyticus* Hbe603 consisted of partial regions of plasmids, while pBEH1, pBEH6, and pBEH7 are the common plasmids shared. Plasmids sizes of draft genome *B. endophyticus* strains were significant smaller than the *B. endophyticus* Hbe603 plasmids (Additional file 4: Table S1). None of the *B. endophyticus* plasmids were similar with the *B. anthracis* pXO1 and pXO2 plasmids.



Virulence, resistance and defense genes

Coding sequences linked to antibiotic- and toxic compound resistances were identified in the *B. endophyticus* strains. Comparative analysis of *B. endophyticus* 3618_1C, 3631_9D, 3631_10C, KCTC 13922, and DSM 13796 showed unique coding sequences that include arsenical-resistance protein *Acr3*, copper resistance protein D for copper homeostasis, multidrug resistance transporter *Bcr/CflA* family and fosfomycin resistance protein *fosB* that are absent in the *B. endophyticus* Hbe603 and 2102 genomes. *B. endophyticus* 3617_2C strain also contained these coding sequences except the *Acr3* and the multidrug resistance transporter *Bcr/CflA* family CDS. The transcriptional regulator *NfxB* was present in *B. endophyticus* 3618_1C and 3617_2C strains (i.e. absent in the other compared *B. endophyticus* strains in this study). This transcriptional regulator is involved in the MexC-MexD-OprJ multidrug efflux system that contributes to antibiotic- or

toxic compounds resistances [18]. The genome analyses of *B. endophyticus* strains confirmed the presence of CDS for the macrolide-specific efflux protein *macA* and permease protein *macB* for the multidrug resistance efflux pumps, except in strains 3618_1C and 2102. The *MacAB-TolC* macrolide efflux transport system has mostly been studied in gram-negative bacteria. The presence of *macA* in the system is known to stimulate the ATPase activity of *macB* to bind macrolides such as erythromycin and azithromycin. Meanwhile the overproduction of *macA* and *macB* results in an increase resistance to the macrolides antibiotics [19]. *B. endophyticus* is regarded as plant-endophytic bacterium that survives high-salt concentration [1, 13]. The sigma-M predicted to response to high concentration of salt [20] was found in the 8 compared *B. endophyticus* genomes in this study. Jia et al. [16] predicted other sigma factors responsible for gene regulation in *B. endophyticus*.

Bacillus endophyticus prophages

PHAGE_Bacill_phBC6A52 was the common intact prophage in strains 3631_9D and 3631_10C. *B. endophyticus* 3631_10C presented additional two partial prophage regions annotated as PHAGE_Lister_B054_NC_009812 and Bacill_1_NC_009737. The latter, PHAGE_Bacill_1_NC_009737, was also present in *B. endophyticus* 3617_2C. About 7 prophage regions were identified in *B. endophyticus* 3618_1C strain (Table 3). This included PHAGE_Bacill_G_NC_023719, PHAGE_Burkho_phi023719, PHAGE_Synech_S_MbCM100_NC_023584, PHAGE_Enterophi92_NC_023693, PHAGE_Escher_vB_EcoM_UFV13_NC_031103, PHAGE_Bacill_SP_15_NC_031245 and PHAGE_Bacil_BM5_NC_029069. The 7 prophages were also identified in *B. endophyticus* DSM_13,796 and KCTC 13922 except for PHAGE_Enterophi92_NC_023693 and PHAGE_Escher_vB_EcoM_UFV13_NC_031103. However, the prophage regions differ in their sizes. Only 4 prophages were determined in the *B. endophyticus* Hbe603 reference strain, whereby most were annotated as hypothetical proteins [16]. In *B. endophyticus* 2102, no prophage sequence regions were identified. Comparative analysis of prophages between *B. anthracis* strains 3631_1C and 20SD [14] and *B. endophyticus* sequenced in this study indicated that the four Lambda Ba prophages remain unique to *B. anthracis*.

PGA biosynthesis complex

The PGA subunits *pgsB*, *pgsC*, *pgsA* and γ -glutamyl transpeptidase (*ggt*), and *pgsE* genes were present in the 4 sequenced *B. endophyticus* strains (3617_2C, 3618_1C,

3631_9D, 3631_10C) and other 4 compared *B. endophyticus* genomes (2102, Hbe603, KCC 13922, DSM 13796) in this study. The PGA subunits of *B. endophyticus* genomes are located in the chromosome compared to the plasmid, pXO2, of *B. anthracis*. In *B. anthracis*, the PGA subunits are presented and annotated as *capBCADE* (Fig. 4). They are associated with the synthesis of the poly- γ -glutamate capsule formation rather than a released PGA. Due to no capsule formation observed in the *B. endophyticus* strains, this suggests that PGA biosynthesis is associated in a released form. *Bacillus* species genomes i.e. *B. subtilis*, and *B. licheniformis* (Fig. 4) consist of *pgs* subunits. The amino acid sequence identities of *cap/pgs* subunits to *B. anthracis* are indicated in Fig. 4 indicating the percentages of amino acid similarities between *B. endophyticus*, *B. anthracis* and *B. subtilis*. *B. endophyticus* and *B. subtilis* synthetic *pgsBCA* genes are homologous to the *capBCA* genes of *B. anthracis*. The study identified a *pgsE* subunit of *B. endophyticus*, which is analog to *capE* in *B. anthracis* (Fig. 4) and also referred to *ywtC* in *B. subtilis*. The subunit *pgsS* (*ywtD*) is present in the *B. subtilis* and *B. licheniformis* PGA synthetic operon and absent from *B. endophyticus* and *B. anthracis* PGA synthetic operon (Fig. 4). The amino acid sequence of *B. endophyticus capC* is 82% similar to *B. anthracis capC*, indicating a high probability of *capC* region primer annealing in either *B. endophyticus* or *B. anthracis* strains. The capsule regulons *acpA* and *acpB* in *B. anthracis* were observed on the same PGA operon. In *B. endophyticus* genomes, none of these two regulons were observed in the PGA complex operon (Fig. 4).

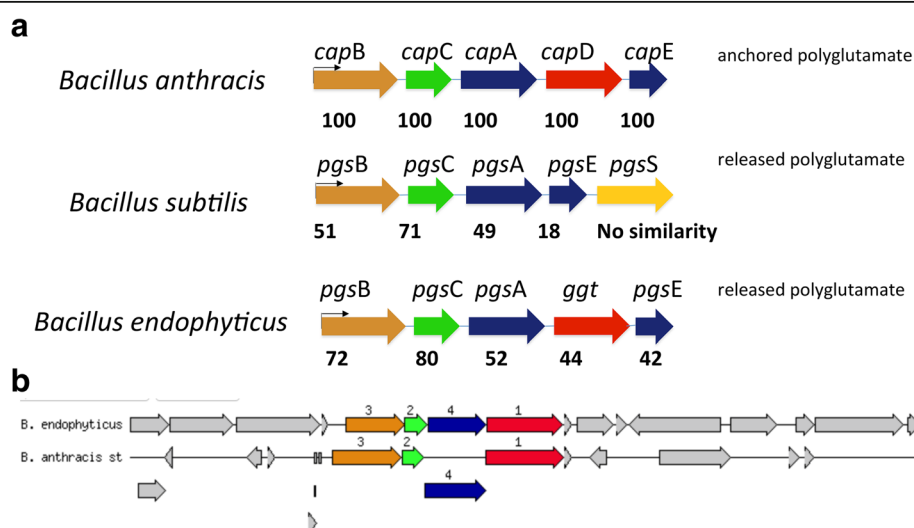


Fig. 4 Comparative structure of the polyglutamate (PGA) subunit genes of the *Bacillus endophyticus* 3631_9D, *B. anthracis* Ames and *B. subtilis* natto IF03336. All *cap/pgs* coding sequence are indicated in colours with (a) representing the comparison of the PGA synthetic operon of *B. anthracis*, *B. subtilis*, and *B. endophyticus*. Numbers indicates amino acid sequence identities (%) of the *cap/pgs* proteins to those of *B. anthracis*. (b) Indicates the annotated sequence based comparison of the *B. endophyticus* 3631_9D and *B. anthracis* Ames PGA genes. Number 1 (red) represents *pgs/capD*, 2 *pgs/capC*, 3 (brown) *pgs/capB*, 4 (blue) *pgs/capA*

Glutamyltranspeptidases (*ggt*)

An open reading frame (ORF) that encodes for γ -glutamyltranspeptidases (GGT) was present in the sequenced *B. endophyticus* strains (Fig. 4b) and other 4 compared *B. endophyticus* strains (2102, Hbe603, KCC 13922, DSM 13796). In this study, the nucleotide sequence analysis of the *ggt* in *B. endophyticus*, *B. anthracis* and other *Bacillus* species showed sequenced *B. endophyticus* strains cluster with the compared *B. endophyticus* strains (Fig. 5). Single nucleotide and amino acid variations were observed between aligned *ggt* of *B. endophyticus* and *B. anthracis* 20SD. The aligned *ggt* amino acid sequences of the reported *B. endophyticus* strains in this study are 44% identical to *B. anthracis* (Fig. 4). The sequenced *B. endophyticus* strains in the study had the same nucleotide identity profile with *B. endophyticus* DSM 13976 and KCTC 13922. *B. endophyticus* 3618_1C grouped separately amongst the other *B. endophyticus* strains, and this was also observed in the heat map (Fig. 2). There was a clear separation between the *ggt* of the *B. endophyticus* strains and the other *Bacillus* species, with the closest being *B. anthracis* Ames ancestor and *B. megaterium* (Fig. 4).

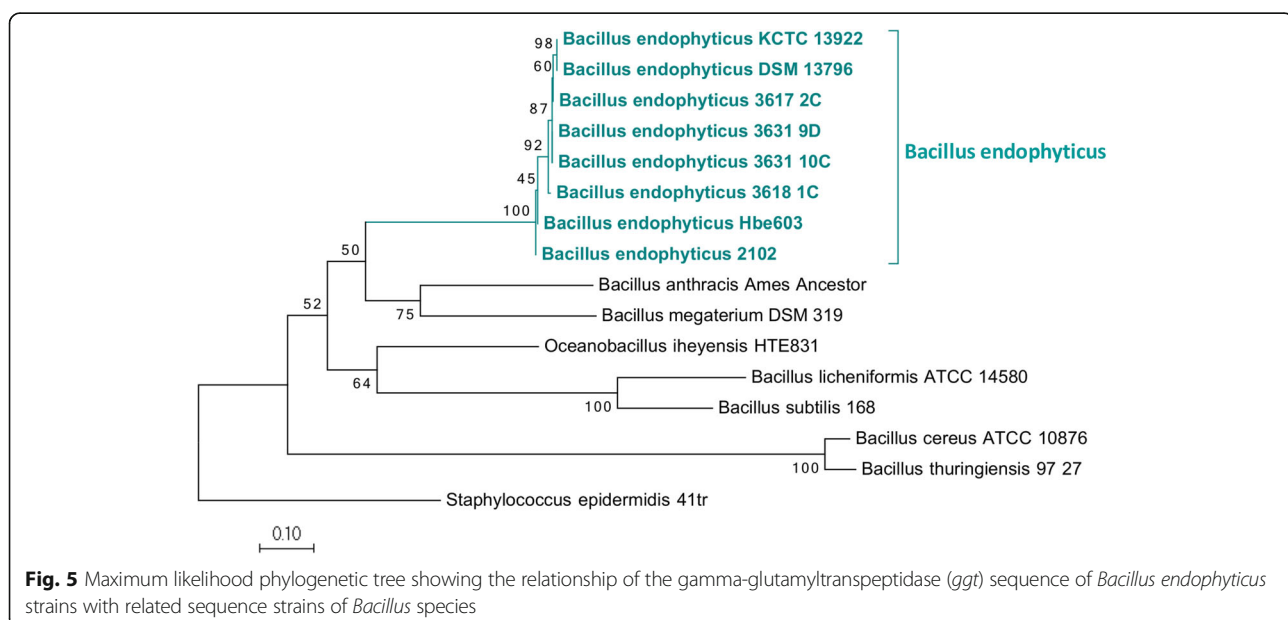
Bacillus endophyticus and *B. anthracis* features

The annotation of *B. endophyticus* strains and *B. anthracis* showed the presence of the import and iron release four-gene cluster (*feuABCD*) and the Fe-bacillibactin (iron carrier) uptake system common in both. The four-gene operon of *feuA-feuB-feuC-feuD* and trilactone hydrolase (bacillibactin) siderophore *YuiI* (*BesA*) was identified in the *B. endophyticus* genomes. Bacillibactin siderophore is synthesized through the alternative non-ribosomal peptide

synthetase pathways and helps the bacterium in iron acquisition from their environment [21]. Genes identified in both *B. endophyticus* and *B. anthracis* also included bacitracin ABC transporters, bacitracin export ATP-binding protein *BceA* and permease protein *BceB*, which confers resistance to bacitracin or stress response as defensive mechanisms.

Discussion

The presence of the PGA subunits *pgs/capA-C* in the South African *B. endophyticus* strains isolated alongside *B. anthracis* strains from the 2009 anthrax outbreak initiated the comparative investigation of these two species. *B. endophyticus* and *B. anthracis* can be differentiated based on sensitivity to the γ -phage [13], which is not a reliable differentiating character as resistance to γ -phage had been reported amongst the normally γ -phage sensitive *B. anthracis* [13, 22]. In this study a more comprehensive approach that included morphology, biochemical as well as WGS were used to compare these two species in order to identify differentiating characteristics for diagnostic purposes. *B. endophyticus* has not been reported to date to be isolated with *B. anthracis*. This and the presence of PGA genes in *B. endophyticus* is noteworthy since capsule genes is an important diagnostic characteristic of *B. anthracis*. *B. anthracis* could be differentiated from *B. endophyticus* based on unique capsulated box-shaped bacilli in long chains (in culture), γ -phage susceptibility characteristic and the presence of the toxin *pag* gene. *B. endophyticus* showed round-edged bacilli present either as single cells or in short chains, γ -phage resistant and the absences of the toxin *pag* gene. Identification of the *pgs/capBCA* genes of the PGA biosynthetic pathways



in both species using WGS comparative analysis shows the value of this approach. The *pgsBCA*, γ -glutamyl-transpeptidase (*ggt*) and *pgsE* open reading frames were identified in the chromosomes of *B. endophyticus* genomes.

The South African *B. endophyticus* strains were differentiated from *B. anthracis* based on γ -phage microbiological characteristics and real-time PCR, whereas 16S rRNA sequences and Omnilog identified the *B. endophyticus* strains [13]. However identification of the *B. cereus sensu lato* group using 16S rRNA gene sequencing is often challenging, since it has been regarded as a single taxon based on similar 16S rRNA sequences [23]. The diagnosis of *B. anthracis* require the use of microbiology characteristics as well as conventional or real-time PCR that detects *B. anthracis* specific chromosomal regions, toxin genes on pXO1 and capsule genes on pXO2 [2]. However, regions similar to the *B. anthracis* plasmids (pXO1 and pXO2) have been reported in other *Bacillus* species [11, 12] as observed with conventional PCR of *B. endophyticus* that amplified *capA*, *capB* and *capC* regions [13].

Previous studies have reported a close relationship between *B. endophyticus* and *B. smithii* [1], which was also demonstrated in this study (Table 2, Additional file 3: Figure S3). They could be differentiated based on capsule, motility and rods morphology appearance [1, 13, 24]. The WGS of *B. endophyticus* strains reported in the study were closely related to *B. megaterium* DSM 319 using RAST as reported in *B. endophyticus* 2102 WGS [25]. However, *B. megaterium* DSM 319 does not contain any plasmids unlike other *B. megaterium* strains [26] and this has the potential to create a bias in RAST annotations [17]. *B. megaterium* bacilli (2.0–5.0 μm) are slightly larger than *B. endophyticus* (2.5–3.5 μm) and both are non-motile (Table 2). Features of *B. megaterium* can be confused with *B. anthracis* as both are non-motile, encapsulated and some *B. megaterium* strains are non-hemolytic [27], but can be differentiated based on penicillin and γ -phage sensitivity [28]. The γ -phage sensitivity is noted in *B. anthracis* strains containing the γ -phage receptor GamR gene [29]. None of the sequenced and compared *B. endophyticus* genomes had this gene. *B. endophyticus* is also non-motile, non-hemolytic, and penicillin sensitive, which do not distinguish it from *B. anthracis*. *B. megaterium*, *B. endophyticus* and *B. anthracis* can be differentiated based on morphology followed by verification of virulence factors and/or prophage region using real-time PCR [30].

None of the lambda prophage regions of *B. anthracis* were found in *B. endophyticus* using WGS comparative analysis. As indicated the prophage regions of *B. anthracis* lambdaBa03 (01–04) accurately distinguished *B. anthracis* from *B. endophyticus* and other related *Bacillus* species

[30]. The *B. endophyticus* strains in this study presented many different prophage regions. The *B. endophyticus* strains 3618_1C shared common prophages with the *B. endophyticus* DSM_13,796 and KCTC 13922. Jia et al. [16] determined four prophage regions in the *B. endophyticus* Hbe603 strain, which were determined as hypothetical proteins that are different from the prophages in *B. endophyticus* strains reported in this study. The shared prophage regions amongst *B. endophyticus* strains can be investigated as more genomes become available that could be used in diagnostic assays.

WGS of the sequenced *B. endophyticus* strains in this study are closely related to *B. endophyticus* DSM 13796 and KCTC 13922 based on the average nucleotide identity (Fig. 2). The overrepresentation of the COG in the core cluster analysis might show that *B. endophyticus* has a high number of highly conserved genes and that horizontal gene transfer does not necessarily play a major role in its evolution. One key feature of *B. endophyticus* identified through WGS is the bacillibactin-associated genes for biosynthesis that are also present in *B. anthracis* and many other members of the *B. cereus sensu lato* group [21]. The bacitracin cluster of genes identified in *B. endophyticus* and *B. anthracis* are known to be a peptide antibiotic that is non-ribosomally synthesized in some strains of *Bacillus* [31], especially in *B. subtilis*. It has the ability to disrupt the cell wall and peptidoglycan synthesis of the gram-positive and gram-negative bacteria. However, bacillibactin and bacitracin cannot be used as differentiating features of *B. endophyticus* strains since they are also present in *B. anthracis* strains.

B. endophyticus Hbe603 consists of one chromosome and 8 plasmids that belong to the members of *Bacillus* group [16]. The function or the role of the plasmids has not yet been studied. Sequence comparison revealed no similarities between *B. endophyticus* and *B. anthracis* plasmids. The PGA complex is present in most *Bacillus* species including *B. licheniformis* [32], *B. subtilis* [10], *B. anthracis* [4] and *B. cereus sensu lato* group including *B. cereus* biovar *anthracis* [12, 33]. In this study, the PGA biosynthesis operon was also identified in the *B. endophyticus* genomes. The PGA subunits are located in the chromosome of the *B. endophyticus* strains unlike in the plasmid of *B. anthracis*.

The polyglutamate depolymerase *capD* is present in the *B. anthracis* [7] and belongs to the γ -glutamyltransferase (GGT) family. This gene is responsible for the covalent anchoring of the capsule to peptidoglycan and act as a depolymerase in *B. anthracis* [7]. *B. anthracis capD* gene is related to *B. subtilis* natto *ywrD* and *B. licheniformis* DSM13 *ggt*. However, *ywrD* or *ggt* is located in the chromosome and reside in a locus distant from the *pgsBCA* subunits. The *ggt* and *capD* subunits were present in both *B. endophyticus* and *B. anthracis* genomes

respectively (Fig. 4). The *ggt* is located on a locus adjacent to the *pgsBCA* subunit genes in the chromosome of the sequenced *B. endophyticus* (3631_9D, 3618_1C, 3631_10C, 3617_2C) and other compared *B. endophyticus* genomes (2102, Hbe603, KCC 13922, DSM 13796). The *ggt* identified in the *B. endophyticus* has different nucleotide and amino acid variations with *B. anthracis* and *B. subtilis*. Annotation of this subunit in *B. endophyticus* strains showed that is not linked with the attachment of the PGA to peptidoglycan, however it's associated with the PGA biosynthesis. The identified γ -glutamyltransferase in *B. endophyticus* genomes may suggest that it hydrolyses PGA biosynthesis as suggested for *B. subtilis* *ggt* that hydrolyses the PGA in an exo-type manner [34]. In *B. subtilis* NAFM5, the GGT was shown to have hydrolysed γ -D-L PGA from the D- and L-glutamate during stationary phase through transcriptional activation [35].

The *pgsE* subunit is known to stimulate the PGA production in the presence of zinc [4]. However in *B. subtilis*, high concentrations of *pgsB*, *pgsC*, and *pgsA* were determined to form PGA in the absence of *pgsE* [36]. There is a small ORF present in *B. endophyticus* (Fig. 4) strains annotated as hypothetical protein, which has the same nucleotide size (144 bp) than *B. anthracis* *capE*. Protein alignment of *B. endophyticus* *pgs/capE* is 42% identical to *capE* of *B. anthracis*. This ORF may be important for the PGA biosynthesis and act as an analog *pgs/capE* since *B. anthracis* *capE* is required for PGA biosynthesis [4]. The small ORF is found after the *ggt/capD* in both the *B. endophyticus* and *B. anthracis* PGA operon (Fig. 4). The *B. subtilis* *pgsS* is an exo- γ -glutamyl hydrolase that is linked to the release of PGA in the environment [4]. The γ -D-L-glutamyl hydrolase *pgsS* lies immediately downstream of *pgsBCA* genes in *B. subtilis* chromosome [37]. This subunit encode enzyme that cleaves the glutamyl bond between D- and L-glutamic acids of PGA. The *pgsS* subunit was not identified in the *B. endophyticus* genomes. An ORF was identified in the PGA operon of *B. endophyticus* genomes, annotated as a putative esterase/lipase, that lies immediately downstream after the *pgsE*. This putative extracellular esterase belongs to the hydrolase enzymes family that might also be involved in the hydrolysis of PGA, but this hypothesis needs further investigation. The regulatory genes, *acpA*, *acpB* and *atxA* (located in pXO1) are known to control the expression of *B. anthracis* capsule PGA biosynthesis operon *capBCADE* [5]. The two regulons *acpA* and *acpB* located in the pXO2 were observed in the *B. anthracis* 20SD PGA biosynthesis operon, which is absent in the *B. endophyticus* PGA operon.

The exo-polysaccharide biosynthesis ORF was identified in the *B. endophyticus* genomes. It consisted of manganese-dependent protein-tyrosine phosphatase,

tyrosine-protein kinase transmembrane modulator *epsC*, and tyrosine-protein kinase *epsD*. The tyrosine-protein kinase transmembrane modulator *EpsC* and tyrosine-protein kinase *EpsD* are found in the same operon. Extracellular polysaccharides (EPS) are polymers that consist of different simple sugars. They are produced by variety of bacteria and may be assembled as capsular polysaccharides (CPS) tightly associated with the cell surface or they may be liberated into the growth medium. In *E. coli* and *B. subtilis* the *epsC* and *epsD* are reported to control UDP-glucose dehydrogenase activity [38, 39]. In *B. subtilis* strains, cells are held together by EPS and amyloid-like fibers for biofilm formation [40]. In *B. endophyticus* genomes, in the same operon of exo-polysaccharide, the UDP-glucose dehydrogenase and hyaluronan synthase enzymes were identified. Hyaluronan synthase is membrane bound enzyme that is used to produce the glycosaminoglycan hyaluronan at the cell surface through the membrane. Hyaluronan synthesis in most bacteria is associated with protecting the bacteria against host and environmental factors, which may be detrimental to survival [41]. The hyaluronic acid polysaccharide capsule was found in the *Streptococcus pyrogenes* [41]. In order for *S. pyrogenes* to synthesize a HA capsule, at least three different genes must be present and arranged in an operon designated the HA synthesis manner [41]. This includes the HA synthase and two sugar precursors (UDP-glucose dehydrogenase and UDP-glucose-pyrophosphorylase). In *B. endophyticus* genomes only one sugar precursor UDP-glucose dehydrogenase, and the hyaluronan synthase are present. The role of HA needs further investigation in the *B. endophyticus* strains.

Conclusion

B. endophyticus is a gram-positive, non-motile, non-hemolytic, rod-shaped bacterium which is endospore forming, penicillin sensitive but γ -phage resistant. *B. anthracis* has all these characteristics in common with *B. endophyticus* with the exception that it is γ -phage sensitive bacterium. *Bacillus* species, which include *B. anthracis*, *B. megaterium*, *B. endophyticus* and *B. smithii* can be differentiated based on their morphological appearances and other microbiological features. However, most of these microbiological features (biochemical tests i.e. the presence of lecithinase, starch, VP test motility and other tests) are not routinely used for identification and characterization of *Bacillus* species. Molecular techniques such as real-time PCR targeting species-specific chromosomal markers, virulence genes and 16S rRNA sequencing, should continuously be used to identify or distinguish related *Bacillus* species. This can further be supplemented with specific prophages of the bacterium or other specific genes present in the genome. *B. endophyticus* is considered as industrial important due to biotechnology properties like the production of antibiotics such as fosfomycin and bacitracin.

B. endophyticus can easily be differentiated from *B. anthracis* based on the morphology appearance but confirmation of virulence factors like capsule genes identified in *B. endophyticus* could complicate anthrax diagnostics. Whole genome sequencing identified and differentiated *B. anthracis* and *B. endophyticus* PGA capsule genes. *B. anthracis* and *B. endophyticus* PGA biosynthesis subunits were determined to be located in the pXO2 and chromosome respectively. The *B. endophyticus* strains couldn't synthesize a surface associated γ -PGA, suggesting that PGA helps the bacteria to survive under adverse conditions. Therefore *B. endophyticus* is a non-capsulated bacterium that survives at high salt concentrations. Prophage regions have emerged as key markers in distinguishing *B. anthracis* and eliminating other related *Bacillus* species. The study highlights the significance of using whole genome shotgun sequencing to identify virulence and other important genes that might be present amongst unknown samples from natural outbreaks.

Methods

Isolates

The *B. endophyticus* and *B. anthracis* isolates included in this study were isolates collected during the 2009 anthrax outbreak in the Northern Cape Province (NCP) of South Africa. These isolates included a *B. endophyticus* and *B. anthracis* isolate from the same animal. The *B. endophyticus* were isolated from blood collected from animal carcasses whereas *B. anthracis* isolates were isolated from soil below the carcass as well as blood collected from animal carcasses (Table 1). The *B. endophyticus* isolates exhibited some similar phenotypic and genetic similarities to those of *B. anthracis* [13] and therefore we characterized these isolates to enhance and contribute to the diagnosis of *B. anthracis*. The incubation condition for *B. endophyticus* range from 10 to 55 °C although the optimum growth temperature is between 28 and 30 °C, but this study used conditions specific for anthrax diagnosis as described by the International protocols for anthrax [42].

Phenotypic characterization

In this study we focused mainly on capsule characterization of *B. endophyticus* strains to enhance the phenotypic characterization previously done on South African *B. endophyticus* and *B. anthracis* strains [13] as well as summarizing phenotypic characterizations of related *Bacillus* species. Four *B. endophyticus* and three *B. anthracis* strains isolated from animal anthrax cases in NCP in South Africa available at Agricultural Research Council–Onderstepoort Veterinary Institute (ARC-OVI) were used in this study (Table 1). The *B. endophyticus* and *B. anthracis* isolates were collected from 2009 anthrax outbreaks in the NCP of South Africa (Table 1). The

samples were processed at the ARC-OVI reference laboratory (Onderstepoort, South Africa), where *B. anthracis* suspected cases are confirmed. Pure cultures were grown on 5% SBTA, followed by incubation at 37 °C for 24 h for the observation of colony morphology and to determine hemolytic activity [42]. Colony morphology was observed on nutrient agar containing 0.8% sodium bicarbonate following incubation in the presence of 5% CO₂ at 37 °C for 24–48 h in the dark to induce capsule formation. The capsules from strains incubated on 0.8% sodium bicarbonate supplemented nutrient agar were stained using India ink, Giemsa and copper sulphate followed by visualization using light microscopy [42, 43]. Each culture was also transferred to blood serum and incubated under both aerobic and anaerobic conditions at 37 °C for 24 h to determine the formation of a capsule [42]. Blood smears were stained using Rapi-Diff and visualized by light microscopy. The positive control for the capsule production included *B. anthracis* 3618_2D (cap⁺, virulent strain) [13] whereas the negative controls included *B. licherniformis* ATCC 12759 (cap⁻) and *B. anthracis* Sterne (cap⁻) strains. Phenotypic properties of *B. endophyticus* and *B. anthracis* were compared with those from published literature including *B. megaterium* and *B. cereus* as shown in Table 2 ([1, 24, 27, 42, 44], <http://www.tgw1916.net>).

Genomic DNA extraction

B. endophyticus and *B. anthracis* strains (Table 1) were inoculated in 2 ml nutrient broth, followed by overnight incubation at 37 °C. The cells were harvested by centrifugation at 5000 xg for 10 min. Genomic DNA was extracted from the harvested cells using the DNAeasy Tissue kit (Qiagen, Germany) according to the manufacturer's instructions. The isolated DNA was then quantified using the Qubit[®] fluorometric method (Life Technologies, USA) according to the manufacturer's instructions. The DNA integrity was monitored through electrophoreses using a 0.8% agarose gel pre-stained with ethidium bromide and visualized on a UV transilluminator.

High-throughput sequencing

Shotgun library preparation of four *B. endophyticus* (Table 1) strains was performed using the Nextera DNA Sample Preparation kit (Illumina, USA). Clusters generation and the sequencing were performed using the TruSeq™ PE Cluster kit v2-cBot-HS and TruSeq SBS v3-HS (200 cycle) kit respectively (Illumina, USA). The sequencing was performed on the HiScan SQ sequencer (Illumina, USA).

Genome assembly and annotation

Sequence data quality was assessed using FastQC software v 0:10.1 [45]. Ambiguous nucleotide sequence and sequence adapters were trimmed using CLC Genomic Workbench 7.5 (Denmark). The de novo assemblies

were performed using the CLC Genomic Workbench 7.5. The *B. endophyticus* strains contigs were further extracted and analyzed with BLASTn [46] using *B. endophyticus* Hbe603 (Genbank accession no: CP011974) as reference genome. MAUVE tool [47] was used to order the sequence of *B. endophyticus* reported in the study using *B. endophyticus* Hbe603 as a reference. The assembled contigs were annotated using NCBI prokaryotic genome automatic annotation pipeline (PGAAP) and rapid annotation using subsystem technology [48] annotation server for subsystems and functional annotation [17]. The presence of prophage sequence regions in the 8 *B. endophyticus* genomes (3631_9D, 3631_10C, 3618_1C, 3617_2C, Hbe603, 2102, KCC 13922, and DSM 13796) were determined using PHAge Search Tool (PHAST) [49].

16S rRNA gene phylogenetic analysis

The 16S rRNA sequence region consisting of approximately 1500 bases was extracted from the assembled genomes of *B. endophyticus* strains (3631_9D, 3618_1C, 3631_10C and 3617_2C). These sequences were further aligned and compared with the 16S rRNA gene sequences of *Bacillus* species available in NCBI (<http://www.ncbi.nlm.nih.gov>). NCBI BLAST homology searches of the 16S rRNA gene sequences were performed to assess homologous hits to sequences available in NCBI. Multiple alignments of the gene sequences extracted from assembled genomes and from those mined from NCBI were performed using MAFFT [50]. Maximum likelihood analysis of the *B. endophyticus* 16S rRNA nucleotide sequences and related *Bacillus* group sequences were performed using 1000 bootstrap iterations in MEGA 6.0.

Average nucleotide identities, pan-genome analyses and functional classification of orthologous genes

The CDSs (coding domain sequences) of *B. endophyticus* sequenced strains were subsequently compared against each using pair-wise BLASTn, to allow for calculations of average nucleotide identities. The pan-genome homology of all the 8 *B. endophyticus* (3631_9D, 3631_10C, 3618_1C, 3617_2C, Hbe603, 2102, KCC 13922, and DSM 13796) were computed using the get homologues tool [51] with default parameters. In brief, the tool conducted similarity searches between the CDSs of all 8 genomes using pair-wise BLASTp [46], and these were subsequently clustered into the different pan-genomic categories using OrthoMCL [52]. The analysis resulted in four clusters, and these were defined as: core-genes present in all genomes; softcore-genes present in 95% of the genomes; shell-genes present in few but not all genomes; and the cloud – genes present in two or less of the genomes. The core and the softcore represent sets of conserved or house keeping genes. The softcore clusters

were included in the analysis because the sequenced draft genomes of *B. endophyticus* strains in this study might be missing some of the essential genes. Both the shell and the cloud compose of accessory genes that play a role towards the lifestyle and adaptation characteristics of an organism to its particular environment.

The four clusters determined for the 8 genomes were searched for shared pattern similarities against a conserved domain database of cluster of orthologous groups using rps-blast with $-E < 1e-3$. Genes with shared pattern similarities were assigned classes that were later categorized into the COG (clusters of orthologous group) subgroups to determine their distributions for all the clusters.

Polyglutamate (PGA) subunit genes analysis

The presences of the PGA synthesis genes were determined for the 8 *B. endophyticus* strains (3631_9D, 3631_10C, 3618_1C, 3617_2C, Hbe603, 2102, KCC 13922, and DSM 13796) using analyses on the RAST server with the annotated draft genomes [17]. The PGA-capsule subunits were extracted from the annotated contigs of the *B. endophyticus* genomes using comparative analysis of RAST. *B. anthracis* PGA capsule subunits were compared with the *B. endophyticus* PGA subunits using the same annotation system. BLASTp [46] was used to compare the PGA proteins of *B. anthracis*, *B. endophyticus* and *B. subtilis*. Phylogenetic tree analysis of the subunit gene *capD/pgsD* of *B. endophyticus*, *B. anthracis* and other closely related species was constructed using maximum-likelihood. Multiple alignments of the gene sequences were constructed using multiple sequence alignment based on fast fourier (MAFFT) [50]. Alignment of the corresponding amino acid sequences was performed using CLC Genomic Workbench 7.5. MEGA 6.0 was used to construct the phylogenetic tree using 1000 bootstrap iterations.

Genome sequences and accession numbers

The four sequenced genomes of *B. endophyticus* were deposited in the Genbank genome database under accession numbers: *B. endophyticus* 3631_9D LVYL000000000, *B. endophyticus* 3631_10C LVYK000000000, *B. endophyticus* 3618_1C LWAI000000000 and *B. endophyticus* 3617_2C LWAG000000000. The additional four genomes that were used in the comparative analysis of *B. endophyticus* strains were retrieved from NCBI genebank. Accession numbers: *B. endophyticus* Hbe603 GCA_000972245.3, *B. endophyticus* 2102 GCA_000283255.1, *B. endophyticus* DSM_13796 GCA_900115845.1 and *B. endophyticus* KCTC 13922 GCA_001590825.1. The sequenced *B. endophyticus* genome sequences in the study were further compared to the South African *B. anthracis* 20SD and 3631_1C strains (Genbank accession nr LGCC000000000 and LGCD000000000).

Additional files

Additional file 1: Figure S1. (1) Colony morphology of (a) *Bacillus endophyticus* that is small circular, wet and non-mucoid and (b) *B. anthracis* appear circular, mucoid on nutrient agar supplemented with sodium bicarbonate at 5% CO₂ after incubation at 37 °C. Colony morphology of *B. endophyticus* and *B. anthracis* on sheep blood agar incubated at 37 °C. *B. anthracis* shows the characteristic shiny, rough with ground-glass appearance compared to the white slimy and smooth colonies of *B. endophyticus*. (TIFF 2652 kb)

Additional file 2: Figure S2. Phenotypic electron microscopic examination of the morphology of *B. endophyticus* strains after 24 h incubation on nutrient agar containing 0.8% sodium bicarbonate stained using copper sulphate. (TIFF 4206 kb)

Additional file 3: Figure S3. Phylogenetic tree of 16S ribosomal RNA sequence of the *Bacillus endophyticus* 3618_1C, 3631_9D, 3617_2C and 3631_10C strains with related *Bacillus* species using maximum likelihood. *Geobacillus thermoglucosidarius* was used as an out-group. Bootstrap values > 60 are indicated at the internodes. (TIFF 219 kb)

Additional file 4: Table S1. Plasmid comparison of the four sequenced *Bacillus endophyticus* strains (3618_1C, 3631_9D, 3631_10C, 3617_2C) with *B. endophyticus* Hbe603 strain. (DOC 18 kb)

Abbreviations

ANI: Average nucleotide identity; ARC-OVI: Agricultural Research Council–Onderstepoort Veterinary Institute; BLAST: Basic local alignment search tools; BLASTn: Basic local alignment search tool nucleotide; BLASTp: Basic local alignment search tool protein; CAP: Capsule; CDSs: Coding domain sequences; COG: Clusters of orthologous group; CPS: Capsular polysaccharides; EPS: Extracellular polysaccharides; GGT-gamma: Glutamyltranspeptidases; MAFFT: Multiple sequence alignment based on fast fourier; MAUVE: Multiple alignment of conserved genomic sequence with rearrangements; NA: Not available; NaCl: Sodium chloride; NCBI: National centre for biotechnology information; NCP: Northern Cape province; NGS: Next generation sequencing; ORF: Open reading frame; PCR: Polymerase chain reaction; PGA: Polyglutamate; PGAAP: Prokaryotic genome automatic annotation pipeline; PGS: Polyglutamate synthase; RAST: Rapid annotations using subsystems technology; WGS: Whole genome sequencing; γ : Gamma

Acknowledgements

We are grateful to the National Research Foundation (NRF), NRF-THRIP, Agricultural Research Council-Biotechnology Platform and AgriSETA for financial support.

Funding

This study was funded by the National Research Foundation (NRF), NRF-THRIP, Agricultural Research Council-Biotechnology Platform and AgriSETA for financial support. NRF, NRF-THRIP and AgriSETA assisted with running of the experiments, while ARC-Biotechnology provided funding for sequencing and bioinformatics support. The funding organizations only supplied funds and did not review or supply any other input.

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Endnotes

Not applicable.

Authors' contributions

KEL participated for running the experiments and writing the manuscript. OKIB assisted in running some of the bioinformatics analysis. JM and JR assisted in design of the project and drafting the manuscript. EM, FC and HvH participated in the design of the study, drafting the manuscript, revising it critically and provided funding. All authors have read and approved the manuscript.

Ethics approval and consent to participate

No animals involved, not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹The Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort 0110, South Africa. ²Department of Veterinary Tropical Diseases, University of Pretoria, Private bag X4, Onderstepoort 0110, South Africa. ³Bacteriology section, Agricultural Research Council-Onderstepoort Veterinary Institute, Private Bag X5, Onderstepoort 0110, South Africa. ⁴College of Agriculture and Environmental Sciences, University of South Africa, Florida Campus, Christiaan De Wet/ Pioneer Dr, P.O. Box X6, Florida 1710, South Africa. ⁵Department of Biochemistry and Microbiology, University of Zululand, Private Bag X1001, KwaDlangezwa 3886, South Africa.

Received: 30 November 2017 Accepted: 20 June 2018

Published online: 09 July 2018

References

- Reva ON, Smirnov VV, Pettersson B, Priest FG. *Bacillus endophyticus* sp. nov., isolated from the inner tissues of cotton plants (*Gossypium* sp.). *Int J Syst Evol Microbiol.* 2002;52(Pt 1):101–7.
- Health WofA. Manual of diagnostic tests and vaccines for terrestrial animals: mammals, birds and bees. In: Biological Standards Commission, World Organization for Animal Health Paris, France; 2012.
- Makino S, Uchida I, Terakado N, Sasakawa C, Yoshikawa M. Molecular characterization and protein analysis of the cap region, which is essential for encapsulation in *Bacillus anthracis*. *J Bacteriol.* 1989;171(2):722–30.
- Candela T, Fouet A. Poly-gamma-glutamate in bacteria. *Mol Microbiol.* 2006; 60(5):1091–8.
- Drysdale M, Bourgogne A, Hilsenbeck SG, Koehler TM. atxA controls *Bacillus anthracis* capsule synthesis via acpA and a newly discovered regulator, acpB. *J Bacteriol.* 2004;186(2):307–15.
- Drysdale M, Heninger S, Hutt J, Chen Y, Lyons CR, Koehler TM. Capsule synthesis by *Bacillus anthracis* is required for dissemination in murine inhalation anthrax. *EMBO J.* 2005;24(1):221–7.
- Candela T, Fouet A. *Bacillus anthracis* CapD, belonging to the gamma-glutamyltranspeptidase family, is required for the covalent anchoring of capsule to peptidoglycan. *Mol Microbiol.* 2005;57(3):717–26.
- Mesnage S, Tosi-Couture E, Mock M, Gounon P, Fouet A. Molecular characterization of the *Bacillus anthracis* main S-layer component: evidence that it is the major cell-associated antigen. *Mol Microbiol.* 1997;23(6):1147–55.
- Candela T, Mock M, Fouet A. CapE, a 47-amino-acid peptide, is necessary for *Bacillus anthracis* polyglutamate capsule synthesis. *J Bacteriol.* 2005;187(22): 7765–72.
- Ashiuchi M, Soda K, Misono H. A poly-gamma-glutamate synthetic system of *Bacillus subtilis* IFO 3336: gene cloning and biochemical analysis of poly-gamma-glutamate produced by *Escherichia coli* clone cells. *Biochem Biophys Res Commun.* 1999;263(1):6–12.
- Rasko DA, Rosovitz MJ, Okstad OA, Fouts DE, Jiang L, Cer RZ, Kolsto AB, Gill SR, Ravel J. Complete sequence analysis of novel plasmids from emetic and periodontal *Bacillus cereus* isolates reveals a common evolutionary history among the *B. Cereus*-group plasmids, including *Bacillus anthracis* pXO1. *J Bacteriol.* 2007;189(1):52–64.
- Klee SR, Brzuszkiewicz EB, Nattermann H, Bruggemann H, Dupke S, Wollherr A, Franz T, Pauli G, Appel B, Liebl W, et al. The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. Cereus* with *B. Anthracis* virulence plasmids. *PLoS One.* 2010;5(7):e10986.
- Lekota KE, Hassim A, Mafofo J, Rees J, Muchadeyi FC, Van Heerden H, Madoroba E. Polyphasic characterization of *Bacillus* species from anthrax

- outbreaks in animals from South Africa and Lesotho. *J Infect Dev Ctries*. 2016;10(8):814–23.
14. Lekota KE, Mafofo J, Madoroba E, Rees J, van Heerden H, Muchadeyi FC. Draft genome sequences of two south African *Bacillus anthracis* strains. *Genome Announc*. 2015;3:6.
 15. Makarova KS, Wolf YI, Koonin EV. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res*. 2013;41(8):4360–77.
 16. Jia N, Du J, Ding MZ, Gao F, Yuan YJ. Genome sequence of *Bacillus endophyticus* and analysis of its companion mechanism in the *Ketogulonigenium vulgare*-*Bacillus* strain consortium. *PLoS One*. 2015;10(8):e0135104.
 17. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. The RAST server: rapid annotations using subsystems technology. *BMC Genomics*. 2008;9:75.
 18. Gotoh N, Tsujimoto H, Tsuda M, Okamoto K, Nomura A, Wada T, Nakahashi M, Nishino T. Characterization of the MexC-MexD-OprJ multidrug efflux system in *DeltamexA-mexB-oprM* mutants of *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother*. 1998;42(8):1938–43.
 19. Xu Y, Sim SH, Song S, Piao S, Kim HM, Jin XL, Lee K, Ha NC. The tip region of the MacA alpha-hairpin is important for the binding to TolC to the *Escherichia coli* MacAB-TolC pump. *Biochem Biophys Res Commun*. 2010;394(4):962–5.
 20. Horsburgh MJ, Moir A, Sigma M, an ECF RNA polymerase sigma factor of *Bacillus subtilis* 168, is essential for growth and survival in high concentrations of salt. *Mol Microbiol*. 1999;32(1):41–50.
 21. Hotta K, Kim CY, Fox DT, Koppisch AT. Siderophore-mediated iron acquisition in *Bacillus anthracis* and related strains. *Microbiology*. 2010;156(Pt 7):1918–25.
 22. Odendaal MW, Pieterse PM, de Vos V, Botha AD. The biochemical, morphological and virulence profiles of *Bacillus anthracis* isolated in the Kruger National Park. *Onderstepoort J Vet Res*. 1991;58(1):21–6.
 23. Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolsto AB. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. *Appl Environ Microbiol*. 2000;66(6):2627–30.
 24. Sumathi C, Jayashree S, Sekaran G. Screening and evaluation of antimicrobial activity of tannery fleshing utilizing fish gut bacteria. 2012.
 25. Lee YJ, Lee SJ, Kim SH, Lee SJ, Kim BC, Lee HS, Jeong H, Lee DW. Draft genome sequence of *Bacillus endophyticus* 2102. *J Bacteriol*. 2012;194(20):5705–6.
 26. Eppinger M, Bunk B, Johns MA, Edirisinghe JN, Kutumbaka KK, Koenig SS, Creasy HH, Rosovitz MJ, Riley DR, Daugherty S, et al. Genome sequences of the biotechnologically important *Bacillus megaterium* strains QM B1551 and DSM319. *J Bacteriol*. 2011;193(16):4199–213.
 27. Beesley CA, Vanner CL, Helsel LO, Gee JE, Hoffmaster AR. Identification and characterization of clinical *Bacillus* spp. isolates phenotypically similar to *Bacillus anthracis*. *FEMS Microbiol Lett*. 2010;313(1):47–53.
 28. Turnbull, PC. Definitive identification of *Bacillus anthracis*—a review. *J Appl Microbiol*. 1999;87(2):237–40.
 29. Davison S, Couture-Tosi E, Candela T, Mock M, Fouet A. Identification of the *Bacillus anthracis* (gamma) phage receptor. *J Bacteriol*. 2005;187(19):6742–9.
 30. Agren J, Hamidjaja RA, Hansen T, Ruuls R, Thierry S, Vigre H, Janse I, Sundstrom A, Segerman B, Koene M, et al. In silico and in vitro evaluation of PCR-based assays for the detection of *Bacillus anthracis* chromosomal signature sequences. *Virulence*. 2013;4(8):671–85.
 31. Johnson BA, Anker H, Meleney FL. Bacitracin: a new antibiotic produced by a member of the *B. Subtilis* group. *Science*. 1945;102(2650):376–7.
 32. McLean RJ, Beauchemin D, Clapham L, Beveridge TJ. Metal-binding characteristics of the gamma-Glutamyl capsular polymer of *Bacillus licheniformis* ATCC 9945. *Appl Environ Microbiol*. 1990;56(12):3671–7.
 33. Hoffmaster AR, Ravel J, Rasko DA, Chapman GD, Chute MD, Marston CK, De BK, Sacchi CT, Fitzgerald C, Mayer LW, et al. Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *Proc Natl Acad Sci U S A*. 2004;101(22):8449–54.
 34. Uchida I, Makino S, Sasakawa C, Yoshikawa M, Sugimoto C, Terakado N. Identification of a novel gene, dep, associated with depolymerization of the capsular polymer in *Bacillus anthracis*. *Mol Microbiol*. 1993;9(3):487–96.
 35. Kimura K, Tran LS, Uchida I, Itoh Y. Characterization of *Bacillus subtilis* gamma-glutamyltransferase and its involvement in the degradation of capsule poly-gamma-glutamate. *Microbiology*. 2004;150(Pt 12):4115–23.
 36. Yamashiro D, Yoshioka M, Ashiuchi M. *Bacillus subtilis* pgsE (formerly ywtC) stimulates poly-gamma-glutamate production in the presence of zinc. *Biotechnol Bioeng*. 2011;108(1):226–30.
 37. Ashiuchi M, Misono H. Biochemistry and molecular genetics of poly-gamma-glutamate synthesis. *Appl Microbiol Biotechnol*. 2002;59(1):9–14.
 38. Grangeasse C, Obadia B, Mijakovic I, Deutscher J, Cozzzone AJ, Doublet P. Autophosphorylation of the *Escherichia coli* protein kinase Wzc regulates tyrosine phosphorylation of Ugd, a UDP-glucose dehydrogenase. *J Biol Chem*. 2003;278(41):39323–9.
 39. Elsholz AK, Wacker SA, Losick R. Self-regulation of exopolysaccharide production in *Bacillus subtilis* by a tyrosine kinase. *Genes Dev*. 2014;28(15):1710–20.
 40. Branda SS, Gonzalez-Pastor JE, Dervyn E, Ehrlich SD, Losick R, Kolter R. Genes involved in formation of structured multicellular communities by *Bacillus subtilis*. *J Bacteriol*. 2004;186(12):3970–9.
 41. Weigel PH. Functional characteristics and catalytic mechanisms of the bacterial hyaluronan synthases. *IUBMB Life*. 2002;54(4):201–11.
 42. Organization WH, Epizootics IOo: Anthrax in humans and animals: World Health Organization; 2008.
 43. Markey BK, Leonard FC, Archambault M, Cullinane A, Maguire D. *Clinical Veterinary*. In: *Microbiology*, 2nd edition. Edinburgh: Elsevier; 2013. p. 901.
 44. Buck CA, Anacker RL, Newman FS, Eisenstark A. Phage isolated from lysogenic *Bacillus anthracis*. *J Bacteriol*. 1963;85:1423–30.
 45. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.
 46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
 47. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004;14(7):1394–403.
 48. Gierczynski R, Kaluzewski S, Rakin A, Jagielski M, Zasada A, Jakubczak A, Borkowska-Opacka B, Rastawicki W. Intriguing diversity of *Bacillus anthracis* in eastern Poland—the molecular echoes of the past outbreaks. *FEMS Microbiol Lett*. 2004;239(2):235–40.
 49. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res*. 2011;39(Web Server issue):W347–52.
 50. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
 51. Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol*. 2013;79(24):7696–701.
 52. Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, Shanmugam D, Roos DS, Stoekert CJ Jr. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr Protoc Bioinformatics*. 2011;Chapter 6(Unit 6 12):11–9.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

