Article

Discovery Medicine 2024; 36(184): 1030–1040
https://doi.org/10.24976/Discov.Med.202436184.96

Whole Genome-Sequencing and Phylogenetic Analysis of *Bacillus anthracis* from 2019–2023 in Ningxia Hui Autonomous Region, China

Ran Xian[1,2,†], Cong Yang[2,†], Jianwei Gao[2], Xiang Liu[2], Qian Chen[1,2], Lei Gao[1,2], Yu Zhao[1], Wenhe Kuai[2], Jiangtao Ma[2,*]

[1]School of Public Health, Ningxia Medical University, 750001 Yinchuan, Ningxia, China
[2]Ningxia Hui Autonomous Region Center for Disease Control and Prevention, 750004 Yinchuan, Ningxia, China
*Correspondence: nxcdcmjt@sina.com (Jiangtao Ma)
[†]These authors contributed equally.
Published: 20 May 2024

Background: Since 2019, the incidence of anthrax in the Ningxia Hui Autonomous Region has increased significantly compared with previous years, so in this situation the anthrax in the Ningxia region not only had a detrimental impact on public health, but also inflicted significant economic repercussions. Therefore, we conducted a molecular epidemiological study of 20 strains from 2019–2023 isolates. This study investigated the origin of *Bacillus anthracis* and its genetic diversity.
Methods: We conducted canonical single-nucleotide polymorphisms (CanSNPs) typing and whole genome sequencing based on the extracted nucleic acid of *Bacillus anthracis*. Based on the whole genome drafts, we studied the genomic characteristics of 20 isolates. Meanwhile, we performed phylogenetic studies based on genome-wide core single-nucleotide polymorphisms (SNPs) using MEGA's Maximum Likelihood (ML) method and core-genome-based multilocus sequence typing (cgMLST) of the core genomes of these strains using BioNumerics' minimum spanning tree (MST) model.
Results: The 20 isolates were categorized into sub-lineages A.Br.001/002, and comparative genomic analyses of these strains with other isolates from other parts of the world showed that the strains from Ningxia were correlated with isolates from Europe, Indonesia, Georgia (USA), and Beijing (China). For the 20 isolates in Ningxia, the genetic relationship of the isolates isolated from the same year or region was relatively close.
Conclusion: The A.Br.001/002 subgroup was the dominant endemic strain in Ningxia. The genetic relationship and phylogenesis between isolates from Ningxia and strains from Europe and Indonesia suggest that anthrax spread around the globe through ancient trade routes.

Keywords: *Bacillus anthracis*; whole-genome sequence (WGS); CanSNPs; cgMLST; genome-wide core SNPs; phylogenetic hypothesis

Introduction

Bacillus anthracis is the causative agent of anthrax [1]. It is a gram-positive bacterium belonging to the *Bacillus cereus sensu lato* [2]. As we all know, it often infects herbivores, causing acute and fatal diseases [3,4]. *Bacillus anthracis* produces spores that can survive for extended periods in various environments [5], threatening global spread and a stable epidemic focus. In China, anthrax has been present for a long time, especially in poor rural areas [6]. According to previous reports, *Bacillus anthracis* contamination has been detected in up to 6 percent of environmental samples [7]. Therefore, an anthrax outbreak is more likely if the outside environment is suitable. According to a study [8], from 1955 to 2022, the western region of China had a high anthrax incidence. The incidence of anthrax increased year by year in Ningxia Hui Autonomous Region from 2019 to 2023, respectively $0.026/10^4$ (2019), $0.014/10^4$ (2020), $0.032/10^4$ (2021), $0.037/10^4$ (2022) and $0.07/10^4$ (2023); nevertheless, from 2004 to 2018, the average annual reporting rate of anthrax in Ningxia was only $0.011/10^4$. This dramatic increase represents a significant public health concern.

Globally, *Bacillus anthracis* is divided into three main categories: A, B, and C, and is further characterized into 12 subclasses [9]. Over time, the anthrax subgroup expansion, such as A.Br.008/009, was divided into A.Br.008/011 and A.Br.011/009 [10]. While in China, *Bacillus anthracis* is mainly divided into five sub-lineages in the A lineage: A.Br.001/002, A.Br.Ames, A.Br.008/009, A.Br.Australia94, A.Br. Vollum [11]. Previous reports have observed *Bacillus anthracis* in China as correlated with isolates from Europe, South Asia, the USA, etc. [7]. However, this study neglected Ningxia, where anthrax has had a high incidence in recent years. Concerningly, no systematic molecular epidemiology of anthrax has been carried out. Therefore, the study on the origin and distribution of anthrax isolates in Ningxia is significant for preventing and controlling the anthrax epidemic.

Bacillus anthracis has a low evolution rate and a high degree of homogeneity [12]. The diversity between different strains is shallow compared to other bacteria; therefore, high-resolution molecular characterization techniques are needed to distinguish strains effectively [13, 14]. Modern genotyping methods are canonical single-nucleotide polymorphisms (CanSNPs) [15] and multilocus variable-number tandem repeat (VNTR) analysis (MLVA). CanSNPs use a series of representative cladistic single-nucleotide polymorphisms (SNPs) to define phylogenetic clades and determine the phylogenetic location of strains. For the existing MLVA typing technology, genetic relationships are determined by analyzing 8 [16], 15 [9], 20 [17], 25 [18] and 31 [19] VNTR sites. Although MLVA has a high resolution, it can easily cause homogeneity problems, so this result can lead to incorrect assignments in phylogenetic analyses (This problem is caused by the use of marker genes with high mutation rates.) [15]. These problems can be avoided by using single-nucleotide polymorphisms (SNPs) via whole genome sequencing [10,20,21] and core-genome-based multilocus sequence typing (cgMLST) [22, 23]. These approaches can subtype the pathogen and provide an ideal marker for high-precision diagnostic analysis.

This study determined the whole-genome sequence (WGS) of *Bacillus anthracis* isolates prevalent in 2019–2023. Firstly, CanSNPs analysis was used to determine its position in the global evolution of *Bacillus anthracis*. Then, whole genome core SNPs and cgMLST were used to reconstruct the phylogenetic tree of *Bacillus anthracis* so that we can elucidate the pathogenic characteristics of *Bacillus anthracis* in this area and provide a theoretical basis for scientific prevention and control, respectively.

Materials and Methods

DNA Extraction

Firstly, 20 strains of *Bacillus anthracis* isolated from 2019–2023 stored in our laboratory were resuscitated and cultivated by nutrient agar (NA) for 24 hours. We extracted nucleic acids in the Biosafety Level 2+ Laboratory (BSL-2+) from a QIAGEN DNA Mini Kit (175023515, QIAGEN, Dusseldorf, North Rhine-Westphalia, Germany). Finally, the extracted nucleic acids were inoculated on NA for 24 hours to confirm bio-safety. Nucleic acid concentration was determined by Qubit™ 3 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), and nucleic acid concentrations higher than 2.0 µmol/L were retained. High-quality nucleic acids were used for whole genome sequencing.

Draft Whole-Genome Sequence (WGS)

Illumina MiSeq (Illumina, San Diego, CA, USA) was used to index sequencing; different joint combinations were used to distinguish different strain sequences. Nextera XT DNA Library Prep (Illumina, San Diego, CA, USA) was used to build a sequencing library. Miseq Reagen Kit V2,300-cycles (Illumina, San Diego, CA, USA) was used when the sample sequencing started. After this step, short-read data sets were exported using FastQ format. Finally, quality control was performed using CLC Genomic Workbench 21.0.5 (QIAGEN, Redwood City, CA, USA). After editing ambiguous and joint sequences, the original *Bacillus anthracis* strain (Ames Ancestor: NC_007530.2) was used as the reference genome to splice the edited data.

CanSNPs Genotyping

The TaqMan method, a previously validated protocol, was used to detect the sub-lineage of strains [9]. Primers/Probes (Sangon Biotech, Shanghai, China) were synthesized according to Van, and the nucleotide type was determined by the QuantStudioTM 7 Flex Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA) [9]. At this stage, we could confirm the basic position of *Bacillus anthracis* on the evolutionary tree.

Pan-Genomic Functional Analysis

Reference Genome Sequence Acquisition

In this study, the sequence of reference (included 39 strains of *Bacillus anthracis*), original genome (Ames Ancestor: NC_007530.2), and *Bacillus cereus* (AH820: NC_011773.1) were derived from NCBI RefSeq library (https://www.ncbi.nlm.nih.gov/refseq/). All strains were listed in **Supplementary Table 1**.

Pan-Genomic Characterization

Firstly, Prokka (v 1.14.6) (Melbourne, Victoria, Australia) was used to annotate the genome sequence of the above strains [24]. Secondly, to evaluate the relationship between the 20 isolates and anthrax reference strain Ames Ancestor, the following procedures were performed: Average nucleotide identity (ANI) was compared by fastANI (v 1.1) (Atlanta, GA, USA) [25]. Comparem (v 0.1.2) (Halifax, Nova Scotia, Canada) [26] was used to compare Average amino acid identity (AAI) (https://github.com/dparks1134/CompareM). The Tetra-Nucleotide characteristics (TETRA) were compared by pyani (v 0.2.12) (Aberdeen, Aberdeenshire, UK) [27]. Thirdly, to visualize the above result, the pheatmap (https://CRAN.R-project.org/package=pheatmap) package was used, which comes from R 4.3.1 (https://www.R-project.org/), to output the above heatmaps. Fourthly, the genetic distance was calculated by Mash (v 2.3) (Bethesda, MD, USA) [28,29]. Peppan (v 1.0.6) (Coventry, UK) [30] performed genome-wide reconstruction of 20 isolates. A Python script was used to output the previous step. Namely, PEPPAN.gff was used to obtain the dilution curve file of the isolates' pan-genomic. The last dilution curves were then drawn by pandas (v 1.3.5) and matplotlib (v 3.3.5) [31].

Table 1. Basic information of *Bacillus anthracis* strains and genomes in Ningxia, 2019–2023.

Strains	Location (County/District)	CanSNPs group	Genome size (bp)	GC content (%)	CDS	RNAs	Repeat region	Prophage regions	Plasmids	Mash
CHN-NX-BA-2019-01	HeLan	A.Br.001/002	5,485,169	35.09	5781	52	0	1	2	0.00152214
CHN-NX-BA-2019-02	HeLan	A.Br.001/002	5,485,762	35.09	5782	59	0	1	2	0.00152214
CHN-NX-BA-2020-01	HeLan	A.Br.001/002	5,445,748	35.09	5733	55	0	1	2	0.00136596
CHN-NX-BA-2020-02	YongNing	A.Br.001/002	5,485,088	35.09	5788	59	0	1	2	0.00152214
CHN-NX-BA-2021-01	YongNing	A.Br.001/002	5,445,559	35.09	5731	62	0	2	2	0.00139189
CHN-NX-BA-2021-02	LiTong	A.Br.001/002	5,446,556	35.09	5730	65	0	2	2	0.00139189
CHN-NX-BA-2021-03	LiTong	A.Br.001/002	5,444,530	35.09	5733	70	0	1	2	0.00139189
CHN-NX-BA-2021-04	YongNing	A.Br.001/002	5,445,938	35.09	5729	67	0	2	2	0.00139189
CHN-NX-BA-2021-05	YongNing	A.Br.001/002	5,444,552	35.09	5733	64	0	1	2	0.00136596
CHN-NX-BA-2022-01	YongNing	A.Br.001/002	5,450,739	35.09	5728	54	0	1	2	0.00139189
CHN-NX-BA-2022-02	LiTong	A.Br.001/002	5,451,599	35.11	5739	56	0	1	2	0.00136596
CHN-NX-BA-2022-03	JinFeng	A.Br.001/002	5,447,952	35.09	5727	58	0	2	2	0.00136596
CHN-NX-BA-2022-04	YongNing	A.Br.001/002	5,446,473	35.09	5765	48	0	2	2	0.00157453
CHN-NX-BA-2022-05	YongNing	A.Br.001/002	5,454,991	35.16	5729	47	0	1	2	0.00136596
CHN-NX-BA-2023-04	YongNing	A.Br.001/002	5,448,576	35.10	5727	70	1	2	2	0.00139189

Table 1. Continued.

Strains	Location (County/District)	CanSNPs group	Genome size (bp)	GC content (%)	CDS	RNAs	Repeat region	Prophage regions	Plasmids	Mash
CHN-NX-BA-2023-05	YongNing	A.Br.001/ 002	5,444,856	35.10	5723	53	0	1	2	0.00136596
CHN-NX-BA-2023-06	XiXia	A.Br.001/ 002	5,443,814	35.10	5710	54	0	2	2	0.00136596
CHN-NX-BA-2023-07	YongNing	A.Br.001/ 002	5,435,364	35.10	5696	52	0	1	2	0.00146992
CHN-NX-BA-2023-08	HeLan	A.Br.001/ 002	5,445,197	35.09	5714	57	1	1	2	0.00136596
CHN-NX-BA-2023-09	YuanZhou	A.Br.001/ 002	5,439,972	35.09	5705	38	0	2	2	0.00139189

In this table, Strains were the naming serial number of *Bacillus anthracis* isolated from Ningxia in 2019–2023 Location was the specific geographical location of *Bacillus anthracis* in Ningxia Hui Autonomous Region when it was isolated, which was accurate to every district and county. Genome size was the size of the sequenced draft genome. GC content was the ratio of G and C bases in the genome. CanSNPs group was the specific subgroup of *Bacillus anthracis* in the evolutionary tree. CDS was the number of coding sequences in the genome. RNAs contained the numbers of rRNA, tRNA, and tmRNA. Repeat-region regions were the number of Repeat-sequence areas. Prophage regions were the number of Prophage sequence areas. Plasmids consisted of pXO1 and pXO2. Mash resulted from a reference genome Ames_Ancestor compared with 20 isolates of *Bacillus anthracis* (The distance between genomes.). CanSNPs, canonical single-nucleotide polymorphisms.

In this stage, if ANI >95%, AA1 >95%, TETRA >99%, and MASH <0.5, the two genomes are considered members of the same species group (The MASH means Genome distance that Fast genome and metagenome distance estimation using MinHash.).

Genome-Wide Phylogenetic Analysis

Whole Genome SNP Extraction and Phylogenetic Analysis

All the genome sequences of *Bacillus anthracis* were uploaded to PubMLST (https://pubmlst.org/), which aimed to obtain allele data of the core genome in the whole genome [22,32]. BioNumerics 7.6 (Applied-Maths, Sint-Martens-Latem, Belgium) was used to construct phylogenetic trees using complete linkage and minimum spanning tree (MST). This step determined the relationship and origin between the 20 isolates and other strains based on the established minimum spanning tree.

Core Genome Extraction and Phylogenetic Analysis

First, sinppy (v 4.6.0) (Melbourne, Victoria, Australia) (https://github.com/tseemann/snippy) was used to extract the core SNP sequences, including 20 *Bacillus anthracis* strains, Bacillus cereus AH820, and 40 reference strains. Secondly, gubbins (v 2.4.1) (Hinxton, Cambridge, UK) were used to remove the parts of the core SNP sequences that belonged to gene recombination [33]. Finally, SNP sites (v 2.5.1) (Hinxton, Cambridge, UK) were used for core SNP analysis [34]. Strains of AH820 and the final dataset containing 18078 SNPs were Bootstrap 1000 times using the Tamura-Nei model of MEGA 11 to infer phylogenetic relationships by the maximum likelihood method [35]. This step explores genetic diversity among 20 isolates and other *Bacillus anthracis* strains based on established phylogenetic trees.

Results

Basic Features of the Pan-Genome

In this study, all isolates were preserved in the laboratory, isolated from different Ningxia province regions in 2019–2023. All strains were analyzed using the CanSNPs analysis method proposed by Van [9] to determine where *Bacillus anthracis* is located in the global *Bacillus anthracis* phylogeny. All the strains were clustered in the A.Br.001/002 group in this study (see Table 1).

The sequenced whole genome sketches of the 20 isolates were between 5,400,000 and 5,500,000 bp, similar to the original strain of Ames Ancestor *Bacillus anthracis* (5,227,419 bp). Their genome coverage reached more than 98.9%. Its GC content was similar to the reference GC content (35.38%), and it carried two plasmids (pXO1 and pXO2) [2,36].

Pan-Genomic Characterization

The pairwise comparison of 20 isolates with the Ames Ancestor genome revealed that all ANI >99.965%, AAI >99.98%, and TETRA >99.97% (see Fig. 1). Mash comparisons (Ames Ancestor compared with isolates) were less than 0.01 (see Table 1). These results indicated that all of the genomes were extremely similar and belonged to members of the same genomic species group.

In Fig. 1A, they were clustered in the same category in the heatmap, which were Ames Ancestor, isolates CHN-NX-BA-2023-07/08/09 and CHN-NX-BA-2022-04, showing similar genetic relationship, while the remaining strains formed new sub-branches. In Fig. 1B, these genomes revealed that they were highly similar, the results of the pairwise comparison of amino acid identities. However, they were still divided into two broad categories in which strains with 100% similarity existed clustered in the same subclass. According to the comparison of Tetra Nucleotides consistency, we can observe the results in which all strains were classified into two major groups, including Ames Ancestor and 20 isolates. In contrast, between the 20 isolates, CHN-NX-BA-2022-05 was classified into two groups, with the remaining 19 strains in Fig. 1C. Ames Ancestor belongs to the A.Br.Ames sub-lineage in the global phylogeny. The remaining 20 isolates are of the A.Br.001/002 sub-lineage, in which A.Br.Ames was formed by a single nucleotide mutation at the A.Br.002 locus in the A.Br.001/002 sub-lineage [7], which explained the very high degree of similarity in the heatmap.

In Fig. 1D, as the number of sequenced strains increased, the total number of pan-genomic genes increased slowly and then increased rapidly when the number of sequenced strains was close to between 2 and 4. The next increase slowly plateaued, and finally, the total number of genes no longer increased when the number of sequenced strains was close to 17. In contrast, the total number of core genomic genes decreased with the number of sequenced strains and then stabilized when the number was close to 17. According to Fig. 1D, it can be seen that although the genome of anthrax was open, the whole genome would soon undergo closure, which indicated that in different environments, the species was able to acquire new genes by exchanging genetic material with other species in various ways or by self-evolution. Still, its ability to obtain foreign genes would weaken as the species evolves. *Bacillus anthracis* is also often present in the external environment as spores, so its genetic evolution is relatively well conserved for this bacterium.

Genome-Wide Phylogenetic Analysis

Phylogenetic Analysis of the Core Genome of Bacillus Anthracis

Fig. 2 shows the minimum spanning tree constructed from the core genomes of the 20 isolates with the remaining 39 comparison strains as well as the reference strain (3803

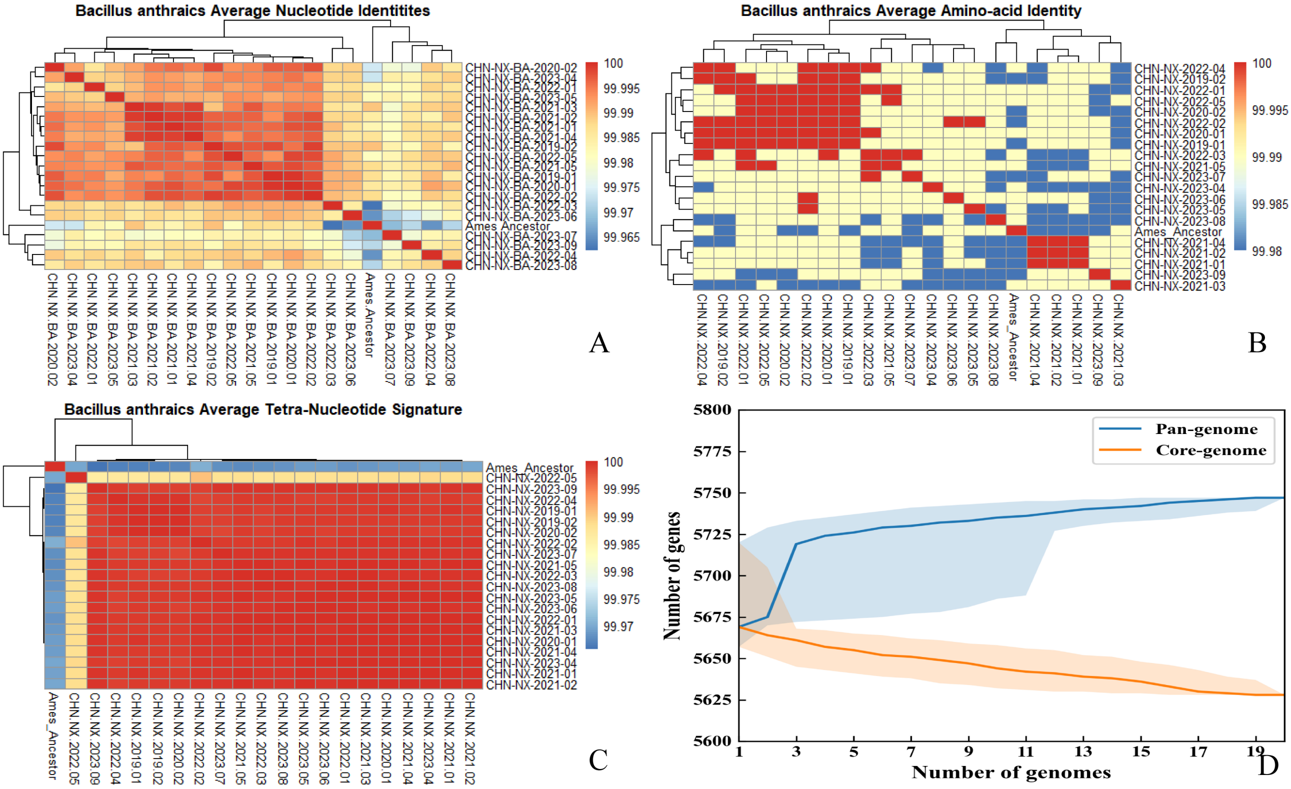

Fig. 1. The Genome characteristic of 20 isolates. (A) The results of pairwise comparisons of mean nucleotide identity. (B) The results of pairwise comparisons of mean amino acid identity. (C) The results of pairwise comparisons of tetranucleotide identity. (D) The dilution curves of the genomic sequences of the 20 isolates, where the pan-genomic curve is the blue curve, and the orange curve is the core-genomic curve, where the shaded portion on the curves is the 95% confidence interval.

MLST housekeeping genes in total) and a phylogenetic tree constructed from (4138 MLST housekeeping genes) using *Bacillus cereus* (AH820) as an outgroup (See **Supplementary Fig. 1**).

In Fig. 2, the evolutionary divergence of the French strain 99-100 (A.Br.011/009) gave rise to the sub-lineage A.Br.Australia94 by radiation. Thereout, the evolutionary divergence of this sub-lineage (via A.Br.002 mutation) gave rise to the sub-lineage A.Br.001/002, and the Ames subgroup was mutated by A.Br.001/002 (via A.Br.001 mutation). In Fig. 2, it can be found that strain CHN-NX-BA-2021-01 had a rapid evolution to produce a clonal group of two strains of CHN-NX-BA-2021-02/04 that had highly similar core genomic information (noted as group 01). From this clonal group (group 01) strains, radiation rapidly occurred, giving rise to the Ames sub-group, Steren subgroup, and the remaining five branches. The rest also produced a highly clonal group of strains with CHN-NX-BA-2019-01, CHN-NX-BA-2020-02, and CHN-NX-BA-2021-02 (denoted as group 02), from which radiation produced most of the isolates in this study (11 strains), one isolate from Indonesia was found to have evolved from this clonal group. Among the remaining three evolutionary branches, CHN-NX-BA-2021-03 was genetically very close to group 01, while CHN-NX-BA-2023-08/09 was genetically distant from clonal group 01. In Fig. 2, the strains from Ningxia (CHN-NX-BA-2021-01) are located at the bottom of A16, A16R, A0389, Ames, Sterne, and the other 19 strains of *Bacillus anthracis* from Ningxia. This indicated that the strains (CHN-NX-BA-2021-01) from Ningxia were the origin.

Phylogenetic Analysis of Whole Genome Core SNPs of Bacillus Anthracis

Fig. 3 shows the phylogenetic tree constructed based on the maximum likelihood method and the Tamura-Nei model for the 20 isolates with the remaining 39 comparator strains and the reference strain, which included 5896 SNPs, and the phylogenetic tree with the *Bacillus cereus* (AH820) as outgroups (18,078 SNPs), shown in **Supplementary Fig. 2**.

In Fig. 3, the phylogenetic tree constructed by genome-wide core SNPs has a similar structure as the phylogenetic tree constructed by cgmlst, in which it was found in the tree constructed by core SNPs that CHN-NA-BA-2022-01/02/03/04 clustered in the same subcategory, and at the same time CHN-NX-BA-2023-09 also belonged to a same category in this phylogenetic tree with the above strains. Simultaneously, CHN-NX-BA-2023-08 clustered in one category with A16 and A16R

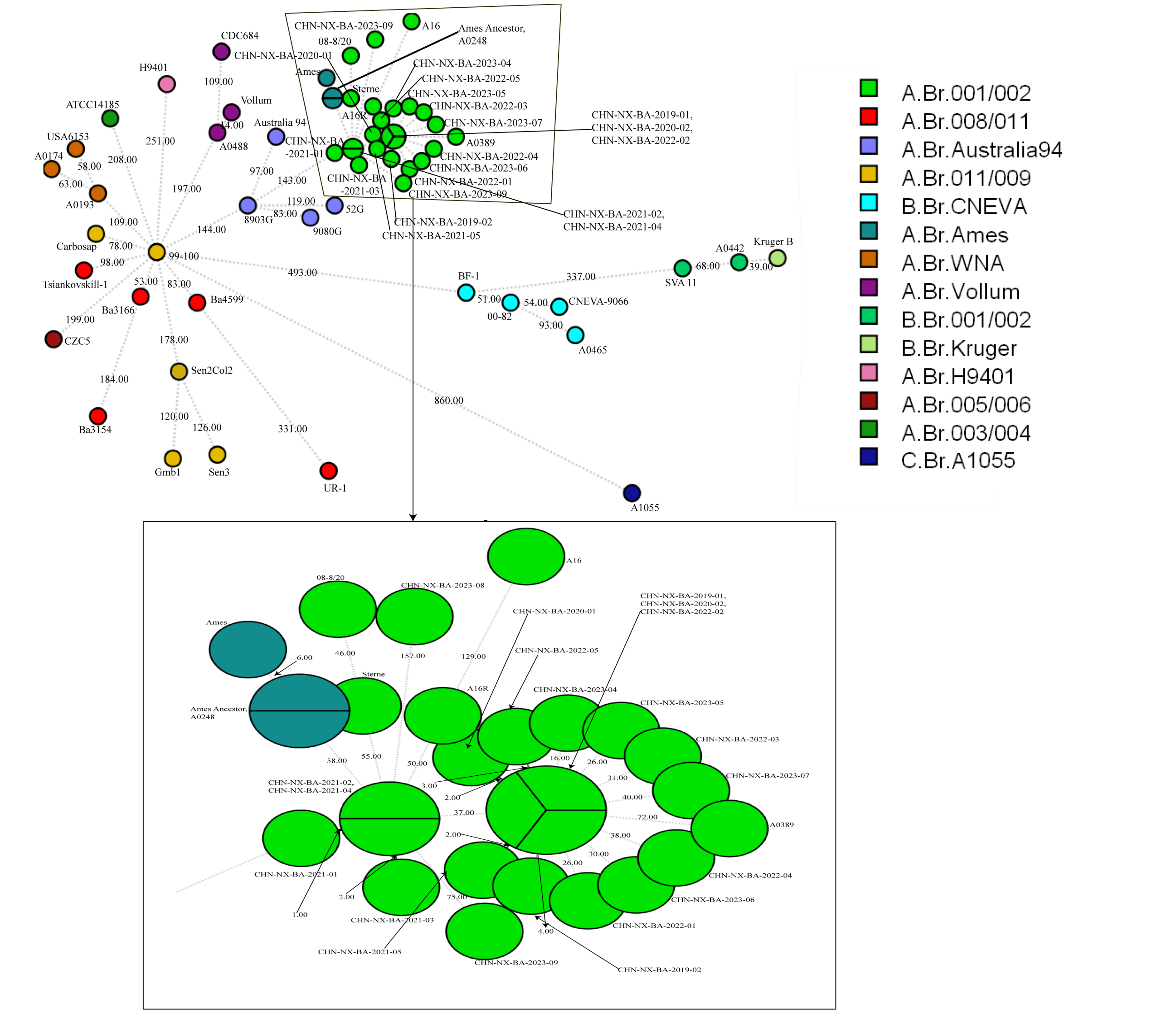

Fig. 2. A minimum spanning tree of 20 *Bacillus anthracis* isolates and 38 other contrast isolates with 1 reference strain. Different colors were used to characterize 14 distinct sub-lineages of *Bacillus anthracis* that were most dominant on the evolutionary tree, which belonged to the A, B, and C groups. The phylogenetic tree consists of 59 core genomes of *Bacillus anthracis*.

strains. While the rest of the isolates were clustered in the same general group with strain A0389 where from Indonesia, Fig. 3 showed that strains CHN-NX-BA-2019-01/02 (recorded as SNP01 group), CHN-NX-BA-2022-01/03 (recorded as SNP02 group), CHN-NX-BA-2022-04/05 (recorded as SNP03 group), CHN-NX-BA-2023-04/06 strain (recorded as SNP04 group) were having close affinities, respectively. So, on this premise, the SNP01 and SNP02 groups were closer to each other, the SNP04 group was closer to CHN-NX-BA-2023-07, and the rest of the strains were clustered under the same broad category with these strains.

These results demonstrate that strains from the same year or region are genetically more closely related than strains from different years or regions.

The evolutionary history was inferred using the Maximum Likelihood method and the Tamura-Nei model [37]. The tree with the highest log likelihood (–39,577.41) is shown. The percentage of trees in which the associated taxa clustered is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model and selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 59 nucleotide sequences. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 5896 positions in the final dataset. Evolutionary analyses were conducted in MEGA 11 [35].

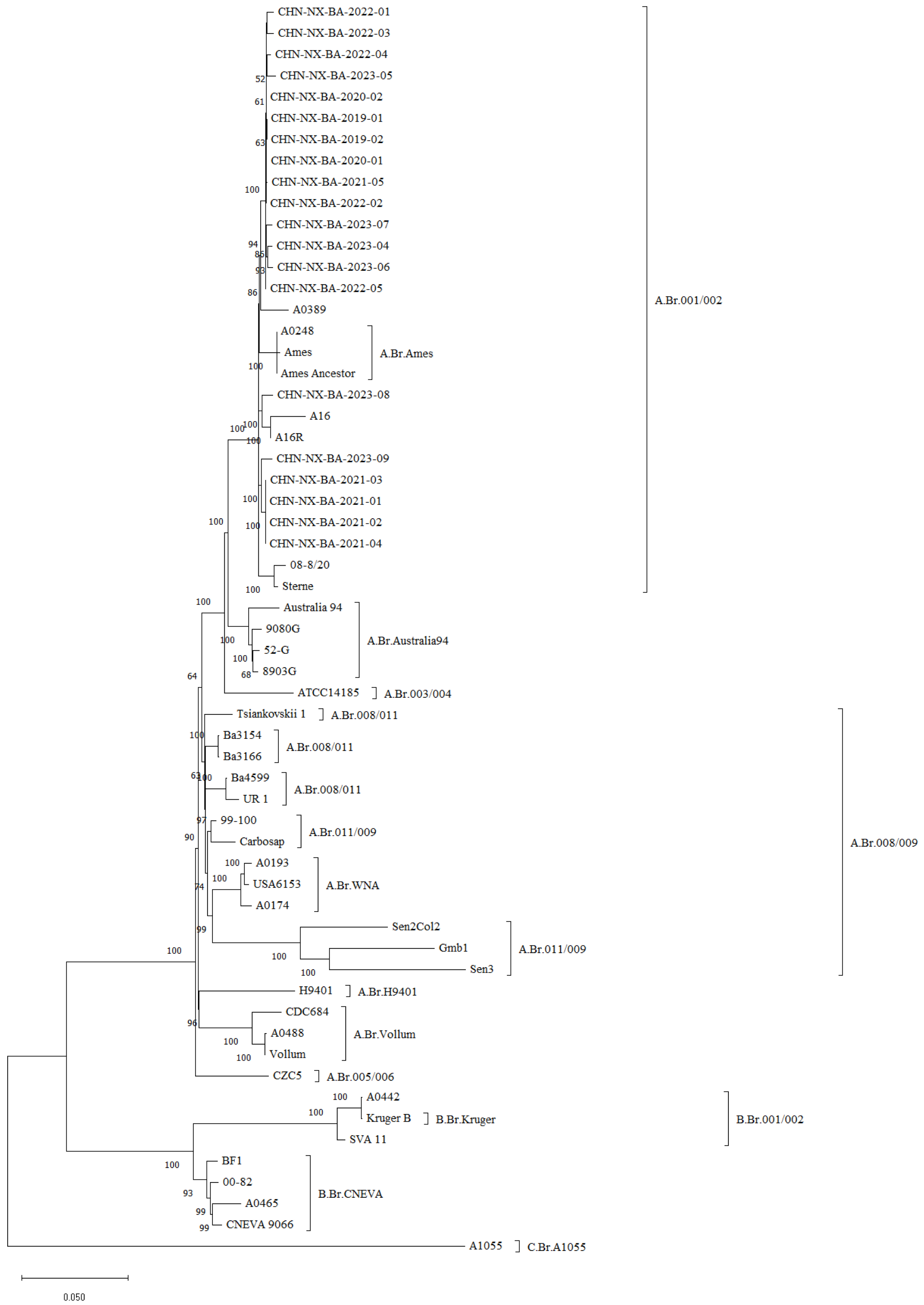

Fig. 3. Evolutionary analysis using Maximum Likelihood method.

Discussion

Whole-genome sequence (WGS) has been established for bacterial subtyping and is regularly used to study pathogen transmission, investigate outbreaks, and perform routine surveillance [22,23,38], therefore, the combined application of cgMLST and Genome-wide SNPs technology developed according to WGS technology can be rapid and harmonized interlaboratory comparisons of *Bacillus anthracis*, essential for global surveillance and outbreak analysis. Although WGS technology is more complex and cumbersome than MLVA, Single Nucleotide Repeats (SNR) and other molecular typing techniques, WGS technology can fully reflect the differences between closely related strains, and we can easily get the information of WGS [15,22,39]. So, in this study we carried out the research with WGS technology.

The 20 isolates of *Bacillus anthracis* isolated from Ningxia province from 2019 to 2023 belong to the A.Br.001/002sub-lineage, of which A.Br.001/002 sub-lineage was the most widely distributed and far-reaching in China and global world [9,11]. This type of *Bacillus anthracis* was the most widespread in China and spread in a number of ways in the country.

Analysis of the core genome of *Bacillus anthracis* makes it evident that the existence of an evolutionary route such as this one that strain 99-100 (France, A.Br.011/009) was first related to A.Br.Australia94 strain 8903G (Arizona, USA), and then to the A.Br.001/002. The subline isolate of this study, CHN-NX-BA-2019-01, was interrelated, and although these strains are genetically distant from each other, MST found that three strains were linked through two evolutionary divergences. This suggests that there may be a common ancestor between these strains that evolved step-by-step over time, and this evolutionary route, which we can confirm in the phylogenetic tree of *Bacillus anthracis*, arises through the mutation of several major SNP sites [9]. Additionally, the existence of such an evolutionary route from CHN-NX-BA-2021-02/04 to Sterne to 08-8/20 (France) [40] (all A.Br.001/002 sub-lineage strains) reinforces the rule that there is a close correlation between isolates from Ningxia and strains from the European continent. As a neighboring province to Xinjiang, where the main anthrax groups present are A.Br.008/009, A.Br.Australia94, and A.Br.Vollum sub-lineages have been revealed to be closely related to strains in Western Europe and South Asia in related studies [7]. Therefore, according to this study, *Bacillus anthracis* in China is related to *Bacillus anthracis* in Europe and the South Asian subcontinent. Furthermore, the *Bacillus anthracis* found in Ningxia is also related to *Bacillus anthracis* in Europe. We can infer that *Bacillus anthracis* spread among each other through ancient trade routes. As a neighboring province of Xinjiang, the Ningxia region has had reciprocal trade relations with the Xinjiang region through the “Hexi Corridor” since ancient times. Therefore, isolates from the Ningxia region can communicate with isolates from other parts of Eurasia. In the evolutionary tree constructed according to the genomic core SNPs, it can be found that the genetic relationship of strain 08-8/20 from France was close to the strains in Ningxia, which further supports the hypothesis. A16, A16R, and Ames Ancestor, etc., generated from the evolutionary divergence of group 01, have a common ancestor from Inner Mongolia [7], which is also a neighboring province of and adjacent to Ningxia, so there is a hypothesis that group 01 isolates may have been come from or influenced by Inner Mongolia. Therefore, some strains in Ningxia may have come from Inner Mongolia. The origin of the isolate A0389 (Indonesia) [9] of group 02 evolutionary differentiation may be related to the maritime Silk Road, which originated from the Qin and Han Dynasties, through which China established contacts with many regions in Southeast Asia and resulted in the exchange of *Bacillus anthracis* strains, and different evolutionary divergences took place under the effect of time. Existing evidence has proved that A16 is related to Ames subgroup strains [7], and in this study, strains A16, 08-8/20, A0389, and other strains were found in the evolutionary branch of group 01. This phenomenon once again confirmed the reliability of the transmission of anthrax through ancient trade.

The core genome MST analysis shows that the core genomes of CHN-NX-BA-2021-01 and CHN-NX-BA-2021-02/04 are almost the same. The latter (group 01) for the former compared to the core genomes occurred in a very low degree of evolution (high genomic similarity), and the genetic relationship between the latter (group 01) and CHN-NX-2021-03 is also. It was similar to group 01 after evolutionary divergence; strain CHN-NX-BA-2023-09 was differentiated. Combined with epidemiological data, it was found that strains CHN-NX-BA-2021-01 and CHN-NX-BA-2021-04 originated from the same county in the Ningxia region. In contrast, CHN-NX-BA-2021-02/03 (same county) was located in neighboring counties. Although CHN-NX-BA-2023-08 and strain CHN-NX-BA-01 in subgroup 01 were isolated from neighboring districts, the genetic distances between the strains are very far apart, thus indicating that the two strains may have evolved from different ancestors in different regions. In the phylogenetic tree construction by core SNPs, it was found that strain CHN-NX-BA-2023-08 was extremely closely related to strains A16 and A16R. This finding again supports another source of *Bacillus anthracis* in Ningxia, i.e., some strains may have evolved from strains from Inner Mongolia. A16 and A16R strains have extremely similar affinities, and this finding again supports another source of *Bacillus anthracis* in Ningxia, i.e., the evolution of strains from Inner Mongolia may have generated some strains [7]. At the same time, CHN-NX-BA-2023-09 comes from relatively more distant districts and counties, and thus, the response shows relatively distant affinities in the phylogenetic tree. Combined with the genome-wide core SNPs, the evolutionary tree found also reflected similar phylogenetic associations.

Combined with the core genome MST analysis, group 02 strains were found to be from strains from different years and districts, suggesting a cryptic transmission chain that links these regions together. Except for CHN-NX-BA-2022-03, which didn’t have geographical origin from the same location compared with group 02, other isolates evolved by group 02 had the same geographical origin as the stains from group 02. The evolutionary tree constructed by core SNPs also showed a similar genetic relationship, and the difference with the evolutionary tree constructed by MST was that the strains in group 02 failed to show the same genetic relationship, which may be because accessory genes and plasmids were not used in the evolutionary tree constructed by MST [22].

Combined with the genome-wide core SNPs phylogenetic tree, we can find that isolates from the same year in Ningxia can be clustered together in the phylogenetic tree and combined with the isolation locations of the strains. We can see that strains from the same or neighboring districts have a more advanced affinity. In contrast, strains from neighboring districts with close affinities confirm that *Bacillus anthracis* can affect neighboring districts through the hidden chain of transmission. In contrast, strains from different years and the same districts that have close affinities, proving that there may be a possible anthrax epidemic focus.

Therefore, based on the above conclusions, we found the interaction pathways and possible sources of *Bacillus anthracis* between Ningxia and other parts of the world and successfully established the molecular epidemiological baseline of *Bacillus anthracis* in Ningxia, providing a possibility for rapid tracing of the source of anthrax outbreaks in the future.

Conclusion

The present study found that the A.Br.001/002 subline strain is prevalent in Ningxia. In addition, there are different routes for the origin and spread of *Bacillus anthracis* in Ningxia and are related to strains in Europe and Indonesia, so if we want to reveal the geographic origin of *Bacillus anthracis* deeply, which needs to be addressed by collecting the whole genome sequencing information of strains from all over the world and obtaining more isolates locally from Ningxia, to geographically establish the epidemiological links and hypotheses for the global spread of anthrax.

Availability of Data and Materials

All data have been supporting to National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov). And everyone who needs can open access. Everyone can get the ID of the NCBI in **Supplementary Table 2**.

Author Contributions

RX: Conceptualization, Investigation, Formal analysis, Writing — original draft, Visualization. CY: Investigation, Resources, Project administration, Funding acquisition, Supervision. JG: Investigation, Resources, Project administration, Funding acquisition, Supervision. XL: Investigation, Formal analysis. QC: Investigation, Formal analysis. LG: Investigation, Formal analysis. YZ: Supervision, Formal analysis, Investigation. WK: Project administration, Supervision, Formal analysis. JM: Investigation, Resources, Writing — review & editing, Project administration, Funding acquisition, Supervision. All authors have been involved in revising it critically for important intellectual content. All authors gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This research was funded by the National Natural Science Foundation of China (81960607). We also received funding from the Natural Science Foundation of Ningxia Province, China (2022AAC03712 and 2023AAC02075).

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.24976/Discov.Med.202436184.96.

References

[1] Datta KK, Singh J. Anthrax. Indian Journal of Pediatrics. 2002; 69: 49–56.

[2] Pilo P, Frey J. Bacillus anthracis: molecular taxonomy, population genetics, phylogeny and patho-evolution. Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases. 2011; 11: 1218–1224.

[3] Katani R, Schilling MA, Lyimo B, Eblate E, Martin A, Tonui T, *et al*. Identification of Bacillus anthracis, Brucella spp., and Coxiella burnetii DNA signatures from bushmeat. Scientific Reports. 2021; 11: 14876.

[4] Sidwa T, Salzer JS, Traxler R, Swaney E, Sims ML, Bradshaw P, *et al*. Control and Prevention of Anthrax, Texas, USA, 2019. Emerging Infectious Diseases. 2020; 26: 2815–2824.

[5] Carlson CJ, Getz WM, Kausrud KL, Cizauskas CA, Blackburn JK, Bustos Carrillo FA, *et al*. Spores and soil from six sides: interdisciplinarity and the environmental biology of anthrax (Bacillus anthracis). Biological Reviews of the Cambridge Philosophical Society. 2018; 93: 1813–1831.

[6] Swartz MN. Recognition and management of anthrax–an update. The New England Journal of Medicine. 2001; 345: 1621–1626.

[7] Simonson TS, Okinaka RT, Wang B, Easterday WR, Huynh L, U'Ren JM, *et al*. Bacillus anthracis in China and its relationship to worldwide lineages. BMC Microbiology. 2009; 9: 71.

[8] Li Y, Yin W, Hugh-Jones M, Wang L, Mu D, Ren X, *et al*. Epidemiology of Human Anthrax in China, 1955-2014. Emerging Infectious Diseases. 2017; 23: 14–21.

[9] Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, Ravel J, *et al*. Global genetic population structure of Bacillus anthracis. PloS One. 2007; 2: e461.

[10] Price EP, Seymour ML, Sarovich DS, Latham J, Wolken SR, Mason J, *et al*. Molecular epidemiologic investigation of an anthrax outbreak among heroin users, Europe. Emerging Infectious Diseases. 2012; 18: 1307–1313.

[11] Wang D, Wang B, Zhu L, Wu S, Lyu Y, Feng E, *et al*. Genotyping and population diversity of Bacillus anthracis in China based on MLVA and canSNP analysis. Microbiological Research. 2020; 233: 126414.

[12] Keim P, Gruendike JM, Klevytska AM, Schupp JM, Challacombe J, Okinaka R. The genome and variation of Bacillus anthracis. Molecular Aspects of Medicine. 2009; 30: 397–405.

[13] Derzelle S, Thierry S. Genetic diversity of Bacillus anthracis in Europe: genotyping methods in forensic and epidemiologic investigations. Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science. 2013; 11: S166–S176.

[14] Rasko DA, Worsham PL, Abshire TG, Stanley ST, Bannan JD, Wilson MR, *et al*. Bacillus anthracis comparative genome analysis in support of the Amerithrax investigation. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108: 5027–5032.

[15] Keim P, Van Ert MN, Pearson T, Vogler AJ, Huynh LY, Wagner DM. Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases. 2004; 4: 205–213.

[16] Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, *et al*. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within Bacillus anthracis. Journal of Bacteriology. 2000; 182: 2928–2936.

[17] Sahl JW, Pearson T, Okinaka R, Schupp JM, Gillece JD, Heaton H, *et al*. A Bacillus anthracis Genome Sequence from the Sverdlovsk 1979 Autopsy Specimens. mBio. 2016; 7: e01501–16.

[18] Lista F, Faggioni G, Valjevac S, Ciammaruconi A, Vaissaire J, le Doujet C, *et al*. Genotyping of Bacillus anthracis strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. BMC Microbiology. 2006; 6: 33.

[19] Beyer W, Bellan S, Eberle G, Ganz HH, Getz WM, Haumacher R, *et al*. Distribution and molecular evolution of bacillus anthracis genotypes in Namibia. PLoS Neglected Tropical Diseases. 2012; 6: e1534.

[20] Antwerpen M, Beyer W, Bassy O, Ortega-García MV, Cabria-Ramos JC, Grass G, *et al*. Phylogenetic Placement of Isolates Within the Trans-Eurasian Clade A.Br.008/009 of *Bacillus anthracis*. Microorganisms. 2019; 7: 689.

[21] Chiaverini A, Abdel-Glil MY, Linde J, Galante D, Rondinone V, Fasanella A, *et al*. Whole Genome Sequencing for Studying *Bacillus anthracis* from an Outbreak in the Abruzzo Region of Italy. Microorganisms. 2020; 8: 87.

[22] Abdel-Glil MY, Chiaverini A, Garofolo G, Fasanella A, Parisi A, Harmsen D, *et al*. A Whole-Genome-Based Gene-by-Gene Typing System for Standardized High-Resolution Strain Typing of Bacillus anthracis. Journal of Clinical Microbiology. 2021; 59: e0288920.

[23] Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. Clinical Microbiology and Infection: the Official Publication of the European Society of Clinical Microbiology and Infectious Diseases. 2018; 24: 350–354.

[24] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics (Oxford, England). 2014; 30: 2068–2069.

[25] Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications. 2018; 9: 5114.

[26] Parks DH, Beiko RG. Measures of phylogenetic differentiation provide robust and complementary insights into microbial communities. The ISME Journal. 2013; 7: 173–183.

[27] Pritchard L, Glover RH, Humphris S, Elphinstone J.G, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical Methods. 2016; 8: 12–24.

[28] Ondov BD, Starrett GJ, Sappington A, Kostic A, Koren S, Buck CB, *et al*. Mash Screen: high-throughput sequence containment estimation for genome discovery. Genome Biology. 2019; 20: 232.

[29] Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, *et al*. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biology. 2016; 17: 132.

[30] Zhou Z, Charlesworth J, Achtman M. Accurate reconstruction of bacterial pan- and core genomes with PEPPAN. Genome Research. 2020; 30: 1667–1679.

[31] Hunter JD. Matplotlib: A 2D graphics environment. Computing in Science & Engineering. 2007; 9: 90–95.

[32] Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Research. 2018; 3: 124.

[33] Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, *et al*. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Research. 2015; 43: e15.

[34] Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, *et al*. *SNP-sites*: rapid efficient extraction of SNPs from multi-FASTA alignments. Microbial Genomics. 2016; 2: e000056.

[35] Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution. 2021; 38: 3022–3027.

[36] Duesbery NS, Vande Woude GF. Anthrax toxins. Cellular and Molecular Life Sciences: CMLS. 1999; 55: 1599–1609.

[37] Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution. 1993; 10: 512–526.

[38] Girault G, Thierry S, Cherchame E, Derzelle S. Application of high-throughput sequencing: discovery of informative SNPs to subtype Bacillus anthracis. Advances in Bioscience and Biotechnology. 2014; 5: 669–677.

[39] Kenefic LJ, Beaudry J, Trim C, Huynh L, Zanecki S, Matthews M, *et al*. A high resolution four-locus multiplex single nucleotide repeat (SNR) genotyping system in Bacillus anthracis. Journal of Microbiological Methods. 2008; 73: 269–272.

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