



Article

Molecular Epidemiology and Phylogeography of Domestically Established *Bacillus anthracis* Isolates Collected in the United States from 1982 to 2013

Chung K. Marston ^{1,*}, Christopher A. Gulvik ¹ , Cari A. Beesley ¹, Mili Sheth ², Mark Burroughs ², Amy K. Swinford ³ , Matthew M. Erdman ^{4,†}, Kristin A. Clothier ⁵, Rajesh Maganbhai Parmar ⁶, Maureen M. Sullivan ⁷, Marianna E. Martinez ⁸, Kristy L. Pabilonia ⁹, Zachary P. Weiner ¹ and Alex R. Hoffmaster ¹

¹ Bacterial Special Pathogens Branch, Division of High-Consequence Pathogens and Pathology, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30329-4018, USA

² Biotechnology Core Facility Branch, Division of Core Laboratory Services and Response, Office of Laboratory Systems and Response, Centers for Disease Control and Prevention, Atlanta, GA 30329-4018, USA

³ Texas A&M Veterinary Medical Diagnostic Laboratory, College Station, TX 77843-4471, USA

⁴ National Veterinary Services Laboratories, United States Department of Agriculture, Ames, IA 50010, USA

⁵ California Animal Health and Food Safety Laboratory, School of Veterinary Medicine, University of California-Davis, Davis, CA 95616, USA

⁶ Animal Disease Research and Diagnostics Laboratory, South Dakota State University, Brookings, SD 57007, USA

⁷ Public Health Laboratory, Minnesota Department of Health, St. Paul, MN 55155, USA

⁸ Oregon State Public Health Laboratory, Public Health Division, Oregon Health Authority, Hillsboro, OR 97124, USA

⁹ College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO 80523, USA

* Correspondence: cdk5@cdc.gov

† Current address: Center for Veterinary Biologics, United States Department of Agriculture, Ames, IA 50010, USA.

Abstract

Bacillus anthracis is endemic in the United States causing periodic outbreaks in wildlife and domestic animals. Currently, human anthrax cases in the U.S. are rare but were common in the 1950s–1960s due to industrial work with imported *B. anthracis*-contaminated animal products. Multiple-locus variable-number tandem repeat analysis (MLVA) initially differentiated *B. anthracis* into 89 genotypes and two major clades. Recently, whole-genome sequencing (WGS) was implemented to differentiate *B. anthracis* which allows for higher resolution and can eliminate risk of homoplasy. To assess the molecular diversity of U.S.-established isolates, we performed MLVA and WGS on 81 *B. anthracis* isolates from domestic animals or soil. By MLVA, most isolates ($n = 58$, 72%) were in the Western North America (WNA)/A1.a cluster. Isolates were also observed in the Ames (A3.b), Vollum (A4), and Group B clusters. Using WGS, two major clades (A and B) and four clusters (WNA, Ames, Vollum, Group B) were identified. The four WGS clusters correlated with previously established MLVA clusters (A1a, A3b, A4, and B1, respectively). Further differentiation of the WNA cluster showed that isolates collected from the same state generally clustered together and more broadly by region (west, central, Texas). In the current study, we provide an update on the genetic diversity of domestically established *B. anthracis* strains using MLVA and WGS. WGS was able to provide additional differentiation, particularly within the WNA cluster, which can lend assistance in epidemiological investigations.



Academic Editor: Eleonora Cella

Received: 23 March 2026

Revised: 15 May 2026

Accepted: 20 May 2026

Published: 23 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

Keywords: *B. anthracis*; MLVA; genotyping; whole-genome sequencing; SNPs

1. Introduction

Bacillus anthracis, the causative agent of anthrax, is a spore-forming, Gram-positive bacterium. *B. anthracis* that subsists primarily in the soil, infecting mainly wild and domesticated herbivores and, occasionally, humans [1]. Presently, *B. anthracis* can be isolated worldwide with outbreaks occurring periodically on most continents [2].

The first reported United States (U.S) anthrax case was in 1834 in Louisiana, with subsequent epizootic transmission events in various states by the end of the century [3]. While earlier historical accounts suggested introduction of anthrax into the U.S. during the French settlement of the Mississippi Delta in the mid-1700s [3], later genomic studies proposed competing hypotheses for the origin of the predominant Western North American (WNA) lineage, including a pre-Columbian introduction from Asia via Northern North America [4,5] and, alternatively, a more recent post-Columbian, trade-mediated introduction from Western Europe, with France proposed as a likely source [6,7].

With the introduction of a vaccination program for livestock in the late 1800s to early 1900s, the frequency of outbreaks was reduced, although certain regions of the U.S. (e.g., California, Louisiana, North and South Dakota, Nebraska and Texas) continued to have activity throughout the 20th century [3,8,9]. Many human cases of anthrax in the 1950s and 1960s were attributed to the handling and processing of animal products containing *B. anthracis* spores in textile mills and other industrial settings [10]. Furthermore, *B. anthracis* became a cause for concern as a biowarfare agent during the Cold War era and was used as a bioterror agent in the 2001 postal mailings incident in the U.S. [11,12].

Molecular subtyping methods allow for the ability to distinguish bacterial isolates of the same species from one another. While *B. anthracis* has been difficult to differentiate due to its monomorphic nature, Keim et al. developed multiple-locus variable-number tandem repeat analysis (MLVA) which differentiated *B. anthracis* into 89 genotypes and two major clades (A and B) [13]. While different MLVA schemes have been developed over time and an additional clade has been reported (clade C), for the past twenty years, MLVA has been used consistently to genotype *B. anthracis* [5,14–18]. The CDC performed MLVA-8 on all isolates received in the reference laboratory as part of a standard isolate workup from 2000 to 2023. MLVA genotypes of strains from the U.S. have been identified throughout the entire A clade and the B1 cluster [13]. However, it is difficult to ascertain which isolates were collected as a result of importation (e.g., from an imported animal product) or definitively originated in the U.S.

In order to obtain higher resolution genotyping, genome sequencing of isolates has become more frequent, and a variety of genome-wide differentiation methods have been applied to *B. anthracis* for different purposes, including evaluating the phylogenetic diversity of *B. anthracis* isolates in historical collections or domestically collected in other countries [19–22]. To date, there are limited reports analyzing the diversity of *B. anthracis* isolates originating from the U.S. using whole-genome sequencing. However, after the development of a canonical single-nucleotide polymorphism (canSNP) genotyping assay where 14 canonical SNPs were used for phylogenetic analysis identifying 12 canSNP lineages or subgroups, it was reported that the majority of strains from North America belong to the Western North America (WNA) lineage [4]. Limited reports have been published using canSNP and other SNP-based assays to confirm that the Ames strain and Ames-like strains persist in Texas [5,23].

Unlike bioterrorism events where extensive analysis would be performed to determine precise geographic origin and engineering detection, generally, public health response often begins with rapidly grouping a new isolate with its nearest neighbors and to determine if the isolate has been previously observed. Precise ancestral detection with genome sequences of *B. anthracis* uses well-established read-mapping methods (e.g., NASP [24] and SPANDx [25]) where nucleotide identities at various positions in reference to the Ames strain define certain lineages. Currently, no software exists to label these genomic samples with lineage names, leaving users to do manual searches which can be error prone. Read-mapping approaches can provide high-resolution phylogenetic placement [21,22], whereas assembly-based comparison methods offer a practical and rapid approach for initial grouping of closely related isolates [24]. We therefore used an assembly-based method to quickly differentiate the U.S. isolates and then interpreted the resulting clusters in the context of established *B. anthracis* lineages. The lineage names of *B. anthracis* are not well-known to the majority of public health officials and therefore add confusion, not clarification, to responses. Additionally, excluding all the strains that help assign lineage names gives faster computational and response times.

To expand and update the knowledge of isolates currently circulating in the U.S., we performed MLVA and whole-genome sequencing on *B. anthracis* isolates from animals infected in the U.S. or soil samples collected following a domestic outbreak. A collection of isolates archived at the Centers for Disease Control and Prevention (CDC) from various sources (animal and environmental) and geographic origins (i.e., U.S. states) collected over the last four decades was studied. We also used an assembly-based method for analysis to quickly differentiate the isolates.

2. Results

2.1. MLVA Genotyping

MLVA genotyping was performed on all 81 isolates (Table 1). A genotype (GT) was assigned to 75 of the isolates. Six isolates (one from Nevada and five from Texas) lacked one or both *B. anthracis* virulence plasmids and, thus, a genotype could not be assigned. However, when only the chromosomal loci were used for clustering, previously reported MLVA clusters could be determined for the isolates that were not assigned a GT [13]. Nine different GTs were observed among the 75 for which a GT could be assigned. For consistency, clades and clusters were named using well-known, previously published designations. By MLVA, the isolates separated into two clades (A and B) with clade A containing three clusters (WNA, Ames, and Vollum) and clade B containing one cluster (Group B). Most isolates ($n = 58$, 72%) were in the WNA (A1.a) cluster. Six isolates from Texas were in the Ames (A3.b) cluster (GTs 62, 90). Eight isolates were in the Vollum (A4) cluster (all GT 78), one isolate originated from Oklahoma, with the remaining isolates coming from Texas. In addition, three isolates from California and five soil isolates from Texas were in the Group B cluster. No isolates included in the study were found to be in the C clade.

Table 1. *Bacillus anthracis* isolates ($n = 81$) collected in the United States from 1976 to 2014.

State	No. of Isolates	Source	MLVA-8 Cluster; GT(s) ¹	WGS Cluster	CanSNP Sub-Lineage ⁴
California	18	Cow	WNA ³ (A1.a), Vollum (A4), Group B (B1); 3, 83	WNA, Vollum, Group B	A.Br.WNA, A.Br.Vollum, B.Br.001/002
South Dakota	11	Cow, Sheep	WNA; 3	WNA	A.Br.WNA

Table 1. Cont.

State	No. of Isolates	Source	MLVA-8 Cluster; GT(s) ¹	WGS Cluster	CanSNP Sub-Lineage ⁴
Minnesota	8	Cow, Horse	WNA; 3	WNA	A.Br.WNA
Texas	24	Sheep, Deer, Cow, Aoudad (barbary sheep), Goat, Horse, Soil	WNA, Ames, Vollum, Group B; 6, 62, 78, 90, NG ²	WNA, Ames, Vollum, Group B	A.Br.WNA, A.Br.Ames, A.Br.Vollum, B.Br.001/002
Montana	5	Cow, Bison	WNA; 3, 4	WNA	A.Br.WNA
Nevada	5	Cow	WNA; 3, 160, NG	WNA	A.Br.WNA
Colorado	3	Cow	WNA; 3	WNA	A.Br.WNA
New Mexico	2	Cow	WNA; 3	WNA	A.Br.WNA
Oregon	2	Cow	WNA; 3	WNA	A.Br.WNA
Mississippi	1	Cow	WNA; 1	WNA	A.Br.WNA
Nebraska	1	Cow	WNA; 3	WNA	A.Br.WNA
Oklahoma	1	Sheep	Vollum; 78	Vollum	A.Br.Vollum

¹ GTs 1-83 as described by Keim et al. [13]. ² NG, not genotyped. MLVA-8 genotype not assigned because one or both virulence plasmids (pXO1 and pXO2) were not detected. ³ WNA, Western North America. ⁴ CanSNP lineages as described by Van Ert et al. [5].

The most common GT among the isolates was GT 3 ($n = 48$) resulting in 59% of the isolates in the study being indistinguishable by this MLVA method. The state with the most diversity was Texas with isolates observed in all four MLVA clusters (Table 1). One isolate from Mississippi was GT 1 which is still within the WNA cluster. Two isolates from Nevada were assigned GT 160, which is a newly identified GT within the WNA cluster.

2.2. Initial Genome Clustering

To place these newly sequenced isolate genomes into a global phylogenetic context, we first confirmed all genomes' group well within established *B. anthracis* canSNP lineages (Figure 1). Mean silhouette coefficients (silhouette scores) were used to initially identify the optimal number of clusters for each genome distance method. Both Parsnp methods (with ML and MP phylogenies) indicated the *B. anthracis* datasets were best separated into two major clades (A and B) with four clusters (WNA, Ames, Vollum, and Group B) of isolates (Figure 2). The WGS clusters were equivalent to previously established MLVA clusters A1.a, A3.b, A4, and B1, respectively as well as canSNP lineages.

2.3. Genome Diversity

The majority of the isolates occurred in the WNA cluster ($n = 59$) while the Ames cluster contained the fewest ($n = 6$) (Figures 1 and 2A,B). The Vollum and Group B clusters contained nine and eight isolates each (respectively). The WNA cluster was the most numerous and widespread geographically (11 states) representing 72% of all U.S. isolates. The Ames cluster ($n = 6$) contained only two MLVA GTs with all isolates originating from a single state (Texas). The Vollum cluster included two MLVA genotypes from three different states. The Group B cluster had only one MLVA GT, although several isolates could not be genotyped due to lacking virulence plasmids. The Group B cluster isolates were from two states (California and Texas) (Figure S1).

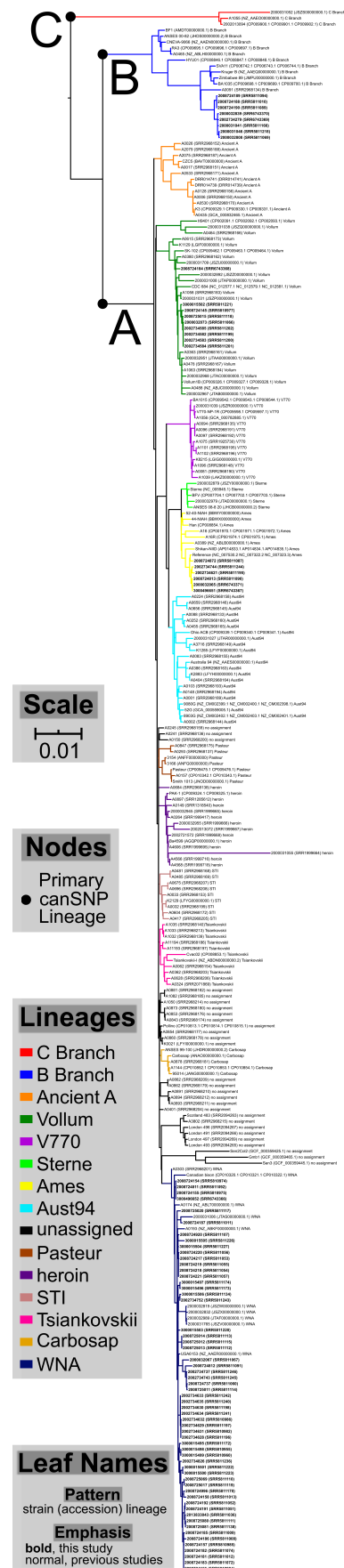


Figure 1. Global phylogenetic tree of *B. anthracis* isolate genomes included in this study (indicated in bold) and additional global *B. anthracis* isolates. Branch colors correspond with traditional canSNP lineages. Scale bar, substitutions per core SNP site.

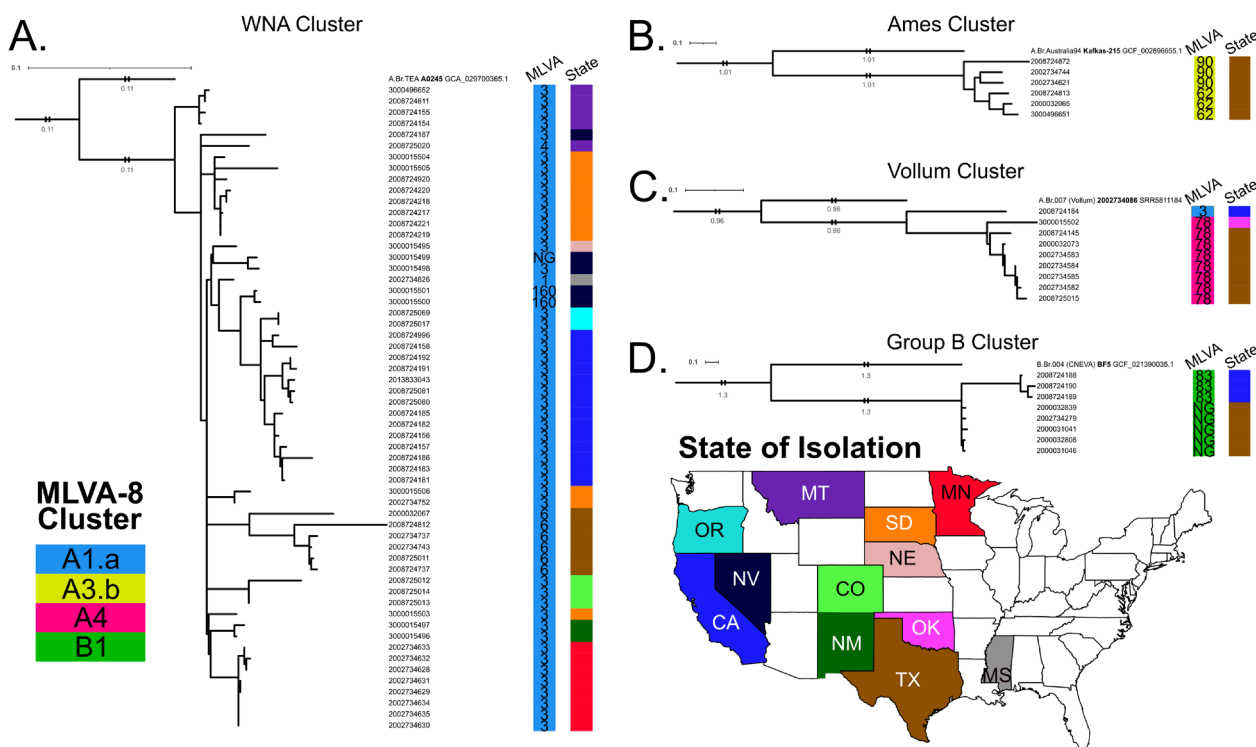


Figure 2. Phylogenetic tree based on whole-genome sequencing (WGS) and corresponding MLVA results. Alignment fractions to its largest genome size in each sample set were (A) 76%, (B) 91%, (C) 89%, and (D) 86%. WGS clusters are denoted by separate figure panels (A–D). MLVA genotypes are color-coded beside tree leaves as well as state origin of isolation. Scale bar, substitutions per core SNP site. Select branches have been artificially pruned denoted with double lines through the branch and lengths overlaid below. Tree roots of nearest neighbors for each panel followed strain (in bold) references described in [7].

Overall, our genome clustering results agreed well with the previously reported MLVA clades A and B. With one exception, individual MLVA GTs were only observed in a single WGS cluster (Figure 2). MLVA GT 3 was represented in two clusters; most isolates were in the WNA cluster with one isolate from California in the Vollum cluster. As with genotyping by MLVA, Texas harbored the most diverse *B. anthracis* strains with representatives in all four major genome clusters by WGS. Isolates from California occurred in three different clusters (WNA, Ames, Vollum). Isolates from the midwestern states were exclusively found in the WNA cluster.

2.4. WNA Analysis

The WNA cluster contained the majority of the isolates and was observed in all states except Oklahoma (Figure 2A). Since MP is a faster algorithm than the ML inference, and because they had identical isolates clustered in each group (Supplemental Figure S2), we used only the Parsnp MP method. Isolates from eleven states were represented in the WNA lineage and were at least 30 SNPs apart from each other (Figure 3). The initial SNP analysis identified just 4,049,258 bp as core, and the subsequent focused analysis of only the WNA lineage expanded SNP comparisons across 4,179,019 bp of their core genome sizes. Such high resolution enabled isolate grouping according to geographic region (west, central, and Texas), and for nine of 11 states, isolates from the same state grouped with each other. The two exceptional states were South Dakota and Nevada.

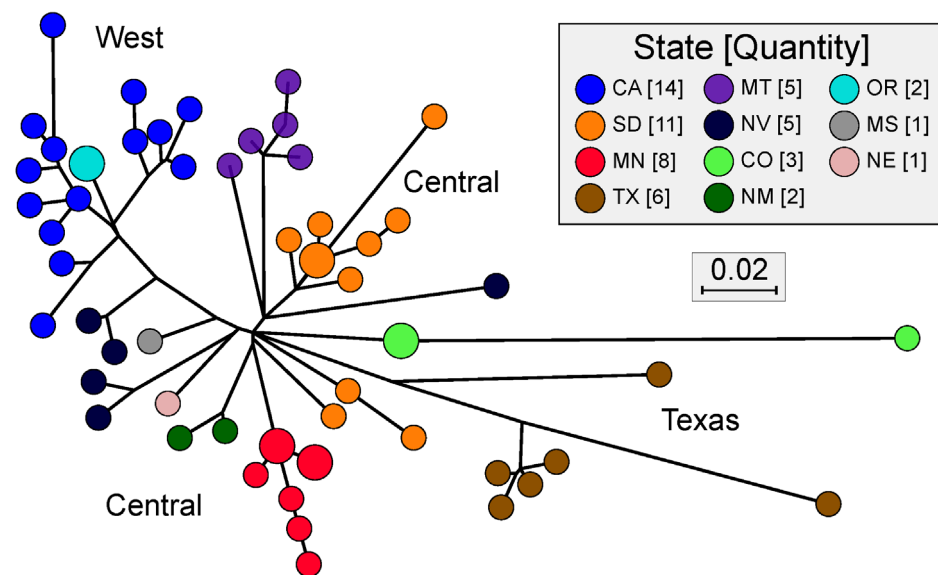


Figure 3. Minimum spanning tree of the isolates from the WNA lineage. Alignment fraction of this subset was 76% of the largest genome reference (5.439 Mbp in isolate 2002734737). Isolates are colored according to state origin of isolation with colors corresponding to Figure 1. Isolates with no distance apart are shown with an enlarged circle size proportional to quantity. Scale bar, substitutions per core SNP site.

3. Discussion

In the current study, we performed MLVA and WGS to ascertain the genetic diversity of *B. anthracis* isolates collected in the U.S. over a 38-year period. To obtain a true representation of geography linked with the genetic diversity of *B. anthracis* in the U.S., we included strains from animals (with limited geographic movement) that were infected in the U.S., thus eliminating importation as an issue for determining geographic source. Additionally, the five soil isolates included were recovered from soil samples collected following an outbreak in cattle from Texas. Using MLVA, we observed nine genotypes in four clusters among the 81 strains included. In 2000, a study of 32 isolates from the U.S. reported 16 different MLVA GTs, but source information was not reported so it is unclear whether all GTs are established in the U.S. [13]. Kenefic et al. described two isolates from anthrax outbreaks in Texas in 2006 and 2007 reporting that these isolates were the Ames genotype and that this genotype continued to persist in the state [23], a result we reiterate in the current study. Sue et al. also reported on MLVA GTs of *B. anthracis* isolates collected in the U.S. with the majority of these isolates coming from environmental sources [17]. Many of these isolates were recovered from swabs or other samplings of industrial sites where contaminated imported animal products were processed and were not domestically established strains.

While MLVA provides some insight into the genetic diversity of U.S. *B. anthracis* isolates, WGS allows for increased resolution while still correlating well with previous reports using MLVA. Like MLVA clustering, using WGS, the majority of the isolates (72%) are in a single cluster (WNA). Previous studies on *B. anthracis* from the U.S. corroborate this result by reporting the MLVA cluster A1.a and the Western North American (A.Br.WNA) canSNP lineage as most common in the U.S. [8]. Isolates in the Ames cluster contained isolates that were the same GT as the well-known Ames strain, which is in the A.Br.Ames canSNP lineage.

A minimum spanning tree (MST) is illustrated for fine-scale subclustering in Figure 3. Inadequacies in phylogenetic trees for molecular epidemiology [26] were avoided by using the parsimony phylogeny in GrapeTree. Polytomies occur in *B. anthracis* phylogenies due to

its limited genetic diversity and are also a typical feature of MST-based typing approaches. While not inherently problematic for public health response, such unresolved branching patterns can limit the resolution of transmission inferences when distinguishing among closely related isolates. Unlike MLVA, WGS allowed for further differentiation within the WNA cluster (Figure 3). Figure 3 firmly establishes state-level resolution is possible with WGS whereas with MLVA such resolution is not possible (i.e., not differentiable). Clustering by region and furthermore by states may allow for identification of source locations when a patient does not have a clear exposure or has potential exposures in multiple states. For neighboring states, some isolates group together (although they remain distinguishable from one another) such as California and Oregon. Initial automated clustering performs well and allows for researchers to spend more time evaluating within lineage relatedness, but subclustering is less useful when fewer isolates are more easily manageable to visualize for interpretation. As more diverse genotypes are identified within a lineage, we anticipate subclustering might become a more valuable approach, but currently it only appears modestly useful in ruling out certain groups when attempting to place a new incoming isolate.

The clustering of isolates by MLVA and WGS agreed with one exception. Although isolate 2008724184 from California was originally assigned MLVA GT3, WGS placed it within the Vollum lineage, where it also formed a distinct branch in the broader genomic comparison (Figure 1). We therefore consider this an unresolved discordance between the MLVA result and WGS-based phylogenetic placement, rather than evidence that GT3 isolates generally occur in both lineages. However, inspection of the electropherogram data showed no evidence of mixed-allele signals in the MLVA. To confirm these results, the genomic extraction, MLVA, and WGS were repeated producing the same results. This isolate showed single variants of alleles, suggestive of an axenic (pure) culture).

Most isolates included in the study were collected in Texas and California ($n = 42$, 52%). The isolates collected within these two U.S. states exhibited the most diversity. Isolates from California were observed in both MLVA clades and three WGS clusters. Isolates from Texas were even more diverse with isolates in all clades and clusters. Interestingly, while this study reiterates previous reports of clade B isolates in California [5,8], we also report on isolates from Texas in clade B. These Texas soil isolates were collected as part of an outbreak investigation in Foard and Cottle Counties in 1976 [27]. No additional outbreaks or cases have been reported in this region of Texas, and whether the spores associated with this outbreak persist in the soil, along with the potential threat to public health, remain unknown. WGS analysis determined these isolates were in the Group B cluster with the California “B” strains but split the strains into two distinct groups by state (Figure 2D). Broader WGS analysis of these California and Texas “B” isolates with global strains determined that they are within the B1 cluster (Supplemental Figure S2) [14,28].

Although clade A isolates have been found worldwide and the geographic origin of clade C isolates is unknown due to lack of strain information, in general, clade B isolates are rare in the U.S. and worldwide having been observed previously in parts of Africa, Europe, and more recently, in Bhutan and the Caucasus region [5,8,29,30]. Sue et al. [17] also reported a clade B environmental isolate from Alaska, which is a geographically distant state and not connected to the contiguous U.S. However, this isolate’s original geographic source is questionable since it was recovered from a swab at an industrial site. While the limited number of sequenced isolates in this clade make it difficult to ascertain whether the two geographically distant groups of “B” isolates in this study were introduced separately into the U.S., this is the most likely scenario given their geographic separation. Additionally, outbreaks associated with clade B strains could have been caused by imported

products and not locally, environmentally established strains since they do not seem to be recurring. An epidemiologic investigation of the 1976 Texas outbreak was undertaken by CDC to determine the source of infection. The two counties in which the outbreak occurred did not have any previously reported anthrax cases although some anecdotal evidence may exist of outbreaks occurring in 1934 and 1941. Based on limited sampling and interviews with ranchers, investigators concluded that since no probable source was discovered and the affected counties were “in the midst of an endemic area,” the outbreak was most likely caused by spores already present in the soil and not from an outside source (unpublished internal CDC report). Since molecular characterization was not available at the time, the investigators were unaware that these strains were very distantly related to known genotypes typically observed in Texas and, thus, would not have known to continue with the investigation and additional sampling. Therefore, the possibility of importation cannot be completely ruled out. To our knowledge, no studies have been performed to determine the source of the clade B strains from California.

Previous studies of *B. anthracis* from the U.S. similarly indicate that most domestically established isolates belong to the Western North America (WNA; A.Br.WNA) lineage, which is the predominant North American lineage. Current phylogeographic evidence indicates that A.Br.WNA is a lineage that became established and diversified in North America following its introduction in the 16th or 17th century [5,7]. By contrast, A.Br.Ames does not appear to be of North American origin. Instead, it is derived from the A.Br.001/002–A.Br.081(Ames) complex of likely East Asian/Inner Asian origin, with the strongest support pointing to China, particularly Inner Mongolia and surrounding regions [5,7,31]. Lineages detected among U.S. isolates, such as A.Br.Ames, are better interpreted as introduced Old World lineages rather than lineages that emerged in North America.

The animal isolates in this study were collected in the Central and Western U.S. from the 1980s to 2010s. In the past, anthrax outbreaks among livestock were reported on the east coast of the U.S. [9]. While the factors resulting in the gradual reduction in outbreaks in the Eastern U.S. remains unknown, two factors that most likely played a role are effective vaccination programs and urbanization of former rural areas [8]. The isolate collected the furthest east in the current study was the lone strain collected in Mississippi. This strain was collected from a county in Mississippi that borders the river and, thus, would not generally be considered isolated from the Eastern U.S. The geographic origins of the isolates in the current study continue to reiterate the notion that *B. anthracis* circulates in previously noted “anthrax districts” in the U.S. [9] as the majority of isolates were collected in California, South Dakota, and Texas.

In the current study, we provide and update the genetic diversity of domestically established *B. anthracis* isolates by comparing WGS to a previously established MLVA method. The use of WGS allowed for increased resolution and correlation of strains to their state or region of origin. These molecular epidemiology tools and analysis can be very useful in public health investigations in attempting to identify sources of infection with unknown exposures or with multiple possible exposures in multiple states. The current study is limited by the number isolates included with many collected during the same year within a U.S. state. Increasing the number of geographically and temporally diverse isolates that are available for WGS is likely to increase our knowledge of genetic diversity of *B. anthracis* in the U.S. However, federal regulations overseeing the use, possession, and transfer of select agents, such as *B. anthracis*, and the increased cost of shipping select agents make it difficult to obtain isolates from other laboratories within the U.S.

4. Materials and Methods

4.1. *B. anthracis* Isolates

Seventy-six isolates were collected from 1982 to 2014 from 12 U.S. states and seven different animal species [cow (*Bos taurus*), bison (*Bison bison*), goat (*Capra aegagrus hircus*), deer (*Odocoileus virginianus*), horse (*Equus caballus*), aoudad (*Ammotragus lervia*), and sheep (*Ovis aries*)] (Table 1). Five soil isolates associated with a 1976 anthrax outbreak in Texas cattle were also included [27]. Two strains from Texas (2002734621 and 2002734744) were previously described by Kenefic et al. [23]. One strain from Colorado (2008725012) was previously described by Yang et al. [32].

4.2. Extraction of Genomic DNA

Cultures were grown overnight on trypticase soy agar with 5% sheep blood (TSA II, Becton Dickinson and Co., Ltd., Sparks, MD, USA). DNA was extracted using one of two methods: (1) QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA, USA) or (2) Maxwell 16 Instrument (Promega, Madison, WI, USA). For the QIAamp extraction, one 10 µL loopful of cells from overnight growth on TSA II was inoculated in heart infusion broth (Remel, Lenexa, KS, USA) for 18–24 h. Cells were pelleted by centrifugation for 10 min at 5000 × *g*. After removal of broth, DNA was extracted from the remaining cells using the Qiagen QIAampDNA Blood Mini Kit following the manufacturer's protocol for isolating Gram-positive bacteria. For extractions on the Maxwell instrument, four to five colonies of overnight growth from TSA II were mechanically disrupted by vortexing for 2 min in a suspension of silica beads and TE buffer. The suspension was centrifuged for 30 s at 10,000 × *g*. In total, 300 µL of the resulting supernatant was used for DNA extraction following the manufacturer's protocol for blood and cells. All manipulation of cultures and subsequent extractions were performed in a BSL3 laboratory.

4.3. MLVA

MLVA was performed as previously described on an ABI 3130xL or ABI 3500xL sequencer (Applied Biosystems, Waltham, MA, USA) [13,17]. Briefly, eight loci were amplified: six chromosomal (*vrnA*, *vrnB1*, *vrnB2*, *vrnC1*, *vrnC2*, and CG3) and two plasmid-associated loci (pXO1-aat and pXO2-at). Fragment analysis was performed using GeneMapper v5.0 (Applied Biosystems, Waltham, MA, USA). Genotypes were assigned based on previously established fragment sizes [13,17]. UPGMA clustering analysis was performed using BioNumerics v7.6 (Applied Maths, Austin, TX, USA).

4.4. Whole-Genome Sequencing

Paired-end genomic DNA libraries were generated with NEBNext Ultra DNA kits (New England BioLabs Inc., Ipswich, MA, USA) and sequenced with either an Illumina MiSeq or Illumina HiSeq (Illumina Inc., San Diego, CA, USA). Sequence data were archived and accessions for each isolate are listed in Supplemental Table S1.

4.5. Genome Analyses

Illumina reads were cleaned with BBDOUK 35.92 to remove PhiX and Trimmomatic 0.35 to clip off adapters and discard low fidelity (<Q30) basecalls [33]. Cleaned reads at least 50 bp in length were used to generate genome assemblies with SPAdes 3.13.0 [34], and contigs with at least 5x coverage and 1 kbp in length were used for subsequent analyses.

For assembly-based SNP calling, we used ParSNP 1.3 [35] and VCFtools 0.1.17 [36] with the Ames ancestor chromosome (NC_007530.2) and plasmids (NC_007322.2 and NC_007323.3) as a reference [37]. FastTreeMP 2.1.3 [38] and RAxML 8.2.11 [39] were used to construct maximum parsimony (MP) and maximum likelihood (ML) trees. Distance

trees were visualized in iTOL 5.5.1 [40], FigTree 1.4.3, and GrapeTree 0.0.8 [41] and edited in Inkscape 0.92.4 [42].

Tree comparisons were performed with the dendextend 1.1.2 library in R [43]. Phylogenetic trees were imported into dendropy 4.2.0 to use phylogenetic distances among samples for clustering. K-means clustering was performed with scikit-learn 0.22.2 [44,45]. Individual silhouette coefficients describing the Euclidean distance between each sample from the closest sample outside of the cluster were calculated by comparing the average distance to a sample's nearest cluster (separation) and its average distance within its cluster (cohesion) with the silhouette_score function in scikit-learn's metrics class. The mean of silhouette coefficients was used as a silhouette score for each k-value evaluated. The optimal k-value (cluster quantity) was selected according to the smallest k-value with a local maximum silhouette score (mean) determined with NumPy 1.16.1 and SciPy 1.3.2 [46].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jgbg1020009/s1>, Figure S1: State origin with whole-genome sequencing cluster attribution (A) throughout the United States, by county in the state of California (B), and by county in the state of Texas (C). WGS, whole-genome sequencing. WNA, Western North America. Figure S2: Comparison of maximum parsimony (MP) and maximum likelihood (ML) phylogenies for SNPs identified in Parsnp.; Table S1: List of isolates included in this study. Table S2: Variant call formatted (VCF) loci for each tree illustrated in this study. Each worksheet tab is organized for each tree figure.

Author Contributions: Conceptualization: C.K.M. and A.R.H.; Methodology: C.K.M., C.A.G. and A.R.H.; Formal analysis: C.K.M. and C.A.G.; Investigation: C.K.M., C.A.B., M.S., M.B. and C.A.G.; Resources: C.K.M., A.K.S., M.M.E., K.A.C., R.M.P., M.M.S., M.E.M., K.L.P. and A.R.H.; Writing—original draft preparation: C.K.M., C.A.G. and A.R.H.; Writing—review and editing: C.K.M., C.A.G., C.A.B., M.S., M.B., A.K.S., M.M.E., K.A.C., R.M.P., M.M.S., M.E.M., K.L.P., Z.P.W. and A.R.H.; Visualization: C.K.M., Z.P.W., C.A.G. and A.R.H.; Supervision: Z.P.W. and A.R.H.; Project administration: C.K.M. and A.R.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA). Accession numbers are listed in Supplemental Table S1.

Acknowledgments: We thank Michael Frace (Centers for Disease Control and Prevention) for his contributions in the beginning of this study. Additionally, we thank the following for their efforts in contributing isolates for this study: Richard Walker (University of California-Davis), Stephen Stokke (Montana Department of Livestock), Daphne Ware (Mississippi Public Health Laboratory), Texas A&M Veterinary Medical Diagnostic Laboratory, National Veterinary Services Laboratories, University of California-Davis, South Dakota State University, Minnesota Department of Health, Colorado State University, and Oregon State Public Health Laboratory. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC. The mention of company names or products does not constitute endorsement by the CDC.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MLVA	Multiple-locus variable-number tandem repeat analysis
SNP	Single-nucleotide polymorphism
WGS	Whole-genome sequencing
ANI	Average nucleotide identity
WNA	Western North America
canSNP	Canonical single-nucleotide polymorphism

References

1. Beyer, W.; Turnbull, P.C. Anthrax in animals. *Mol. Asp. Med.* **2009**, *30*, 481–489. [[CrossRef](#)] [[PubMed](#)]
2. Carlson, C.J.; Kracalik, I.T.; Ross, N.; Alexander, K.A.; Hugh-Jones, M.E.; Fegan, M.; Elkin, B.T.; Epp, T.; Shury, T.K.; Zhang, W.; et al. The global distribution of *Bacillus anthracis* and associated anthrax risk to humans, livestock and wildlife. *Nat. Microbiol.* **2019**, *4*, 1337–1343. [[CrossRef](#)] [[PubMed](#)]
3. Klemm, D.M.; Klemm, W.R. A history of anthrax. *J. Am. Vet. Med. Assoc.* **1959**, *135*, 458–462. [[PubMed](#)]
4. Kenefic, L.J.; Pearson, T.; Okinaka, R.T.; Schupp, J.M.; Wagner, D.M.; Hoffmaster, A.R.; Trim, C.B.; Chung, W.K.; Beaudry, J.A.; Jiang, L.; et al. Pre-Columbian origins for North American anthrax. *PLoS ONE* **2009**, *4*, e4813. [[CrossRef](#)] [[PubMed](#)]
5. Van Ert, M.N.; Easterday, W.R.; Huynh, L.Y.; Okinaka, R.T.; Hugh-Jones, M.E.; Ravel, J.; Zanecki, S.R.; Pearson, T.; Simonson, T.S.; U'Ren, J.M.; et al. Global genetic population structure of *Bacillus anthracis*. *PLoS ONE* **2007**, *2*, e461. [[CrossRef](#)] [[PubMed](#)]
6. Vergnaud, G.; Girault, G.; Thierry, S.; Pourcel, C.; Madani, N.; Blouin, Y. Comparison of French and Worldwide *Bacillus anthracis* Strains Favors a Recent, Post-Columbian Origin of the Predominant North-American Clade. *PLoS ONE* **2016**, *11*, e0146216. [[CrossRef](#)] [[PubMed](#)]
7. Vergnaud, G.; Antwerpen, M.H.; Grass, G. *Bacillus anthracis* Phylogeography: Origin of the East Asian Polytoymy and Impact of International Trade for Its near Global Dispersal. *Pathogens* **2025**, *14*, 1041. [[CrossRef](#)] [[PubMed](#)]
8. Kenefic, L.J.; Okinaka, R.T.; Keim, P. Population Genetics of *Bacillus*: Phylogeography of Anthrax in North America. In *Bacterial Population Genetics in Infectious Disease*; Wiley Online Library: Online, 2010; pp. 167–180.
9. Stein, C.D. The history and distribution of anthrax in livestock in the United States. *Vet. Med.* **1945**, *40*, 340–349.
10. Brachman, P.S.; Fekety, F.R. Industrial anthrax. *Ann. N. Y. Acad. Sci.* **1958**, *70*, 574–584. [[CrossRef](#)] [[PubMed](#)]
11. Jernigan, D.B.; Raghunathan, P.L.; Bell, B.P.; Brechner, R.; Bresnitz, E.A.; Butler, J.C.; Cetron, M.; Cohen, M.; Doyle, T.; Fischer, M.; et al. Investigation of bioterrorism-related anthrax, United States, 2001: Epidemiologic findings. *Emerg. Infect. Dis.* **2002**, *8*, 1019–1028. [[CrossRef](#)] [[PubMed](#)]
12. Rasko, D.A.; Worsham, P.L.; Abshire, T.G.; Stanley, S.T.; Bannan, J.D.; Wilson, M.R.; Langham, R.J.; Decker, R.S.; Jiang, L.; Read, T.D.; et al. *Bacillus anthracis* comparative genome analysis in support of the Amerithrax investigation. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5027–5032. [[CrossRef](#)] [[PubMed](#)]
13. Keim, P.; Price, L.B.; Klevytska, A.M.; Smith, K.L.; Schupp, J.M.; Okinaka, R.; Jackson, P.J.; Hugh-Jones, M.E. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J. Bacteriol.* **2000**, *182*, 2928–2936. [[CrossRef](#)] [[PubMed](#)]
14. Lienemann, T.; Beyer, W.; Pelkola, K.; Rossow, H.; Rehn, A.; Antwerpen, M.; Grass, G. Genotyping and phylogenetic placement of *Bacillus anthracis* isolates from Finland, a country with rare anthrax cases. *BMC Microbiol.* **2018**, *18*, 102. [[CrossRef](#)] [[PubMed](#)]
15. Lista, F.; Faggioni, G.; Valjevac, S.; Ciammaruconi, A.; Vaissaire, J.; le Doujet, C.; Gorge, O.; De Santis, R.; Carattoli, A.; Ciervo, A.; et al. Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. *BMC Microbiol.* **2006**, *6*, 33. [[CrossRef](#)] [[PubMed](#)]
16. Rume, F.I.; Ahsan, C.R.; Biswas, P.K.; Yasmin, M.; Braun, P.; Walter, M.C.; Antwerpen, M.; Grass, G.; Hanczaruk, M. Unexpected genomic relationships between *Bacillus anthracis* strains from Bangladesh and Central Europe. *Infect. Genet. Evol.* **2016**, *45*, 66–74. [[CrossRef](#)] [[PubMed](#)]
17. Sue, D.; Marston, C.K.; Hoffmaster, A.R.; Wilkins, P.P. Genetic diversity in a *Bacillus anthracis* historical collection (1954 to 1988). *J. Clin. Microbiol.* **2007**, *45*, 1777–1782. [[CrossRef](#)] [[PubMed](#)]
18. Thierry, S.; Tourterel, C.; Le Fleche, P.; Derzelle, S.; Dekhil, N.; Mendy, C.; Colaneri, C.; Vergnaud, G.; Madani, N. Genotyping of French *Bacillus anthracis* strains based on 31-loci multi locus VNTR analysis: Epidemiology, marker evaluation, and update of the internet genotype database. *PLoS ONE* **2014**, *9*, e95131. [[CrossRef](#)] [[PubMed](#)]
19. Derzelle, S.; Girault, G.; Kokotovic, B.; Angen, O. Whole Genome-Sequencing and Phylogenetic Analysis of a Historical Collection of *Bacillus anthracis* Strains from Danish Cattle. *PLoS ONE* **2015**, *10*, e0134699. [[CrossRef](#)] [[PubMed](#)]

20. Okutani, A.; Inoue, S.; Morikawa, S. Comparative genomics and phylogenetic analysis of *Bacillus anthracis* strains isolated from domestic animals in Japan. *Infect. Genet. Evol.* **2019**, *71*, 128–139. [[CrossRef](#)] [[PubMed](#)]
21. Rondinone, V.; Serrecchia, L.; Parisi, A.; Fasanella, A.; Manzulli, V.; Cipolletta, D.; Galante, D. Genetic characterization of *Bacillus anthracis* strains circulating in Italy from 1972 to 2018. *PLoS ONE* **2020**, *15*, e0227875. [[CrossRef](#)] [[PubMed](#)]
22. Zhang, H.; Zhang, E.; He, J.; Li, W.; Wei, J. Genetic characteristics of *Bacillus anthracis* isolated from northwestern China from 1990 to 2016. *PLoS Neglected Trop. Dis.* **2018**, *12*, e0006908. [[CrossRef](#)] [[PubMed](#)]
23. Kenefic, L.J.; Pearson, T.; Okinaka, R.T.; Chung, W.K.; Max, T.; Trim, C.P.; Beaudry, J.A.; Schupp, J.M.; Van Ert, M.N.; Marston, C.K.; et al. Texas isolates closely related to *Bacillus anthracis* Ames. *Emerg. Infect. Dis.* **2008**, *14*, 1494–1496. [[CrossRef](#)] [[PubMed](#)]
24. Sahl, J.W.; Lemmer, D.; Travis, J.; Schupp, J.M.; Gillece, J.D.; Aziz, M.; Driebe, E.M.; Drees, K.P.; Hicks, N.D.; Williamson, C.H.D.; et al. NASP: An accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. *Microb. Genom.* **2016**, *2*, e000074. [[CrossRef](#)] [[PubMed](#)]
25. Sarovich, D.S.; Price, E.P. SPANDx: A genomics pipeline for comparative analysis of large haploid whole genome re-sequencing datasets. *BMC Res. Notes* **2014**, *7*, 618. [[CrossRef](#)] [[PubMed](#)]
26. Teixeira, A.S.; Monteiro, P.T.; Carriço, J.A.; Ramirez, M.; Francisco, A.P. Not Seeing the Forest for the Trees: Size of the Minimum Spanning Trees (MSTs) Forest and Branch Significance in MST-Based Phylogenetic Analysis. *PLoS ONE* **2015**, *10*, e0119315. [[CrossRef](#)] [[PubMed](#)]
27. CDC. Animal Anthrax-Texas. *MMWR Morb. Mortal. Wkly. Rep.* **1976**, *25*, 293.
28. Lekota, K.E.; Hassim, A.; Rogers, P.; Dekker, E.H.; Last, R.; de Klerk-Lorist, L.; van Heerden, H. The reporting of a *Bacillus anthracis* B-clade strain in South Africa after more than 20 years. *BMC Res. Notes* **2018**, *11*, 264. [[CrossRef](#)] [[PubMed](#)]
29. Eremenko, E.I.; Riazanava, A.G.; Tsygankova, O.I.; Tsygankova, E.A.; Buravtseva, N.P.; Kulichenko, A.N. Genotype diversity of the *Bacillus anthracis* strains isolated from the Caucasus region. *Mol. Genet. Mikrobiol. I Virusol.* **2012**, *27*, 74–78. [[CrossRef](#)]
30. Thapa, N.K.; Tenzin; Wangdi, K.; Dorji, T.; Migma; Dorjee, J.; Marston, C.K.; Hoffmaster, A.R. Investigation and control of anthrax outbreak at the human-animal interface, Bhutan, 2010. *Emerg. Infect. Dis.* **2014**, *20*, 1524–1526. [[CrossRef](#)] [[PubMed](#)]
31. Simonson, T.S.; Okinaka, R.T.; Wang, B.; Easterday, W.R.; Huynh, L.; U'Ren, J.M.; Dukerich, M.; Zanecki, S.R.; Kenefic, L.J.; Beaudry, J.; et al. *Bacillus anthracis* in China and its relationship to worldwide lineages. *BMC Microbiol.* **2009**, *9*, 71. [[CrossRef](#)] [[PubMed](#)]
32. Yang, A.; Mullins, J.C.; Van Ert, M.; Bowen, R.A.; Hadfield, T.L.; Blackburn, J.K. Predicting the Geographic Distribution of the *Bacillus anthracis* A1.a/Western North American Sub-Lineage for the Continental United States: New Outbreaks, New Genotypes, and New Climate Data. *Am. J. Trop. Med. Hyg.* **2020**, *102*, 392–402. [[CrossRef](#)] [[PubMed](#)]
33. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
34. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)] [[PubMed](#)]
35. Treangen, T.J.; Ondov, B.D.; Koren, S.; Phillippy, A.M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **2014**, *15*, 524. [[CrossRef](#)] [[PubMed](#)]
36. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)] [[PubMed](#)]
37. Ravel, J.; Jiang, L.; Stanley, S.T.; Wilson, M.R.; Decker, R.S.; Read, T.D.; Worsham, P.; Keim, P.S.; Salzberg, S.L.; Fraser-Liggett, C.M.; et al. The complete genome sequence of *Bacillus anthracis* Ames “Ancestor”. *J. Bacteriol.* **2009**, *191*, 445–446. [[CrossRef](#)] [[PubMed](#)]
38. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)] [[PubMed](#)]
39. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]
40. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **2007**, *23*, 127–128. [[CrossRef](#)] [[PubMed](#)]
41. Zhou, Z.; Alikhan, N.F.; Sergeant, M.J.; Luhmann, N.; Vaz, C.; Francisco, A.P.; Carrico, J.A.; Achtman, M. GrapeTree: Visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* **2018**, *28*, 1395–1404. [[CrossRef](#)] [[PubMed](#)]
42. Contributors, I. *Inkscape*, Version 0.91; Inkscape Project: Mountain View, CA, USA, 2015. Available online: <https://inkscape.org/release/0.91/> (accessed on 8 December 2017).
43. Galili, T. dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **2015**, *31*, 3718–3720. [[CrossRef](#)] [[PubMed](#)]

44. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; et al. API design for machine learning software: Experiences from the scikit-learn project. *arxiv* **2013**, arXiv:1309.0238.
45. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
46. Oliphant, T. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 10–20. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.