BMC Infectious Diseases

RESEARCH

Open Access

Pan-genome analysis reveals novel chromosomal markers for multiplex PCR-based specific detection of *Bacillus anthracis*

Tuvshinzaya Zorigt[1,2*], Yoshikazu Furuta[1,2], Atmika Paudel[1,2,3], Harvey Kakoma Kamboyi[1,2], Misheck Shawa[1,2], Mungunsar Chuluun[1,2], Misa Sugawara[1], Nyamdorj Enkhtsetseg[4], Jargalsaikhan Enkhtuya[5], Badgar Battsetseg[6], Musso Munyeme[7], Bernard M. Hang'ombe[8] and Hideaki Higashi[1,2]

Abstract

Background *Bacillus anthracis* is a highly pathogenic bacterium that can cause lethal infection in animals and humans, making it a significant concern as a pathogen and biological agent. Consequently, accurate diagnosis of *B. anthracis* is critically important for public health. However, the identification of specific marker genes encoded in the *B. anthracis* chromosome is challenging due to the genetic similarity it shares with *B. cereus* and *B. thuringiensis*.

Methods The complete genomes of *B. anthracis*, *B. cereus*, *B. thuringiensis*, and *B. weihenstephanensis* were *de novo* annotated with Prokka, and these annotations were used by Roary to produce the pan-genome. *B. anthracis* exclusive genes were identified by Perl script, and their specificity was examined by nucleotide BLAST search. A local BLAST alignment was performed to confirm the presence of the identified genes across various *B. anthracis* strains. Multiplex polymerase chain reactions (PCR) were established based on the identified genes.

Result The distribution of genes among 151 whole-genome sequences exhibited three distinct major patterns, depending on the bacterial species and strains. Further comparative analysis between the three groups uncovered thirty chromosome-encoded genes exclusively present in *B. anthracis* strains. Of these, twenty were found in known lambda prophage regions, and ten were in previously undefined region of the chromosome. We established three distinct multiplex PCRs for the specific detection of *B. anthracis* by utilizing three of the identified genes, *BA1698*, *BA5354*, and *BA5361*.

Conclusion The study identified thirty chromosome-encoded genes specific to *B. anthracis*, encompassing previously described genes in known lambda prophage regions and nine newly discovered genes from an undefined gene region to the best of our knowledge. Three multiplex PCR assays offer an accurate and reliable alternative method for detecting *B. anthracis*. Furthermore, these genetic markers have value in anthrax vaccine development, and understanding the pathogenicity of *B. anthracis*.

Keywords *Bacillus anthracis*, Pan-genome, Anthrax, Specific genes, PCR

*Correspondence:
Tuvshinzaya Zorigt
tuvshinzaya@czc.hokudai.ac.jp
Full list of author information is available at the end of the article

Background

Bacillus anthracis is a Gram-positive, spore-forming bacterium that can cause a fatal infection called anthrax in animals and humans [1]. It is the most likely bioterrorism agent due to its history of being intentionally released against civilians, and yet spores are easily found in nature and can be utilized for large-scale production and dissemination [2]. Depending on the route of exposure, *B. anthracis* infection in humans manifests in three forms; cutaneous, gastrointestinal, and inhalational. Once the disease reaches a critical stage, treatment becomes ineffective, and the patient eventually dies from bacteremia and toxic shock. This is due to *B. anthracis* releasing various toxins, including edema factor (EF), lethal factor (LF), and protective antigen (PA), which are encoded on a plasmid pXO1. In addition, proteins (CapB, CapC, and CapA) for capsule biosynthesis encoded on another plasmid called pXO2 are required for the complete virulence of *B. anthracis* [3]. *B. anthracis* Sterne 34F2 strain, lacking the pXO2, serves as a commercial vaccine strain utilized in anthrax immunization for animals [4].

B. anthracis is a well-known member of the *Bacillus cereus sensu lato* group, which also includes other species such as *B. thuringiensis*, *B. mycoides*, *B. weihenstephanesis*, and *B. cereus*. Among the species in the *B. cereus s.l group*, *B. anthracis*, *B. cereus*, and *B. thuringiensis* share a close genetic relationship and are challenging to distinguish from one another [5]. However, their phenotypes and the diseases they cause differ significantly. While *B. anthracis* causes anthrax, *B. cereus* causes foodborne illnesses in humans, often associated with the consumption of contaminated food. In contrast, *B. thuringiensis*, primarily influences insects, acting as an effective biological pesticide. Given the public health threat posed by anthrax, it is essential to identify *B. anthracis*-specific genes and develop rapid diagnostics capable of accurately distinguishing it from other *B. cereus s.l* species.

Two main diagnostic methods are commonly employed in the differentiation of these species. The first method involves phenotypic traits such as colony characteristics, motility, antibiotic susceptibility, and hemolytic activity on blood agar [6–8]. *B. anthracis* colonies exhibit a characteristic appearance, often described as Medusa head with a concentrated center and indistinct borders. Additionally, unlike *B. cereus* and *B. thuringiensis*, *B. anthracis* is non-motile, penicillin-susceptible, and does not cause hemolysis. However, these bacteriological methods necessitate the use of specific blood agar plates and overnight incubation for colony formation. It also requires proper training to judge colony morphology accurately.

The second method utilizes genetic markers for molecular diagnostics. The two virulence plasmids have been the main target of various molecular diagnostics to detect *B. anthracis* and differentiate it from other closely related species [9]. However, it is important to note that *B. anthracis* strains can lose either one or both plasmids, as described previously [10]. In addition, unusual *B. cereus* strains have emerged that cause anthrax-like disease in humans and animals by acquiring virulence plasmids highly similar to pXO1 and pXO2 of *B. anthracis* [11]. Consequently, identifying *B. anthracis*-specific chromosomal markers is crucial to distinguish it from closely related *B. cereus* group species and remains a focal point of research for many scientists. However, certain chromosomal markers (e.g., *BA813*, *gyrA*, *gyrB*, and *rpoB*) initially identified as specific to *B. anthracis* were later found in some *B. cereus* strains, leading to false positive results [12–14]. *B. anthracis*, *B. cereus*, and *B. thuringiensis* all possess the *plcR* gene, a transcriptional activator that is active in *B. cereus* and *B. thuringiensis* but inactivated in *B. anthracis* due to thymine insertion at nucleotide position 640. This mutation enables the use of *plcR* to distinguish *B. anthracis* from its close relatives. However, detecting this single nucleotide mutation using standard methods is challenging and often results in false positives [15, 16]. Although the melt curve analysis with the mismatch amplification mutation assay [17], improved detection accuracy, there remains a need for a new alternative marker suitable for cost-effective, standard PCR applications.

This study aimed to find *B. anthracis*-specific chromosomal marker sequences or genes using pan-genome analysis, which captures the entire range of genetic variation within or between species, reduces the bias in genetic analysis and develops rapid diagnostic tools applicable in the field.

Methods

Pan-genome analysis for identifying genes potentially unique to *B. anthracis*

A total of 151 complete genomes were downloaded from the National Center for Biotechnology Information (NCBI) for analysis (Additional file 1). This dataset included 50 genomes, which were selected randomly from each of the following species: *B. anthracis*, *B. cereus*, and *B. thuringiensis.* Additionally, one complete genome of *B. weihenstephanensis* was included as an outgroup control species. The genomes were *de novo* annotated with Prokka version 1.11 [18]. Roary version 3.13.0 [19] was used to deduce the pan-genome of the dataset. The Roary utilized the Prokka annotations as input, generating a gene-presence-absence spreadsheet (Additional file 2). Finally, the Perl script was employed to count the genes present in *B. anthracis* strains but absent in the genomes of *B. cereus* and *B. thuringiensis* strains.

Heap's law

Heap's law was used to determine the pan-genome status (openness or closedness) of *B. anthracis* ($n = 115$) in comparison to its genetically close relatives, *B. cereus* ($n = 142$) and *B. thuringiensis* ($n = 93$), utilizing all available complete sequences for each species from the GenBank database (Additional file 1).

The total number of unique genes identified through sequencing additional genomes can be modeled by Heap's law ($n = kN^{\gamma}$), which follows the power law function [20]. Where n is the number of genes; N is the number of genomes; k is intercept, represents the initial number of unique genes when one genome is considered; and γ is a parameter that characterizes the rate at which new genes are added as more genomes are included. If γ is closer to 1, it indicates an open pangenome, meaning the number of unique genes continues to increase significantly with additional genomes. Conversely, if γ is closer to 0, it suggests a closed pangenome, where most genetic diversity is captured with the current dataset.

Nucleotide BLAST search and local BLAST alignment

To ensure the specificity of the genes identified from the initial pan-genome comparison among 151 genomes, each gene was submitted to a nucleotide BLAST (BLASTn) search against the NCBI database, excluding *B. anthracis*, to verify that the identified genes were not present in other organisms.

Additionally, to confirm the presence and consistency of the identified genes across various *B. anthracis* strains, a local BLAST alignment was performed incorporating the chromosomally complete 132 genomes of *B. anthracis* strains currently available in GenBank (Additional file 3).

String analysis

To predict potential physical and functional protein-protein interactions among the proteins encoded by the identified genes derived from non-prophage regions in *B. anthracis*, the STRING v.12.0 database was employed [21]. STRING integrates known and predicted protein-protein association data for a large number of organisms using various sources of information, including genomic context, high-throughput experiments, co-expression, and literature mining. The interactions were analyzed using default settings, with a confidence score cutoff of 0.7 to ensure high-confidence interactions. The results were visualized and interpreted to elucidate the potential roles of the proteins in the biological pathways and processes relevant to *B. anthracis*.

Bacterial strains

A total of 62 bacterial strains originating from diverse sources, were utilized to evaluate the efficacy and specificity of multiplex PCR. This collection of strains included 17 strains of *B. anthracis*, encompassing one commercial vaccine strain and 16 virulent strains previously isolated from wildlife, livestock, and humans infected with anthrax in Zambia and Mongolia, countries known for their endemic anthrax status in Africa and Asia, respectively. These strains were selected to represent the temporal and geographic diversity of the respective countries. In Zambia, nine strains were isolated from Lower Zambezi National Park, South Luangwa National Park, and Western province over a decade from 2011 to 2021. In Mongolia, seven strains were collected from five distinct provinces, namely Selenge, Khuvsgul, Uvurkhangai, Khentii, and Ulaanbaatar over a period of years from 2001 to 2015. Only genomic DNA of virulent *B. anthracis* strains was transported from Zambia and Mongolia to Japan. Additionally, the collection included 29 *B. cereus* strains formerly isolated from nosocomial infection cases in Japanese hospitals [22] or obtained from reputable repositories such as the American Type Culture Collection and the Biodefense and Emerging Infections Research Resources Repository. Moreover, the strains extended to include *Bacillus* species, namely *B. thuringiensis*, *B. licheniformis*, and *B. subtilis*, as well as Gram-positive and negative non-*Bacillus* species. The bacterial strains used in the present study and their sources are given in Table 4.

Genomic DNA extraction

Glycerol stocks of bacterial cultures were inoculated on Lysogeny broth (LB) agar and grown overnight at 37°C with aeration. Genomic DNA was isolated using the QIAamp PowerFecal DNA Kit (Qiagen, Germany) according to the manufacturer's protocol. The concentration and purity of extracted genomic DNA were assessed by NanoDrop One^{C} (Thermo Fisher Scientific, USA) spectrophotometer.

Oligonucleotide primers and multiplex PCR conditions

Oligonucleotide primers targeting *B. anthracis*-specific genes encoded on its chromosome and two common marker genes, *pag* and *capA*, on pXO1 and pXO2 plasmids, respectively, were designed using SnapGene 5.0.8. Nucleotide sequences of sets of primers, along with the expected size of the PCR products, are shown in Table 1. A common primer set was utilized for amplifying the *pag* gene across all three multiplex PCRs, while for the *capA* gene, two different primer

Table 1 List of primers used in newly developed multiplex PCR assays and previously developed PCR tests for *B. anthracis* detection

Primer name	Target gene	Primer sequence 5′ – 3′	PCR product (bp)	Reference
MPCR[a]-1 for detecting *B. anthracis* chromosome-encoded gene and plasmids				
TZ_3	*BA1698*	CGAGTTTATTAAAAGAAGTTCCAGA	293	this study
TZ_4 – 1		CGAATATGGACCTGATGTTG		
TZ_1	*pag*	CCAGACCGTGACAATGATGG	612	this study
TZ_2		TGGTAACACGTTGTAGATTGGAG		
TZ_7	*capA*	CACGTAAATGATATTGCACAAGTGC	152	this study
TZ_8		GGCAGATATTATTGTGGGACACC		
MPCR[a]-2 for detecting *B. anthracis* chromosome-encoded gene and plasmids				
TZ_9	*BA5354*	ATGAATGAAGATAAGCATTGTAA	192	this study
TZ_10		CTATTCTTCATCAGCAATTTGAA		
TZ_1	*pag*	CCAGACCGTGACAATGATGG	612	this study
TZ_2		TGGTAACACGTTGTAGATTGGAG		
TZ_7 – 1	*capA*	AGTTGTTGTCTCCACTGATAC	100	this study
TZ_8 – 1		AAATCATAAACATGTTATTGAA		
MPCR[a]-3 for detecting *B. anthracis* chromosome-encoded gene and plasmids				
TZ_11	*BA5361*	CCCCAATATATTAACAATAGAC	392	this study
TZ_12		GTTTTTGAATGTTTTTTGGTTC		
TZ_1	*pag*	CCAGACCGTGACAATGATGG	612	this study
TZ_2		TGGTAACACGTTGTAGATTGGAG		
TZ_7	*capA*	CACGTAAATGATATTGCACAAGTGC	152	this study
TZ_8		GGCAGATATTATTGTGGGACACC		
Duplex PCR assay for detecting pXO1 and pXO2 plasmids, as recommended by WHO[b] guideline				
CAP1234	*cap*	CTGAGCCATTAATCGATATG	846	[23]
CAP1301		TCCCACTTACGTAATCTGAG		
PA5	*pag*	TCCTAACACTAACGAAGTCG	596	[24]
PA8		GAGGTAGAAGGATATACGGT		
A single PCR assay for targeting *B. anthracis* chromosome encoded gene, as recommended by WHO[b] guideline				
S-layer-Upper	*sap*	CGCGTTTCTATGGCATCTCTTCT	639	[25]
S-layer-Lower		TTCTGAAGCTGGCGTTACAAAT		
B. anthracis PCR Detection kit by Takara[c]				
MO11	*cap*	GACGGATTATGGTGCTAAG	591	[26]
MO12		GCACTGGCAACTGGTTTTG		
PA7	*pag*	ATCACCAGAGGCAAGACACCC	211	[24]
PA6		ACCAATATCAAAGAACGACGC		

[a] MPCR, Multiplex PCR

[b] WHO, World Health Organization

[c] Takara, commercial kit

sets were designed, considering the size of co-amplified gene products within each multiplex PCR to ensure distinct band sizes for clear differentiation on agarose gel electrophoresis.

Multiplex PCR assays were performed using a TaKaRa Ex Taq Hot Start Version (TaKaRa, Japan), which consists of PCR buffer, dNTPs, and Ex taq DNA polymerase separately. PCR reactions were carried out using AB Applied Biosystem 2720 thermal cycler (Applied Biosystem, USA) in 50 µl volumes with genomic DNA as template and containing 1 µM of each primer. A gradient PCR with different annealing temperatures was performed to optimize assay conditions.

The conditions of multiplex PCR (MPCR) assays differ depending on the sets of primers. The thermocycling condition of MPCR-1 for amplifying *BA1698*, *pag*, and *capA* consisted of initial denaturation at 98°C for 2 min, followed by 25 cycles of denaturation at 98° for 5 s, annealing at 59°C for 30 s, extension at 72°C for 30 s, and final extension at 72°C for 8 min. The condition of MPCR-2 for amplifying *BA5354*, *pag*, and *capA* included an initial denaturation at 98°C for 2 min, followed by 25 cycles of denaturation at 98° for 5 s,

annealing at 52°C for 30 s, extension at 70°C for 30 s, and final extension at 70°C for 8 min. The reaction of MPCR-3 for amplifying *BA5361*, *pag*, and *capA* goes an initial denaturation at 98°C for 2 min, followed by 25 cycles of denaturation at 98° for 5 s, annealing at 56°C for 30 s, extension at 72°C for 30 s, and final extension at 72°C for 8 min. The amplified PCR products were observed on a 1.5% agarose gel by electrophoresis. PCR products were purified using MinElute PCR Purification Kit (Qiagen, Germany), and Sequencing PCR was performed using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA), followed by Sanger sequencing using a 3130 Genetic Analyzer (Applied Biosystems, USA). In addition, previously developed PCR assays with specific primers (Table 1) for the molecular detection of *B. anthracis* were used for comparison with our newly developed multiplex PCRs to confirm their specificity. The reactions were conducted under specified conditions described in previous studies [23–25] or according to the manufacturer's instructions (Takara Bio Inc., Japan). The commercial PCR kit for *B. anthracis* detection manufactured by Takara Bio Inc. targets *pag* encoded on pXO1 and *cap* encoded on pXO2 plasmids. The expected size of the amplified PCR product is 591 bp for *cap* and 211 bp for *pag*. Additionally, this PCR system consists of control bands (409 bp for *pag*, 98 bp for *cap*) which would be visible after agarose gel electrophoresis.

To determine the detection limit of multiplex PCR assays, serial dilutions (640, 320, 160, 80, 40, 20, and 10 pg/µl) of genomic DNA of *B. anthracis* CZC5 were examined with MPCRs or the previous PCR methods and analyzed on agarose gel electrophoresis. Sterile water served as a negative control.

Furthermore, multiplex PCRs were performed directly using a colony of *B. anthracis* Sterne 34F2 strain grown on LB agar to evaluate the practical usage of the assays.

Results

Pan-genome analysis and identification of *B. anthracis*-specific genes

The pan-genome status, whether open or closed, reflects the likelihood of discovering a new gene or gene family when a new genome sequence is added to the analysis. An open pan-genome is likely to grow with new gene discoveries, whereas a closed pan-genome is unlikely to add more. The status of the pan-genome is modeled by Heap's law estimation, of which exponent parameter γ value between 0 and 1 reflects the openness of a given pan-genome [20]. The pan-genome of *B. anthracis*, which includes genomes of 115 diverse strains, yielded a γ value of 0.09 (Fig. 1A), indicating a more closed pan-genome. In contrast, the pan-genomes of *B. cereus* ($n=142$) and *B. thuringiensis* ($n=93$) showed more open characteristics, with γ values of 0.51 and 0.52, respectively (Fig. 1B, C). Additionally, unlike *B. cereus* and *B. thuringiensis*, where accessory genes constitute the majority, the pan-genome of *B. anthracis* is primarily composed of core genes.

Furthermore, to investigate the *B. anthracis*-specific genes, we choose 50 strains from each species, *B. anthracis*, *B. cereus*, and *B. thuringiensis.* Additionally, one complete genome of *B. weihenstephanensis* was incorporated in the combined dataset as an outgroup control. The pan-genome of the 151 strains consisted of 66,052 genes, of which 1,490 were core genes present in at least 99% of the sampled genomes, and 686 were soft-core genes present in at least 95% of all genomes analyzed. The non-core genomes were further divided into 6,221 shell genes, present in 15-95% of the genomes, and 57,655 accessory genes, present in less than 15% of the genomes.

Further, based on the strains clustering on the phylogenetic tree, we defined three groups of *B. cereus s.l* group: Bce1, Bce2, and Ban (Fig. 2 and Additional file 4). The Bce1 group consists of *B. thuringiensis* and *B. cereus* strains. Bce2 group also comprised *B. cereus* and *B. thuringiensis* strains, however, this group included unusual *B. cereus* strains such as G9241 and *B. thuringiensis* 97 − 27, which were previously known to cause anthrax-like diseases in humans and animals due to acquisition of plasmids highly similar to *B. anthracis* pXO1 and pXO2 [11]. All *B. anthracis* strains were categorized into one group named Ban.

A heatmap displays the core and accessory genes. The core-genome phylogeny is categorized into three distinct groups: the Ban group, which comprises *B. anthracis strains*; the Bce1 group, including common

(See figure on next page.)

Fig. 1 Pan-genome analysis of *B. anthracis* and its closely related species, *B. cereus* and *B. thuringiensis*, with Heap's law estimation. **A** The pan-genome of 115 *B. anthracis* strains and the corresponding Heap's law estimation. **B** The pan-genome of 142 *B. cereus* strains, along with its Heap's law estimation. **C** The pan-genome of 93 *B. thuringiensis* strains with Heap's law estimation. Each Heap's law estimation was conducted with 1000 permutations. The blue dots in the Heaps law graphic represent the observed number of unique genes as genomes are incrementally added. The red line represents the fitted Heap's law curve based on the observed data. The x-axis represents the number of genomes. The y-axis represents the count of unique genes. k is intercept, representing the initial number of unique genes when only a few genomes are considered. γ is an exponent parameter that characterizes the rate at which new genes are added as more genomes are included in the analysis.If γ is closer to 1, it indicates an open pangenome, meaning the number of unique genes continues to increase significantly with additional genomes. Conversely, if γ is closer to 0, it suggests a closed pangenome, where the most genetic diversity is captured with the current dataset

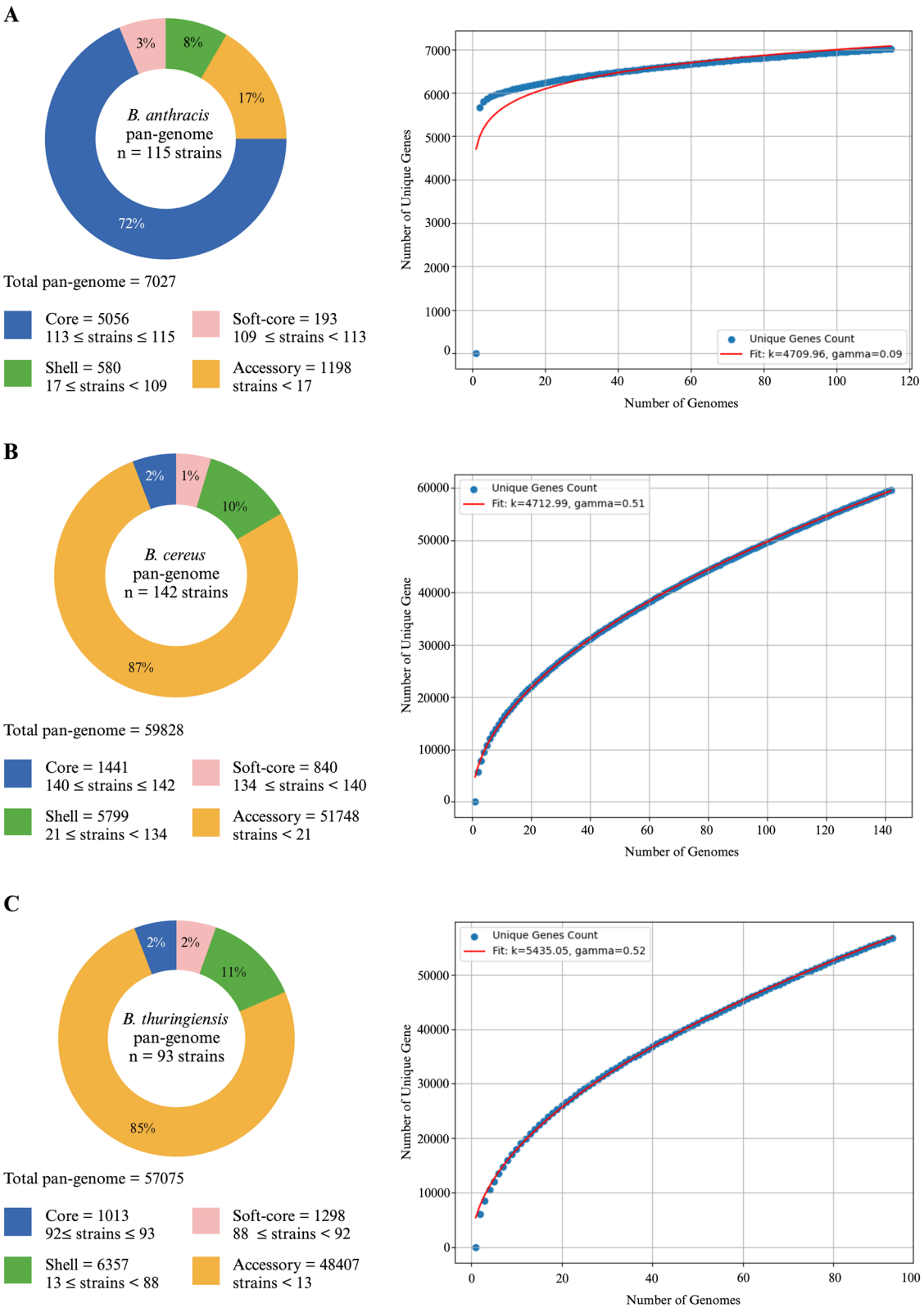

Fig. 1 (See legend on previous page.)

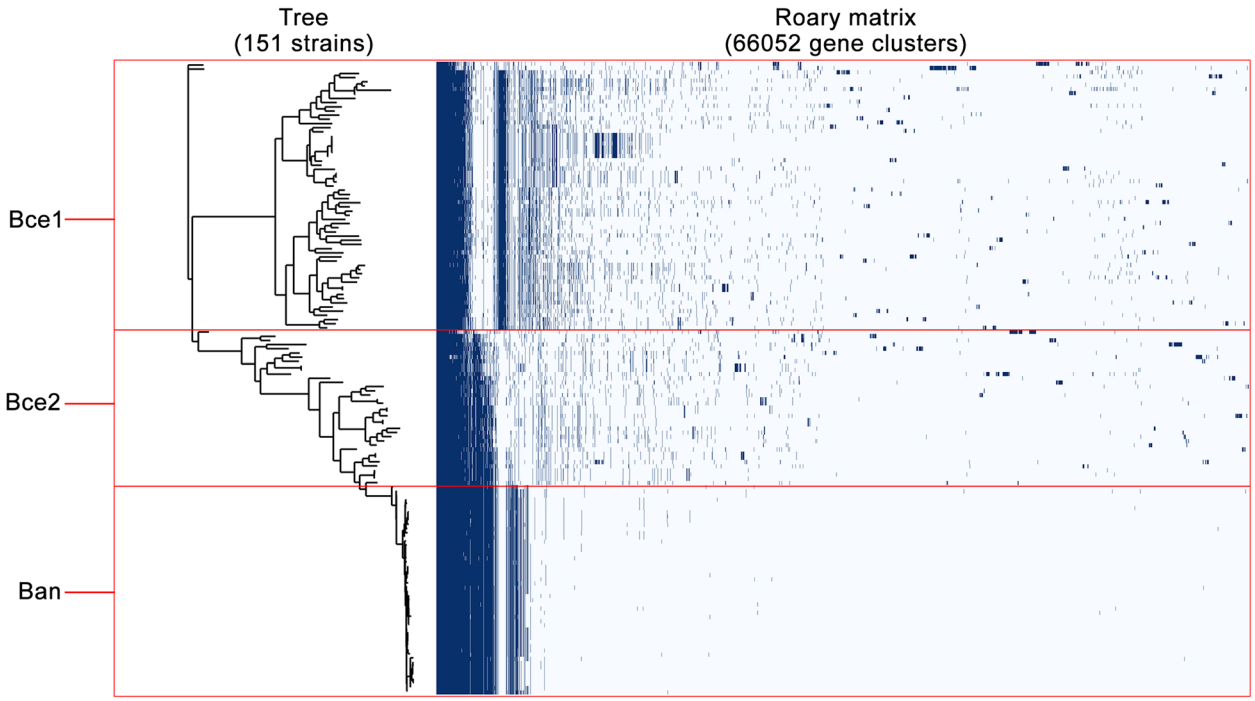

Fig. 2 The pan-genome analysis illustrates the clustered presence or absence of genes and the distribution of accessory genes among the 151 *Bacillus* strains

Table 2 Number of group-specific genes

Ban[a]	Bce1[a]	Bce2[a]	Number of genes[b]
+	+	+	1060
+	+	-	0
+	-	+	17
+	-	-	136
-	+	+	0

[a] If the group is included in the calculation of group-specific genes; +, If not -,

[b] The number of genes found in groups corresponding to conditions

B. cereus and *B. thuringiensis* strains; and the Bce2 group, consisting of *B. cereus* and *B. thuringiensis* strains closely related to *B. anthracis*.

A Perl script was employed to extract the genes that are exclusively present in *B. anthracis* strains and absent in the other two groups of genomes. This initial analysis identified a total of 136 genes that are potentially exclusive to the genomes of *B. anthracis* strains (Table 2).

Among these, 88 genes corresponded to previously identified four prophage regions [27], and the remaining 48 genes were located outside these regions (Additional file 5). Notably, we observed a gene region spanning coordinates between 1,596,297 and 1,605,500 that was previously undefined (Fig. 3). This region contains genes *BA1693-BA1699*, which encode hypothetical proteins belonging to the glycosyltransferase family and *BA1701*, coding for a hypothetical protein in the 5'-monophosphate dehydrogenase family. Additionally, two transposases, *BA1703* and *BA1704*, were found within this region (Fig. 4).

To verify the specificity of the 136 genes identified from the pan-genome comparative analysis, each gene was used as a query and subjected to a BLASTn search with an expected threshold of 0.05 (E-value = 0.05) against the NCBI database. Genes that matched only with *B. anthracis* or its phage and not with other bacterial species were considered specific or partially specific to *B. anthracis* if the query coverage of the nucleotide sequence alignment was less than 50% (Table 3).

Local BLAST alignment further confirmed that 127 out of 132, chromosomally complete genomes of *B. anthracis* strains available in the GenBank database, possess all the identified genes listed in Table 3. Five *B. anthracis* strains, MCCC 1A02161, MCCC 1A01412, HDZK-BYSB7, CMF9, and Mn106-1 head 2chi found lacking any of those genes. These five strains have been reported to be misidentified as *B. anthracis*, due to their lack of several conserved *B. anthracis*-specific SNPs [32].

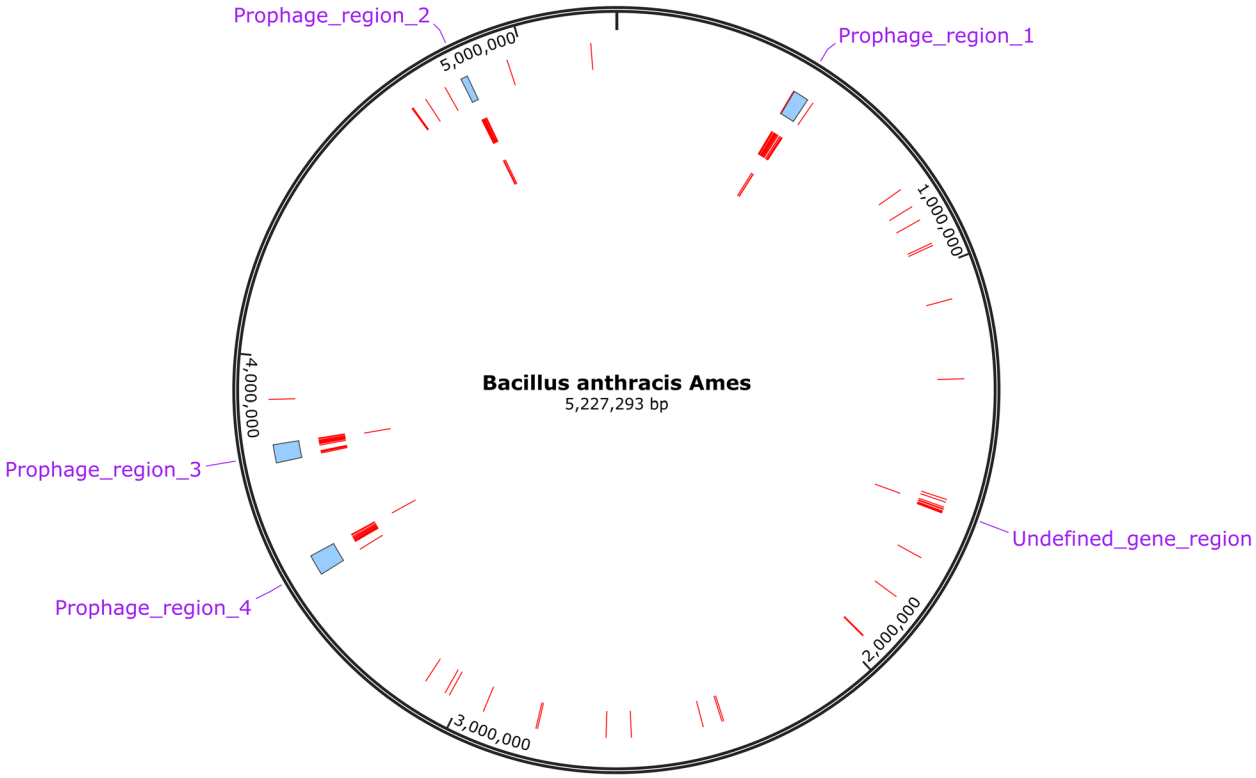

Fig. 3 The localization of genes identified from pan-genome analysis on the chromosome of *B. anthracis* Ames. Red lines indicate the genes identified in this study, while blue boxes represent the lambda prophage regions described in a previous study [27]. An additional new region was identified between the coordinates 1,596,297 and 1,605,500

Development of Multiplex PCR using the identified *B. anthracis*-specific genes

We developed three distinct multiplex PCR assays (MPCR-1–3), each targeting a specific chromosome-encoded gene and two known markers, *pag*, and *capA*, encoded on the pXO1 and pXO2, respectively, plasmids of *B. anthracis*. Due to the predominance of prophage genes and genes coding glycosyltransferases, we proportionally selected *BA1698*, *BA5354*, and *BA5361* as representatives of the chromosome-encoded genes, which were identified as specific to *B. anthracis* in the present study. A total of 62 bacterial strains, listed in Table 4, were utilized to evaluate the efficacy and specificity of multiplex PCR.

The efficacy of multiplex PCRs was examined using virulent *B. anthracis* strains previously isolated from various sources, including humans, livestock, and wildlife in endemic regions of Zambia and Mongolia. Additionally, a non-virulent *B. anthracis* Sterne 34F2 strain was included in the examination. All *B. anthracis* strains generated distinct and bright PCR products corresponding to the *BA1698*, *BA5354*, and *BA5361* genes found on the chromosome, as well as the *pag* and *capA* genes encoded on the pXO1 and pXO2 plasmids, respectively, in Figs. 5, 6 and 7. Sterile distilled water instead of template DNA in the reaction mixture served as a negative control, which resulted in no PCR product.

Moreover, the specificity of multiplex PCRs was tested using 35 other *Bacillus* spp. and 10 Gram-positive and negative non-*Bacillus* species. None of these bacterial species generated a PCR product for the chromosome-encoded genes that were identified as specific to *B. anthracis*. Notably, an atypical *B. cereus*, specifically the G9241 strain (Sample ID 38), showed a PCR product for the *pag* gene of *B. anthracis*, which is encoded on its pXO1 plasmid (Figs. 5, 6 and 7). This strain of *B. cereus* was found to harbor a plasmid that resembles the pXO1 of *B. anthracis*, carrying genes for major toxins responsible for anthrax, namely PA, LF, and EF [11].

The sensitivity of multiplex PCRs was determined using genomic DNA from *B. anthracis* CZC5 strain. MPCR-1 and 3 yielded detectable amplicons on agarose gel at a minimum DNA concentration of 80 pg/µl, while MPCR-2 exhibited a lower limit of 160 pg/µl for (Fig. 8).

Additionally, colony PCR performed directly using *B. anthracis* Sterne 34F2 colony without the need for DNA extraction, effectively yielded the desired amplicons for chromosome-encoded genes and the *pag* gene encoded on pXO1 plasmid (Additional file 6).

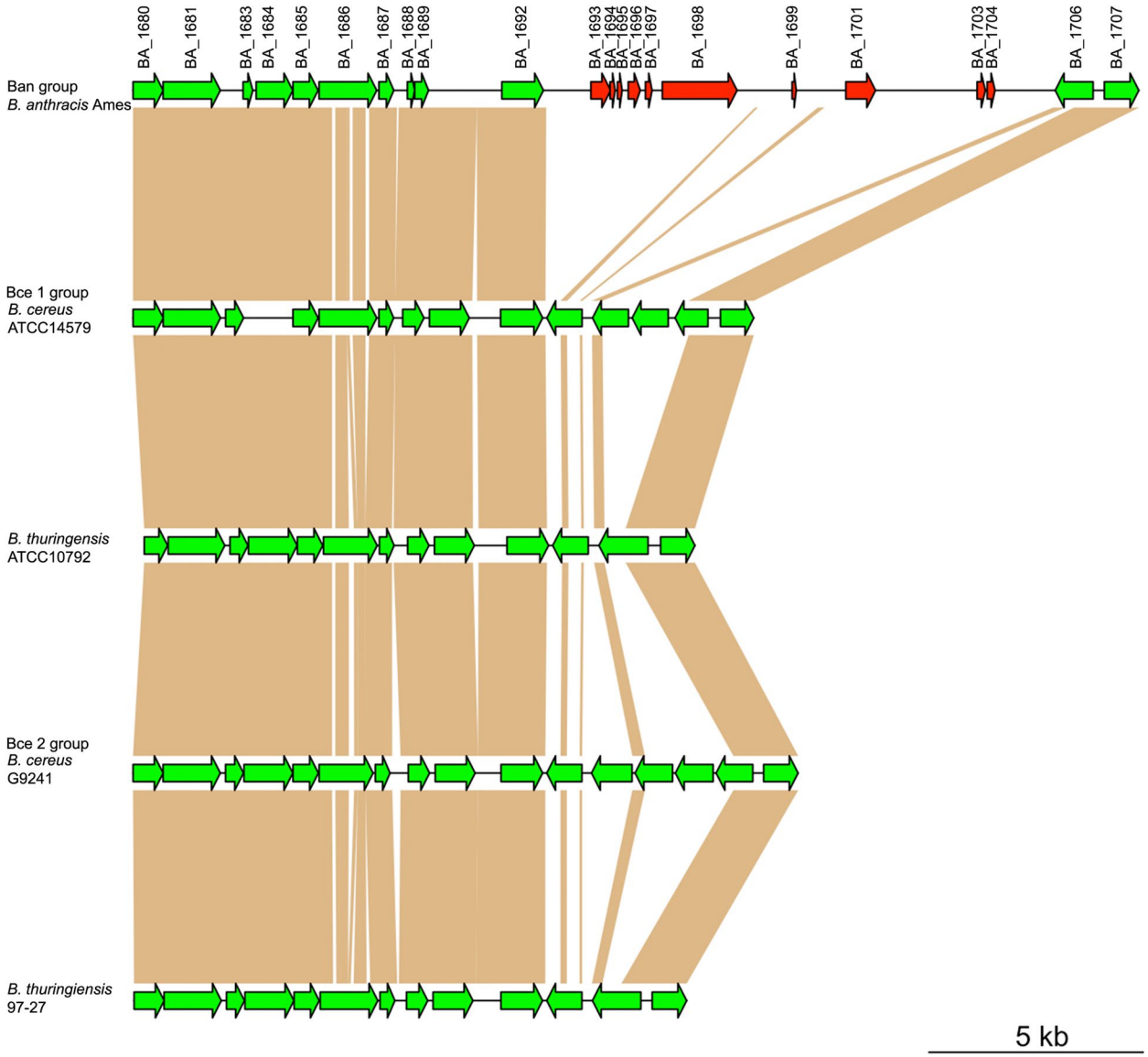

Fig. 4 Undefined specific gene region in the *B. anthracis* chromosome. *B. anthracis* Ames represents the Ban group; *B. cereus* ATCC14579 and *B. thuringiensis* ATCC10792 represent the Bce1 group; while *B. cereus* G9241 and*B. thuringiensis* 97-27 represent the Bce2 group, as categorized in this study. Green arrows indicate genes that are common among the bacterial species and red arrows indicate genes exclusive to *B. anthracis*

Further, to verify the specificity and overall ability of the newly developed multiplex PCR assays to accurately detect the presence of *B. anthracis*, the panel of *Bacillus* species, as well as other common environmental bacterial species listed in Table 4 were also tested by previous PCR assays recommended by the World Health Organization [23–25] and commercially available PCR kit (Takara Bio Inc., Japan) for *B. anthracis* detection. Consistent with our developed assays, each of these PCR methods detected *pag* and *cap* genes encoded on pXO1 and pXO2 plasmids in *B. anthracis* strains (Sample ID 1–10) and detected *pag* in atypical *B. cereus* G9241 (Sample ID 38) (Additional file 7, 8, and 9). Although the sensitivity of these methods was relatively high, capable of detecting DNA concentration 10 pg/ml or less (Additional file 10), PCR targets *B. anthracis* chromosome encoded *sap* gene was cross-reactive with *B. cereus* strains (Sample ID 33 and 39) and produced non-specific band (Sample ID 36) (Additional file 8). Despite the internal controls, the commercial kit also produced non-specific amplifications with non-target organisms (Additional file 9), which may indicate cross-reactivity with closely related bacterial species or the presence of similar sequences in non-target organisms.

Table 3 Genes identified as specific or partially specific to *B. anthracis* through a BLASTn search against the NCBI database

Gene regions*,#	Locus tag[a, b,1–4]	Coding product
Undefined gene region*	*BA1693*[b]	Glycosyltransferase, group 2 family
	BA1694[a]	Hypothetical protein, glycosyltransferase family protein
	BA1695[a]	Hypothetical protein, glycosyltransferase family protein
	BA1696[a]	Hypothetical protein, glycosyltransferase family protein
	BA1697[a]	Hypothetical protein, glycosyltransferase family protein
	BA1698[a,1]	TPR/glycosyltransferase domain protein
	BA1699[a]	Hypothetical protein
	BA1701[a]	Hypothetical protein, inosine 5′-monophosphate dehydrogenase family protein
	BA1703[b]	Transposase
	BA1704[b]	Transposase
Prophage region 1#	*BA0444*[a]	Hypothetical protein
	BA0454[a]	Hypothetical protein
	GBAA_RS02495[a]	Hypothetical protein
	BA0455[b]	DUF3983 domain-containing protein
	BA0463[a]	Hypothetical protein
	BA0479[a,2]	Prophage tail endopeptidase domain-containing protein
	BA0480[a]	Hypothetical protein
Prophage region 2#	*BA5340*[a]	Conserved domain protein
	BA5344[a]	HNH Endonuclease
	BA5345[a,3]	Glycosyl-4,4′-diaponeurosporenoate acyltransferase
	BA5346[b]	DUF2513 domain-containing protein
	BA5347[a]	Conserved domain protein
	BA5351[b]	Hypothetical protein
	BA5354[a]	Hypothetical protein
	BA5355[a]	Terminase
	BA5357[a,4]	Conserved domain protein
	BA5361[a]	Phage major capsid protein
	BA5362[a]	Conserved domain protein
Prophage region 3#	*BA4071*[a]	Hypothetical protein
	BA4098[b]	Conserved domain protein

[a] Genes without matches in non-*Bacillus* spp, and other bacteria

[b] Genes partially specific to *B. anthracis*, if the query coverage of the nucleotide sequence alignment was less than 50%

[*] Previously undefined gene region described in this study

[#] Lambda prophage region previously described in reference [27]

[1] Chromosomal markers have been described in reference [28]

[2] Chromosomal markers have been described in reference [29]

[3] Chromosomal markers have been described in reference [30]

[4] Chromosomal markers have been described in reference [31]

Discussion

B. anthracis is highly pathogenic and can cause fatal infections in humans and animals. Due to its potential threat to national security, public health, and socioeconomic stability driven by a consequential loss of livestock, it is classified as a high-priority biological agent. *B. anthracis* has a significant overlap in genetic content with *B. cereus* and *B. thuringiensis*, making it challenging to distinguish from the other species. These *Bacillus* species are widespread and naturally occurring in nature. Identifying the genetic differences specific to *B. anthracis* is crucial for understanding its pathogenicity and designing accurate DNA-based detection methods.

We investigated chromosome-encoded genes specific to *B. anthracis*, employing whole genome comparison analyses that incorporated pan-genome of the three most

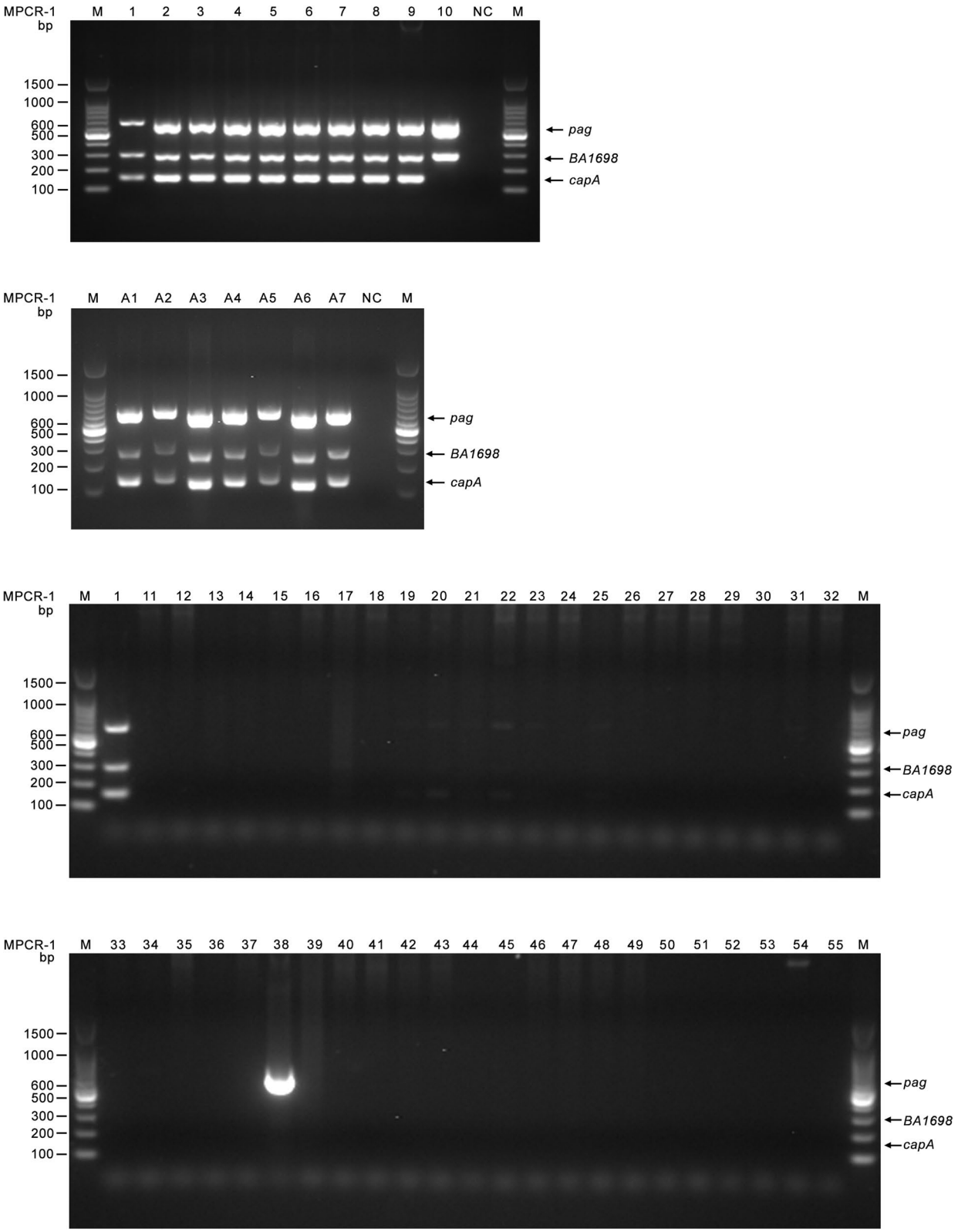

Fig. 5 MPCR-1 targets *BA1698*,*pag,* and *capA*. M is a 100 bp DNA ladder used as a marker. Lanes 1–10, and A1-A7 are amplicon results of *B. anthracis* strains (Strain ID in Table 4, 1–10, A1-A7), lane NC no-template negative control. Lanes 11–55 are amplicon results of non-*B. anthracis* strains (Strain ID in Table 4, 11–55). We used 1.5% agarose gel

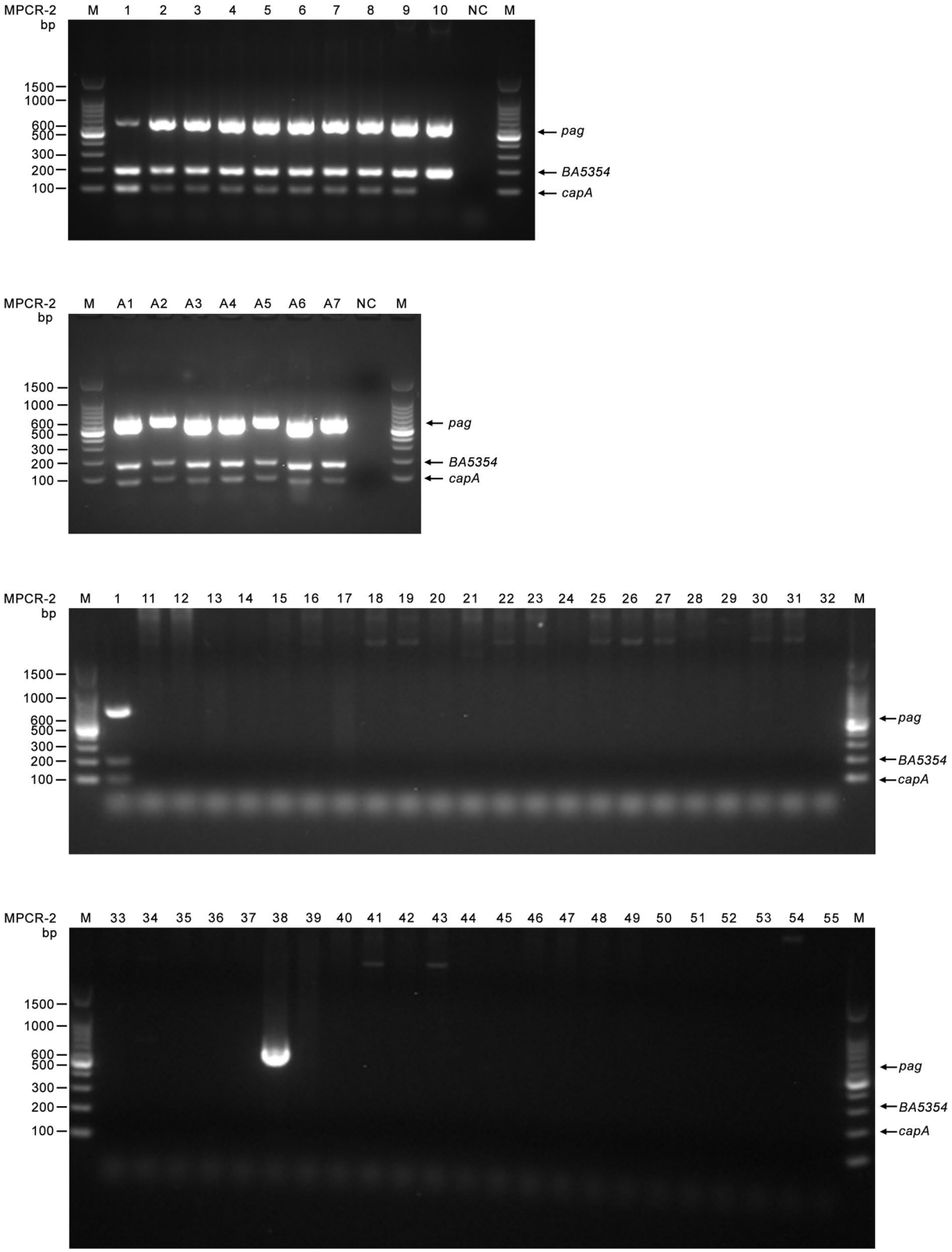

Fig. 6 MPCR-2 targets *BA5354*,*pag,* and *capA*. M is a 100 bp DNA ladder used as a marker. Lanes 1–10, and A1-A7 are amplicon results of *B. anthracis* strains (Strain ID in Table 4, 1–10, A1-A7), lane NC no-template negative control. Lanes 11–55 are amplicon results of non-*B. anthracis* strains (Strain ID in Table 4, 11–55). We used 1.5% agarose gel

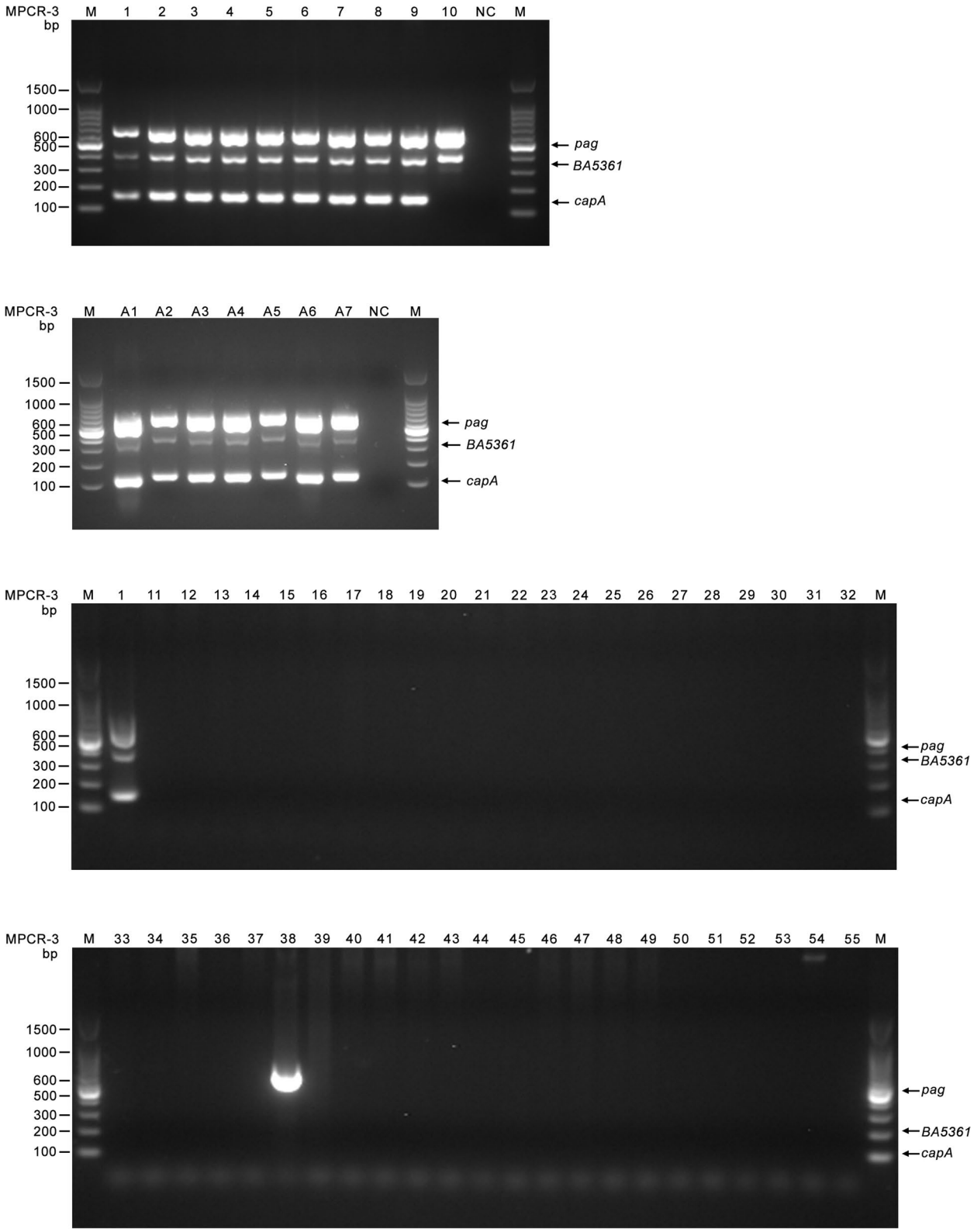

Fig. 7 MPCR-3 targets *BA5361*,*pag,* and *capA*. M is a 100 bp DNA ladder used as a marker. Lanes 1–10, and A1-A7 are amplicon results of *B. anthracis* strains (Strain ID in Table 4, 1–10, A1-A7), lane NC no-template negative control. Lanes 11–55 are amplicon results of non-*B. anthracis* strains (Strain ID in Table 4, 11–55). We used 1.5% agarose gel

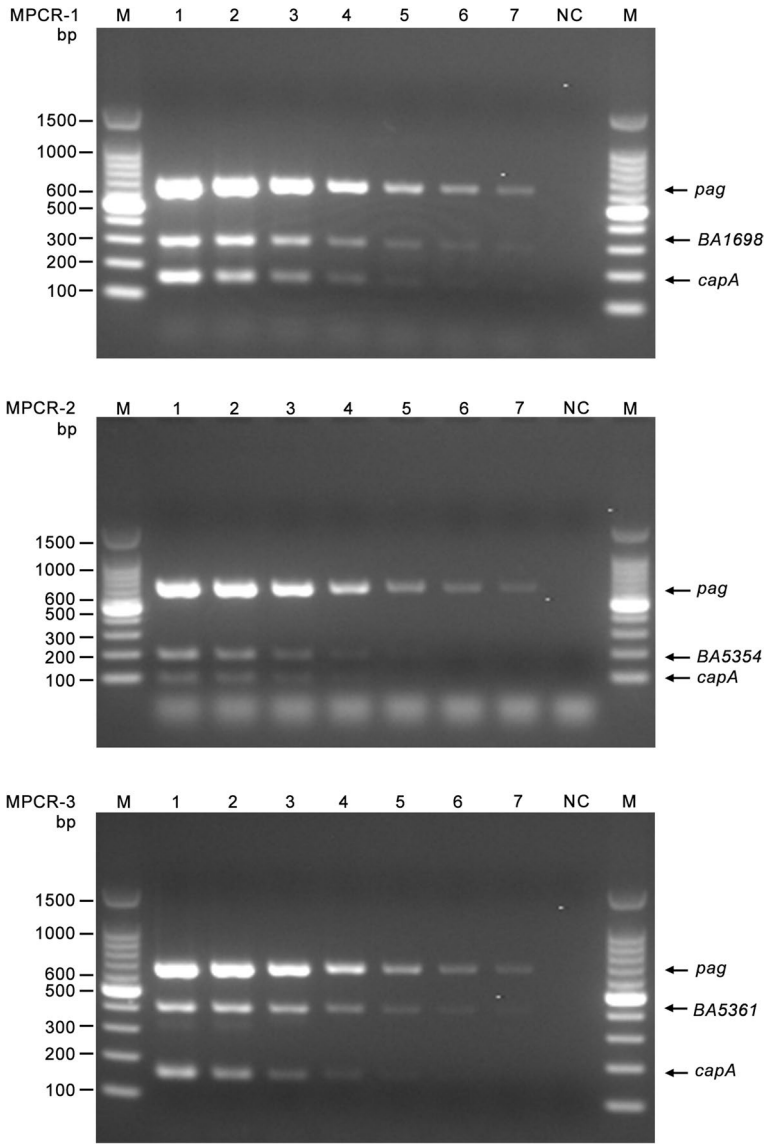

Fig. 8 Sensitivity of multiplex PCRs analyzed by agarose gel electrophoresis. M is a 100 bp DNA ladder used as a marker. Lanes 1–7 are amplicon results of *B. anthracis* CZC5 strain (640, 320, 160, 80, 40, 20, 10 pg/µl), lane NC no-template negative control

genetically related species in the *B. cereus s.l* group: *B. cereus*, *B. anthracis*, and *B. thuringiensis*. Here, we demonstrated chromosome-encoded genes specific to *B. anthracis* and developed a PCR method utilizing some of these genes for the detection of *B. anthracis*.

Over the years, the increasing number of sequenced bacterial genomes has made it possible to capture genomic variation both among diverse strains within the same species and between different species. In this study, we analyzed the complete genomes of 115 diverse *B. anthracis* strains to further characterize their pan-genome. The core genes predominate in the pan-genome of *B. anthracis*, and they are well-conserved among strains. This high conservation of core genes and low genetic variability may reflect a stable evolutionary history and a lack of significant changes or adaptations in this species. Heap's law estimation supports this conclusion, showing that the pan-genome of *B. anthracis* is closed; most genes have already been captured, and the discovery rate of new genes approaches zero (Fig. 1A). This aligns with the earlier work of Tettelin et al., [34], who analyzed the genomes of eight *B. anthracis* strains and similarly concluded that its pan-genome is closed, suggesting a highly clonal species with low genomic variability. This consistent genetic variability in *B. anthracis* could be advantageous for diagnosing this species, as it

Table 4 Bacterial strains used in this study and their possessions of *B. anthracis* chromosome-encoded genes and two other plasmid markers

Bacterial species	Sample ID	Strains	Sources (References[a, b])	Target genes for PCR amplification				
				BA1698[c]	***BA5354***[c]	***BA5361***[c]	***pag***[c]	***capA***[c]
Bacillus anthracis	1	CZC5	Hippo[a]	+	+	+	+	+
	2	shH20	Human	+	+	+	+	+
	3	LZH18	Hippo	+	+	+	+	+
	4	SLI21	Impala	+	+	+	+	+
	5	SLH21	Hippo	+	+	+	+	+
	6	LZH20b	Hippo	+	+	+	+	+
	7	WPC18	Cattle	+	+	+	+	+
	8	WPC20	Cattle	+	+	+	+	+
	9	LZH20a	Hippo	+	+	+	+	+
	10	Sterne 34F2	Commercial vaccine strain	+	+	+	+	
	A1	Selenge01	Cattle	+	+	+	+	+
	A2	Khuvsgul07	Goat	+	+	+	+	+
	A3	Uvurkhangai08	Cattle	+	+	+	+	+
	A4	Ulaanbaatar09	Cattle	+	+	+	+	+
	A5	Khentii10	Cattle	+	+	+	+	+
	A6	Khentii13	Cattle	+	+	+	+	+
	A7	Khuvsgul15	Horse	+	+	+	+	+
Bacillus cereus	11	MRY14-0045	Patient blood[b]					
	12	MRY14-0057	Patient blood[b]					
	13	MRY14-0060	Patient blood[b]					
	14	MRY14-0074	Hospital towel[b]					
	15	MRY14-0075	Hospital towel[b]					
	16	MRY14-0079	Hospital towel[b]					
	17	MRY14-0100	Hospital towel[b]					
	18	MRY14-0105	Hospital microwave oven[b]					
	19	J1	Infusion[b]					
	20	J2	Infusion[b]					
	21	J7	Patient skin[b]					
	22	J10	Hospital sheet[b]					
	23	J39	Patient blood[b]					
	24	J51	Patient blood[b]					
	25	J62	Hospital linen room washed towel[b]					
	26	J75	Hospital linen room washed towel[b]					
	27	30040	Patient blood[b]					
	28	30043	Patient blood[b]					
	29	30048	Patient blood[b]					
	30	30052	Patient blood[b]					
	31	30075	Patient blood[b]					
	32	30077	Patient blood[b]					
	33	30090	Patient blood[b]					
	34	30102	Patient blood[b]					
	35	14579	ATCC					
	36	10987	ATCC					
	37	Yonago40	Patient blood[b]					
	38	G9241	BEI Resources				+	
	39	E33L	BEI Resources					

Table 4 (continued)

Bacterial species	Sample ID	Strains	Sources (References[a, b])	Target genes for PCR amplification				
				BA1698[c]	*BA5354*[c]	*BA5361*[c]	*pag*[c]	*capA*[c]
Bacillus thuringiensis	40	4A10	BGSC					
	41	4A12	BGSC					
	42	4D1	BGSC					
	43	4D24	BGSC					
Bacillus licheniformis	44	10716	ATCC					
Bacillus subtilis	45	168	BGSC					
Staphylococcus aureus	46	RN4220	BEI Resources					
	47	USA300-0114	BEI Resources					
	48	71080	BEI Resources					
	49	F003B2N-C	BEI Resources					
Enterococcus faecalis	50	S613	BEI Resources					
	51	V587	BEI Resources					
	52	UAA714	BEI Resources					
	53	UAA945	BEI Resources					
Pseudomonas aeruginosa	54	PA14	BEI Resources					
Escherichia coli	55	MG1655	NEB					

ATCC American Type Culture Collection, *BEI Resources* The Biodefense and Emerging Infections Research Resources Repository, *BGSC* The Bacillus Genetic Stock Center, *NEB* New England BioLabs

[a] Strains isolated in reference [33]

[b] Strains isolated in reference [22]

[c] If target genes were amplified; +, if not empty

reduces the likelihood of significant genetic variations affecting diagnostic markers. On the other hand, the pan-genome of *B. cereus* and *B. thuringiensis* present a high percentage of accessory genes and a much smaller proportion of core genes. Their pan-genome statuses are open (Fig. 1B, C), implying that substantial genetic diversity continues to expand as more strains are sequenced, increasing the number of unique genes. This result aligns with previous study [35] that reported the openness of *B. cereus s. l.* group pan-genome. Identifying strain-specific genes among diverse *B. cereus* and *B. thuringiensis* strains could be valuable for predicting pathogenicity, enhancing diagnostic accuracy, and treatment strategies.

Studies have been searching for genomic differences between *B. anthracis* and its closest relative in the *B. cereus* group using different approaches. Read et al., reported that four lambda prophage regions are absent in the chromosome of *B. anthracis* Ames compared to its 19 close neighbors via comparative genomic hybridization [27]. These regions contain various specific and non-specific genes, and further studies determined species-specific genetic markers for *B. anthracis* by designing multiple primers targeting prophage regions [29], suggesting *BA0479*, *BA5356*, *BA4094*, and *BA3805* as possible gene markers for identifying *B. anthracis*. Radnedge et al., revealed *BA5345*, located in prophage region 2, through amplified fragment length polymorphism (AFLP), and suppression subtractive hybridization (SSH) methods [30]. In addition, *BA5357* [31] and *BA5358* [36] also encoded in prophage region 2, and *BA1698* [28] were previously described as specific to *B. anthracis* by comparative genome analysis. In agreement with these former studies, our pan-genome analysis also recognized all those previously reported specific genes and further expanded the number of *B. anthracis*-specific genes in the prophage regions and beyond (Additional file 5 and Table 3). This underscores the robustness and reliability of our method in *B. anthracis*-specific gene identification. We identified 136 genes that are present in the genomes of *B. anthracis* strains and absent in the *B. cereus* and *B. thuringiensis* strains included in our dataset for pan-genome comparison. Eighty eight out of 136 genes locate in prophage regions and 48 genes were found from other locations in the *B. anthracis* chromosome. This number was further reduced to 30 after global BLASTn in NCBI, because to ensure the gene specificity, we removed the genes that matched any organism other than *B. anthracis* or its phage. We found 20 genes in prophage regions, including those formerly reported genes. As prophage regions contain both specific and non-specific genes that can complicate the development of diagnostic tools, our study addresses this challenge by

precisely indicating each specific gene with its unique locus tag. Also, similar to the previous studies [29], we observed that each of the four lambda prophage regions incorporates genes encoding site-specific recombinases (*BA5363*, *BA4075*, *BA3832*, and *BA0427*), which catalyze the recombination event therefore might facilitate the phage integration into the genome of *B. anthracis* (Additional file 5).

Further, our approach has enabled us to discover an additional unique gene region range from 1,596,297–1,605,500, which include *BA1693-BA1699*, encoding hypothetical proteins belonging to the glycosyltransferase family, *BA1701*, encoding a putative inosine 5′-monophosphate dehydrogenase, and two transposases, *BA1703* and *BA1704* (Fig. 4). To elucidate the potential role of proteins encoded by these genes in *B. anthracis*, we employed STRING v.12.0 [21], which predicts potential physical and functional protein-protein interactions. A significant interaction was found only for glycosyltransferase encoded by *BA1698* and it was predicted to interact with enzymes involved in various biological processes, namely, the 4-alpha-glucan branching enzyme, which is key in glycogen formation; dTDP-4-dehydrorhamnose 3,5-epimerase, dTDP-glucose 4,6-dehydratase, and glucose-1-phosphate thymidylyltransferase, crucial for the biosynthesis of rhamnose, an essential component of the bacterial cell wall, and UDP-glucose 6-dehydrogenase, integral to the formation of the antiphagocytic capsule formation, respectively (Additional file 11). In fact, previous research has shown that the glycosyl residue composition of cell walls vary between different clades of *B. cereus* strains and is distinct from that of *B. anthracis* [37]. However, *B. cereus* strains (G9241, 03BB102, 03BB87), which are phylogenetically closest to *B. anthracis*, exhibit glycosyl compositions that closely resemble the cell walls of *B. anthracis* strains [38]. Considering this, the hypothetical proteins belonging to glycosyltransferases found only in *B. anthracis* may be associated with distinct cell-surface characteristics and virulence of the bacterium. Furthermore, the predicted interactions with those of enzymes and proteins not only underline the role of the glycosyltransferase encoded by *BA1698* in synthesizing and modifying essential polysaccharides for cell wall and capsule formation but also suggest a broader involvement in the organism's energy accumulation and stress response via glycogen metabolism. Glycogen, as a primary storage form of glucose, its metabolism, and accumulation might be critical for *B. anthracis*, enabling it to remain viable during the dormant spore state, which can last for decades. Like other dormant bacterium [39], *B. anthracis* could strategically conserve energy to transition back to an active, vegetative state from dormancy. Thus, understanding the role of the glycosyltransferases found to be unique to *B. anthracis* may have important implications in improving our knowledge regarding this bacterium's survival. Further experimental studies, such as gene knockout, enzyme activity, sporulation, and virulence tests in animal models, would be necessary to conclusively determine the biological role and importance of the specific glycosyltransferases in *B. anthracis*. Such studies could also reveal potential targets for drugs or antigens for vaccines against anthrax.

The main criteria for designing a DNA-based method to detect *B. anthracis* is to prevent false-positive results from its closely related species and ensure the presence of specific genes or sequences in all *B. anthracis* isolates to avoid false-negative outcomes. We confirmed the presence of all genes identified from our analysis among the complete genomes of 132 *B. anthracis* strains through local BLAST alignment. However, five strains, MCCC 1A02161, MCCC 1A01412, HDZK-BYSB7, CMF9, and Mn106-1 head 2chi lacked any of the genes we identified (Additional file 3). This agrees with Lyu et al., who indicated that strains MCCC 1A02161 and MCCC 1A01412 were misidentified as *B. anthracis* in the NCBI genome database, due to the absence of several conserved *B. anthracis*-specific SNPs in their genome [32]. Accordingly, we verified the absence of thymine at nucleotide position 640 of the *plcR* gene, a chromosomal signature for distinguishing *B. anthracis* from its closely related neighbors, thus confirming that these strains are not *B. anthracis*. Altogether, these results indicate that the genes identified in prophage regions and newly identified gene region are highly conserved among globally diverse *B. anthracis* strains.

Further, we used three of the identified genes to establish three multiplex PCR assays for detecting *B. anthracis*. Each multiplex PCR targets one chromosome-encoded gene and two plasmid-encoded markers. Despite the presence of plasmid-encoded markers, *pag*, and *capA* in unusual *B. cereus* and *B. thuringiensis* strains, we included these markers in the multiplex PCRs as they are clinically relevant targets for anthrax diagnosis. Moreover, to account for the potential emergence of atypical *Bacillus* species that possess one or more chromosome-encoded genes our assays targeted, we developed three distinct PCR assays. These assays enable cross-validation and ensure reliable detection and differentiation of *B. anthracis* from atypical strains when necessary.

Due to the high pathogenicity of *B. anthracis* and strict international regulations governing its use and transportation, obtaining a diverse set of *B. anthracis* strains was challenging. Nevertheless, we obtained 16 virulent strains from two different anthrax endemic regions, Africa, and Asia through our collaborative partners. From Zambia,

we received genomic DNA of nine virulent strains previously isolated from wildlife, livestock, and humans infected with anthrax, and seven from Mongolia. These strains were carefully selected to reflect the temporal and geographic diversity from each country, as they were isolated from various locations at different times. Additionally, one vaccine strain was examined to assess the efficacy of the multiplex PCR assays.

All *B. anthracis* strains were accurately detected by the multiplex PCRs. This supports the effectiveness of our assay across a wider range of genetic variants but also demonstrates its potential applicability in diverse geographical settings.

Furthermore, we tested the specificity of these newly developed PCR assays against 45 bacterial stains, including *B. anthracis* closely related *Bacillus* species, and common environmental bacteria (Table 4). This collection included *B. cereus* G9241, and E33L, which were previously defined to be very closely related to *B. anthracis* [40]. Moreover, the phylogenetic tree (Additional file 4) constructed based on the core-genome alignment of strains of the three *Bacillus* species, showed that *B. cereus* strains that have previously been associated with nosocomial infection (e.g., MRY14-0074, J2, 30052) exhibit even higher genetic resemblance to *B. anthracis* [41]. Despite their close genetic similarities, none of these strains yielded false-positive results in our PCR assays, underscoring the specificity of our methods. Based on the comparative ratio between the number of true negatives ($n=45$) and the sum of true negatives ($n=45$) and false positives ($n=0$), the specificity of the multiplex PCRs was determined to be 99.9%. All true positive samples ($n=17$) were detected within the detection limit of DNA concentration (Fig. 8).

Our experimental results, *in silico* analysis, and previous studies have indicated that *B. anthracis* strains are genetically highly monomorphic [42, 43], which could reduce the likelihood that our assay would fail to detect genetically divergent strains. However, the current study has a limitation in sample size; therefore, further validation is required to assess reproducibility and consistency by incorporating more *B. anthracis* strains isolated from different geographical regions and clinical backgrounds to ensure the robustness of these assays. So far, multiplex PCRs developed in this study offer alternative specific and practical tools for early diagnosis of anthrax. Particularly, we established a standard multiplex PCR assay, recognizing the urgent need for anthrax diagnosis in developing countries such as Zambia and Mongolia, where resources are limited, and anthrax prevalence is high. Our assay provides a cost-effective, accurate, and easy-to-implement solution, suitable for settings with limited laboratory infrastructure. It only requires standard PCR equipment, which is widely available in these regions, thereby enhancing its diagnostic utility. In the future, our current assays would be developed to include isothermal amplification techniques to further facilitate the diagnostic process.

Conclusion

The chromosome-encoded specific genes of *B. anthracis* identified in this study offer an advantage over previously defined marker genes, which often result in false-positive results. Our pan-genome analysis successfully captured all known specific genes in previously described lambda prophage regions and further expanded the number of specific genes within these regions.

Additionally, our research has revealed a previously undefined gene region coding ten *B. anthracis*-specific genes, of which nine were reported for the first time, to our knowledge. These findings contribute to our current knowledge of the unique genetic attributes of *B. anthracis* in comparison to its genetically similar neighboring species, and provide distinct markers that distinguish *B. anthracis* from other closely related *Bacillus* species.

In the future, in-depth analyses of these genes unique to *B. anthracis* could potentially lead to unveil metabolic pathways that enable the bacterium to survive in prolonged nutrient-poor conditions, as well as their roles in the sporulation and germination life cycle. This research could serve as a foundation for exploring potential vaccine candidates and drug targets for the treatment of anthrax.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-024-09817-9.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.
Supplementary Material 9.
Supplementary Material 10.
Supplementary Material 11.
Supplementary Material 12.

Acknowledgements

We thank the Microbiology Unit, Paraclinical Studies University of Zambia, Hokkaido University Center for Zoonosis Control in Zambia, and Institute for Veterinary Medicine Mongolia for their cooperation in providing genomic DNA of *Bacillus anthracis*. We would like to acknowledge the support of BEI Resources in providing the bacterial strains necessary for our study. The following reagents were obtained through BEI Resources, NIAID, NIH: *Enterococcus*

faecalis, Strain V587, NR-31979; *Enterococcus faecium*, Strain UAA714, NR-32065; *Enterococcus faecium*, Strain UAA945, NR-32094; *Pseudomonas aeruginosa*, Strain PA14, NR-50573; *Staphylococcus aureus*, Strain F003B2N-C, NR-30546; *Bacillus cereus*, Strain G9241, NR-9564; *Bacillus cereus*, Strain E33L, NR-12264. The following reagent was obtained through BEI Resources, NIAID, NIH as part of the Human Microbiome Project: *Enterococcus faecalis*, Strain S613, HM-334. The following reagents were provided by the Network on Antimicrobial Resistance in *Staphylococcus aureus* (NARSA) for distribution by BEI Resources, NIAID, NIH: *Staphylococcus aureus*, Strain RN4220, NR-45946; *Staphylococcus aureus*, Strain USA300-0114, NR-46070; *Staphylococcus aureus*, Strain 71080, NR-46418.

Authors' contributions
T.Z., Y.F., and H.H. conceptualized the study. T.Z. and Y.F. analyzed the data. T.Z. and M.Sh. conducted the experiments. A.P., J.E., N.E., B.B., H.K.K., M.M., and B.M.H. provided resources. M.Ch. and M.S. contributed with technical assistance. T.Z. wrote the manuscript, which was corrected and approved by all other co-authors. All authors read and approved the final manuscript.

Funding
This work was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan and Japan Society for the Promotion of Science (JSPS) under Grants-in-Aid for Scientific Research (KAKENHI) to T.Z. (Grant Number 23K19460), H.H. (Grant Number 18K19436), and A.P. (Grant Number 21K15430)., The Japan Program for Infectious Diseases Research and Infrastructure (JIDRI) to H.H. (Grant Number JP23wm0125008) from the Japan Agency for Medical Research and Development (AMED). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials
Complete genome datasets of bacterial strains used in the pan-genome analysis and *B. anthracis* strains used for local BLAST alignment were downloaded from NCBI and their accession numbers were listed in Additional file 1 and Additional file 3.

Declarations

Ethics approval and consent to participate
Ethical approval was granted by Excellence in Research Ethics Converge with reference number 2021-Nov-024.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Division of Infection and Immunity, International Institute for Zoonosis Control, Hokkaido University, Sapporo, Japan. [2]Graduate School of Infectious Diseases, School of Veterinary Medicine, Hokkaido University, Sapporo, Japan. [3]GenEndeavor LLC, 26219 Eden Landing Rd, Hayward, CA, USA. [4]Laboratory of Infectious Diseases and Immunology, Institute of Veterinary Medicine, Mongolian University of Life Sciences, Ulaanbaatar, Mongolia. [5]Laboratory of Food Safety and Hygiene, Institute of Veterinary Medicine, Mongolian University of Life Sciences, Ulaanbaatar, Mongolia. [6]Laboratory of Molecular Genetics, Institute of Veterinary Medicine, Mongolian University of Life Sciences, Ulaanbaatar, Mongolia. [7]Public Health Unit, Disease Control Studies, School of Veterinary Medicine, University of Zambia, Lusaka, Zambia. [8]Microbiology Unit, Paraclinical Studies, School of Veterinary Medicine, University of Zambia, Lusaka, Zambia.

Received: 29 January 2024 Accepted: 27 August 2024
Published online: 09 September 2024

References

1. Carlson CJ, Kracalik IT, Ross N, Alexander KA, Hugh-Jones ME, Fegan M, Elkin BT, Epp T, Shury TK, Zhang W, et al. The global distribution of Bacillus anthracis and associated anthrax risk to humans, livestock and wildlife. Nat Microbiol. 2019;4(8):1337–43.
2. Kyriacou DN, Adamski A, Khardori N. Anthrax: from antiquity and obscurity to a front-runner in bioterrorism. Infect Dis Clin North Am. 2006;20(2):227–51, viii.
3. Mock M, Fouet A. Anthrax. Annu Rev Microbiol. 2001;55(1):647–71.
4. Turnbull PC. Anthrax vaccines: past, present and future. Vaccine. 1991;9(8):533–9.
5. Baldwin VM. You can't B. cereus – a review of Bacillus cereus strains that cause anthrax-like disease. Front Microbiol. 2020;11: 1731.
6. Logan NA, Carman JA, Melling J, Berkeley RC. Identification of bacillus anthracis by API tests. J Med Microbiol. 1985;20(1):75–85.
7. Mohammed MJ, Marston CK, Popovic T, Weyant RS, Tenover FC. Antimicrobial susceptibility testing of Bacillus anthracis: comparison of results obtained by using the national committee for clinical laboratory standards broth microdilution reference and Etest agar gradient diffusion methods. J Clin Microbiol. 2002;40(6):1902–7.
8. Odendaal MW, Pieterson 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.
9. Ramisse V, Patra G, Garrigue H, Guesdon JL, Mock M. Identification and characterization of Bacillus anthracis by multiplex PCR analysis of sequences on plasmids pXO1 and pXO2 and chromosomal DNA. FEMS Microbiol Lett. 1996;145(1):9–16.
10. Turnbull PCB, Hutson RA, Ward MJ, Jones MN, Quinn CP, Finnie NJ, Duggleby CJ, Kramer JM, Melling J. Bacillus anthracis but not always anthrax. J Appl Bacteriol. 1992;72(1):21–8.
11. 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.
12. Qi Y, Patra G, Liang X, Williams LE, Rose S, Redkar RJ, DelVecchio VG. Utilization of the rpoB gene as a specific chromosomal marker for real-time PCR detection of Bacillus anthracis. Appl Environ Microbiol. 2001;67(8):3720–7.
13. Hurtle W, Bode E, Kulesh DA, Kaplan RS, Garrison J, Bridge D, House M, Frye MS, Loveless B, Norwood D. Detection of the Bacillus anthracis gyrA gene by using a minor groove binder probe. J Clin Microbiol. 2004;42(1):179–85.
14. Ramisse V, Patra G, Vaissaire J, Mock M. The Ba813 chromosomal DNA sequence effectively traces the whole Bacillus anthracis community. J Appl Microbiol. 1999;87(2):224–8.
15. Easterday WR, Van Ert MN, Simonson TS, Wagner DM, Kenefic LJ, Allender CJ, Keim P. Use of single nucleotide polymorphisms in the plcR gene for specific identification of Bacillus anthracis. J Clin Microbiol. 2005;43(4):1995–7.
16. Easterday WR, Van Ert MN, Zanecki S, Keim P. Specific detection of Bacillus anthracis using a TaqMan mismatch amplification mutation assay. Biotechniques. 2005;38(5):731–5.
17. Birdsell DN, Pearson T, Price EP, Hornstra HM, Nera RD, Stone N, Gruendike J, Kaufman EL, Pettus AH, Hurbon AN, et al. Melt analysis of mismatch amplification mutation assays (Melt-MAMA): a functional study of a cost-effective SNP Genotyping Assay in bacterial models. PLoS ONE. 2012;7(3): e32866.
18. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.
19. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691–3.
20. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol. 2008;11(5):472–7.
21. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(D1):D638-46.
22. Akamatsu R, Suzuki M, Okinaka K, Sasahara T, Yamane K, Suzuki S, Fujikura D, Furuta Y, Ohnishi N, Esaki M, et al. Novel sequence type in Bacillus cereus strains associated with nosocomial infections and bacteremia, Japan. Emerg Infect Dis. 2019;25(5):883–90.
23. Hutson RA, Duggleby CJ, Lowe JR, Manchee RJ, Turnbull PC. The development and assessment of DNA and oligonucleotide probes for the specific detection of Bacillus anthracis. J Appl Bacteriol. 1993;75(5):463–72.
24. Beyer W, Glöckner P, Otto J, Böhm R. A nested PCR method for the detection of Bacillus anthracis in environmental samples collected from former tannery sites. Microbiol Res. 1995;150(2):179–86.

25. World Health Organization. Guidelines for the Surveillance and Control of Anthrax in Human and animals. 3rd ed. Geneva: WHO; 1998.
26. Makino SI, Cheun HI, Watarai M, Uchida I, Takeshi K. Detection of anthrax spores from the air by real-time PCR. Lett Appl Microbiol. 2001;33(3):237–40.
27. Read TD, Peterson SN, Tourasse N, Baillie LW, Paulsen IT, Nelson KE, Tettelin H, Fouts DE, Eisen JA, Gill SR, et al. The genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria. Nature. 2003;423(6935):81–6.
28. 홍석현.: A new Bacillus anthracis specific gene and its diagnostic application. Seoul National University Graduate School. 2016.
29. Sozhamannan S, Chute MD, McAfee FD, Fouts DE, Akmal A, Galloway DR, Mateczun A, Baillie LW, Read TD. The Bacillus anthracis chromosome contains four conserved, excision-proficient, putative prophages. BMC Microbiol. 2006;6(1): 34.
30. Radnedge L, Agron PG, Hill KK, Jackson PJ, Ticknor LO, Keim P, Andersen GL. Genome differences that distinguish Bacillus anthracis from Bacillus cereus and Bacillus thuringiensis. Appl Environ Microbiol. 2003;69(5):2755–64.
31. Létant SE, Murphy GA, Alfaro TM, Avila JR, Kane SR, Raber E, Bunt TM, Shah SR. Rapid-viability PCR method for detection of live, virulent Bacillus anthracis in environmental samples. Appl Environ Microbiol. 2011;77(18):6570–8.
32. Lyu Y, Wang D, Yuan L, Feng E, Zhu L, Pan C, Guo Y, Liu X, Wang H. Rapid Identification of bacillus anthracis in silico and on-site using novel single-nucleotide polymorphisms. Microbiol Spectr. 2022;10(3):e0228521.
33. Ohnishi N, Maruyama F, Ogawa H, Kachi H, Yamada S, Fujikura D, Nakagawa I, Hang'ombe MB, Thomas Y, Mweene AS, et al. Genome sequence of a Bacillus anthracis outbreak strain from Zambia, 2011. Genome Announc. 2014;2(2):e0016-0014.
34. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, et al. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial pan-genome. Proc Natl Acad Sci U S A. 2005;102(39):13950–5.
35. Bazinet AL. Pan-genome and phylogeny of Bacillus cereus sensu lato. BMC Evol Biol. 2017;17(1):176.
36. Wielinga PR, Hamidjaja RA, Ågren J, Knutsson R, Segerman B, Fricker M, Ehling-Schulz M, de Groot A, Burton J, Brooks T, et al. A multiplex real-time PCR for identifying and differentiating B. Anthracis virulent types. Int J Food Microbiol. 2011;145:S137-44.
37. Forsberg LS, Choudhury B, Leoff C, Marston CK, Hoffmaster AR, Saile E, Quinn CP, Kannenberg EL, Carlson RW. Secondary cell wall polysaccharides from Bacillus cereus strains G9241, 03BB87 and 03BB102 causing fatal pneumonia share similar glycosyl structures with the polysaccharides from Bacillus anthracis. Glycobiology. 2011;21(7):934–48.
38. Leoff C, Saile E, Sue D, Wilkins P, Quinn CP, Carlson RW, Kannenberg EL. Cell wall carbohydrate compositions of strains from the Bacillus cereus group of species correlate with phylogenetic relatedness. J Bacteriol. 2008;190(1):112–21.
39. Klotz A, Georg J, Bučinská L, Watanabe S, Reimann V, Januszewski W, Sobotka R, Jendrossek D, Hess WR, Forchhammer K. Awakening of a dormant cyanobacterium from nitrogen chlorosis reveals a genetically determined program. Curr Biol. 2016;26(21):2862–72.
40. Han CS, Xie G, Challacombe JF, Altherr MR, Bhotika SS, Brown N, Bruce D, Campbell CS, Campbell ML, Chen J, et al. Pathogenomic sequence analysis of Bacillus cereus and Bacillus thuringiensis isolates closely related to Bacillus anthracis. J Bacteriol. 2006;188(9):3382–90.
41. Furuta Y, Tsujinouchi M, Shawa M, Zorigt T, Miyajima Y, Paudel A, Suzuki S, Higashi H. Complete genome sequences of 24 strains of Bacillus cereus Isolated from nosocomial infection and bacteremia cases in Japan. Microbiol Resour Announc. 2022;11:e0120321.
42. Kolstø AB, Tourasse NJ, Økstad OA. What sets Bacillus anthracis apart from other Bacillus species? Annu Rev Microbiol. 2009;63:451–76.
43. Timofeev V, Bakhteeva I, Khlopova K, Mironova R, Titareva G, Goncharova Y, Solomentsev V, Kravchenko T, Dyatlov I, Vergnaud G. New Research on the Bacillus anthracis genetic diversity in Siberia. Pathogens 2023;12(10).

Publisher's note