

Article

Investigating Anthrax-Associated Virulence Genes among Archival and Contemporary *Bacillus cereus* Group Genomes

Susanna J. Sabin ¹, Cari A. Beesley ², Chung K. Marston ², Taylor K. Paisie ³, Christopher A. Gulvik ², Gregory A. Sprenger ³, Jay E. Gee ², Rita M. Traxler ², Melissa E. Bell ², John R. McQuiston ² and Zachary P. Weiner ^{2,*}

¹ Laboratory Leadership Service Fellow Assigned to the National Center for Emerging and Zoonotic Infectious Diseases, CDC, Atlanta, GA 30329, USA

² Centers for Disease Control and Prevention, National Center for Emerging and Zoonotic Infectious Diseases, Division of High-Consequence Pathogens and Pathology, Bacterial Special Pathogens Branch, 1600 Clifton Rd, Atlanta, GA 30329, USA

³ Oak Ridge Institute for Science and Education, Oak Ridge, TN 37830, USA

* Correspondence: xxd7@cdc.gov

Abstract: *Bacillus anthracis* causes anthrax through virulence factors encoded on two plasmids. However, non-*B. anthracis* organisms within the closely related, environmentally ubiquitous *Bacillus cereus* group (BCG) may cause an anthrax-like disease in humans through the partial adoption of anthrax-associated virulence genes, challenging the definition of anthrax disease. To elucidate these phenomena and their evolutionary past, we performed whole-genome sequencing on non-*anthracis* BCG isolates, including 93 archival (1967–2003) and 5 contemporary isolates (2019–2023). We produced annotated genomic assemblies and performed a pan-genome analysis to identify evidence of virulence gene homology and virulence gene acquisition by linear inheritance or horizontal gene transfer. At least one anthrax-associated virulence gene was annotated in ten isolates. Most homologous sequences in archival isolates showed evidence of pseudogenization and subsequent gene loss. The presence or absence of accessory genes, including anthrax-associated virulence genes, aligned with the phylogenetic structure of the BCG core genome. These findings support the hypothesis that anthrax-associated virulence genes were inherited from a common ancestor in the BCG and were retained or lost across different lineages, and contribute to a growing body of work informing public health strategies related to anthrax surveillance and identification.

Keywords: anthrax; *Bacillus anthracis*; *Bacillus cereus* group; genomics; pseudogenes; gene decay



Citation: Sabin, S.J.; Beesley, C.A.; Marston, C.K.; Paisie, T.K.; Gulvik, C.A.; Sprenger, G.A.; Gee, J.E.; Traxler, R.M.; Bell, M.E.; McQuiston, J.R.; et al. Investigating Anthrax-Associated Virulence Genes among Archival and Contemporary *Bacillus cereus* Group Genomes. *Pathogens* **2024**, *13*, 884.

<https://doi.org/10.3390/pathogens13100884>

Academic Editor: Leonardo A. Sechi

Received: 16 August 2024

Revised: 26 September 2024

Accepted: 2 October 2024

Published: 10 October 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bacillus anthracis is the causative agent of anthrax, an acute disease in humans and non-human mammals. It is a globally distributed soil-dwelling bacterium that can persist in the environment for long periods through sporulation. Anthrax in humans is contracted via four primary exposure routes. Cutaneous and gastrointestinal anthrax results from contact with or the consumption of contaminated animal products and are the most common forms. The mortality rate without treatment can reach 30% for cutaneous anthrax and 74% for gastrointestinal anthrax without treatment [1]. Injection anthrax, linked to intravenous drug use, can lead to systemic disease, with its overall mortality rate reaching 25% or exceeding 90% if meningitis develops [1,2]. Inhalation anthrax has typically been associated with the processing or handling of hides or hair from contaminated animals, with a mortality rate close to 72% [1]. This is also the mode of transmission historically associated with the use of *B. anthracis* as a bioweapon, as demonstrated during the 2001 intentional distribution of *B. anthracis* spores through the U.S. Postal Service. Of the 22 resulting infections, 11 were inhalation anthrax, which resulted in five deaths [1,3,4]. It is thus classified by the United

States Departments of Agriculture and Health and Human Services as a Tier 1 select agent due to its potential to pose a severe threat to animal and human health [5].

The virulence of *B. anthracis* is largely due to two large plasmids, pXO1 and pXO2, which encode the genes for a tripartite toxin and antiphagocytic capsule. The pXO1 replicon contains genes encoding for the lethal factor (*lef*), the edema factor (*cya*), and the protective antigen (*pagA*), while pXO2 contains the genes which code for the characteristic poly-γ-D-glutamic acid (PGA) capsule (*capBCADE*). Molecular tests for *B. anthracis* generally rely on the detection of these virulence factors, and PCR methods amplify small segments of DNA found on pXO1, pXO2, and the *B. anthracis* chromosome [6,7].

Taxonomically, *B. anthracis* is within a group of closely related, Gram-positive, low-GC-content, sporulating bacteria referred to here as the *Bacillus cereus* group (BCG) (also known as *Bacillus cereus sensu lato*) [8,9]. Bacteria within this group, such as *B. cereus sensu stricto* and *B. thuringiensis*, tend to be environmentally ubiquitous and opportunistically cause disease in humans. *B. cereus* is typically associated with diarrheal and emetic food poisoning, with variation in virulence across different strains, though it has emerged as a source of hospital-acquired infections [10]. *B. thuringiensis* is most known as an insect pathogen and used as a bioinsecticide [8]. However, in the genomics era, numerous other species have been introduced to the BCG and, simultaneously, the lines between BCG species have been increasingly blurred.

Close genomic relatedness and phenotypic variability within the BCG have challenged traditional taxonomic classifications, including delineations between *B. anthracis* and the rest of the BCG [11,12]. BCG taxonomy was further complicated by the discovery of new anthrax-causing bacteria that killed wild great apes in the Côte d'Ivoire and Cameroon [13,14]. *Bacillus cereus* biovar *anthracis* differed from traditional *B. anthracis* in phenotype (motility and gamma phage susceptibility) and had chromosomal DNA more like *B. cereus* isolates [15,16]. However, it harbored two virulence plasmids, pBCXO1 and pBCXO2, that showed high similarity to the pXO1 and pXO2 of *B. anthracis* [16,17]. Additional isolates, identified as *B. cereus* with *B. anthracis* toxin genes, were identified in metalworkers in Louisiana and Texas suffering from severe pneumonia, referred to as welder's anthrax [18–23]. These isolates expressed the pXO1 toxin virulence genes but expressed different capsules and lacked pXO2 homologs [18,24]. One additional cutaneous case caused by a similar atypical *Bacillus* bacterium [25] was not associated with metalworking or a laboratory-acquired infection [26]. As emerging data continuously challenge the lines between species in the BCG and the definition of anthrax disease, it is of critical importance in clinical and public health contexts to correctly classify these organisms and understand their potential pathogenicity [1,27–30]. Also important is understanding how and when anthrax-associated virulence genes may occur in different BCG lineages.

Here, we present whole-genome sequencing data from 93 non-*anthracis* BCG isolates archived at the Centers for Disease Control and Prevention in Atlanta (CDC), Georgia, USA (1967–2003) and 5 contemporary (2019–2023) non-*anthracis* BCG isolates from clinical cases sent to the CDC. The archival strains represent bacteria characterized as *B. cereus* or *B. thuringiensis* through traditional microbiological techniques isolated from clinical or environmental contexts not associated with food poisoning or food contamination, while the contemporary strains were referred to the Zoonoses and Select Agent Laboratory at the CDC for *B. anthracis* rule-out and identification. We leveraged this unique, time-structured dataset to investigate the occurrence of anthrax-associated virulence genes in the BCG and whether the genes were introduced to non-*anthracis* bacteria by horizontal gene transfer or linear inheritance through multiple methods of genomic inference. These findings contribute to our understanding of the evolution of anthrax-associated virulence in the BCG.

2. Methods

2.1. Origin and Handling of Archival Isolates

Since the 1960s, the Centers for Disease Control and Prevention has received clinical specimens and bacterial isolates from healthcare providers, state and local public health laboratories, universities, and governmental researchers. “Genus cards” for submitted isolates identified as *Bacillus cereus* and *Bacillus thuringiensis*, which recorded an isolate identification number, standardized biochemical findings, and the source of the isolate, were reviewed. Isolates not indicated as deriving from a food source or clinical source suggesting food poisoning were selected for inclusion in the archival isolate set for whole-genome sequencing. The isolates were stored at $-70\text{ }^{\circ}\text{C}$ as blood stocks.

2.2. Origin and Handling of Contemporary Isolates

The contemporary isolates (2019–2023) included in this project were submitted to the Centers for Disease Control and Prevention for *Bacillus anthracis* rule-out following the receipt of a positive result for 1 out of 3 targets in *B. anthracis* real-time PCR [6]. The contemporary isolates were stored at $-70\text{ }^{\circ}\text{C}$ as spore suspensions in deionized water containing 25% glycerol.

2.3. DNA Preparation

Subcultures of the archival isolate stocks were incubated overnight on trypticase soy agar with 5% sheep blood at $35\text{ }^{\circ}\text{C}$. Using a $10\text{ }\mu\text{L}$ loop, one to three loopfuls of cells were collected from each fresh plate and used to inoculate a 2 mL, screw-top tube of $\sim 350\text{ }\mu\text{L}$ of 0.1 mm Zirconia/Silica beads and $650\text{ }\mu\text{L}$ Tris-EDTA (0.01 M, pH 8.0) buffer. Each inoculum was vortexed horizontally at maximum speed for 2 min to perform cell lysis by bead beating. The tubes were then centrifuged at $10,000\times g$ for 30 s. Genomic DNA was extracted from the bead-beaten supernatant using the automated Promega Maxwell CSC instrument and Cultured Cells DNA kit (Promega, Madison, WI, USA). Following DNA extraction, the eluate was filter-inactivated using centrifugation ($8000\times g$ for 2 min) in Millipore Ultrafree 0.1 μM filter tubes (Millipore Sigma, Burlington, MA, USA) [31]. The filtered genomic DNA extract was then stored at $-20\text{ }^{\circ}\text{C}$ for real-time PCR (rtPCR) and genome sequencing.

2.4. Real-Time PCR

An rtPCR assay for *B. anthracis* was performed for all isolates, as reported elsewhere [6]. A set of three primer pairs and probes was used, targeting pXO1, pXO2, and the *B. anthracis* chromosome, which are intended to produce amplicons approximately 98, 137, and 96 base pairs in length, respectively. We used the PerfeCTa Multiplex qPCR SuperMix with Low Rox (QuantaBio, Beverly, MA, USA) and performed rtPCR on the filtered DNA extracts using an ABI 7500 Fast Dx (Applied BioSystems, Waltham, MA, USA) instrument with 40 amplification cycles. A plasmid containing all three gene targets was used as the positive control, and molecular-grade water was used for the negative, non-template controls.

2.5. Illumina Library Preparation and Sequencing

A subset of 39 extracts was prepared and sequenced at the CDC. A filtered genomic DNA extract was quantified using a Qubit (Invitrogen, Carlsbad, CA, USA) fluorometer, and $\sim 400\text{ ng}$ of DNA per isolate was prepared for sequencing using the Illumina DNA Prep kit (Illumina, San Diego, CA, USA) per the manufacturer’s instructions. The libraries were dual-indexed using the Illumina CD Indexes kit (Illumina, San Diego, CA, USA). Amplified, purified libraries were quantified on a Qubit (Invitrogen, Carlsbad, CA, USA). The contemporary isolates were sequenced on the iSeq 100 platform with iSeq 100 v2 reagent kits (150-cycles, paired-end) (Illumina, San Diego, CA, USA). Thirty-four of the archival isolates were sequenced on the MiSeq platform (Illumina, San Diego, CA, USA). Of those, 9 were sequenced using 300-cycle kits, and 25 were sequenced using 600-cycle kits. For the MiSeq, equimolar pools of 6–10 libraries were combined per sequencing run.

A subset of 59 filtered archival genomic DNA extracts were sent to the Translational Genomics Research Institute (TGen, Phoenix, AZ, USA) for sequencing. DNA was quantified using a Qubit (Invitrogen, Carlsbad, CA, USA) and each extract was normalized to 100 ng using the Integra ASSIST PLUS pipetting robot (Integra Biosciences Corp., Hudson, NH, USA). Libraries were prepared with Watchmaker's DNA Library Prep Kit with Fragmentation (Watchmaker Genomics, Boulder, CO, USA) according to the manufacturer's instructions with slight modifications to the resuspension volumes (42 μ L post-ligation and 32 μ L post-amplification). Truncated, Illumina-compatible adapters were used at a concentration of 15 μ M (Integrated DNA Technologies, Coralville, IA, USA), and unique dual-index primers were added from a single-use plate (Integrated DNA Technologies, Coralville, IA, USA). Amplified libraries were quantified using a Qubit (Invitrogen, Carlsbad, CA, USA) and quality was assessed using a D5000 ScreenTape on a TapeStation 4200 System (Agilent Technologies, Santa Clara, CA, USA). The libraries were assigned to pools based on approximate genome sizes and automated equimolar pooling of the final libraries occurred on the Integra ASSIST PLUS Platform. The pools were quantified with a Qubit (Invitrogen, Carlsbad, CA, USA) and fragment size was determined with a D5000 ScreenTape (Agilent Technologies, Santa Clara, CA, USA) regional analysis. Sequencing was performed on a NovaSeq X Plus system (Illumina, San Diego, CA, USA) using the NovaSeq X Series 10B Reagent Kit (300 cycles, paired-end) following standard Illumina protocols.

2.6. Genome Assembly

The following processing steps were applied to paired sequencing data from all isolates, contemporary and archival, through the nf-core [32] wf-paired-end-illumina-assembly workflow (v. 2.3.0) [33] written in Nextflow (domain specific language [DSL2] v. 22.04.3, Seqera, Barcelona, Spain) [34]. Raw sequencing reads were scrubbed of human DNA and broken read pairs were discarded [35–37]. The remaining reads were then downsampled to approximately 100 \times coverage to mitigate any inadvertent amplification of sequencing errors. Genome size, read count, and original sequencing depth were calculated using KMC (v. 3.2.2) [38] and Seqtk (v. 1.4-r122) [39]. Read subsampling was completed using Seqtk (v. 1.3-r106) [39]. Any remaining PhiX sequences were removed using a BBDuk (v. 38.94) [37] adapter and quality trimming was completed using Trimmomatic (v. 0.39) [40], and overlapping read pairs were merged using FLASH (v. 1.2.11) [41]. The cleaned reads were assembled de novo using SPAdes [42], and the resulting contigs were filtered and corrected using Biopython (v. 1.68, python v 2.7.18) [43], BWA (v. 0.7.17-r1188) [44], SAMtools (v. 1.9) [45], and Pilon (v. 1.23) [46]. QUAST (v. 5.0.2) [47] and CheckM2 (v. 1.0.1) [48] were used to generate assembly metrics. Depth of assembly coverage was calculated using BEDTools (v. 2.29.2) [49]. MLST genotyping was performed on each assembly using mlst (v. 2.23.0) [50,51] and assemblies were annotated using Prokka (v. 1.14.5) [52]. Sequences of 16S rRNA were extracted from each assembly using Biopython (v. 1.68, python v 2.7.18) [43] and Barrnap (v. 0.8) [53]. The extracted 16S rRNA gene sequences underwent naïve Bayesian classification using RDP Classifier [54] for genus-level detection. The 16S rRNA gene sequences were also aligned against a database of 16S gene type strain sequences (Bioprojects PRJNA33175 and PRJNA33317) using BLASTn+ (v. 2.15.0+) and the best bitscore alignment was selected [55]. In addition, cleaned reads were classified using both Kraken (v. 1.1.1) [56] and Kraken2 (v. 2.1.3) [57] to assess the sequences for potential contamination.

2.7. Whole-Genome Comparison with *Bacillus cereus* Group Type Strains

The isolate genome assemblies were compared with *Bacillus cereus* group (BCG) type strains spanning the diversity of the BCG (Supplementary Table S1) [58–78].

Bidirectional average nucleotide identity (ANIb) was conducted using the wf-ani workflow [79], written in Nextflow (domain specific language [DSL2] v. 22.04.3, Seqera, Barcelona, Spain) [34], and whole-genome digital DNA:DNA hybridization (dDDH) was performed on the Type (Strain) Genome Server (TYGS) [80]. Isolates which TYGS called as potential new species due to low dDDH values with any available type strain were classified according to the closest available type strain by the dDDH d_4 metric and ANIb.

The only discordance between TYGS identification and ANIb occurred for isolates with the highest similarity to the *B. thuringiensis* type strain. In these cases, TYGS identified the isolates as belonging to *B. cereus*, as *B. thuringiensis* could be interpreted as a sub-species within the *B. cereus sensu stricto* (s.s.) genomospecies clade [81]. The isolates were annotated as *B. cereus/thuringiensis*, in contrast to those of highest similarity to the *B. cereus* type strain, which were only annotated as *B. cereus*. The closest type strains and dDDH/ANIb statistics for each isolate are listed in Supplementary Table S2.

2.8. Annotation

Prokka (v. 1.14.5) [52] was used to annotate the genome assemblies separately from the assembly pipeline to include a custom database of translated virulence and regulator genes known to occur in virulent *B. anthracis* and the welder's anthrax isolate G9241 in addition to the 'Bacteria' database (see Supplementary Table S3). Prokka was run with an e-value threshold of 1×10^{-8} . The resulting GenBank files were screened for genes of interest in each isolate. The genes of interest coding for virulence factor proteins typically found in *B. anthracis* were *pagA* (protein sequence AJI08141.1, protective antigen), *lef* (protein sequence AJI08174.1, lethal factor), *cya* (protein sequence AJI08175.1, edema factor), *capA* (protein sequence WP_001253155.1, capsule polyglutamate synthetase), *capB* (protein sequence WP_003159943.1, capsule biosynthesis protein CapB), *capC* (protein sequence WP_000468007.1, capsular polyglutamate amide ligase/translocase PgsC), *capD* (protein sequence WP_071735531.1, capsule biosynthesis protein CapD), *capE* (protein sequence AAT28992.2, capsule biosynthesis protein CapE), *inhA* (protein sequences YP_017301 for *inhA1* and YP_017909 for *inhA2*, M6-family metalloprotease immune inhibitors InhA1 and InhA2), and *atxA* (protein sequences AJI08122.1 for *atxA1* and AJI07941.1 for *atxA2*, anthrax toxin expression trans-acting positive regulators AtxA1 and AtxA2). We also investigated the presence of genes found in welder's anthrax isolate G9241 thought to contribute to its virulence. These genes were the following: *certhrax* (protein sequence EAL15945.1, lethal factor precursor); *bpsX* (protein sequence EAL15985.1, transcriptional regulator, 2C LytR family); *bpsA* (protein sequence EAL15984.1, chain length determinant protein); *bpsB* (protein sequence EAL15983.1, exopolysaccharide biosynthesis protein); *bpsC* (protein sequence EAL15982.1, UTP-glucose-1-phosphate uridylyltransferase); *bpsD* (protein sequence EAL15981.1, glycosyl transferase, 2C putative); *bpsE* (protein sequence EAL15980.1, sialic acid synthase, 2C putative); *bspF* (protein sequence EAL15979.1, UDP-N-acetylglucosamine 2-epimerase, 2C putative); *bspG* (protein sequence EAL15978.1, CMP-sialic acid synthetase, 2C putative); *hasA* (protein sequence EAL12809.1, glycosyl transferase, 2C group 2 family protein domain protein); *hasB* (protein sequence EAL12810.1, UTP-glucose-1-phosphate uridylyltransferase); and *hasC* (protein sequence EAL12811.1, UDP-glucose/GDP-mannose dehydrogenase family, 2C NAD binding domain family). Lastly, we included virulence genes seen in more virulent strains of *B. cereus*, including *nheABC* (protein sequences YP_018530, YP_018531, WP_001172020; non-hemolytic enterotoxin), *cytK* (protein sequence NP_830896, cytotoxin K), and *hblABC* (protein sequences NP_832844, NP_832845, and NP_832847; hemolysin BL binding component precursor, Hemolysin BL lytic component L2).

2.9. Chromosome and Plasmid Alignments

To assess whether gene annotations corresponding to a known plasmid (e.g., pXO1, pXO2) were indicative of the presence of the entire plasmid replicon, we aligned the sequencing data to reference sequences for each plasmid of interest. The reference sequences included pXO1 from the Ames Ancestor *B. anthracis* (NC_007322.2), pBCXO1 from the welder's anthrax isolate G9241 (NZ_CP009592.1), pBCXO1 from the welder's anthrax isolate 033B87 (NZ_CP009940.1), pXO2 from the Ames Ancestor *B. anthracis* (NC_007323.3), and pBC218 from G9241 (NZ_CP009591.1). Cleaned sequence reads were aligned to the reference sequences using bowtie2 (v. 2.5.1) [82], and the alignments were sorted with samtools (v. 1.18) [45]. VCF files were generated from the alignments using bcftools

(v. 1.10.2) [83]. The depth of coverage and nucleotide identity were plotted across the reference sequence using matplotlib (v. 3.5.3) [84] in Python (v. 3.9.13).

2.10. Pan-Genome Characterization

Prior to using the pan-genome tool, we generated a Mash-distance-based tree of the isolate and type strain genome assemblies (see “Type strain similarity assessment” above) using Mashtree (v. 1.4.6) [85]. A group of type strains were clustered separately from any of the archival or contemporary isolates in the Mash distance tree and consisted of the following: *B. clarus*, *B. cytotoxicus*, *B. gaemokensis*, *B. bingmayonensis*, *B. arachidis*, *B. pseudomycooides*, *B. manliponensis*, and *B. rhizoplanae*. They were excluded from further analyses in addition to one archival isolate, B2797, which was found to belong to the genus *Lysinibacillus*.

The GFF files generated by Prokka were used as the input for Roary (v. 3.11.2) [86], a pan-genome characterization tool. Roary was run using the -e and n non-default options to generate a multiple-FASTA core genome alignment using MAFFT [87].

2.11. Gene Annotation Inspection

Gene annotations were interrogated using the gene_presence_absence.csv output from Roary to explore protein clusters and cluster quality. Protein clusters were based on a minimum of 95% identity and a Markov cluster algorithm inflation value of 1.5 [86]. Additionally, amino acid reference sequences for each gene of interest were queried against the genome assemblies using tBLASTn [55] to provide amino acid identity metrics between the annotations and reference sequences. When required, annotated sequences were aligned with the gene reference and visually inspected in MEGA-X [88].

2.12. Core Genome Maximum Likelihood Phylogeny

Uniform sites were removed from the core genome alignment using SNP sites (v. 2.3.3) [89], and the resulting alignment was used to generate a maximum likelihood (ML) tree using RAxML-ng (v. 1.0.1-master) [90]. On the SNP (variable) sites, we used the general time-reversible gamma substitution model with the “Lewis” model of ascertainment bias control (“GTR + G + ASC_LEWIS”). The resulting binary file was used as the input for ML tree inference with bootstrapping (“bs-metric fmb, tbe”).

2.13. Comparison of ML and Gene Presence/Absence Tree

The optimal ML phylogeny from the core genome and the accessory gene presence/absence tree was visualized and explored in FigTree (v. 1.4.4) [91] and compared with a tanglegram. The tanglegram was generated in R using the *ape* (v. 5.7-1) [92] and *Dendextend* (v. 1.17.1) [93] packages. Both trees were rooted with the *B. paramycooides* type strain. The correlation between the dendrograms was evaluated with two parameters, as described previously [94]: Baker’s Gamma association index [95] and cophenetic distance [96].

2.14. Clock-like Signal Test for BCG

The core genome ML phylogeny was supplied to TempEst (v. 1.5.3) [97] to determine whether the phylogeny had a clock-like signal. Each isolate was annotated with the year of isolation. The type strains were annotated with the year of isolation as noted on the associated NCBI BioSample entry or inferred from their BioSample registration dates or primary publications (see Supplementary Table S1). The “Best-fitting root” option was selected to optimize the tree.

2.15. Clock-like Signal Test in Genomespecies Partitions

In addition to testing for a clock-like signal in the full BCG ML phylogeny, we tested for a clock-like signal within a selection of BCG genomespecies that were well represented by our data. We constructed the core genome maximum likelihood (ML) phylogeny for the *B. cereus*, *sensu stricto* (*s.s.*) and *B. mosaicus* genomespecies and analyzed them using

TempEst (v. 1.5.3) [97]. Each isolate was annotated with its year of collection, and the “Best-fitting root” option was selected to optimize the phylogeny. Type strains were excluded from this analysis.

3. Results

3.1. The Sequenced Isolates Represented Numerous Species/Subgroups within the *Bacillus cereus* Group

The archival isolates were initially identified through traditional microbiological methods as either *B. cereus* or *B. thuringiensis* and were last classified between 1967 and 2003. Since 2003, there have been substantial increases in newly described species within the BCG, as well as proposed taxonomic changes [8,9,69,81,98]. Based on whole-genome comparison between the archival sequences and current BCG type strains using ANIb and dDDH metrics (Supplementary Table S2), we re-classified the isolates into 12 species. One isolate did not show high similarity to any BCG type strain and was classified as *Lysinibacillus capisici* in TYGS [80]. This isolate, B2797, was excluded from further analyses. The other 11 species were classified as *B. albus*, *B. anthracis*, *B. cereus*, *B. mobilis*, *B. mycooides*, *B. pacificus*, *B. paranthracis*, *B. pretiosus*, *B. cereus/thuringiensis*, *B. toyonensis*, and *B. tropicus* (Figure 1). Eleven archival isolates were noted as potential new species by TYGS due to their low d_4 values across all tested type strains. These isolates were classified according to the type strain which received the highest d_4 value and ANIb (Supplementary Table S2). The contemporary isolates were identified as *B. paranthracis* and *B. nitratireducens* (Figure 1). According to the 2020 proposed genomospecies classification scheme for the BCG [81], the isolates represent *B. cereus, s.s.*; *B. mosaicus*; *B. mycooides*; and *B. toyonensis*.

3.2. Pan-Genome Inference

Roary identified a total of 66,408 genes, 1647 of which were present in at least 99% of the genomes and represented the core genome of the dataset. There were 328 additional genes present in 95–99% of the genomes, 5728 in 15–95% of the genomes, and 58,705 genes in less than 15% of the genomes. The number of unique genes and the size of the pan-genome proceeded on a nearly linear positive trajectory as more genomes were added to the analysis, while the size of the core genome remained stable and close to the 1647 count from approximately the addition of the 28th genome (Supplementary Figure S1). These trends agree with previous findings suggesting the BCG has an “open” pan-genome [99].

3.3. Five Archival Isolates and Five Contemporary Isolates Received Annotations for at Least One Anthrax-Associated Virulence Gene

We screened all archival and contemporary isolate genome assemblies for a suite of virulence genes and virulence gene regulators important for pathogenesis in the BCG (see Section 2). Archival isolates B0818, B4510, E5429, and G9898 all received annotations for *pagA* from Prokka (Table 1, Figure 1). We explored the annotations more deeply by investigating the protein clusters identified by Roary as part of the pan-genome characterization and by aligning reference protein sequences to the isolate genomes with tBLASTn [55]. The *pagA* annotation for B0818 appeared to be spurious due to low amino acid identity between a protective antigen reference sequence (AJI08141.1) and the annotated region and being one of only two genes identified on a single, short contig (Supplementary Table S4). B4510 and E5429 had identical sequences annotated for *pagA* (Supplementary Table S4). However, these sequences were only 16% the expected length of the gene, with 380 bp compared to the expected 2294 bp (Table 1, e.g., NC_010934.1). G9898 had two annotations for *pagA*, one of which shared a protein cluster with the *B. anthracis* type strain and shared 100% amino acid identity with the protective antigen reference sequence. The second shared 61% identity. The two annotations were on separate contigs.

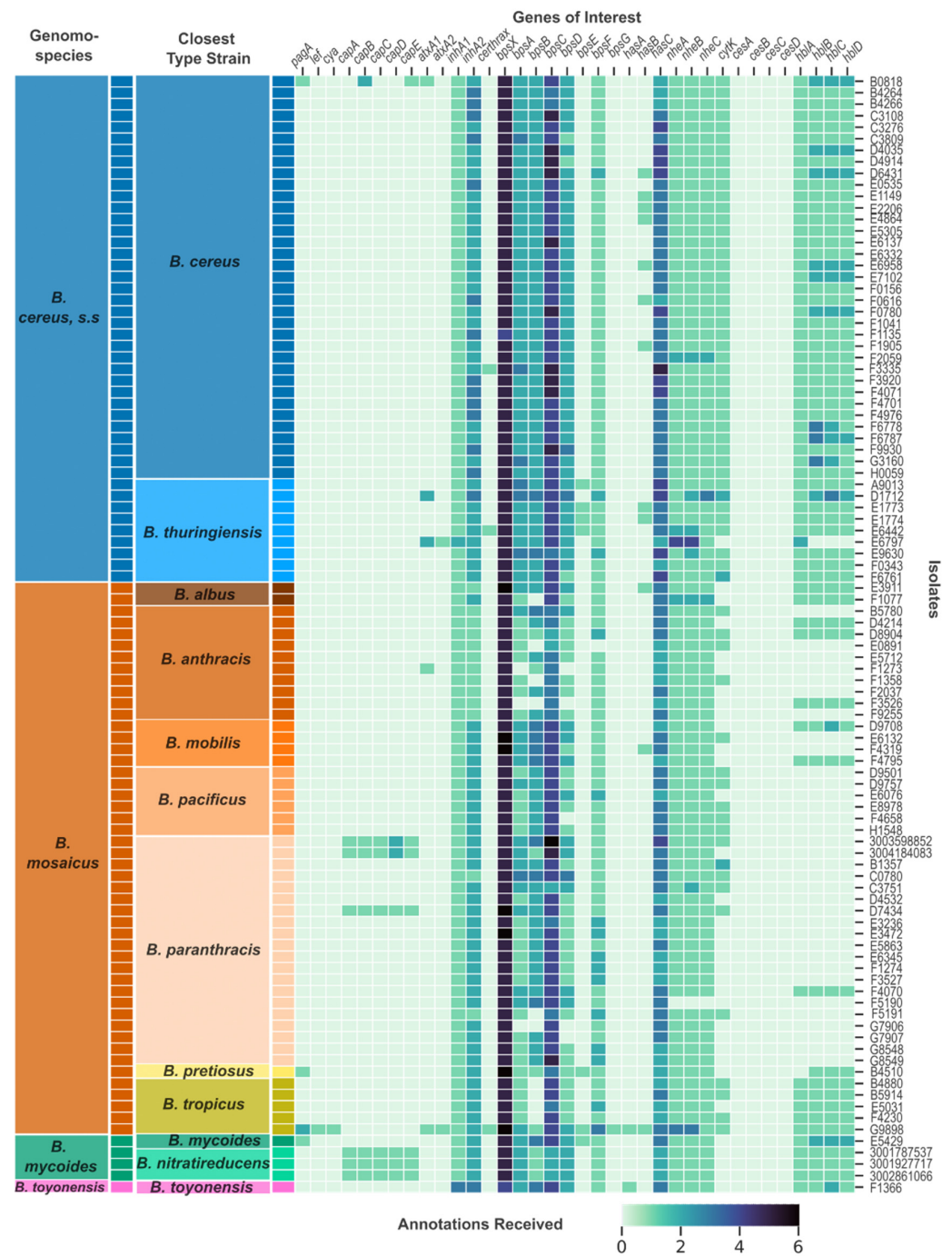


Figure 1. Presence and number of annotations received by virulence genes across the *Bacillus cereus* group. Each row represents a unique isolate genome, and each column represents a virulence-related gene from the *Bacillus cereus* group (BCG). The darkness of each cell represents the number of annotations that an isolate’s genomic assembly received for the corresponding gene from Prokka. The two columns on the left side of the heatmap represent the closest species type strain to that isolate as determined by ANI and dDDH (“Closest Type Strain” column) and the genomespecies to which that type strain belongs according to the 2020 genomespecies organization of the BCG (“Genomespecies” column). The genomespecies are colored differently and labeled in the “Genomespecies” column. The closest type strains are colored in different hues similar to their corresponding genomespecies color and labeled in the “Closest Type Strain” column.

Table 1. Isolates with annotations for pXO1 and pXO2 anthrax-associated virulence genes. Entries in the “Gene Length (bp)” column are bolded for annotations that were truncated compared to the reference sequence gene length.

	Isolate Name	Isolation Source	Isolation Location (Country: State)	Isolation Year	Closest Species Type Strain	Gene	Amino Acid Identity (%)	Gene Length (bp)	Reference Gene Length (bp)
Archival	B0818	Pustule	USA: NM	1968	<i>B. cereus</i>	<i>pagA</i> *	32.35	1628	2294
						<i>capB_1</i>	80.33	389	1395
						<i>capB_2</i>	73.81	623	1395
						<i>capE</i>	52.4	140	143
	B4510	Ulcer	USA: MA	1970	<i>B. pretiosus</i>	<i>pagA</i>	34.71	380	2294
	D7434	Blood	USA: NY	1976	<i>B. paranthracis</i>	<i>capB</i>	90.52	1130	1395
						<i>capC</i>	84.5	449	449
						<i>capA</i>	60.26	1199	1235
						<i>capD</i>	57.17	1615	1587
	E5429	Blood	USA: CO	1979	<i>B. mycoides</i>	<i>pagA</i>	34.71	380	2294
	G9898	Sputum	USA: LA	1996	<i>B. tropicus</i> †	<i>pagA_1</i>	100	2294	2294
						<i>pagA_2</i>	61.31	2282	2294
						<i>lef</i>	100	2429	2429
						<i>cya</i>	100	2402	2402
Contemporary	3001787537	sputum	USA: IL	2019	<i>B. nitratireducens</i>	<i>capB</i>	89.88	1181	1395
						<i>capC</i>	91.95	449	449
						<i>capA</i>	87.08	1211	1235
						<i>capD</i>	82.52	1586	1587
	3001927717	urine	USA: ND	2020	<i>B. nitratireducens</i>	<i>capE</i>	82.6	143	143
						<i>capB</i>	89.88	1181	1395
						<i>capC</i>	91.95	449	449
						<i>capA</i>	87.08	1211	1235
	3002861066	urine	USA: CA	2022	<i>B. nitratireducens</i>	<i>capD</i>	82.52	1586	1587
						<i>capE</i>	82.6	143	143
						<i>capB</i>	89.88	1190	1395
						<i>capC</i>	91.95	449	449
	3003598852	blood	USA: TN	2022	<i>B. paranthracis</i>	<i>capA</i>	87.86	1255	1235
						<i>capD_1</i>	79.69	1043	1587
<i>capD_2</i>						83.3	431	1587	
<i>capE</i>						84.78	143	143	
3004184083	blood	USA: AR	2023	<i>B. paranthracis</i>	<i>capB</i>	90.15	1130	1395	
					<i>capC</i>	91.95	449	449	
					<i>capA</i>	87.86	1255	1235	
					<i>capD_1</i>	79.69	1043	1587	
<i>capD_2</i>	83.3	431	1587						
<i>capE</i>	84.78	143	143						

* The *pagA* annotation for B0818 was likely spurious; † G9898 was noted as a potential new species by TYGS due to its low dDDH values with all BCG type strains. However, G9898 had the highest similarity by ANI and dDDH values to the *B. tropicus* type strain.

G9898 was the only isolate in the dataset to receive annotations for *lef* and *cya*—the toxin genes that accompany *pagA* on the pXO1 plasmid. The G9898 annotated sequences

shared 100% identity with both the lethal factor (AJI08174.1) and edema factor reference sequences (AJI08175.1). This genome lacked the *capBACDE* operon but did receive annotations for two complete capsule operons (Figure 1). *bpsX-H*, which encodes an exopolysaccharide capsule, and *hasACB*, which encodes a hyaluronic acid capsule, are thought to be required for anthrax-like pathogenesis in the absence of the PGA capsule [24]. G9898 was the only isolate in this dataset with complete exopolysaccharide capsule and hyaluronic capsule operons (Figure 1).

Archival isolates B0818 and D7434 received two and five capsule gene annotations, respectively. B0818 received two sequential *capB* annotations (loci 05385 and 05386). Both annotated sequences shared relatively high amino acid identity with the CapB reference sequence (73.8% and 80.3%, respectively) (WP_003159943.1) but were shorter than the expected 1395 bp, with locus 05385 consisting of 389 bp and locus 05386 consisting of 623 bp (Table 1). Upon inspection of the B0818 sequences aligned with the reference sequence in MEGAX [87], it appeared that the full gene was interrupted by a stop codon. The *capE* annotation for B0818 had only 52% amino acid identity with the reference (WP_003159943.1) but was of appropriate length (140 bp compared to the 143 bp reference sequence) and was on the same contig as the *capB* annotations for this isolate (Table 1). Archival isolate D7434 received annotations for a complete capsule operon, *capBCADE*, as did all five contemporary isolates. All six isolates had *capB* sequences group together in a single protein cluster (Figure 2). D7434's sequences had their own cluster for *capCADE* (Figure 2). The sequences for *capA* and *capD* both fell to 60% and 57% amino acid identity with the reference sequences (Table 1). This was a deviation from the other *cap* annotations in D7434 and all *cap* annotations in the contemporary isolates, which ranged from 79 to 92% (Table 1). All contemporary isolates shared a single protein cluster for *capC* (Table 1). The contemporary *B. nitratireducens* and *B. paranthracis* isolates formed separate protein clusters for *capADE* (Figure 2). The contemporary *B. paranthracis* isolates had two *capD* annotations each, which formed their own clusters (Figure 2). Upon closer inspection, *capD* was split identically in both isolates. There was no evidence of a split in the archival *B. paranthracis*, which was isolated in 1968. There were no geographic patterns reflected in these data.

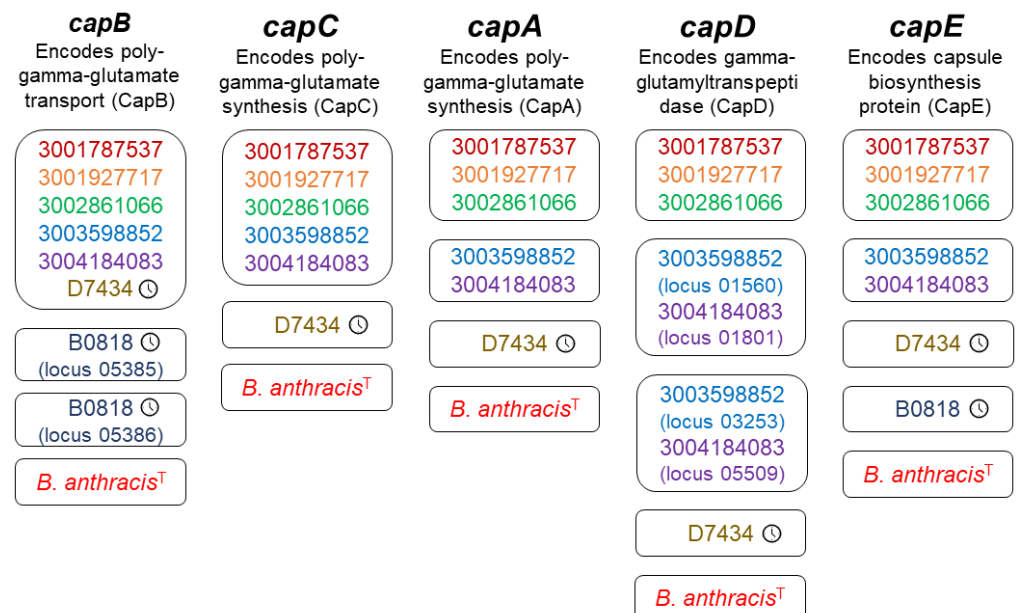


Figure 2. Protein clusters for each poly-y-D-glutamic acid capsule operon gene. Each box represents a protein cluster identified by Roary for each poly-y-D-glutamic acid (PGA) capsule operon gene, represented by each column of protein cluster boxes. Shared protein clusters indicate a minimum of 95% shared nucleotide identity in the annotated sequences. Each strain/isolate is represented by a

different color. The clock icon to the right of the genome ID indicates an archival isolate. The “T” superscript indicates the *B. anthracis* type strain. All other genomes are contemporary isolates. Isolates that received multiple annotations for the same gene which formed different protein clusters are annotated with the locus of the annotation in the genome assembly.

3.4. *atxA* and *Certhrax* Annotations

In addition to the anthrax-associated virulence genes listed above, we also screened the genomes for *atxA* and *Certhrax* genes. The transcriptional activator gene *atxA* regulates virulence expression in *B. anthracis* and has been identified in atypical BCG strains that have caused anthrax-like diseases. Two distinct copies of this gene have been found in atypical BCG organisms, with the second, *atxA2*, sharing 79% nucleotide identity with *atxA1* [30]. Five genomes received at least one *atxA* annotation from Prokka (Figure 1). Of these, G9898 and B0818 also received anthrax-associated virulence gene annotations. Upon investigation at the protein level, G9898 was the only genome to share >70% amino acid identity with the AtxA1 and AtxA2 reference sequences (AJI08122.1 and AJI07941.1), with 100% amino acid identity for each gene. The other genomes shared very low amino acid identity with the reference sequences, ranging from 23 to 33% (B0818: 25.5%, D1712: 26.5%, E6797: 24% amino acid identity with AJI08122.1; B0818: 24.6%, D1712: 24.7%, E6797: 23.3% amino acid identity with AJI07941.1). The *Certhrax* gene encodes a toxin first identified in the welder’s anthrax isolate G9241 that shares structural similarity with the anthrax lethal factor [30]. Three genomes received annotations for *Certhrax*, though again, G9898 was the only one to share substantial amino acid identity with the reference sequence (G9898: 100%, F3335: 32.5%, E6442: 34% amino acid identity with EAL15945.1).

3.5. Four Archival Isolates and All Contemporary Isolates That Received Virulence Gene Annotations Did Not Show Evidence of Complete *Bacillus Anthracis* Virulence Plasmids

We assessed whether there was evidence for complete or partial plasmid homologs in the isolates with virulence gene annotations by aligning the cleaned and trimmed sequencing data from each isolate to reference sequences for the plasmids of interest from *B. anthracis* and two previously characterized welder’s anthrax isolates, G9241 and 033B87. Despite full and partial virulence gene homologs being present in this dataset, only G9898 shows evidence of containing a complete *B. anthracis* plasmid homolog (Figure 2). G9898 further shows complete alignment to the pBC218 plasmid sequence from G9241 (Figure 3). The isolates containing the complete PGA capsule operon do not show evidence of having a pXO2 homolog plasmid (Figure 3). It is possible the genes are located on the chromosome or plasmid that differs dramatically from the Ames Ancestor pXO2.

3.6. Isolates with Virulence-Related Genes Were Distributed across the *B. cereus* Group

We utilized type strain and genomospecies classification (Figure 1) with core genome ML phylogeny (Figure 4) to investigate whether there was a phylogenetic pattern in the occurrence of anthrax-associated virulence genes. The five archival isolates that received annotations for anthrax-associated virulence genes came from five different species as determined by type strain comparisons across the *B. cereus*, *s.s.*; *B. mycooides*; and *B. mosaicus* genomospecies. Notably, B4510 and E5429, which had identical truncated *pagA* sequences, came from two different genomospecies (Figure 1). The contemporary isolates came from two type species, *B. nitratreducens* and *B. paranthracis*, across two genomospecies, *B. mycooides* and *B. mosaicus*. The PGA operon genotypes of the contemporary isolates clustered together according to taxon (Supplementary Table S4). The ML phylogeny indicates the *B. nitratreducens* isolates have nearly identical core genomes, while the contemporary *B. paranthracis* isolates with virulence genes were different and distributed among archival *B. paranthracis* isolates with and without virulence gene annotations (Figure 4).

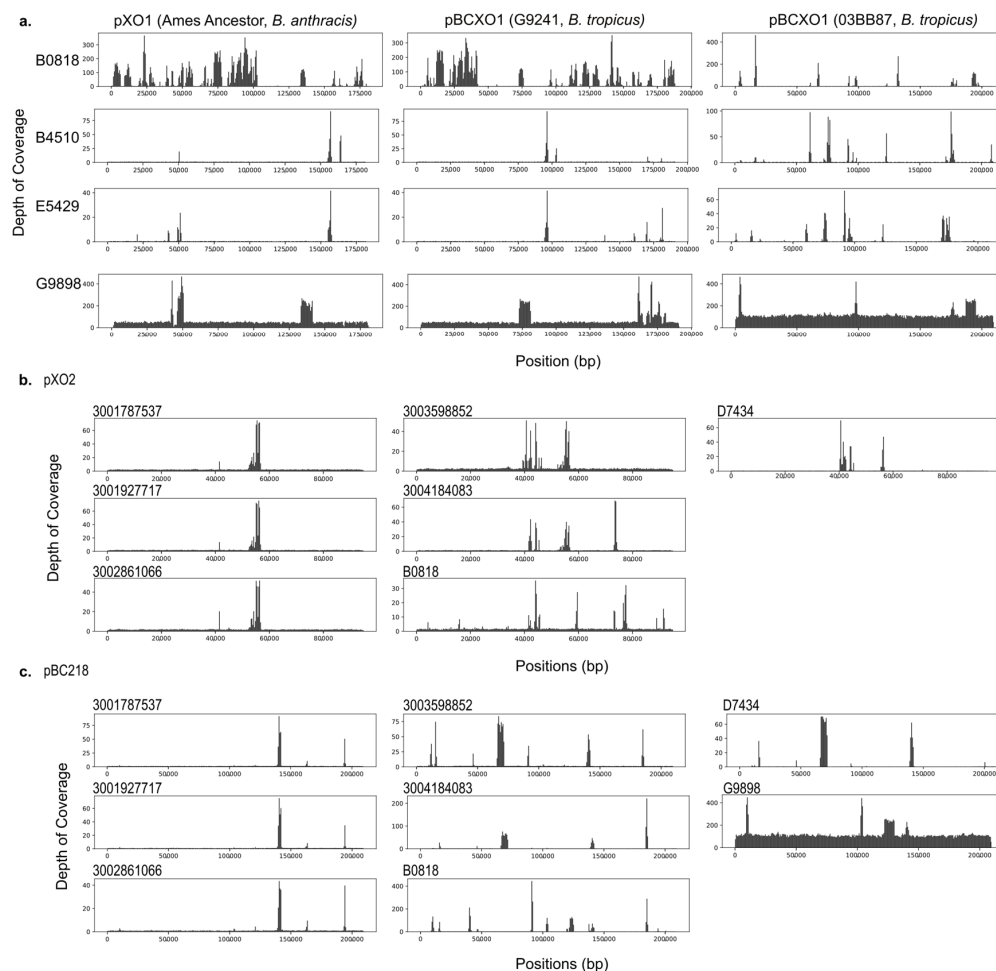


Figure 3. Depth of coverage for sequencing reads aligned to reference sequences. Cleaned sequence reads were aligned to reference sequences with bowtie2 to visualize the breadth (horizontal) and depth (vertical) of coverage for inferring replicon presence/absence. (a) pXO1 from *B. anthracis* Ames Ancestor, pBCXO1 from G9241, and pBCXO1 from 03BB87. Only G9898 (bottom row) indicates the presence of a complete pXO1 homologous plasmid. (b) pXO2 from *B. anthracis* Ames Ancestor. The contemporary *B. nitratireducens* (3001787537, 3001927717, and 3002861066), contemporary *B. paranthracis* (3003598852 and 3004184063), archival *B. cereus* genome B0818, and archival *B. paranthracis* genome D7434 did not indicate the presence of pXO2 despite receiving annotations for the PGA capsule operon genes. (c) pBC218 from G9241. In the G9241 welder’s anthrax BCG genome, the *bps* exopolysaccharide capsule genes are housed on the pBC218 plasmid. Again, G9898 was the only genome showing evidence of containing a pBC218-like plasmid.

3.7. The Core Genome and Accessory Gene Presence/Absence Tree Topologies Are Congruent

To determine whether the presence and absence of accessory genes across the dataset were correlated with the core genome phylogeny, we compared the binary accessory gene (presence/absence) tree topology with the core genome ML tree using a tanglegram (Figure 5). The core genome phylogeny is represented by 1647 core genes, which are shared by all isolates in the dataset. The accessory gene tree is based on the presence/absence of 64,761 accessory genes. In general, the trees had congruent topologies. There was no reorganization of leaves between the genomospecies, and notably there was no clustering of isolates in which virulence genes were detected within the accessory gene presence/absence tree. The most notable difference is the reorganization of the *B. luti* type strain between the two trees. This genomospecies is represented by a single type strain, and none of the isolates sequenced for this project clustered with *B. luti*. We quantified the strength of the correlation between the trees using Baker’s gamma index (also known as the Goodman–Kruskal–gamma index)

correlation [95] and the cophenetic Spearman correlation [96]. The Baker’s gamma correlation and the cophenetic Spearman correlation were both 0.96, indicating the presence or absence of accessory genes does not supersede the core genome phylogeny.

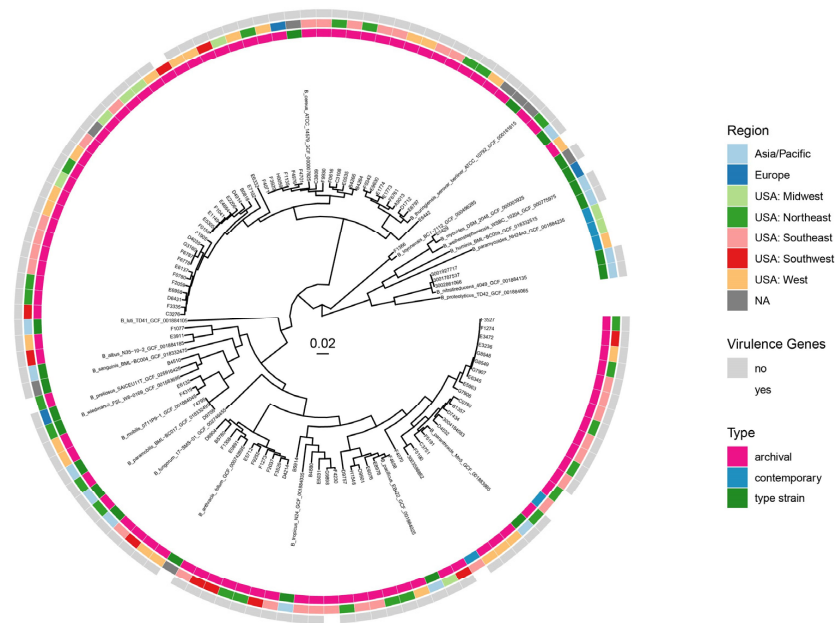


Figure 4. Core genome ML phylogeny. An ML phylogeny generated from the core genome alignment of 92 archival isolates, 5 contemporary isolates, and 21 BCG type strains. Three colored rings combine the genomic data with epidemiological data of each *Bacillus* isolate (n = 118). The tree is rooted with the *B. paramycooides* type strain (n = 345,224 total sites). The scale bar indicates substitutions per site. Bootstrap values were removed for ease of visualization. A tree with bootstrap values can be found in the Supplementary Information (Supplementary Figure S2).

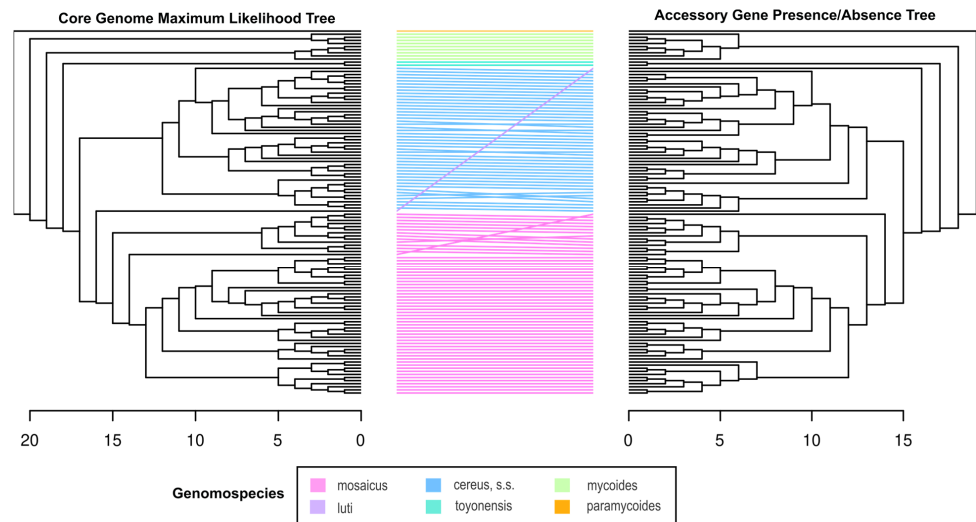


Figure 5. Tanglegram comparing the core genome phylogeny and accessory gene presence/absence tree. The colors of the lines connecting the tree leaves represent the genospecies classification of each clade according to the 2020 genospecies classification scheme for the *B. cereus* group [81] as guided by the core genome phylogeny and the distribution of type strains. The scale bars indicate the height of each tree (the largest number of edges between the tree leaves and the root). While there is entanglement within the genospecies clades, there is no entanglement between clades, indicating the accessory gene presence/absence tree largely recapitulates the topology of the core genome phylogeny.

3.8. There Is No Evidence of a Clock-like Signal in the Time-Structured *B. cereus* Group Dataset

Given the time-structured nature of the dataset, we tested whether there was a temporal signal in the core genome phylogeny by determining if there was a correlation between the phylogenetic distance from the estimated root of the tree to the leaves and ages of the leaves. Under the heuristic residual mean squared function and with the best fitting root, the R^2 (2.9×10^{-3}) and correlation coefficient (5.3×10^{-2}) were extremely small, indicating there was no correlation between the isolation date and the genomic distance from the root (Supplementary Figure S3).

3.9. There Is a Weak Clock-like Signal in the *B. cereus*, *Sensu Stricto* and *B. mosaicus* Genomospecies Partitions

Despite current taxonomic ambiguity, the BCG is made up of multiple species; we chose to also assess clock-like signal within the most well-represented genomospecies in our dataset according to the most recent proposed genome-based classification scheme from Carroll and colleagues [81]: *B. cereus*, *s.s.*, and *B. mosaicus* (Figure 1). Under the heuristic residual mean squared function and using the best fitting root, the *B. cereus*, *s.s.*, partition had an R^2 of 0.019 and correlation coefficient of 0.138 (Supplementary Figure S3). The *B. mosaicus* partition had a slightly stronger clock-like signal with an R^2 of 0.083 and correlation coefficient of 0.288 (Supplementary Figure S3). Though weak, these statistics indicate an association between genomic distance from the root and the isolation date in contrast to the whole BCG.

4. Discussion

Here, we produced 98 genomes from archival and contemporary isolates collected between 1967 and 2023, 97 of which were confirmed to be from the BCG. In 10 isolates, we identified DNA sequences homologous to virulence genes found in classically defined *B. anthracis*. Though these genomes were spread throughout the BCG phylogeny, the older isolates with homologous sequences showed evidence of pseudogenization and gene decay. Our pan-genome analysis indicated no meaningful sharing of accessory genes between lineages in the BCG. Though the temporal signals we identified in the BCG, *B. cereus*, *s.s.*, genomospecies, and *B. mosaicus* genomospecies were weak or absent, there were many factors aside from recombination that could have confounded the clock signal. Ultimately, we found no strong evidence of the virulence genes (intact or otherwise) being distributed throughout the BCG by gene transfer.

The most striking genome assessed in this study, G9898, was the only isolate analyzed that possessed anthrax toxin genes and full exopolysaccharide and hyaluronic acid capsule operons. This isolate was derived from a case of fatal pneumonia in a Louisiana welder in 1996 [100], and was previously shown to express a capsule [101]. As the only *B. tropicus* isolate in the dataset with anthrax-associated virulence genes, it fits into the growing set of welder's anthrax isolates identified as *B. tropicus* [18,23].

The *pagA* sequences identified in B4510/E5429 and the *capB* sequences identified in B0818 were likely non-functional due to their interruption by stop-codons. Though the *cap* operon annotated in D7434 was complete, the amino acid identity for *capACD* compared to the reference sequences was below 80%, suggesting the capsule may have differed phenotypically for this strain. Additionally, the *capD* sequences from the contemporary *B. paranthracis* isolates appear to be interrupted in a similar fashion to the *capB* sequence in B0818. These findings may indicate a larger pattern of anthrax-associated virulence gene pseudogenization.

Though plasmid transfer has been demonstrated between different lineages of the BCG in experimental and environmental contexts through transduction and conjugation [8], we did not see evidence in this dataset of natural, large-scale gene exchange between lineages. We only see evidence of a full pXO1 homolog plasmid in G9898, the welder's anthrax case. The remainder of the isolates, including the contemporary isolates encoding the full *cap* operon, did not show evidence of possessing complete homolog plasmids. The genes and

gene fragments identified here are either on the chromosome or a different plasmid. This could be determined definitively in the future through long-read sequencing. We furthermore found that while we do see anthrax-associated virulence genes spread throughout different lineages of the BCG, the presence and absence of accessory genes largely recapitulates the topology of the core genome phylogeny. This would suggest there is not frequent gene transfer between the BCG genomospecies lineages represented in the dataset such that they would meaningfully override the core genome phylogeny. This study recapitulates previous findings in support of the linear inheritance of anthrax-associated virulence genes rather than acquisition by gene transfer, particularly by plasmid transfer [94].

However, one line of evidence presented here indicates gene transfer occurs in the BCG in general. Despite the congruence between the core genome and accessory gene presence/absence trees, the pan-genome analysis did recapitulate previous findings suggesting that the BCG pan-genome is open [98] and members of the BCG are regularly acquiring new and unique genes (Supplementary Figure S1). The most likely source of this observed gene diversity is gene transfer [102,103].

The lack of a temporal signal in the whole BCG core genome SNP-based ML phylogeny with estimated isolation dates spanning 1872–2023 (Supplementary Tables S1 and S3) could arguably indicate gene transfer. However, we found weak temporal signals when the BCG was separated into genomospecies partitions for *B. cereus*, *s.s.*, and *B. mosaicus*. Recombination and hypermutation are the two nucleotide-level biological processes known to disrupt strict clock-like patterns in phylogenies due to variant mixing between lineages [97]. It is also possible the BCG life history and lifestyle of sporulation and intermittent dormancy contributes to the lack of a clock-like signal due to extended and/or irregular generation times [104]. Different populations within and between lineages may enter dormancy for varying amounts of time and reproduce at different rates, contributing to a seemingly stochastic phylogenetic signal.

Our findings regarding the apparent pseudogenization of anthrax-associated virulence gene homologs suggest they were not acquired by any lineage in which they are present in recent history. Rather, they may be an illustration of gene deletion in progress, which acts as a balance to the open pan-genome for maintaining a reasonable genome size [105]. When we consider the pseudogenization of the anthrax-associated virulence genes together with the lack of evidence for transfer between lineages by means of gene transfer, the most likely explanation for their presence throughout the BCG is linear inheritance from a common ancestor. While the genes featured here are crucial for the pathogenesis of anthrax disease, most BCG bacteria are not obligate pathogens and must survive and compete in diverse ecological niches [8,69,70]. It is plausible that most descendant lineages of this common ancestor, possessing the complete set of toxin and PGA capsule genes, diversified into niches where the virulence genes were not advantageous. While the virulence genes may not have been deleterious, they did not confer a selective advantage. Mutations or deletions may have accumulated in them through genetic drift, causing their pseudogenization and eventual loss in most of the BCG. In contrast to the rest of the BCG, the highly conserved pathogenic *B. anthracis* depends on infecting a mammalian host to emerge from its spore state and reproduce [94], and has a “closed” pan-genome, meaning the number of unique genes added to the *B. anthracis* pan-genome with each genome added decreases to zero [106]. The phenomenon of genomic reduction by pseudogenization and reduced gene transfer on the evolutionary path from environmental microbe/opportunistic pathogen to obligate pathogen has been well described [107–109], and gene loss as a signal of ecological diversification in bacteria has been shown recently among *Ruegeria* [110] and *Bordetella* [111].

This study and the time-structured set of BCG isolates have some limitations and offer opportunities for several further analyses that were beyond the scope of this manuscript. While we leveraged whole-genome sequencing data and a pan-genome approach to investigate the possibility of gene transfer within the BCG regarding specific virulence genes, we did not pursue a formal inference of recombination within this dataset. A formal inference

of recombination may be fruitful to explore using subsets of the BCG data presented here in the future. This study also depended on whole-genome short-read sequencing data. Supplementing the short-read data with long-read data would help resolve the genomic structure(s) found in each isolate and the locations of the anthrax-associated virulence genes we identified. We were also unable to leverage detailed clinical data to understand the disease/infection from which each isolate was derived due to a lack of detailed records. Detailed information on clinical presentation, as well as laboratory-based phenotypic analyses, such as capsule expression and characterization, would also contribute to our knowledge of how anthrax-associated virulence genes, when present, are being expressed. Phenotypic analyses of a selection of the isolates sequenced here will be presented in future studies. As we identified ten BCG genomes by whole-genome nucleotide identity (dDDH and ANI) as *B. anthracis* that lacked any pXO1 or pXO2 virulence genes or pseudogenes, it would be fruitful to investigate the gene content of these isolates and their phylogenetic relationships to previously characterized virulent *B. anthracis* strains. Lastly, this dataset is not representative of the whole diversity of the BCG. While we do not find evidence of horizontal gene transfer in the bacterial populations represented in this dataset, we cannot rule out this mechanism of virulence gene acquisition across all populations in the BCG. It is possible the scenario of how anthrax-associated virulence spread across the known BCG is more complex than that presented here. The continued characterization of environmental and clinical BCG isolates is crucial to further enhance our understanding of virulence in the complex.

In this study, we share evidence for homology to and fragmentation of anthrax-associated virulence genes across multiple species and/or genomospecies of the BCG that cannot be classified as *B. anthracis*. This study, along with a growing body of clinical case reports and microbiological research, supports the etiology of anthrax disease being toxin-mediated, rather than being a narrowly taxonomically defined bacterium.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/pathogens13100884/s1>, Figure S1: Pan-genome plots from Roary; Figure S2: Maximum likelihood tree with bootstrapping scores; Figure S3: Tip-to-root distance vs. time plot; Table S1: Reference type strain information; Table S2: All isolates with metadata, including year, location, and sequencing information (coverage, dDDH, ANIb, GC content, N50, and # of contigs); Table S3: Custom virulence gene database for Prokka and BLAST; Table S4: Protein clusters.

Author Contributions: Conceptualization, Z.P.W., C.A.B., R.M.T. and M.E.B.; data curation, S.J.S., C.A.B., C.A.G. and T.K.P.; formal analysis, S.J.S., C.A.G. and T.K.P.; investigation, S.J.S., C.A.B., C.K.M., R.M.T., J.E.G. and M.E.B.; project administration, Z.P.W. and S.J.S.; resources, Z.P.W. and J.R.M.; software, C.A.G., G.A.S. and T.K.P.; supervision, Z.P.W. and J.R.M.; visualization, S.J.S. and T.K.P.; writing—original draft, S.J.S.; writing—review and editing, Z.P.W., C.A.G., C.A.B. and T.K.P. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for this project was provided by the Centers for Disease Control and Prevention (CDC) and its Advanced Molecular Detection (AMD) initiative.

Institutional Review Board Statement: Not applicable. Found to be exempt as this research did not involve human subjects.

Informed Consent Statement: Not applicable.

Data Availability Statement: The 16S rRNA reference sequences came from NCBI BioProjects PRJNA33175 and PRJNA33317. Virulence gene reference sequence accessions are in Supplementary Table S3. Type strain genome and BioSample accessions are in Supplementary Table S1. Plasmid reference sequences can be found under the following accessions: pXO1 from the Ames Ancestor *B. anthracis* (NC_007322.2), pBCXO1 from the welder's anthrax isolate G9241 (NZ_CP009592.1), pBCXO1 from the welder's anthrax isolate 033B87 (NZ_CP009940.1), pXO2 from the Ames Ancestor *B. anthracis* (NC_007323.3), and pBC218 from G9241 (NZ_CP009591.1). The genome sequence data produced in this study are available in NCBI BioProject PRJNA1122551.

Acknowledgments: We thank the team at TGen North (Flagstaff, Arizona, USA) for their assistance with library preparation and sequencing, particularly Madelyn Ford. The conclusions, findings, and opinions expressed by the authors do not necessarily reflect the official position of the U.S. Department of Health and Human Services, the Public Health Service, the Centers for Disease Control and Prevention (CDC), or the authors' affiliated institutions. The use of trade names is for identification only and does not imply endorsement by any of the groups named above.

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

References

1. Bower, W.A.; Hendricks, K.A.; Vieira, A.R.; Traxler, R.M.; Weiner, Z.; Lynfield, R.; Hoffmaster, A. What Is Anthrax? *Pathogens* **2022**, *11*, 690. [CrossRef] [PubMed]
2. Jernigan, J.A.; Stephens, D.S.; Ashford, D.A.; Omenaca, C.; Topiel, M.S.; Galbraith, M.; Tapper, M.; Fisk, T.L.; Zaki, S.; Popovic, T.; et al. Bioterrorism-Related Inhalational Anthrax: The First 10 Cases Reported in the United States. *Emerg. Infect. Dis.* **2001**, *7*, 933–944. [CrossRef] [PubMed]
3. Bush, L.M.; Perez, M.T. The Anthrax Attacks 10 Years Later. *Ann. Intern. Med.* **2012**, *156*, 41–44. [CrossRef] [PubMed]
4. Gursky, E.; Inglesby, T.V.; O'Toole, T. Anthrax 2001: Observations on the Medical and Public Health Response. *Biosecurity Bioterrorism Biodefense Strategy Pract. Sci.* **2003**, *1*, 97–110. [CrossRef] [PubMed]
5. Public Health. 42 CFR Part 73 2005. Available online: <https://www.ecfr.gov/current/title-42/chapter-I/subchapter-F/part-73> (accessed on 20 September 2024).
6. Hoffmaster, A.R.; Meyer, R.F.; Bowen, M.D.; Marston, C.K.; Weyant, R.S.; Thurman, K.; Messenger, S.L.; Minor, E.E.; Winchell, J.M.; Rassmussen, M.V.; et al. Evaluation and Validation of a Real-Time Polymerase Chain Reaction Assay for Rapid Identification of *Bacillus anthracis*. *Emerg. Infect. Dis.* **2002**, *8*, 1178–1182. [CrossRef]
7. Ellerbrok, H.; Nattermann, H.; Ozel, M.; Beutin, L.; Appel, B.; Pauli, G. Rapid and sensitive identification of pathogenic and apathogenic *Bacillus anthracis* by real-time PCR. *FEMS Microbiol. Lett.* **2002**, *214*, 51–59. [CrossRef]
8. Ehling-Schulz, M.; Lereclus, D.; Koehler, T.M. The *Bacillus cereus* Group: *Bacillus* Species with Pathogenic Potential. *Microbiol. Spectr.* **2019**, *7*, 2018. [CrossRef]
9. Carroll, L.M.; Cheng, R.A.; Wiedmann, M.; Kovac, J. Keeping up with the *Bacillus cereus* group: Taxonomy through the genomics era and beyond. *Crit. Rev. Food Sci. Nutr.* **2022**, *62*, 7677–7702. [CrossRef]
10. Bottone, E.J. *Bacillus cereus*, a volatile human pathogen. *Clin. Microbiol. Rev.* **2010**, *23*, 382–398. [CrossRef]
11. Miller, R.A.; Jian, J.; Beno, S.M.; Wiedmann, M.; Kovac, J. Intraclade Variability in Toxin Production and Cytotoxicity of *Bacillus cereus* Group Type Strains and Dairy-Associated Isolates. *Appl. Environ. Microbiol.* **2018**, *84*, e02479-17. [CrossRef]
12. Beesley, C.A.; Vanner, C.L.; Helsen, L.O.; Gee, J.E.; Hoffmaster, A.R. Identification and characterization of clinical *Bacillus* spp. isolates phenotypically similar to *Bacillus anthracis*. *FEMS Microbiol. Lett.* **2010**, *313*, 47–53. [CrossRef] [PubMed]
13. Leendertz, F.H.; Ellerbrok, H.; Boesch, C.; Couacy-Hymann, E.; Matz-Rensing, K.; Hakenbeck, R.; Bergmann, C.; Abaza, P.; Junglen, S.; Moebius, Y.; et al. Anthrax kills wild chimpanzees in a tropical rainforest. *Nature* **2004**, *430*, 451–452. [CrossRef] [PubMed]
14. Leendertz, F.H.; Yumlu, S.; Pauli, G.; Boesch, C.; Couacy-Hymann, E.; Vigilant, L.; Junglen, S.; Schenk, S.; Ellerbrok, H. A new *Bacillus anthracis* found in wild chimpanzees and a gorilla from West and Central Africa. *PLoS Pathog.* **2006**, *2*, e8. [CrossRef]
15. Klee, S.R.; Ozel, M.; Appel, B.; Boesch, C.; Ellerbrok, H.; Jacob, D.; Holland, G.; Leendertz, F.H.; Pauli, G.; Grunow, R.; et al. Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. *J. Bacteriol.* **2006**, *188*, 5333–5344. [CrossRef] [PubMed]
16. Klee, S.R.; Brzuszkiewicz, E.B.; Nattermann, H.; Bruggemann, H.; Dupke, S.; Wollherr, A.; Franz, T.; Pauli, G.; Appel, B.; Liebl, W.; et al. The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids. *PLoS ONE* **2010**, *5*, e10986. [CrossRef]
17. Brezillon, C.; Haustant, M.; Dupke, S.; Corre, J.P.; Lander, A.; Franz, T.; Monot, M.; Couture-Tosi, E.; Jouvion, G.; Leendertz, F.H.; et al. Capsules, toxins and AtxA as virulence factors of emerging *Bacillus cereus* biovar anthracis. *PLoS Negl. Trop. Dis.* **2015**, *9*, e0003455. [CrossRef]
18. Carroll, L.M.; Marston, C.K.; Kolton, C.B.; Gulvik, C.A.; Gee, J.E.; Weiner, Z.P.; Kovac, J. Strains Associated with Two 2020 Welder Anthrax Cases in the United States Belong to Separate Lineages within *Bacillus cereus* sensu lato. *Pathogens* **2022**, *11*, 856. [CrossRef]
19. Dawson, P.; Salzer, J.S.; Schrodtt, C.A.; Feldmann, K.; Kolton, C.B.; Gee, J.E.; Marston, C.K.; Gulvik, C.A.; Elrod, M.G.; Villarma, A.; et al. Epidemiologic Investigation of Two Welder's Anthrax Cases Caused by *Bacillus cereus* Group Bacteria: Occupational Link Established by Environmental Detection. *Pathogens* **2022**, *11*, 825. [CrossRef]
20. Hoffmaster, A.; Ravel, J.; Rasko, D.A.; Chapman, G.D.; Chute, M.D.; Marston, C.K.; De, B.K.; Sacchi, C.T.; Fitzgerald, C.; Mayer, L.W.; et al. Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8449–8454. [CrossRef]

21. Hoffmaster, A.R.; Hill, K.K.; Gee, J.E.; Marston, C.K.; De, B.K.; Popovic, T.; Sue, D.; Wilkins, P.P.; Avashia, S.B.; Drumgoole, R.; et al. Characterization of *Bacillus cereus* isolates associated with fatal pneumonias: Strains are closely related to *Bacillus anthracis* and harbor *B. anthracis* virulence genes. *J. Clin. Microbiol.* **2006**, *44*, 3352–3360. [[CrossRef](#)]
22. Dawson, P.; Schrodte, C.A.; Feldmann, K.; Traxler, R.; Gee, J.E.; Kolton, C.B.; Marston, C.K.; Gulvik, C.; Antonini, J.M.; Negrón, M.E.; et al. Fatal Anthrax Pneumonia in Welders and Other Metalworkers Caused by *Bacillus cereus* Group Bacteria Containing Anthrax Toxin Genes—U.S. Gulf Coast States, 1994–2020. *MMWR. Morb. Mortal. Wkly. Rep.* **2021**, *70*, 1453–1454. [[CrossRef](#)]
23. Pena-Gonzalez, A.; Marston, C.K.; Rodriguez, R.L.; Kolton, C.B.; Garcia-Diaz, J.; Theppote, A.; Frace, M.; Konstantinidis, K.T.; Hoffmaster, A.R. Draft Genome Sequence of *Bacillus cereus* LA2007, a Human-Pathogenic Isolate Harboring Anthrax-Like Plasmids. *Genome Announc.* **2017**, *5*, 17. [[CrossRef](#)] [[PubMed](#)]
24. Oh, S.Y.; Budzik, J.M.; Garufi, G.; Schneewind, O. Two capsular polysaccharides enable *Bacillus cereus* G9241 to cause anthrax-like disease. *Mol. Microbiol.* **2011**, *80*, 455–470. [[CrossRef](#)] [[PubMed](#)]
25. Marston, C.K.; Ibrahim, H.; Lee, P.; Churchwell, G.; Gumke, M.; Stanek, D.; Gee, J.E.; Boyer, A.E.; Gallegos-Candela, M.; Barr, J.R.; et al. Anthrax Toxin-Expressing *Bacillus cereus* Isolated from an Anthrax-Like Eschar. *PLoS ONE* **2016**, *11*, e0156987. [[CrossRef](#)] [[PubMed](#)]
26. Kaiser, J. UPDATED: University of Chicago Microbiologist Infected from Possible Lab Accident. Science Insider. 2011. Available online: <https://www.science.org/content/article/updated-university-chicago-microbiologist-infected-possible-lab-accident> (accessed on 17 April 2024).
27. Carroll, L.M.; Matle, I.; Kovac, J.; Cheng, R.A.; Wiedmann, M. Laboratory Misidentifications Resulting from Taxonomic Changes to *Bacillus cereus* Group Species, 2018–2022. *Emerg. Infect. Dis.* **2022**, *28*, 1877–1881. [[CrossRef](#)] [[PubMed](#)]
28. Carroll, L.M.; Pierneef, R.; Mathole, A.; Atanda, A.; Matle, I. Genomic Sequencing of *Bacillus cereus* Sensu Lato Strains Isolated from meat and Poultry Products in South Africa Enables Inter- and Intranational Surveillance and Source Tracking. *Microbiol. Spectr.* **2022**, *10*, e0070022. [[CrossRef](#)] [[PubMed](#)]
29. Muigg, V.; Cuénod, A.; Purushothaman, S.; Siegemund, M.; Wittwer, M.; Pflüger, V.; Schmidt, K.M.; Weisser, M.; Ritz, N.; Widmer, A.; et al. Diagnostic challenges within the *Bacillus cereus*-group: Finding the beast without teeth. *New Microbes New Infect.* **2022**, *49–50*, 101040. [[CrossRef](#)]
30. Baldwin, V.M. You Can't *B. cereus*—A Review of *Bacillus cereus* Strains That Cause Anthrax-Like Disease. *Front. Microbiol.* **2020**, *11*, 1731. [[CrossRef](#)]
31. Dauphin, L.A.; Bowen, M.D. A simple method for the rapid removal of *Bacillus anthracis* spores from DNA preparations. *J. Microbiol. Methods* **2009**, *76*, 212–214. [[CrossRef](#)]
32. Ewels, P.A.; Peltzer, A.; Fillinger, S.; Patel, H.; Alneberg, J.; Wilm, A.; Garcia, M.U.; Di Tommaso, P.; Nahnsen, S. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **2020**, *38*, 276–278. [[CrossRef](#)]
33. Sprenger, G.; Gulvik, C.; Paisie, T. wf-Paired-End-Illumina-Assembly, 2.3.0. 2024. Available online: <https://github.com/gregorysprenger/wf-paired-end-illumina-assembly> (accessed on 20 September 2024).
34. Di Tommaso, P.; Chatzou, M.; Floden, E.W.; Barja, P.P.; Palumbo, E.; Notredame, C. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **2017**, *35*, 316–319. [[CrossRef](#)] [[PubMed](#)]
35. sra-human-scrubber. 2023. Available online: <https://github.com/ncbi/sra-human-scrubber> (accessed on 20 September 2024).
36. Constantinides, B.; Hunt, M.; Crook, D.W. Hostile: Accurate decontamination of microbial host sequences. *Bioinformatics* **2023**, *39*, btad728. [[CrossRef](#)] [[PubMed](#)]
37. Bushnell, B. BBMap. 2014. Available online: <https://sourceforge.net/projects/bbmap/> (accessed on 20 September 2024).
38. Kokot, M.; Dlugosz, M.; Deorowicz, S. KMC 3: Counting and manipulating k-mer statistics. *Bioinformatics* **2017**, *33*, 2759–2761. [[CrossRef](#)] [[PubMed](#)]
39. Li, H. seqtk. 2023. Available online: <https://github.com/lh3/seqtk> (accessed on 20 September 2024).
40. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
41. Magoc, T.; Salzberg, S.L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **2011**, *27*, 2957–2963. [[CrossRef](#)]
42. 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](#)]
43. Cock, P.J.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423. [[CrossRef](#)]
44. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
45. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
46. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [[CrossRef](#)]

47. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)] [[PubMed](#)]
48. Chklovski, A.; Parks, D.H.; Woodcroft, B.J.; Tyson, G.W. CheckM2: A rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods* **2023**, *20*, 1203–1212. [[CrossRef](#)] [[PubMed](#)]
49. Quinlan, A.R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinform.* **2014**, *47*, 11.12.1–11.12.34. [[CrossRef](#)]
50. Seemann, T. mlst. 2022. Available online: <https://github.com/tseemann/mlst> (accessed on 20 September 2024).
51. Jolley, K.A.; Maiden, M.C. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinform.* **2010**, *11*, 595. [[CrossRef](#)]
52. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)]
53. Seemann, T. Barnap. 2018. Available online: <https://github.com/tseemann/barnap> (accessed on 20 September 2024).
54. Wang, Q.; Garrity, G.M.; Tiedje, J.M.; Cole, J.R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **2007**, *73*, 5261–5267. [[CrossRef](#)]
55. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
56. Wood, D.E.; Salzberg, S.L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **2014**, *15*, R46. [[CrossRef](#)]
57. Wood, D.E.; Lu, J.; Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 257. [[CrossRef](#)]
58. Chen, Y.; Li, Y.; Shen, J.; Liu, Q.; Liu, Y.; Chu, Y.; Xiao, Z. *Bacillus arachidis* sp. nov., Isolated from Peanut Rhizosphere Soil. *Curr. Microbiol.* **2022**, *79*, 231. [[CrossRef](#)]
59. Cohn, F. Untersuchungen über Bakterien. In *Beiträge zur Biologie der Pflanzen 1*; J.U. Kern's Verlag (Max Müller): Wrocław, Poland, 1872; Volume 1, pp. 127–224.
60. Daligault, H.E.; Davenport, K.W.; Minogue, T.D.; Bishop-Lilly, K.A.; Broomall, S.M.; Bruce, D.C.; Chain, P.S.; Coyne, S.R.; Frey, K.G.; Gibbons, H.S.; et al. Twenty Whole-Genome *Bacillus* sp Assemblies. *Genome Announc.* **2014**, *2*, 10–1128. [[CrossRef](#)]
61. Flügge, C. *Die Mikroorganismen: Mit Besonderer Berücksichtigung der Ätiologie der Infektionskrankheiten*; Verlag von F. C. W. Vogel: Leipzig, Germany, 1886.
62. Frankland, G.C.; Frankland, P.F. Studies on some new micro-organisms obtained from air. *R. Soc. Lond. Philos. Trans. B* **1887**, *178*, 257–287. [[CrossRef](#)]
63. Ivanova, N.; Sorokin, A.; Anderson, I.; Galleron, N.; Candelon, B.; Kapatral, V.; Bhattacharyya, A.; Reznik, G.; Mikhailova, N.; Lapidus, A.; et al. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* **2003**, *423*, 87–91. [[CrossRef](#)] [[PubMed](#)]
64. Jeong, H.; Park, S.H.; Choi, S.K. Draft Genome Sequences of Four Plant Probiotic *Bacillus* Strains. *Genome Announc.* **2016**, *4*, 10–1128. [[CrossRef](#)]
65. Jung, M.Y.; Kim, J.S.; Paek, W.K.; Lim, J.; Lee, H.; Kim, P.I.; Ma, J.Y.; Kim, W.; Chang, Y.H. *Bacillus manliponensis* sp. nov., a new member of the *Bacillus cereus* group isolated from foreshore tidal flat sediment. *J. Microbiol.* **2011**, *49*, 1027–1032. [[CrossRef](#)]
66. Kämpfer, P.; Lipski, A.; McInroy, J.A.; Clermont, D.; Criscuolo, A.; Glaeser, S.P. *Bacillus rhizoplanae* sp. nov. from maize roots. *Int. J. Syst. Evol. Microbiol.* **2022**, *72*, 005450. [[CrossRef](#)]
67. Lapidus, A.; Goltsman, E.; Auger, S.; Galleron, N.; Segurens, B.; Dossat, C.; Land, M.L.; Broussolle, V.; Brillard, J.; Guinebretiere, M.H.; et al. Extending the *Bacillus cereus* group genomics to putative food-borne pathogens of different toxicity. *Chem. Biol. Interact.* **2008**, *171*, 236–249. [[CrossRef](#)]
68. Liu, G.; Liu, B.; Lin, N.; Tang, W.; Tang, J.; Lin, Y. Genome sequence of the aerobic bacterium *Bacillus* sp. strain FJAT-13831. *J. Bacteriol.* **2012**, *194*, 6633. [[CrossRef](#)]
69. Liu, Y.; Du, J.; Lai, Q.; Zeng, R.; Ye, D.; Xu, J.; Shao, Z. Proposal of nine novel species of the *Bacillus cereus* group. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 2499–2508. [[CrossRef](#)]
70. Liu, Y.; Lai, Q.; Du, J.; Shao, Z. Genetic diversity and population structure of the *Bacillus cereus* group bacteria from diverse marine environments. *Sci. Rep.* **2017**, *7*, 689. [[CrossRef](#)]
71. Liu, X.; Wang, L.; Han, M.; Xue, Q.H.; Zhang, G.S.; Gao, J.; Sun, X. *Bacillus fungorum* sp. nov., a bacterium isolated from spent mushroom substrate. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 1457–1462. [[CrossRef](#)] [[PubMed](#)]
72. Mendez Acevedo, M.; Carroll, L.M.; Mukherjee, M.; Mills, E.; Xiaoli, L.; Dudley, E.G.; Kovac, J. Novel Effective *Bacillus cereus* Group Species “*Bacillus clarus*” Is Represented by Antibiotic-Producing Strain ATCC 21929 Isolated from Soil. *mSphere* **2020**, *5*, 10–1128. [[CrossRef](#)] [[PubMed](#)]
73. Miller, R.A.; Beno, S.M.; Kent, D.J.; Carroll, L.M.; Martin, N.H.; Boor, K.J.; Kovac, J. *Bacillus wiedmannii* sp. nov., a psychrotolerant and cytotoxic *Bacillus cereus* group species isolated from dairy foods and dairy environments. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 4744–4753. [[CrossRef](#)] [[PubMed](#)]
74. Robas Mora, M.; Fernández Pastrana, V.M.; Oliva, L.L.G.; Lobo, A.P.; Jiménez Gómez, P.A. Plant growth promotion of the forage plant *Lupinus albus* Var. Orden Dorado using *Pseudomonas agronomica* sp. nov. and *Bacillus pretiosus* sp. nov. added over a valorized agricultural biowaste. *Front. Microbiol.* **2023**, *13*, 1046201. [[CrossRef](#)] [[PubMed](#)]

75. Stelder, S.K.; Mahmud, S.A.; Dekker, H.L.; de Koning, L.J.; Brul, S.; de Koster, C.G. Temperature Dependence of the Proteome Profile of the Psychrotolerant Pathogenic Food Spoiler *Bacillus weihenstephanensis* Type Strain WSBC 10204. *J. Proteome. Res.* **2015**, *14*, 2169–2176. [CrossRef]
76. Tohya, M.; Hishinuma, T.; Watanabe, S.; Shimojima, M.; Ogawa, M.; Tada, T.; Kirikae, T. Three novel species of the *Bacillus cereus* group isolated from clinical samples in Japan. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 004993. [CrossRef]
77. Zwick, M.E.; Joseph, S.J.; Didelot, X.; Chen, P.E.; Bishop-Lilly, K.A.; Stewart, A.C.; Willner, K.; Nolan, N.; Lentz, S.; Thomason, M.K.; et al. Genomic characterization of the *Bacillus cereus* sensu lato species: Backdrop to the evolution of *Bacillus anthracis*. *Genome Res.* **2012**, *22*, 1512–1524. [CrossRef]
78. Jimenez, G.; Urdiain, M.; Cifuentes, A.; Lopez-Lopez, A.; Blanch, A.R.; Tamames, J.; Kampfer, P.; Kolsto, A.B.; Ramon, D.; Martinez, J.F.; et al. Description of *Bacillus toyonensis* sp. nov., a novel species of the *Bacillus cereus* group, and pairwise genome comparisons of the species of the group by means of ANI calculations. *Syst. Appl. Microbiol.* **2013**, *36*, 383–391. [CrossRef]
79. Sprenger, G.; Gulvik, C. wf-ani. 2024. Available online: <https://github.com/ParBLISS/FastANI> (accessed on 20 September 2024).
80. Meier-Kolthoff, J.P.; Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* **2019**, *10*, 2182. [CrossRef]
81. Carroll, L.M.; Weidmann, M.; Kovac, J. Proposal of a Taxonomic Nomenclature for the *Bacillus cereus* Group Which Reconciles Genomic Definitions of Bacterial Species with Clinical and Industrial Phenotypes. *mBio* **2020**, *11*, e02901-19. [CrossRef]
82. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]
83. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *Gigascience* **2021**, *10*, giab008. [CrossRef]
84. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]
85. Katz, L.S.; Griswold, T.; Morrison, S.S.; Caravas, J.A.; Zhang, S.; den Bakker, H.C.; Deng, X.; Carleton, H.A. Mashtree: A rapid comparison of whole genome sequence files. *J. Open Source Softw.* **2019**, *4*, 1762. [CrossRef]
86. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [CrossRef] [PubMed]
87. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef]
88. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
89. Page, A.J.; Taylor, B.; Delaney, A.J.; Soares, J.; Seemann, T.; Keane, J.A.; Harris, S.R. SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* **2016**, *2*, e000056. [CrossRef]
90. Kozlov, A.M.; Darrriba, D.; Flouri, T.; Morel, B.; Stamatakis, A. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **2019**, *35*, 4453–4455. [CrossRef]
91. FigTree, 1.4.4. 2018. Available online: <http://tree.bio.ed.ac.uk/software/figtree/> (accessed on 20 September 2024).
92. Paradis, E.; Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2019**, *35*, 526–528. [CrossRef]
93. Galili, T. dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **2015**, *31*, 3718–3720. [CrossRef] [PubMed]
94. Pena-Gonzalez, A.; Rodriguez, R.L.; Marston, C.K.; Gee, J.E.; Gulvik, C.A.; Kolton, C.B.; Saile, E.; Frace, M.; Hoffmaster, A.R.; Konstantinidis, K.T. Genomic Characterization and Copy Number Variation of *Bacillus anthracis* Plasmids pXO1 and pXO2 in a Historical Collection of 412 Strains. *mSystems* **2018**, *3*, e65-18. [CrossRef] [PubMed]
95. Baker, F.B. Stability of Two Hierarchical Grouping Techniques Case I: Sensitivity to Data Errors. *J. Am. Stat. Assoc.* **1974**, *69*, 440–445. [CrossRef]
96. Sokal, R.R.; Rohlf, F.J. The Comparison of Dendrograms by Objective Methods. *Taxon* **1962**, *11*, 33–40. [CrossRef]
97. Rambaut, A.; Lam, T.T.; Max Carvalho, L.; Pybus, O.G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2016**, *2*, vew007. [CrossRef]
98. Liu, Y.; Lai, Q.; Goker, M.; Meier-Kolthoff, J.P.; Wang, M.; Sun, Y.; Wang, L.; Shao, Z. Genomic insights into the taxonomic status of the *Bacillus cereus* group. *Sci. Rep.* **2015**, *5*, 14082. [CrossRef]
99. Bazinet, A.L. Pan-genome and phylogeny of *Bacillus cereus* sensu lato. *BMC Evol. Biol.* **2017**, *17*, 176. [CrossRef]
100. Miller, J.M.; Hair, J.G.; Hebert, M.; Hebert, L.; Roberts, J.; Floyd, J. Fulminating Bacteremia and Pneumonia Due to *Bacillus cereus*. *J. Clin. Microbiol.* **1997**, *35*, 504–507. [CrossRef]
101. Sue, D.; Hoffmaster, A.R.; Popovic, T.; Wilkins, P.P. Capsule Production in *Bacillus cereus* Strains Associated with Severe Pneumonia. *J. Clin. Microbiol.* **2006**, *44*, 3426–3428. [CrossRef]
102. Ochman, H.; Lawrence, J.G.; Groisman, E.A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **2000**, *405*, 299–304. [CrossRef]
103. Wiedenbeck, J.; Cohan, F.M. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* **2011**, *35*, 957–976. [CrossRef] [PubMed]
104. Weller, C.; Wu, M. A generation-time effect on the rate of molecular evolution in bacteria. *Evolution* **2015**, *69*, 643–652. [CrossRef] [PubMed]

105. Mira, A.; Ochman, H.; Moran, N.A. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **2001**, *17*, 589–596. [[CrossRef](#)] [[PubMed](#)]
106. Tettelin, H.; Maignani, V.; Cieslewicz, M.J.; Donati, C.; Medini, D.; Ward, N.L.; Angiuoli, S.V.; Crabtree, J.; Jones, A.L.; Durkin, A.S.; et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13950. [[CrossRef](#)] [[PubMed](#)]
107. Hershberg, R.; Tang, H.; Petrov, D.A. Reduced selection leads to accelerated gene loss in *Shigella*. *Genome Biol.* **2007**, *8*, R164. [[CrossRef](#)]
108. Moran, N.A. Microbial Minimalism: Genome Reduction in Bacterial Pathogens. *Cell* **2002**, *108*, 583–586. [[CrossRef](#)]
109. Rouli, L.; Merhej, V.; Fournier, P.E.; Raoult, D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* **2015**, *7*, 72–85. [[CrossRef](#)]
110. Chu, X.; Li, S.; Wang, S.; Luo, D.; Luo, H. Gene loss through pseudogenization contributes to the ecological diversification of a generalist *Roseobacter* lineage. *ISME J.* **2021**, *15*, 489–502. [[CrossRef](#)]
111. Linz, B.; Ivanov, Y.V.; Preston, A.; Brinkac, L.; Parkhill, J.; Kim, M.; Harris, S.R.; Goodfield, L.L.; Fry, N.K.; Gorringer, A.R.; et al. Acquisition and loss of virulence-associated factors during genome evolution and speciation in three clades of *Bordetella* species. *BMC Genom.* **2016**, *17*, 767. [[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.