Genomic surveillance of *Bacillus cereus sensu lato* strains isolated from meat and poultry products in South Africa enables inter- and intra-national surveillance and source tracking

Laura M. Carroll[a], Rian Pierneef[b], Aletta Mathole[c], Abimbola Atanda[c] , Itumeleng Matle[c]#

[a]Structural and Computational Biology Unit, EMBL, Heidelberg, Germany

[b]Biotechnology Platform, Agricultural Research Council, Onderstepoort Veterinary Research, Onderstepoort, South Africa

[c]Bacteriology Division, Agricultural Research Council: Onderstepoort Veterinary Research, Onderstepoort, South Africa

[#]Corresponding author: Itumeleng Matle, matlei@arc.agric.za

Running Title: South African *Bacillus cereus* genomic sequencing

Abstract Word Count: 243 words (Abstract) + 147 words (Importance)

Text Word Count: 4,527 words

Abstract

Members of the *Bacillus cereus sensu lato* (*s.l.*) species complex, also known as the *B. cereus* group, vary in their ability to cause illness, but are frequently isolated from foods, including meat products; however, food safety surveillance efforts that employ whole-genome sequencing (WGS) often neglect these potential pathogens. Here, WGS was used to characterize *B. cereus s.l.* strains (n = 25) isolated during surveillance of meat products in South Africa. Strains were collected from beef, poultry, and mixed meat products obtained from (i) retail outlets, processing plants, and butcheries across six South African provinces (n = 15, 7, and 1, respectively), and (ii) imports in cold stores (n = 2). Strains were assigned to *panC* Groups IV, III, II, and V (n = 18, 5, 1, and 1, respectively) and spanned multiple genomospecies, regardless of the taxonomy used. All strains possessed diarrheal toxin-encoding genes, while one sequence type 26 (ST26) strain possessed cereulide (emetic toxin) synthetase-encoding genes. No strains harbored anthrax toxin- or capsule-encoding genes. The 25 strains were partitioned into 15 lineages via *in silico* seven-gene multi-locus sequence typing (MLST), six of which contained multiple strains sequenced in this study, which were identical or nearly identical at the whole-genome scale. Five MLST lineages contained (nearly) identical genomes collected from two or three South African provinces; one MLST lineage contained nearly identical genomes from two countries (South Africa and the Netherlands), indicating that *B. cereus s.l.* can spread intra- and inter-nationally via foodstuffs.

Importance

Nation-wide foodborne pathogen surveillance programs that employ high-resolution genomic methods have been shown to provide vast public health and economic benefits. However, *B. cereus s.l.* are often overlooked during large-scale, routine WGS efforts. Thus, to our knowledge, no studies to date have evaluated the potential utility of WGS for *B. cereus s.l.* surveillance and source

47 tracking in foodstuffs. In this proof-of-concept study, we applied WGS to *B. cereus s.l.* strains
48 collected via South Africa's national surveillance program of domestic and imported meat
49 products, and we provide strong evidence that *B. cereus s.l.* can be disseminated intra- and inter-
50 nationally via the agro-food supply chain. Our results showcase that WGS can be used for source
51 tracking of *B. cereus s.l.* in foods, although future WGS and isolate metadata collection efforts are
52 needed to ensure that *B. cereus s.l.* surveillance initiatives are on par with those of other foodborne
53 pathogens.
54
55
56

INTRODUCTION

Bacillus cereus sensu lato (*s.l.*), also known as the *B. cereus* group, is a complex of closely related, Gram-positive, spore-forming species, which are widespread throughout the environment (1). While some members of *B. cereus s.l.* have important industrial applications or roles (e.g., as biocontrol agents in agricultural settings, as food spoilage organisms) (2-6), others are capable of causing illnesses or death in humans and/or animals (1, 7-9). Illnesses caused by members of *B. cereus s.l.* can range in severity from mild to severe/fatal and include anthrax and anthrax-like illness (8, 10, 11), foodborne emetic intoxication (1, 7, 12-15), foodborne diarrheal toxicoinfection (1, 7, 14, 15), and non-gastrointestinal infections (16, 17). As a foodborne pathogen, "*B. cereus*" is estimated to be responsible for more than 256,000 illnesses globally each year (18), although this is likely an underestimate, due to the relatively mild and self-limiting nature of the symptoms that often accompany foodborne illness caused by *B. cereus s.l.* (1).

Food safety surveillance efforts around the world have identified *B. cereus s.l.* strains in a wide variety of foodstuffs (1, 7), including raw intact, processed, and ready-to-eat (RTE) meat and poultry products (19-27). In South Africa specifically, previous surveillance efforts have identified *B. cereus s.l.* in (i) retail meats sold at supermarkets in the Pretoria area (i.e., Vienna sausages, salami, and poultry) (27) and (ii) biltong (a spiced intermediate moisture RTE meat product) sold at supermarkets, stalls, kiosks, and butcheries in Bloemfontein, Free State (28). Most recently, in a study of over two thousand meat product samples collected from butcheries, processing plants, abattoirs, and retail outlets across all nine South African provinces, *B. cereus s.l.* was present in 4.5 and 2.7% of domestic and imported meat products, respectively (26).

Ongoing surveillance efforts in South Africa have indicated that meat and poultry products can harbor *B. cereus s.l.* and may pose a potential food safety risk to South African consumers.

80 However, it is unclear which *B. cereus s.l.* lineages are present in South African meat and poultry
81 products on a genomic scale. Here, we used whole-genome sequencing (WGS) to characterize 25
82 *B. cereus s.l.* strains isolated from raw intact, processed, and RTE meat and poultry products
83 collected from processing plants, butcheries, and retail outlets across South Africa, as well as
84 imported meat products. By comparing South African strains sequenced here to all publicly
85 available *B. cereus s.l.* genomes (n = 2,887 total genomes), we identified multiple *B. cereus s.l.*
86 species present among South African meat and poultry products, and we detected multiple
87 potential inter-national and inter-provincial *B. cereus s.l.* dissemination events. Overall, our study
88 serves as the first genome-scale study of South African *B. cereus s.l.* in foodstuffs and showcases
89 the utility of WGS for *B. cereus s.l.* surveillance and source tracking.
90
91 **RESULTS**
92 ***B. cereus s.l.* are present among domestic and imported meat and poultry products in South**
93 **Africa.** A total of 25 *B. cereus s.l.* strains were isolated from meat and poultry products, which
94 had been collected across South Africa in 2015 and 2016 (Figure 1, Table 1, and Supplemental
95 Table S1). Overall, 19 strains (76.0%) originated from beef products, including: beef wors, a
96 processed South African sausage (n = 7; 28.0%); beef biltong, a South African spiced intermediate
97 moisture RTE meat product (n = 5; 20.0%); processed beef mince (n = 3; 12.0%); RTE beef
98 sausage emulsion (n = 2; 8.0%); and processed beef patties (n = 2; 8.0%, Table 1 and Supplemental
99 Table S1). Five strains (20.0%) were isolated from poultry products, including raw chicken thighs
100 (n = 2; 8.0%), raw chicken quarter-legs (n = 2; 8.0%), and a frankfurter (n = 1; 4.0%, Table 1 and
101 Supplemental Table S1). One strain (4.0%) was isolated from wors, which had been made from a
102 mix of beef, pork, and lamb (Table 1 and Supplemental Table S1).

The majority of strains (23 of 25, 92.0%) were obtained from domestic meat and poultry products acquired across six South African provinces (*n* = 5, 5, 4, 4, 3, and 2 strains from Gauteng, Limpopo, Free State, North West, Mpumalanga, and Western Cape, respectively; Figure 1, Table 1, and Supplemental Table S1). The remaining two strains (8.0%) were isolated from raw, intact chicken quarter-legs, which had been imported into South Africa from the Netherlands (Figure 1, Table 1, and Supplemental Table S1). Fifteen of the 25 *B. cereus s.l.* strains (60.0%) were isolated from meat or poultry products acquired from retail outlets, while seven (28.0%) were acquired from meat or poultry products obtained from processing plants (Table 1 and Supplemental Table S1). Two strains (from chicken quarter-legs imported from the Netherlands; 8.0%) were derived from poultry in cold stores, while one strain (4.0%) was isolated from the mixed beef-pork-lamb wors acquired from a butchery (Table 1 and Supplemental Table S1).

Multiple species are present among *B. cereus s.l.* from South African meat and poultry products. Species-level taxonomic classification of *B. cereus s.l.* is notoriously challenging (1, 29, 30); to avoid taxonomic ambiguities and maximize interpretability, we applied multiple taxonomic assignment and sequence typing methods to the 25 strains sequenced here (Table 1). One such sequence typing framework relies on the pantoate-β-alanine ligase gene (*panC*) to assign *B. cereus s.l.* strains to one of seven or more major phylogenetic groups, which have been proposed to conceptually serve as "species" (31, 32). Using the adjusted eight-group *panC* typing approach implemented in BTyper3 (33), the 25 *B. cereus s.l.* strains sequenced here encompassed four *panC* phylogenetic groups (i.e., "species"; Figure 2 and Table 1). The majority (*n* = 18 of 25, 72.0%) of the strains sequenced here were assigned to *panC* Group IV (Figure 2 and Table 1). The remaining strains were assigned to *panC* Groups III, II, and V (*n* = 5, 1, and 1 strains, representing 20.0%, 4.0%, and 4.0% of isolates sequenced here, respectively; Figure 2 and Table 1).

126 Using the Genome Taxonomy Database (GTDB) taxonomy, the 25 *B. cereus s.l.* strains
127 sequenced here encompassed eight genomospecies (Figure 2 and Table 1). The 18 *panC* Group
128 IV strains sequenced here encompassed four GTDB genomospecies, while the five *panC* Group
129 III strains spanned two GTDB genomospecies (Figure 2 and Table 1). The *panC* Group II and
130 Group V strains (n = 1 each) were each assigned to separate GTDB genomospecies (Figure 2 and
131 Table 1).
132 Using a standardized genomospecies-subspecies-biovar (GSB) nomenclatural framework
133 proposed for *B. cereus s.l.* in 2020 (34) (referred to hereafter as the "2020 GSB" framework), the
134 25 strains sequenced here encompassed three genomospecies (Figure 2 and Tables 1 and 2). All
135 18 *panC* Group IV strains were assigned to the *B. cereus sensu stricto* (*s.s.*) genomospecies (Figure
136 2 and Tables 1 and 2); genes encoding insecticidal toxins (referred to hereafter as "Bt toxin-
137 encoding genes") were detected within all Group IV *B. cereus s.s.* genomes (using
138 BtToxin_scanner2's "old" gene detection approach), meaning that these 18 strains were predicted
139 to belong to the Thuringiensis biovar (i.e., *B. cereus s.s.* bv. Thuringiensis; Table 2). All five *panC*
140 Group III and the single *panC* Group II strain(s) were assigned to genomospecies *B. mosaicus*
141 within the 2020 GSB framework (Figure 2 and Tables 1 and 2). One *panC* Group III *B. mosaicus*
142 strain (i.e., strain S66) was assigned to PubMLST sequence type 26 (ST26) and possessed
143 cereulide (emetic toxin) synthetase-encoding *cesABCD* and was thus assigned to the *cereus*
144 subspecies and biovar Emeticus (i.e., *B. mosaicus* subsp. *cereus* bv. Emeticus; Table 2). The lone
145 *panC* Group V strain sequenced here was assigned to the *B. toyonensis* genomospecies; Bt toxin-
146 encoding genes were detected in this genome (via BtToxin_scanner2's "old" gene detection
147 approach), and thus this strain was predicted to belong to biovar Thuringiensis (i.e., *B. toyonensis*
148 bv. Thuringiensis; Figure 2 and Tables 1 and 2).

Using a rapid, average nucleotide identity (ANI)-based pseudo-gene flow unit assignment scheme, which attempts to assign *B. cereus s.l.* genomes to taxonomic units that mimic *B. cereus s.l.* "species" previously delineated using recent gene flow (33), the 25 strains sequenced here were assigned to six pseudo-gene flow units (pseudo-GFUs; Figure 2 and Supplemental Table S1). The 18 *panC* Group IV and five *panC* Group III strains sequenced here each spanned two pseudo-GFUs (Figure 2 and Supplemental Table S1). The *panC* Group II and Group V strains (n = 1 each) were each assigned to separate pseudo-GFUs, respectively (Figure 2 and Supplemental Table S1).

As mentioned above, considerable phenotypic diversity was predicted among the stains sequenced here, as one strain harbored cereulide synthetase-encoding genes, and 19 strains possessed Bt toxin-encoding genes (detected using BtToxin_scanner2's "old" gene detection approach; Table 2). No anthrax toxin- or capsule-encoding genes were identified within the genomes of the isolates sequenced here (Table 2).

Overall, regardless of whether the *panC*, GTDB, 2020 GSB, or pseudo-gene flow unit assignment frameworks were used, *B. cereus s.l.* strains isolated from meat and poultry products in South Africa were considerably diverse and represented multiple genomospecies (Figure 2, Tables 1 and 2, and Supplemental Table S1). Additionally, using PubMLST's seven-gene MLST scheme for *B. cereus*, 17 of 25 strains (68.0%) encompassed 11 STs, with eight strains (32.0%) assigned to unknown STs (Table 1 and Supplemental Table S1). Due to the considerable genomic diversity observed among isolates sequenced here, major lineages represented by strains sequenced in this study are discussed individually in detail below, largely within the context of *panC* Groups and/or MLST STs, as these frameworks are well-established (31, 32, 35, 36) and likely the most interpretable to readers.

Several *B. cereus s.l.* lineages within *panC* Group IV are distributed across multiple South African provinces. Within *panC* Group IV, the 18 strains sequenced here were partitioned into ten lineages using MLST (referred to hereafter as "MLST-based lineages"; Table 3). Based on (i) the whole-genome phylogeny, (ii) pairwise core single-nucleotide polymorphisms (SNPs) identified within MLST-based lineages, and (iii) ANI values calculated within and between MLST-based lineages, four of ten *panC* Group IV MLST-based lineages contained South African strains sequenced in this study, which were highly similar to at least one other strain at the whole-genome level (Figures 3 and 4 and Table 3). One of these lineages (denoted in Table 3 as Lineage IVA) was composed of three South African strains sequenced in this study (S65, S85, and S87), which were assigned to GTDB's "*B. bombysepticus*" genomospecies and belonged to an unknown ST (Figures 3 and 4 and Table 3). Despite all three genomes being nearly identical (pairwise core SNP distance = 0, pairwise ANI > 99.99; Figure 3 and Table 3), the three strains were isolated from (i) three different establishments and provinces (a processing plant in Limpopo, a retail outlet in Gauteng, and a processing plant in Free State), and (ii) two different types of meat products (two strains from beef wors and one from a poultry frankfurter; Figure 4 and Table 1).

Similar results were observed for *panC* Group IV ST2721 (Lineage IVB in Table 3), which was also assigned to GTDB's "*B. bombysepticus*" genomospecies and contained three strains sequenced in this study (S56, S79, and New_S84; Figures 3 and 4 and Table 3). The three ST2721 genomes were nearly identical (pairwise core SNP distance = 0, pairwise ANI > 99.99; Table 3), despite the fact that the strains originated from three different meat products (one from each of beef wors, beef biltong, and processed beef patties) obtained from three different establishments (from a processing plant in North West province, a retail outlet in North West province, and a retail outlet in Free State, respectively; Figure 4 and Table 1).

194 A third *panC* Group IV lineage of unknown ST, which was assigned to GTDB's *B. cereus*
195 species (Lineage IVD), contained two isolates sequenced in this study (S81 and S88; Figures 3
196 and 4 and Table 3). One strain (S81) had been isolated from beef wors in a processing plant in the
197 North West province; the other (S88) from RTE beef biltong in a retail outlet in Limpopo (Figure
198 4 and Table 1). Both strains were highly similar on a genomic scale (>99.99 ANI) and differed by
199 three core SNPs (Figure 3 and Table 3). For reference, in a previous point source foodborne
200 outbreak caused by *B. cereus s.l.* (37), outbreak isolates could differ by up to seven core SNPs
201 (using the same SNP calling methodology used here) (38).
202 A fourth *panC* Group IV lineage, ST2668, contained three strains sequenced in this study
203 (S53, S63, and S70), which were assigned to GTDB's "*B. thuringiensis_S*" genomospecies
204 (Lineage IVH; Figures 3 and 4 and Table 3). The three strains sequenced here differed by, at most,
205 a single core SNP (Table 3), even though all had been isolated from different meat products
206 (processed beef patties, beef biltong, and processed beef mince) obtained from different
207 establishments (i.e., from retail outlets in each of Gauteng, Free State, and Limpopo, respectively;
208 Figure 4 and Table 1). One publicly available genome associated with a *B. cereus s.l.* strain isolated
209 from grass in KarieDeshe, Israel (39) was additionally assigned to ST2668 (NCBI RefSeq
210 Accession GCF_005217245.1); however, this strain was not closely related to the three nearly
211 identical South African ST2668 strains sequenced in this study (Table 3).
212 The remaining six MLST-based lineages (i.e., Lineages IVC, IVE, IVF, IVG, IVI, and IVJ,
213 corresponding to an unknown ST, ST177, ST2289, ST24, ST1697, and ST1578, respectively;
214 Table 3) contained isolates sequenced in this study, which were not closely related to any other
215 strains at the whole-genome level (Figure 3 and Table 3). Lineages IVC, IVF, and IVI were
216 singleton lineages, which each contained one genome sequenced in this study (S67, S86, and S78,

217 respectively), shared 98.90-99.35 ANI with their closest publicly available neighbors (Figure 3
218 and Table 3). Two lineages (IVG and IVJ) each contained multiple genomes, but only one genome
219 sequenced in this study (i.e., S77 and S51, respectively; Figure 3 and Table 3); for both lineages,
220 the South African genome sequenced here was not identical to any publicly available genomes.
221 The remaining lineage, Lineage IVE (ST177), contained multiple genomes, as well as multiple
222 genomes sequenced in this study (Figure 3 and Table 3). Notably, the two ST177 strains sequenced
223 in this study (S55 and S80), which were assigned to this lineage, were not identical, and were
224 isolated from beef wors from processing plants in the Limpopo and North West provinces,
225 respectively (Table 1), indicating that WGS can be useful for differentiating closely related *B.*
226 *cereus s.l.* genomes within STs.

227 **A *panC* Group III *B. cereus s.l.* lineage with a novel sequence type was detected in beef and**
228 **poultry products from two South African provinces.** Two *panC* Group III *B. cereus s.l.* strains
229 sequenced here were assigned to an unknown ST (S58 and S64; Figure 5 and Table 1). Despite
230 having been isolated from different products (S58 from a raw chicken thigh, S64 from RTE
231 sausage emulsion) from different establishments (retail outlets in Mpumalanga and Western Cape,
232 respectively), the strains were nearly identical (pairwise core SNP distance = 0, pairwise ANI >
233 99.84; Figures 4 and 5 and Table 3), indicating that these two strains share a common source.

234 **A *panC* Group III *B. cereus s.l.* strain with cereulide synthetase-encoding genes belongs to**
235 **the "high-risk" ST26 lineage.** One *panC* Group III *B. cereus s.l.* strain (S66) was assigned to
236 GTDB's *B. paranthracis* genomospecies and possessed cereulide synthetase-encoding genes
237 (Figure 5 and Table 2). This strain, which had been isolated from processed beef mince from a
238 processing plant in Mpumalanga, was a member of ST26 (Figure 5 and Table 2), the ST to which
239 most cereulide-producing *B. cereus s.l.* strains belong (although it should be noted that ST26

240 strains may be capable of producing enterotoxins and causing diarrheal illness, regardless of
241 whether they produce cereulide or not) (38, 40, 41). While members of ST26 are comparatively
242 closely related (>99.52 pairwise ANI), WGS was able to distinguish the South African strain
243 sequenced here from closely related ST26 strains (pairwise core SNP distance >913 relative to
244 publicly available genomes; Figure 5 and Table 3).

245 **A *panC* Group III *B. cereus s.l.* lineage assigned to ST2413 shows evidence of intercontinental**
246 **dissemination.** Two *panC* Group III *B. cereus s.l.* strains sequenced in this study (S59 and S62)
247 were assigned to ST2413 within GTDB's *B. paranthracis* species (Figure 5 and Table 3). Unlike
248 the ST26 strain, which was also assigned to GTDB's *B. paranthracis* genomospecies, neither
249 ST2413 strain possessed cereulide synthetase-encoding genes (Figure 5 and Table 2). Both strains
250 were isolated from raw chicken; however, S59 was isolated from a chicken quarter-leg that had
251 been imported from the Netherlands, and S62 from a chicken thigh sold at a retail outlet in
252 Mpumalanga (Figure 4 and Table 1). Notably, these strains were nearly identical on a genomic
253 scale (pairwise core SNP distance = 1, pairwise ANI = 100.0; Figure 5 and Table 3), despite
254 originating from different continents (i.e., Europe and Africa; Figure 4 and Table 1).

255 **A *panC* Group II *B. cereus s.l.* strain assigned to ST794 is most closely related to a food-**
256 **associated strain responsible for diarrheal illness in Norway.** One *panC* Group II *B. cereus s.l.*
257 strain was sequenced in this study (S57) and was assigned to ST794 within GTDB's *B. wiedmannii*
258 species (Figure 5 and Table 1). Strain S57 had been isolated from RTE beef biltong sold at a retail
259 outlet in Free State (Table 1). When compared to the two publicly available ST794 genomes, S57
260 shared > 99.9 ANI with both publicly available genomes and differed by 31 and 100 core SNPs
261 (relative to genomes with NCBI RefSeq Assembly Accessions GCF_900094845.1 and
262 GCF_007671965.1, respectively; Figure 5 and Table 3). Notably, beef biltong-associated strain

S57 sequenced here was most closely related to strain NVH 0674-98, a psychrotolerant strain that had been isolated in Norway from mashed swedes, which were reportedly responsible for diarrheal foodborne illness (42). The other ST794 strain, DE0555, was an environmental isolate collected in 2018 from Durham, North Carolina in the United States (NCBI BioSample Accession SAMN11792715).

A *panC* Group V *B. cereus s.l.* strain from South African mixed-meat wors most closely resembles a plant-associated strain from the United States. One *panC* Group V *B. cereus s.l.* strain was sequenced in this study (S72) and was assigned to ST223 within GTDB's *B. toyonensis* species (Figure 6 and Tables 1-3). S72 had been isolated in a butchery in Gauteng, from a processed wors composed of a mix of beef, pork, and lamb (Table 1). Strain S72 was most closely related to a publicly available genome, strain AFS092321 (NCBI RefSeq Accession GCF_002552615.1), which had been isolated in 2014 from a tree leaf in North Carolina, United States (NCBI BioSample Accession SAMN07598612) (43); the two strains shared >99.9 ANI and differed by 33 core SNPs (Figure 6 and Table 3).

DISCUSSION

***B. cereus s.l.* lineages can be disseminated inter- and intra-nationally via the food supply chain.** The movement of commodities (e.g., foods, animals, animal products, agricultural products, consumer products) through inter- and intra-national trade can contribute to the global, regional, and local dissemination of microorganisms, including pathogens (44-48). The international agro-food trade specifically plays an increasingly pivotal role in providing food supplies to communities around the globe but can contribute to the dissemination of foodborne pathogens (47, 48).

Consequently, high-resolution technologies such as WGS are being employed increasingly to monitor the spread of pathogens along the food supply chain (49, 50).

Using WGS, we identified six South African *B. cereus s.l.* lineages, which showcased evidence of inter-regional dissemination via the trade and transport of meat products (Figure 4 and Table 3). Notably, one *B. cereus s.l.* lineage showed evidence of intercontinental spread between Europe and Africa: a *B. cereus s.l.* ST2413 strain isolated from raw chicken sold in retail outlets in Mpumalanga, South Africa was identical to a ST2413 strain isolated from chicken imported from the Netherlands. We may hypothesize that the raw chicken sold in Mpumalanga originated from the Netherlands, as the Netherlands was the second-largest exporter of chicken meat products to South Africa in 2014-2016 (i.e., the timeframe in which the strains sequenced here were collected) (51); however, this is merely a hypothesis, as we were unable to confirm this with the retail outlet.

We additionally identified five *B. cereus s.l.* lineages, which showed evidence of inter-provincial spread within South Africa: four *panC* Group IV lineages and one *panC* Group III lineage were each composed of (nearly) identical strains, which were isolated from meat products in two or more South African provinces (Figure 4 and Table 3). Thus, it is likely that strains within each lineage shared a common source; however, a lack of additional metadata and genomes prevents confirmation of this. Overall, these results showcase the utility of WGS for *B. cereus s.l.* source tracking and surveillance, although future studies relying on additional genomes with detailed metadata are needed.

Nomenclatural frameworks that incorporate both genomic and phenotypic data can improve strain-level *B. cereus s.l.* risk assessment. *B. cereus s.l.* species delineation is notoriously challenging, and numerous *B. cereus s.l.* taxonomic frameworks have been proposed (29).

Phenotypic traits historically used for *B. cereus s.l.* species assignment (e.g., motility, colony morphology, ability to cause illness) have long been known to be inconsistent with genome evolution (29, 31, 32). Taxonomies that rely solely on genomic data, however, may lead to incorrect assumptions of a strain's pathogenic potential, particularly when taxonomic labels have deep roots in medicine or industry (29, 34).

For example, within some ANI-based taxonomic frameworks (e.g., GTDB), the *B. anthracis* genomospecies includes *B. cereus s.l.* strains that, historically, would be referred to as "*B. cereus*" or "Group III *B. cereus*"; these strains possess phenotypic characteristics associated with "*B. cereus*" (e.g., as outlined in the U.S. Food and Drug Administration's Bacteriological Analytical Manual) (52, 53), and like "*B. cereus*", they are incapable of causing anthrax (34). These non-anthrax-causing Group III *B. cereus s.l.* strains have been isolated from diverse environments, including food (e.g., milk, egg, whites, spices), consumer products (e.g., baby wipes), and soil, indicating that these organisms are not uncommon in environmental and industrial settings (34). Thus, as WGS becomes more popular in clinical and industrial settings, it is possible that professionals who rely solely on increasingly popular genomic methods for taxonomic delineation (e.g., GTDB, ANI-based comparisons to species type strains) may incorrectly assume that these strains can cause anthrax, due to the "*B. anthracis*" species labels that some taxonomic classification programs produce (29, 34). Here, during routine surveillance of meat products in South Africa, we identified two *panC* Group III *B. cereus s.l.* strains, which did not possess anthrax toxin- or capsule-encoding genes and did not belong to the classic, "clonal" *B. anthracis* lineage associated with anthrax disease (34, 54). These strains would be classified as "*B. cereus*" or "Group III *B. cereus*" using standard microbiological assays (52, 53); however, these strains were assigned to the "*B. anthracis*" genomospecies using GTDB and similar ANI-based methods (Table 1).

Referring to these strains as "*B. anthracis*" would be misleading, as they cannot cause anthrax; thus, this potential miscommunication could have disastrous public health and industrial consequences.

We additionally isolated three *B. cereus s.l.* strains from beef and poultry products, which were assigned to the "*B. paranthracis*" genomospecies via GTDB and similar ANI-based methods (Table 1). As noted previously, "*B. paranthracis*" was proposed as a "novel" species in 2017 (55) but was later found to encompass the well-known foodborne pathogen "emetic *B. cereus*" within its genomospecies boundary (29, 37, 38). One of the "*B. paranthracis*" strains isolated here indeed possessed cereulide synthetase-encoding genes and belonged to ST26 (Table 2), the ST that encompasses most *B. cereus s.l.* strains capable of producing emetic toxin (38). This strain thus likely poses a food safety threat and would most likely be referred to as "emetic *B. cereus*" in clinical or industrial settings. Referring to this strain as "*B. paranthracis*" could be misleading to researchers, clinicians, and other professionals who are not well-versed and up to date in *B. cereus s.l.* taxonomy (29, 34, 38). However, not all "*B. paranthracis*" strains are capable of producing emetic toxin: here, two ST2413 strains isolated from poultry were assigned to the "*B. paranthracis*" genomospecies but did not possess cereulide synthetase-encoding genes (Table 2), indicating that these strains cannot cause emetic intoxication. Thus, differentiating potentially emetic from non-emetic strains is critical for informing public health and food safety efforts.

Recently, we proposed a standardized nomenclatural framework for *B. cereus s.l.* (i.e., the 2020 GSB framework), which can utilize genomic, genetic, and/or phenotypic information for taxonomic classification (33, 34). Importantly, the 2020 GSB framework relies on a standardized collection of biovars (i.e., biovar Anthracis, Emeticus, and Thuringiensis), which can be applied to individual *B. cereus s.l.* strains to convey phenotypes of clinical and/or industrial importance

(i.e., ability to produce anthrax, emetic, and insecticidal toxins, respectively) (33, 34). Within this framework, the absence of the Anthracis biovar term denotes that *B. cereus s.l.* strains sequenced here *cannot* produce anthrax toxin, while the presence/absence of the Emeticus biovar term differentiates cereulide-producing "*B. paranthracis*" from non-cereulide-producing strains (Table 2). While the 2020 GSB framework provides a standardized set of *B. cereus s.l.* genomospecies (Tables 1 and 2) (33, 34), researchers and other professionals may prefer to use more established names for lineages (e.g., obtained via MLST, *panC* group assignment); biovar terms can thus be appended to *B. cereus s.l.* lineage names (e.g., the ST26 strain sequenced here, which possesses cereulide synthetase, can be referred to as "*B. cereus s.l.* ST26 biovar Emeticus"). Overall, standardized taxonomic frameworks that can incorporate both genomic/genetic and phenotypic information may improve strain-level risk evaluation of *B. cereus s.l.*

WGS may improve *B. cereus s.l.* genomic surveillance, traceback investigations, and source tracking. WGS has been shown to improve surveillance and source tracking efforts for numerous foodborne pathogens, including *Escherichia coli, Salmonella enterica*, and *Listeria monocytogenes* (50, 56). While the amount of publicly available WGS data derived from members of *B. cereus s.l.* is increasing (34), efforts to sequence the genomes of food-associated *B. cereus s.l.* strains are lagging relative to other foodborne pathogens. Here, we showed that WGS can conceptually be used for *B. cereus s.l.* surveillance and source tracking; however, our study is limited by a lack of publicly available (i) genomic data and (ii) corresponding metadata associated with *B. cereus s.l.* strains. For example, we identified two identical *B. cereus s.l.* ST2413 strains, which were present in both Dutch and South African raw poultry. However, due to a lack of additional publicly available ST2413 *B. cereus s.l.* genomes, we were unable to gain additional insights into exactly where this lineage originated. Thus, future *B. cereus s.l.* surveillance and

377 WGS initiatives in clinical, industrial, and environmental settings are needed to improve *B. cereus*
378 *s.l.* source tracking and traceback efforts. Furthermore, it is essential that data and metadata
379 obtained in such future initiatives are made publicly available, as international sharing of WGS
380 data can decrease both the amount of time required to solve foodborne outbreaks and the public
381 health burden caused by foodborne pathogens (57). Overall, the proof-of-concept study detailed
382 here highlights the benefits of WGS for *B. cereus s.l.* surveillance and source tracking, even among
383 closely related lineages, and future studies will benefit from increasingly available publicly
384 available WGS data and metadata.
385
386 **MATERIALS AND METHODS**
387 **Isolate selection and whole-genome sequencing.** A subset of 34 isolates were selected from a
388 total of 79 *B. cereus s.l.* isolates from our previous study (26) using simple random sampling
389 without replacement (58) via random numbers generated in Microsoft Excel. Culturing and
390 genomic DNA extraction was performed as described previously (26), using the High Pure PCR
391 Template preparation kit (Roche, Germany). WGS of selected isolates was performed at the
392 Biotechnology Platform, Agricultural Research Council, Onderstepoort, South Africa. DNA
393 libraries were prepared using TruSeq and Nextera DNA library preparation kits (Illumina, San
394 Diego, CA, USA), followed by sequencing on HiSeq and MiSeq instruments (Illumina, San Diego,
395 CA, USA).
396 **Data pre-processing and quality control.** Paired-end reads associated with each of the 34 isolates
397 were supplied as input to Trimmomatic v0.38 (59), which was used to remove Illumina adapters
398 and leading and trailing low-quality/ambiguous bases (LEADING:3 and TRAILING:3,
399 respectively); reads with average per-base quality scores <15 within a 4 bp sliding window

(SLIDINGWINDOW:4:15) were additionally cut, and reads with length < 36 bp were removed. FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to assess the quality of the resulting trimmed paired-end reads.

SKESA v2.4.0 (60) was used to assemble each genome, using the trimmed paired-end reads as input and default settings. SPAdes v3.13.1 (61) was additionally used to assemble each genome in "careful" mode, using trimmed paired-end reads as input. QUAST v4.5 (62) was used to assess the quality of each resulting assembly, and CheckM v1.1.3 (63) was used to evaluate genome contamination/completeness. MultiQC v1.8 (64) was used to assess genome quality in aggregate. Assemblies produced using SKESA were used in all subsequent steps, as they were of higher quality based on metrics produced by QUAST (e.g., N50, number of contigs). Genomes with (i) < 95% completeness (via CheckM), (ii) > 5% contamination (via CheckM), and/or (iii) N50 < 20 Kbp were considered to be of low quality and were excluded (n = 4), yielding a preliminary set of 30 genomes used in subsequent analyses.

***In silico* typing and taxonomic characterization.** BTyper3 v3.1.1 (33) was used to characterize each assembled genome (see section "Data pre-processing and quality control" above) using: (i) ANI-based genomospecies, (ii) ANI-based subspecies, and (iii) biovar assignment, using a standardized nomenclatural framework for *B. cereus s.l.* (34) and dependencies FastANI v1.31 (54) and BLAST v2.9.0 (65); (iv) ANI-based pseudo-gene flow unit assignment (33) (also via FastANI); (v) *in silico* seven-gene MLST using the PubMLST *B. cereus* database (accessed 25 October 2020); (vi) *panC* phylogenetic group assignment, using an adjusted eight-group (Group I-VIII) framework (33). All aforementioned analyses were performed using default settings, as well as with virulence factor minimum coverage thresholds lowered to 0% (--virulence_coverage 0) to confirm virulence factor absence. Because BTyper3 uses a conservative approach for Bt toxin

423 gene detection, the command-line implementation of BtToxin_scanner v1.0
424 (BtToxin_scanner2.pl) was used to identify Bt toxin genes in each genome using default settings
425 (66).

426 Taxonomic classification of assembled genomes was additionally performed using GTDB
427 Release 05-RS95 (17 July 2020) and GTDB-Tk v. 1.3.0 (i.e., "GTDB R95"), using GTDB-Tk's
428 "classify_wf" workflow (67-69). Notably, five genomes were assigned to species outside of *B.*
429 *cereus s.l.* (i.e., three genomes classified as *Escherichia flexneri,* one as *Escherichia dysenteriae*,
430 and one as *Staphylococcus saprophyticus* via GTDB-Tk) and were thus excluded, yielding a final
431 set of 25 *B. cereus s.l.* genomes used in subsequent analyses (Table 1 and Supplemental Table S1).
432 **Phylogenomic comparison of South African *B. cereus s.l.* genomes to *B. cereus s.l.* species**
433 **type strains.** Prokka v1.14.6 (70) was used to annotate each of the 25 *B. cereus s.l.* genomes
434 sequenced in this study. Protein coding sequences derived from the type strain genomes of each
435 of the 23 validly published and effective *B. cereus s.l.* species (accessed 28 August 2021) were
436 downloaded from NCBI's RefSeq Assembly database (see Table 1 of Méndez Acevedo, et al. for
437 all type strain accessions) (71). OrthoFinder v2.5.2 (72, 73) was used to identify orthologues
438 among protein coding sequences associated with all 47 genomes (25 *B. cereus s.l.* genomes
439 sequenced in this study, plus 23 *B. cereus s.l.* species type strain genomes), using MAFFT v7.475
440 (74, 75) for sequence alignment and RAxML-NG v1.0.2 (76) for phylogeny construction.

441 The resulting amino acid sequence alignment was supplied as input to IQ-TREE v1.5.4
442 (77), which was used to construct a maximum likelihood (ML) phylogeny, using 1,000 replicates
443 of the ultrafast bootstrap approximation (78), plus the optimal amino acid substitution model
444 selected using ModelFinder (i.e., the general matrix model with empirical amino acid frequencies
445 and the FreeRate model with six categories; JTT+F+R6) (79-82). The resulting phylogeny was

446 rooted using effective species "*B. manliponensis*" (i.e., the most distant recognized member of *B.*
447 *cereus s.l.*) (83) and annotated using the bactaxR package (34) in R v4.1.2 (84).

448 **Acquisition of publicly available *B. cereus s.l.* genomes.** All assembled genomes submitted to
449 the National Center for Biotechnology Information (NCBI) RefSeq database (85, 86) as one of 23
450 validly published or effective *B. cereus s.l.* species (i.e., *albus, anthracis,* "*bingmayongensis*",
451 *cereus,* "*clarus*", *cytotoxicus, fungorum,* "*gaemokensis*", *luti,* "*manliponensis*", *mobilis, mycoides,*
452 *nitratireducens, pacificus, paramycoides, paranthracis, proteolyticus, pseudomycoides,*
453 *thuringiensis, toyonensis, tropicus, weihenstephanensis, wiedmannii*) (1, 29, 55, 71, 83, 87-99)
454 were downloaded (n = 2,733; accessed 20 March 2021). QUAST and CheckM were used to assess
455 the quality of each assembled genome (see section "Data pre-processing and quality control"
456 above), and BTyper3 (using default settings) and GTDB-Tk were used for typing and/or taxonomic
457 assignment as described above (see section "*In silico* typing and taxonomic characterization"). The
458 rentrez package (v1.2.3) was used to download metadata associated with each genome's
459 BioSample in R v3.6.1 (84, 100, 101). Publicly available genomes meeting all of the following
460 quality thresholds were used in subsequent analyses (n = 2,664; Supplemental Table S2): (i) > 95
461 % completeness (via CheckM), (ii) < 5% contamination (via CheckM), (iii) N50 > 20 Kbp (via
462 QUAST), (iv) composed of < 1,000 contigs (via QUAST).

463 **Acquisition of genomes from a study of *B. thuringiensis* outbreaks.** All sequencing reads
464 associated with isolates from a previous study of outbreaks caused by *B. thuringiensis* in France
465 (102) were downloaded from NCBI's Sequence Read Archive (SRA) database using the SRA
466 Toolkit v 2.8.2 (103, 104). Genomic data for all 171 isolates were pre-processed, assembled, and
467 taxonomically classified as described above, with genomes assembled using SKESA used in
468 subsequent steps (see sections "Data pre-processing and quality control" and "*In silico* typing and

taxonomic characterization" above). Four genomes did not meet the quality thresholds employed in this study (see section "Acquisition of publicly available *B. cereus s.l.* genomes" above) and were thus excluded, yielding 167 genomes from the study, which were used in subsequent analyses (Supplemental Table S3).

Acquisition of genomes derived from strains isolated in conjunction with a previous outbreak caused by emetic *B. cereus s.l.* The genomes of 33 *B. cereus s.l.* strains isolated in conjunction with a 2016 emetic outbreak in New York State (United States) were downloaded, preprocessed, and assembled as described previously (37). The quality of each of the 33 genomes was assessed as described above (see section "Data pre-processing and quality control" above), and all genomes underwent taxonomic classification and typing as described above (see section and "*In silico* typing and taxonomic characterization" above). Two genomes did not meet the quality thresholds employed in this study (see section "Acquisition of publicly available *B. cereus s.l.* genomes" above) and were thus excluded, yielding 31 genomes from the study, which were used in subsequent analyses (Supplemental Table S4).

Within-group phylogeny construction. The 25 *B. cereus s.l.* strains sequenced here spanned four major phylogenetic groups, based on their *panC* sequence (i.e., *panC* Groups II, III, IV, and V; Table 1). Thus, phylogenies were constructed using all genomes assigned to each of the following major lineages: (i) *panC* Group IV (Figure 3), (ii) *panC* Groups II and III (Figure 5), and (iii) *panC* Group V (Figure 6), which are equivalent to the (i) *B. cereus s.s.*, (ii) *B. mosaicus*, and (iii) *B. toyonensis* genomospecies within the 2020 GSB taxonomic framework (33), respectively (*panC* Group II and III genomes were aggregated, due to the fact that these lineages are closely related and polyphyletic; Figure 5).

For each of the three major lineages, Prokka was used to annotate each genome; the resulting GFF files associated with each genome were supplied as input to Panaroo v1.2.8 (105), which was used to partition genes into core- and pan-genome orthologous gene clusters, using the following parameters (all other parameters were set to their default values): (i) "strict" mode (--clean-mode strict), (ii) core genome alignment using MAFFT (-a core --aligner mafft), (iii) a core genome sample threshold of 95% (--core_threshold 0.95). The resulting core genome (nucleotide) alignment was queried using snp-sites v2.5.1 (106), which was used to identify (i) core SNPs and (ii) constant sites among all genomes in the major lineage. The resulting core SNP alignment was supplied as input to IQ-TREE v1.5.4, which was used to construct a ML phylogeny using the General Time-Reversible (GTR) nucleotide substitution model (107), one thousand replicates of the ultrafast bootstrap approximation (78), and an ascertainment bias correction obtained using constant sites output by snp-sites.

For each of the three major lineages, all aforementioned steps were repeated, with the addition of an outgroup genome: for the *panC* Group IV phylogeny, *panC* Group III *B. anthracis* str. Ames Ancestor was used as an outgroup (NCBI RefSeq Accession GCF_000008445.1); for the *panC* Groups II/III and *panC* Group V phylogenies, *panC* Group IV *B. cereus* str. ATCC 14579 was used as an outgroup (NCBI RefSeq Accession GCF_006094295.1). Additionally, only genomes with detailed metadata (i.e., a reported year of isolation, isolation source, and geographic location) were included in this analysis (Supplemental Tables S1-S4). Each of the three resulting phylogenies were rooted and scaled using LSD2 v1.4.2.2 (108) and the following parameters: (i) tip dates corresponding to the year of isolation associated with each genome; (ii) constrained mode (-c), with the root estimated using constraints on all branches (-r as); (iii) variances calculated using input branch lengths (-v 1); (iv) 1,000 samples for calculating confidence intervals for

514 estimated dates (-f 1000); (v) a sequence length of 5,500,000 (-s 5500000); (vi) rooting along the
515 outgroup genome (-g -k). The resulting phylogenies were annotated using the bactaxR package in
516 R.

517 **Delineation of MLST lineages and identification of closely related and "identical" genomes.**
518 FastANI v1.31 was used to calculate ANI values between each of the 25 *B. cereus s.l.* genomes
519 sequenced in this study (i.e., as a query genome), and all genomes assigned to the *panC* Group of
520 the query genome (*panC* Groups II and III were aggregated); genomes were then grouped into
521 lineages based on STs assigned using seven-gene MLST (see section "*In silico* typing and
522 taxonomic characterization" above). For each of the resulting MLST lineages, FastANI was used
523 to calculate pairwise ANI values between all genomes within the MLST lineage (Table 3).

524 For each MLST lineage, Snippy v4.6.0 (https://github.com/tseemann/snippy) was used to
525 identify core SNPs among all genomes assigned to the respective MLST lineage, using (i) a
526 genome sequenced in this study as a reference genome (Table 3); (ii) paired-end reads associated
527 with each genome as input (for the genomes sequenced in this study, as well as the genomes from
528 the Bonis, et al. study and the New York State outbreak study) (37, 102) and/or assembled genomes
529 as input (for NCBI genomes); (iii) default settings. For MLST lineages with more than four
530 genomes (i.e., ST24, ST26, ST177, ST223, ST1578), Gubbins v3.1.3 (109) was used to remove
531 recombination, and core SNPs were identified within the resulting filtered alignment using snp-
532 sites. For all MLST lineages, pairwise core SNP distances were calculated within the MLST
533 lineage (i) among all genomes, (ii) among genomes sequenced in this study, and (iii) between
534 genomes sequenced in this study and publicly available genomes (Table 3), using the dist.gene
535 function in the ape package (110, 111) in R.

Data availability. Paired-end Illumina reads associated with the 25 *B. cereus s.l.* isolates sequenced in this study have been deposited in NCBI's SRA database under BioProject Accession PRJNA798224. Metadata and quality information for all genomes queried in this study are available in Supplemental Table S1 (the 25 isolates sequenced in this study) and Supplemental Tables S2-S4 (all publicly available genomes).

ACKNOWLEDGMENTS

The following organisations are acknowledged for their contributions:

- Gauteng Department of Agriculture and Rural Development (GDRAD) for project funding and the use of data for this study.
- The authors are grateful to the Agricultural Research Council: Onderstepoort Veterinary Research for providing all research facilities.
- Our collaborator at EMBL
- Figures 1 and 4 were created with BioRender.com.

FUNDING

Funding for this project was provided by the Gauteng Department of Agriculture and Rural Development (GDRAD).

AUTHOR CONTRIBUTIONS

LC designed and carried out all computational analyses. IM and LC conceptualized the study. RP and IM sourced the funding for the project. RP supervised the sequencing of the isolates, while MM and AA performed all culturing work and DNA extractions. LC and IM co-wrote the manuscript, with input from all authors. All authors contributed to the article and approved the submitted version.

582 **REFERENCES**

583 1. Stenfors Arnesen LP, Fagerlund A, Granum PE. 2008. From soil to gut: *Bacillus cereus*
584 and its food poisoning toxins. FEMS Microbiol Rev 32:579-606.

585 2. Jouzani GS, Valijanian E, Sharafi R. 2017. *Bacillus thuringiensis*: a successful
586 insecticide with new environmental features and tidings. Appl Microbiol Biotechnol
587 101:2691-2711.

588 3. Raymond B, Federici BA. 2017. In defence of *Bacillus thuringiensis*, the safest and most
589 successful microbial insecticide available to humanity—a response to EFSA. FEMS
590 Microbiology Ecology 93.

591 4. Azizoglu U. 2019. *Bacillus thuringiensis* as a Biofertilizer and Biostimulator: a Mini-
592 Review of the Little-Known Plant Growth-Promoting Properties of Bt. Curr Microbiol
593 76:1379-1385.

594 5. Elshaghabee FMF, Rokana N, Gulhane RD, Sharma C, Panwar H. 2017. *Bacillus* As
595 Potential Probiotics: Status, Concerns, and Future Perspectives. Front Microbiol 8:1490.

596 6. Kantas D, Papatsiros VG, Tassis PD, Giavasis I, Bouki P, Tzika ED. 2015. A feed
597 additive containing *Bacillus toyonensis* (Toyocerin((R))) protects against enteric
598 pathogens in postweaning piglets. J Appl Microbiol 118:727-38.

599 7. Jovanovic J, Ornelis VFM, Madder A, Rajkovic A. 2021. *Bacillus cereus* food
600 intoxication and toxicoinfection. Compr Rev Food Sci Food Saf 20:3719-3761.

601 8. Baldwin VM. 2020. You Can't *B. cereus* - A Review of *Bacillus cereus* Strains That
602 Cause Anthrax-Like Disease. Front Microbiol 11:1731.

603 9. Ehling-Schulz M, Lereclus D, Koehler TM. 2019. The *Bacillus cereus* Group: *Bacillus*
604 Species with Pathogenic Potential. Microbiol Spectr 7.

605 10. Moayeri M, Leppla SH, Vrentas C, Pomerantsev AP, Liu S. 2015. Anthrax Pathogenesis.
606 Annu Rev Microbiol 69:185-208.

607 11. Pilo P, Frey J. 2018. Pathogenicity, population genetics and dissemination of *Bacillus*
608 *anthracis*. Infect Genet Evol 64:115-125.

609 12. Ehling-Schulz M, Frenzel E, Gohar M. 2015. Food-bacteria interplay: pathometabolism
610 of emetic *Bacillus cereus.* Front Microbiol 6:704.

611 13. Rouzeau-Szynalski K, Stollewerk K, Messelhausser U, Ehling-Schulz M. 2020. Why be
612 serious about emetic *Bacillus cereus*: Cereulide production and industrial challenges.
613 Food Microbiol 85:103279.

614 14. Dietrich R, Jessberger N, Ehling-Schulz M, Martlbauer E, Granum PE. 2021. The Food
615 Poisoning Toxins of *Bacillus cereus.* Toxins (Basel) 13.

616 15. Jessberger N, Dietrich R, Granum PE, Martlbauer E. 2020. The *Bacillus cereus* Food
617 Infection as Multifactorial Process. Toxins (Basel) 12.

618 16. Bottone EJ. 2010. *Bacillus cereus*, a volatile human pathogen. Clin Microbiol Rev
619 23:382-98.

17. Glasset B, Herbin S, Granier SA, Cavalie L, Lafeuille E, Guerin C, Ruimy R, Casagrande-Magne F, Levast M, Chautemps N, Decousser JW, Belotti L, Pelloux I, Robert J, Brisabois A, Ramarao N. 2018. *Bacillus cereus*, a serious cause of nosocomial infections: Epidemiologic and genetic survey. PLoS One 13:e0194346.

18. Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, Dopfer D, Fazil A, Fischer-Walker CL, Hald T, Hall AJ, Keddy KH, Lake RJ, Lanata CF, Torgerson PR, Havelaar AH, Angulo FJ. 2015. World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. PLoS Med 12:e1001921.

19. Konuma H, Shinagawa K, Tokumaru M, Onoue Y, Konno S, Fujino N, Shigehisa T, Kurata H, Kuwabara Y, Lopes CAM. 1988. Occurrence of *Bacillus cereus* in Meat Products, Raw Meat and Meat Product Additives. J Food Prot 51:324-326.

20. Tewari A, Singh SP, Singh R. 2015. Incidence and enterotoxigenic profile of *Bacillus cereus* in meat and meat products of Uttarakhand, India. J Food Sci Technol 52:1796-801.

21. Kong L, Yu S, Yuan X, Li C, Yu P, Wang J, Guo H, Wu S, Ye Q, Lei T, Yang X, Zhang Y, Wei X, Zeng H, Zhang J, Wu Q, Ding Y. 2021. An Investigation on the Occurrence and Molecular Characterization of *Bacillus cereus* in Meat and Meat Products in China. Foodborne Pathog Dis 18:306-314.

22. Smith DP, Berrang ME, Feldner PW, Phillips RW, Meinersmann RJ. 2004. Detection of *Bacillus cereus* on selected retail chicken products. J Food Prot 67:1770-3.

23. Osman KM, Kappell AD, Orabi A, Al-Maary KS, Mubarak AS, Dawoud TM, Hemeg HA, Moussa IMI, Hessain AM, Yousef HMY, Hristova KR. 2018. Poultry and beef meat as potential seedbeds for antimicrobial resistant enterotoxigenic *Bacillus* species: a materializing epidemiological and potential severe health hazard. Sci Rep 8:11600.

24. Zeighami H, Nejad-Dost G, Parsadanians A, Daneshamouz S, Haghi F. 2020. Frequency of hemolysin BL and non-hemolytic enterotoxin complex genes of *Bacillus cereus* in raw and cooked meat samples in Zanjan, Iran. Toxicol Rep 7:89-92.

25. Yu S, Yu P, Wang J, Li C, Guo H, Liu C, Kong L, Yu L, Wu S, Lei T, Chen M, Zeng H, Pang R, Zhang Y, Wei X, Zhang J, Wu Q, Ding Y. 2020. A Study on Prevalence and Characterization of *Bacillus cereus* in Ready-to-Eat Foods in China. Frontiers in Microbiology 10.

26. Madoroba E, Magwedere K, Chaora NS, Matle I, Muchadeyi F, Mathole MA, Pierneef R. 2021. Microbial Communities of Meat and Meat Products: An Exploratory Analysis of the Product Quality and Safety at Selected Enterprises in South Africa. Microorganisms 9.

27. Nortjé GL, Vorster SM, Greebe RP, Steyn PL. 1999. Occurrence of *Bacillus cereus* and *Yersinia enterocolitica* in South African retail meats. Food Microbiology 16:213-217.

28. Shale K, Malebo NJ. 2011. QUANTIFICATION AND ANTIBIOTIC SUSCEPTIBILITY PROFILES OF *STAPHYLOCOCCUS AUREUS* AND *BACILLUS CEREUS* STRAINS ISOLATED FROM BILTONG. Journal of Food Safety 31:559-569.

661 29. Carroll LM, Cheng RA, Wiedmann M, Kovac J. 2021. Keeping up with the *Bacillus*
662 *cereus* group: taxonomy through the genomics era and beyond. Crit Rev Food Sci Nutr
663 doi:10.1080/10408398.2021.1916735:1-26.

664 30. Liu Y, Lai Q, Goker M, Meier-Kolthoff JP, Wang M, Sun Y, Wang L, Shao Z. 2015.
665 Genomic insights into the taxonomic status of the *Bacillus cereus* group. Sci Rep
666 5:14082.

667 31. Guinebretiere MH, Thompson FL, Sorokin A, Normand P, Dawyndt P, Ehling-Schulz M,
668 Svensson B, Sanchis V, Nguyen-The C, Heyndrickx M, De Vos P. 2008. Ecological
669 diversification in the *Bacillus cereus* Group. Environ Microbiol 10:851-65.

670 32. Guinebretiere MH, Velge P, Couvert O, Carlin F, Debuyser ML, Nguyen-The C. 2010.
671 Ability of *Bacillus cereus* group strains to cause food poisoning varies according to
672 phylogenetic affiliation (groups I to VII) rather than species affiliation. J Clin Microbiol
673 48:3388-91.

674 33. Carroll LM, Cheng RA, Kovac J. 2020. No Assembly Required: Using BTyper3 to
675 Assess the Congruency of a Proposed Taxonomic Framework for the *Bacillus cereus*
676 Group With Historical Typing Methods. Frontiers in Microbiology 11.

677 34. Carroll LM, Wiedmann M, Kovac J. 2020. Proposal of a Taxonomic Nomenclature for
678 the *Bacillus cereus* Group Which Reconciles Genomic Definitions of Bacterial Species
679 with Clinical and Industrial Phenotypes. mBio 11:e00034-20.

680 35. Jolley KA, Maiden MC. 2010. BIGSdb: Scalable analysis of bacterial genome variation
681 at the population level. BMC Bioinformatics 11:595.

682 36. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics:
683 BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res
684 3:124.

685 37. Carroll LM, Wiedmann M, Mukherjee M, Nicholas DC, Mingle LA, Dumas NB, Cole
686 JA, Kovac J. 2019. Characterization of Emetic and Diarrheal *Bacillus cereus* Strains
687 From a 2016 Foodborne Outbreak Using Whole-Genome Sequencing: Addressing the
688 Microbiological, Epidemiological, and Bioinformatic Challenges. Front Microbiol
689 10:144.

690 38. Carroll LM, Wiedmann M. 2020. Cereulide Synthetase Acquisition and Loss Events
691 within the Evolutionary History of Group III *Bacillus cereus Sensu Lato* Facilitate the
692 Transition between Emetic and Diarrheal Foodborne Pathogens. mBio 11.

693 39. Bucher T, Keren-Paz A, Hausser J, Olender T, Cytryn E, Kolodkin-Gal I. 2019. An
694 active beta-lactamase is a part of an orchestrated cell wall stress resistance network of
695 *Bacillus subtilis* and related rhizosphere species. Environ Microbiol 21:1068-1085.

696 40. Ehling-Schulz M, Svensson B, Guinebretiere MH, Lindback T, Andersson M, Schulz A,
697 Fricker M, Christiansson A, Granum PE, Martlbauer E, Nguyen-The C, Salkinoja-
698 Salonen M, Scherer S. 2005. Emetic toxin formation of *Bacillus cereus* is restricted to a
699 single evolutionary lineage of closely related strains. Microbiology 151:183-197.

700 41. Glasset B, Herbin S, Guillier L, Cadel-Six S, Vignaud ML, Grout J, Pairaud S, Michel V,
701 Hennekinne JA, Ramarao N, Brisabois A. 2016. *Bacillus cereus*-induced food-borne

outbreaks in France, 2007 to 2014: epidemiology and genetic characterisation. Euro Surveill 21.

42. Guinebretiere MH, Loux V, Martin V, Nicolas P, Sanchis V, Broussolle V. 2017. Draft Genome Sequences of 18 Psychrotolerant and 2 Thermotolerant Strains Representative of Particular Ecotypes in the *Bacillus cereus* Group. Genome Announc 5.

43. Grubbs KJ, Bleich RM, Santa Maria KC, Allen SE, Farag S, AgBiome T, Shank EA, Bowers AA. 2017. Large-Scale Bioinformatics Analysis of *Bacillus* Genomes Uncovers Conserved Roles of Natural Products in Bacterial Physiology. mSystems 2.

44. Stein RA, Chirilã M. 2017. Chapter 3 - Routes of Transmission in the Food Chain, p 65-103. *In* Dodd CER, Aldsworth T, Stein RA, Cliver DO, Riemann HP (ed), Foodborne Diseases (Third Edition) doi:https://doi.org/10.1016/B978-0-12-385007-2.00003-6. Academic Press.

45. Ahn J-W, Rhodes MT. 2021. Examining Pathogen-Based Import Refusals: Trends and Analysis From 2002 to 2019. U.S. Department of Agriculture, Economic Research Service, U.S. Department of Agriculture ERS,

46. Rohr JR, Barrett CB, Civitello DJ, Craft ME, Delius B, DeLeo GA, Hudson PJ, Jouanard N, Nguyen KH, Ostfeld RS, Remais JV, Riveau G, Sokolow SH, Tilman D. 2019. Emerging human infectious diseases and the links to global food production. Nat Sustain 2:445-456.

47. Ristaino JB, Anderson PK, Bebber DP, Brauman KA, Cunniffe NJ, Fedoroff NV, Finegold C, Garrett KA, Gilligan CA, Jones CM, Martin MD, MacDonald GK, Neenan P, Records A, Schmale DG, Tateosian L, Wei Q. 2021. The persistent threat of emerging plant disease pandemics to global food security. Proc Natl Acad Sci U S A 118.

48. Ercsey-Ravasz M, Toroczkai Z, Lakner Z, Baranyi J. 2012. Complexity of the international agro-food trade network and its impact on food safety. PLoS One 7:e37810.

49. Brown E, Dessai U, McGarry S, Gerner-Smidt P. 2019. Use of Whole-Genome Sequencing for Food Safety and Public Health in the United States. Foodborne Pathog Dis 16:441-450.

50. Brown B, Allard M, Bazaco MC, Blankenship J, Minor T. 2021. An economic evaluation of the Whole Genome Sequencing source tracking program in the U.S. PLoS One 16:e0258262.

51. Makgopa M. 2020. South Africa Lifts Ban On Poultry Imports From The Netherlands. United States Department of Agriculture (USDA) Foreign Agricultural Service, Service USDoAUFA,

52. Tallent SM, Knolhoff A, Rhodehamel EJ, Harmon SM, Bennett RW. 2019. Chapter 14: *Bacillus cereus*, Bacteriological Analytical Manual (BAM), 8th ed. Food and Drug Administration.

53. Tallent SM, Kotewicz KM, Strain EA, Bennett RW. 2012. Efficient isolation and identification of *Bacillus cereus* group. J AOAC Int 95:446-51.

741 54. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput
742 ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun
743 9:5114.

744 55. Liu Y, Du J, Lai Q, Zeng R, Ye D, Xu J, Shao Z. 2017. Proposal of nine novel species of
745 the *Bacillus cereus* group. Int J Syst Evol Microbiol 67:2499-2508.

746 56. Rantsiou K, Kathariou S, Winkler A, Skandamis P, Saint-Cyr MJ, Rouzeau-Szynalski K,
747 Amezquita A. 2018. Next generation microbiological risk assessment: opportunities of
748 whole genome sequencing (WGS) for foodborne pathogen surveillance, source tracking
749 and risk assessment. Int J Food Microbiol 287:3-9.

750 57. Pettengill JB, Markell A, Conrad A, Carleton HA, Beal J, Rand H, Musser S, Brown EW,
751 Allard MW, Huffman J, Harris S, Wise M, Locas A. 2020. A multinational listeriosis
752 outbreak and the importance of sharing genomic data. The Lancet Microbe 1:e233-e234.

753 58. Rose S. 2015. Management research : applying the principles, First Edition. ed.
754 Routledge, Taylor & Francis Group, New York.

755 59. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
756 sequence data. Bioinformatics 30:2114-20.

757 60. Souvorov A, Agarwala R, Lipman DJ. 2018. SKESA: strategic k-mer extension for
758 scrupulous assemblies. Genome Biology 19:153.

759 61. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
760 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G,
761 Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its
762 applications to single-cell sequencing. J Comput Biol 19:455-77.

763 62. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for
764 genome assemblies. Bioinformatics 29:1072-5.

765 63. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM:
766 assessing the quality of microbial genomes recovered from isolates, single cells, and
767 metagenomes. Genome Res 25:1043-55.

768 64. Ewels P, Magnusson M, Lundin S, Kaller M. 2016. MultiQC: summarize analysis results
769 for multiple tools and samples in a single report. Bioinformatics 32:3047-8.

770 65. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.
771 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

772 66. Ye W, Zhu L, Liu Y, Crickmore N, Peng D, Ruan L, Sun M. 2012. Mining new crystal
773 protein genes from *Bacillus thuringiensis* on the basis of mixed plasmid-enriched genome
774 sequencing and a computational pipeline. Appl Environ Microbiol 78:4795-801.

775 67. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify
776 genomes with the Genome Taxonomy Database. Bioinformatics
777 doi:10.1093/bioinformatics/btz848.

778 68. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA,
779 Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny
780 substantially revises the tree of life. Nat Biotechnol 36:996-1004.

781 69. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. 2020. A
782 complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol
783 38:1079-1086.

784 70. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics
785 30:2068-9.

786 71. Mendez Acevedo M, Carroll LM, Mukherjee M, Mills E, Xiaoli L, Dudley EG, Kovac J.
787 2020. Novel Effective *Bacillus cereus* Group Species "*Bacillus clarus*" Is Represented by
788 Antibiotic-Producing Strain ATCC 21929 Isolated from Soil. mSphere 5.

789 72. Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome
790 comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:157.

791 73. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for
792 comparative genomics. Genome Biol 20:238.

793 74. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid
794 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res
795 30:3059-66.

796 75. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
797 improvements in performance and usability. Mol Biol Evol 30:772-80.

798 76. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast,
799 scalable and user-friendly tool for maximum likelihood phylogenetic inference.
800 Bioinformatics 35:4453-4455.

801 77. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and
802 effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol
803 Evol 32:268-74.

804 78. Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic
805 bootstrap. Mol Biol Evol 30:1188-95.

806 79. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
807 ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods
808 14:587-589.

809 80. Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data
810 matrices from protein sequences. Bioinformatics 8:275-282.

811 81. Soubrier J, Steel M, Lee MSY, Der Sarkissian C, Guindon S, Ho SYW, Cooper A. 2012.
812 The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular
813 Rates. Molecular Biology and Evolution 29:3345-3358.

814 82. Yang Z. 1995. A space-time process model for the evolution of DNA sequences.
815 Genetics 139:993-1005.

816 83. Jung MY, Kim JS, Paek WK, Lim J, Lee H, Kim PI, Ma JY, Kim W, Chang YH. 2011.
817 *Bacillus manliponensis* sp. nov., a new member of the *Bacillus cereus* group isolated
818 from foreshore tidal flat sediment. J Microbiol 49:1027-32.

819 84. R Core Team. 2019. R: A Language and Environment for Statistical Computing, v3.6.1.
820 R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

85. Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:D61-5.

86. Coordinators NR. 2018. Database resources of the National Center for Biotechnology Information. Nucleic acids research 46:D8-D13.

87. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, McVeigh R, O'Neill K, Robbertse B, Sharma S, Soussov V, Sullivan JP, Sun L, Turner S, Karsch-Mizrachi I. 2020. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford) 2020.

88. Sternbach G. 2003. The history of anthrax. The Journal of Emergency Medicine 24:463-467.

89. Frankland GC, Frankland PF, Lankester ER. 1887. XI. Studies on some new micro-organisms obtained from air. Philosophical Transactions of the Royal Society of London (B) 178:257-287.

90. Liu B, Liu GH, Hu GP, Sengonca C, Lin NQ, Tang JY, Tang WQ, Lin YZ. 2014. *Bacillus bingmayongensis* sp. nov., isolated from the pit soil of Emperor Qin's Terra-cotta warriors in China. Antonie Van Leeuwenhoek 105:501-10.

91. Guinebretiere MH, Auger S, Galleron N, Contzen M, De Sarrau B, De Buyser ML, Lamberet G, Fagerlund A, Granum PE, Lereclus D, De Vos P, Nguyen-The C, Sorokin A. 2013. *Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* Group occasionally associated with food poisoning. Int J Syst Evol Microbiol 63:31-40.

92. Liu X, Wang L, Han M, Xue Q-h, Zhang G-s, Gao J, sun X. 2020. *Bacillus fungorum* sp. nov., a bacterium isolated from spent mushroom substrate. International Journal of Systematic and Evolutionary Microbiology 70:1457-1462.

93. Jung MY, Paek WK, Park IS, Han JR, Sin Y, Paek J, Rhee MS, Kim H, Song HS, Chang YH. 2010. *Bacillus gaemokensis* sp. nov., isolated from foreshore tidal flat sediment from the Yellow Sea. J Microbiol 48:867-71.

94. Laubach CA. 1916. Spore-Bearing Bacteria in Dust. Journal of bacteriology 1:493-505.

95. Nakamura LK. 1998. *Bacillus pseudomycoides* sp. nov. Int J Syst Bacteriol 48 Pt 3:1031-5.

96. Milner RJ. 1994. History of *Bacillus thuringiensis*. Agriculture, Ecosystems & Environment 49:9-13.

97. Jimenez G, Urdiain M, Cifuentes A, Lopez-Lopez A, Blanch AR, Tamames J, Kampfer P, Kolsto AB, Ramon D, Martinez JF, Codoner FM, Rossello-Mora R. 2013. 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 36:383-91.

98. Miller RA, Beno SM, Kent DJ, Carroll LM, Martin NH, Boor KJ, Kovac J. 2016. *Bacillus wiedmannii* sp. nov., a psychrotolerant and cytotoxic *Bacillus cereus* group

861 species isolated from dairy foods and dairy environments. Int J Syst Evol Microbiol
862 66:4744-4753.

863 99. Lechner S, Mayr R, Francis KP, Pruss BM, Kaplan T, Wiessner-Gunkel E, Stewart GS,
864 Scherer S. 1998. *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of
865 the *Bacillus cereus* group. Int J Syst Bacteriol 48 Pt 4:1373-82.

866 100. Winter DJ. 2017. rentrez: An R package for the NCBI eUtils API. The R Journal 9:520-
867 526.

868 101. Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E, Karsch-Mizrachi I,
869 Kimelman M, Pruitt KD, Resenchuk S, Tatusova T, Yaschenko E, Ostell J. 2012.
870 BioProject and BioSample databases at NCBI: facilitating capture and organization of
871 metadata. Nucleic Acids Res 40:D57-63.

872 102. Bonis M, Felten A, Pairaud S, Dijoux A, Maladen V, Mallet L, Radomski N, Duboisset
873 A, Arar C, Sarda X, Vial G, Mistou MY, Firmesse O, Hennekinne JA, Herbin S. 2021.
874 Comparative phenotypic, genotypic and genomic analyses of *Bacillus thuringiensis*
875 associated with foodborne outbreaks in France. PLoS One 16:e0246885.

876 103. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database C.
877 2011. The sequence read archive. Nucleic Acids Res 39:D19-21.

878 104. Kodama Y, Shumway M, Leinonen R, International Nucleotide Sequence Database C.
879 2012. The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids
880 Res 40:D54-6.

881 105. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA,
882 Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J. 2020.
883 Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol
884 21:180.

885 106. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-
886 sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom
887 2:e000056.

888 107. Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA
889 sequences. Lectures on mathematics in the life sciences 17:57-86.

890 108. To T-H, Jung M, Lycett S, Gascuel O. 2015. Fast Dating Using Least-Squares Criteria
891 and Algorithms. Systematic Biology 65:82-97.

892 109. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris
893 SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole
894 genome sequences using Gubbins. Nucleic Acids Res 43:e15.

895 110. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in
896 R language. Bioinformatics 20:289-90.

897 111. Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and
898 evolutionary analyses in R. Bioinformatics 35:526-528.

899

Table 1. *B. cereus s.l.* strains sequenced in this study (n = 25).

Strain	Year[a]	Meat Sample Type	Meat Category[b]	Food Animal	Establishment	Province[c]	*panC*[d] Group	MLST ST[e]	GTDB Species[f]	2020 GSB Species[g]
S57	2015	Beef Biltong	RTE	Beef	Retail Outlets	Free State	II	794	*wiedmannii*	*mosaicus*
S58	2015	Chicken Thigh	Raw Intact	Poultry	Retail Outlets	Mpumalanga	III	NA	*anthracis*[h]	*mosaicus*
S59	2016	Chicken 1/4 Leg	Raw Intact	Poultry	Cold Store	Import (NL)	III	2413	*paranthracis*	*mosaicus*
S62	2015	Chicken Thigh	Raw Intact	Poultry	Retail Outlets	Mpumalanga	III	2413	*paranthracis*	*mosaicus*
S64	2016	Sausage Emulsion	RTE	Beef	Retail Outlets	Western Cape	III	NA	*anthracis*[h]	*mosaicus*
S66	2015	Beef Mince	Processed	Beef	Processing Plant	Mpumalanga	III	26	*paranthracis*	*mosaicus* subsp. *cereus*
S51	2015	Beef Mince	Processed	Beef	Retail Outlets	Limpopo	IV	1578	*thuringiensis*	*cereus s.s.*
S53	2015	Beef Patties	Processed	Beef	Retail Outlets	Gauteng	IV	2668	*thuringiensis_S*	*cereus s.s.*
S55	2015	Beef Wors	Processed	Beef	Processing Plant	Limpopo	IV	177	*cereus*	*cereus s.s.*
S56	2015	Beef Wors	Processed	Beef	Processing Plant	North West	IV	2721	*bombysepticus*	*cereus s.s.*
S63	2015	Beef Biltong	Processed	Beef	Retail Outlets	Free State	IV	2668	*thuringiensis_S*	*cereus s.s.*
S65	2015	Beef Wors	Processed	Beef	Processing Plant	Limpopo	IV	NA	*bombysepticus*	*cereus s.s.*
S67	2016	Chicken 1/4 Leg	Raw Intact	Poultry	Cold Store	Import (NL)	IV	NA	*bombysepticus*	*cereus s.s.*
S70	2015	Beef Mince	Processed	Beef	Retail Outlets	Limpopo	IV	2668	*thuringiensis_S*	*cereus s.s.*
S77	2015	Beef Wors	Processed	Beef	Retail Outlets	Gauteng	IV	24	*cereus*	*cereus s.s.*
S78	2015	Beef Biltong	RTE	Beef	Retail Outlets	Gauteng	IV	1697	*thuringiensis_S*	*cereus s.s.*
S79	2015	Beef Biltong	RTE	Beef	Retail Outlets	North West	IV	2721	*bombysepticus*	*cereus s.s.*
S80	2015	Beef Wors	Processed	Beef	Processing Plant	North West	IV	177	*cereus*	*cereus s.s*
S81	2016	Beef Wors	Processed	Beef	Processing Plant	North West	IV	NA	*cereus*	*cereus s.s.*
New_S84	2015	Beef Patties	Processed	Beef	Retail Outlets	Free State	IV	2721	*bombysepticus*	*cereus s.s.*
S85	2015	Beef Wors	Processed	Beef	Retail Outlets	Gauteng	IV	NA	*bombysepticus*	*cereus s.s.*
S86	2015	Sausage Emulsion	RTE	Beef	Retail Outlets	Western Cape	IV	2289	*cereus*	*cereus s.s.*
S87	2015	Frankfurter	Raw Intact	Poultry	Processing Plant	Free State	IV	NA	*bombysepticus*	*cereus s.s.*
S88	2015	Beef Biltong	RTE	Beef	Retail Outlets	Limpopo	IV	NA	*cereus*	*cereus s.s.*
S72	2015	Beef-Pork-Lamb Wors	Processed	Mixed	Butchery	Gauteng	V	223	*toyonensis*	*toyonensis*

[a]Year of isolation

[b]RTE, Ready-to-Eat

[c]Two strains were isolated from meat products imported from the Netherlands (NL)

[d]*panC* phylogenetic group assigned using BTyper3

[e]Sequence type (ST) assigned using the PubMLST seven-gene multi-locus sequence typing (MLST) scheme for *B. cereus* and BTyper3

[f]Genome Taxonomy Database (GTDB) species assigned using GTDB-Tk
[g]Species and subspecies (where applicable), assigned using the 2020 Genomospecies-Subspecies-Biovar (GSB) nomenclatural framework for *B. cereus s.l.* and BTyper3 (see Table 2 for predicted biovar/phenotypic information)
[h]Despite GTDB assigning a species label of "*B. anthracis*", these strains cannot cause anthrax illness, nor do they belong to the classic "*B. anthracis*" lineage most commonly responsible for anthrax illness (34)

914 **Table 2.** Predicted phenotypes of *B. cereus s.l.* strains sequenced in this study (n = 25).

Strain	*panC*[a] Group	MLST ST[b]	GTDB Species[c]	Anthrax Toxin & Capsule Genes[d,e]	Emetic Toxin *cesABCD*[d]	Diarrheal Toxin *nheABC*[d]	Diarrheal Toxin *hblACD*[d]	Diarrheal Toxin *cytK-2*[d]	Bt Toxin Genes[f]	2020 GSB Taxonomy[g,h]
S57	II	794	*wiedmannii*	-	-	+	+	+	-/-/+	
S58	III	NA	*anthracis*[i]	-	-	+	-	+	-/-/+	
S59	III	2413	*paranthracis*	-	-	+	-	-	-/-/-	*B. mosaicus*
S62	III	2413	*paranthracis*	-	-	+	-	-	-/-/+	
S64	III	NA	*anthracis*[i]	-	-	+	-	+	-/-/+	
S66	III	26	*paranthracis*	-	+	+	-	-	-/-/+	*B. mosaicus* subsp. *cereus* bv. Emeticus; *B. cereus* bv. Emeticus; *B.* Emeticus
S51	IV	1578	*thuringiensis*	-	-	+	+	+	-/+/+	
S53	IV	2668	*thuringiensis*	-	-	+	+	+	-/+/-	
S55	IV	177	*cereus*	-	-	+	+	+	-/+/+	
S56	IV	2721	*bombysepticus*	-	-	+	+	+	-/+/-	
S63	IV	2668	*thuringiensis*	-	-	+	+	+	-/+/+	
S65	IV	NA	*bombysepticus*	-	-	+	+	-	+/+/-	
S67	IV	NA	*bombysepticus*	-	-	+	+	+	-/+/+	
S70	IV	2668	*thuringiensis*	-	-	+	+	+	-/+/-	*B. cereus s.s.* bv. Thuringiensis; *B.* Thuringiensis
S77	IV	24	*cereus*	-	-	+	+	+	-/+/+	
S78	IV	1697	*thuringiensis*	-	-	+	+	-	+/+/+	
S79	IV	2721	*bombysepticus*	-	-	+	+	+	-/+/-	
S80	IV	177	*cereus*	-	-	+	+	+	-/+/+	
S81	IV	NA	*cereus*	-	-	+	+	+	-/+/-	
New_S84	IV	2721	*bombysepticus*	-	-	+	+	+	-/+/+	
S85	IV	NA	*bombysepticus*	-	-	+	+	-	+/+/+	
S86	IV	2289	*cereus*	-	-	+	+	+	- /+/+	
S87	IV	NA	*bombysepticus*	-	-	+	+	-	+/+/-	
S88	IV	NA	*cereus*	-	-	+	+	+	-/+/+	
S72	V	223	*toyonensis*	-	-	+	+	-*	-/+/+	*B. toyonensis* bv. Thuringiensis; *B.* Thuringiensis

915 [a]*panC* phylogenetic group assigned using BTyper3

[b]Sequence type (ST) assigned using the PubMLST seven-gene multi-locus sequence typing (MLST) scheme for *B. cereus* and BTyper3

[c]Genome Taxonomy Database (GTDB) species assigned using GTDB-Tk

[d]Viurlence factors detected in each genome using BTyper3 and default thresholds (70% amino acid identity and 80% coverage); presence and absence of virulence factors are denoted by "+" and "-", respectively

[e]All of *cya, lef, pagA* (toxin), *capABCDE*, *hasABC*, and/or *bpsXABCDEFGH* (capsules)

[f]*B. thuringiensis* (Bt) insecticidal toxin-encoding genes detected using (i) BTyper3 (which uses a conservative BLAST-based approach and minimum default amino acid and coverage thresholds of 50 and 70%, respectively), and BtToxin_scanner2's (ii) "old" and (iii) "new" gene detection approaches (detected using default thresholds), each separated by a slash ("/")

[g]Species, subspecies (where applicable), and biovar (where applicable) assigned using the 2020 Genomospecies-Subspecies-Biovar (GSB) nomenclatural framework for *B. cereus s.l.* and BTyper3; multiple taxonomic labels are listed for strains that can be referenced using shorted subspecies and/or biovar notation (separated by a semi-colon)

[h]Biovar Thuringiensis was assigned to genomes in which Bt insecticidal toxin-encoding genes were detected using BtToxin_scanner2's "old" gene detection approach and default settings (which is less conservative than BTyper3's Bt toxin gene detection approach)

[i]Despite GTDB assigning a species label of "*B. anthracis*", these strains cannot cause anthrax illness, nor do they belong to the classic "*B. anthracis*" lineage most commonly responsible for anthrax illness (34)

Table 3. Genomic distances within multi-locus sequence typing (MLST) lineages, which contain strains sequenced in this study.

			# of Genomes				Pairwise SNP Range within MLST Lineage		
MLST Lineage[a]	GTDB Species[b]	MLST ST[c]	Total	Study	ANI Range (Mean)[d]	Reference Genome[e]	Total (Mean)[f]	Within-Study (Mean)[g]	Study-Public (Mean)[h]
IVA*	*bombysepticus*	NA	3	3	99.99-100.0 (100.0)	S85	0-0 (0)	0-0 (0)	NA
IVB*	*bombysepticus*	2721	3	3	99.99-100.0 (100.0)	S79	0-0 (0)	0-0 (0)	NA
IVC	*bombysepticus*	NA	1	1	NA	NA	NA	NA	NA
IVD*	*cereus*	NA	2	2	100.0-100.0 (100.0)	S81	3	3	NA
IVE	*cereus*	177	11	2	99.70-100.0 (99.92)	S80	1-531 (239.2)	16	10-531 (258.11)
IVF	*cereus*	2289	1	1	NA	NA	NA	NA	NA
IVG	*cereus*	24[i,j]	12	1	99.75-100.0 (99.91)	S77	1-1,075 (328.6)	NA	94-1,054 (290.3)
IVH*	thuringiensis_S	2668	4	3	98.90-100.0 (99.49)	S53	0-23,098 (11,549)	0-1 (0.67)	23,097- 23,098 (23,097)
IVI	thuringiensis_S	1697	1	1	NA	NA	NA	NA	NA
IVJ	*thuringiensis*	1578[j]	6	1	98.49-99.92 (98.81)	S51	2,461- 40,845	NA	12,642- 36,103 (29,344)
IIA	*wiedmannii*	794	3	1	99.95-100.0 (99.97)	S57	31-105 (78.67)	NA	31-100 (65.50)
IIIA*	*anthracis*[k]	NA	2	2	99.84-100.0 (99.92)	S64	0	0	NA
IIIB*	*paranthracis*	2413	2	2	100.0-100.0 (100.0)	S59	1	1	NA
IIIC	*paranthracis*	26	77	1	99.52-100.0 (99.84)	S66	0-12,716 (4,797)	NA	913-11,406 (3,494)
VA	*toyonensis*	223[j]	41	1	98.95-99.99 (99.51)	S72	14-10,967 (6,600)	NA	33-10,882 (4,971)

[a]Lineage identifiers (IDs) pertain to clusters displayed in Figures 3 (Group IV), 5 (Groups II and III), and 6 (Group V); IDs with an asterisk contain two or more strains sequenced in this study, which were highly similar on a genomic scale; NA, not available/applicable

[b]Genome Taxonomy Database (GTDB) species assigned using GTDB-Tk

[c]Sequence type (ST) assigned using the PubMLST seven-gene MLST scheme for *B. cereus* and BTyper3

[d]FastANI average nucleotide identity (ANI) range and mean values between all genomes in the lineage (excludes self-comparisons)

[e]Strain sequenced in this study, which was used as a reference genome for single-nucleotide polymorphism (SNP) calling among all genomes within the lineage via Snippy

[f]Pairwise SNP distances calculated among all genomes within the lineage (excludes self-comparisons)

[g]Pairwise SNP distances calculated among all genomes within the lineage that were sequenced in this study (excludes self-comparisons)

[h]Pairwise SNP distances calculated between genomes sequenced in this study and public genomes within the same lineage (excludes self-comparisons)

[i]For polyphyletic ST24, one outlier ST24 genome was excluded (NCBI RefSeq Accession GCF_010580595.1)

[j]ST was polyphyletic

[k]Despite GTDB assigning a species label of "*B. anthracis*", these strains cannot cause anthrax illness, nor do they belong to the classic "*B. anthracis*" lineage most commonly responsible for anthrax illness (34)

953 **FIGURE LEGENDS**

954 **Figure 1.** Geographic origins of *B. cereus s.l.* strains sequenced in this study (*n* = 25).

955

956 **Figure 2.** Maximum likelihood (ML) phylogeny constructed using amino acid sequences derived
957 from the 25 *B. cereus s.l.* isolate genomes sequenced in this study (tip labels colored by
958 geographic origin), plus type strain/species representative genomes of 23 published and effective
959 *B. cereus s.l.* species (gray tip labels). The heatmap to the right of the phylogeny showcases
960 species assignments obtained within the following taxonomic frameworks (from left to right): (i)
961 Genome Taxonomy Database (GTDB) Release 05-RS95 and GTDB-Tk (GTDB R95); (ii)
962 pseudo-gene flow units (GFUs) assigned using BTyper3 (Pseudo-GFU); (iii) genomospecies of
963 the 2020 standardized *B. cereus s.l.* genomospecies/subspecies/biovar (GSB) framework and
964 BTyper3 (2020 GSB); (iv) *panC* Group (I-VIII), assigned using BTyper3 (*panC* Group). The
965 phylogeny was constructed using IQ-TREE, using core orthologs identified among all genomes
966 via OrthoFinder as input. Branch lengths are reported in substitutions per site. Branch labels
967 correspond to branch support percentages obtained using 1,000 replicates of the ultrafast
968 bootstrap approximation (selected for readability). The type strain genome of effective *B. cereus*
969 *s.l.* species "*B. manliponensis*" (the most distant recognized member of *B. cereus s.l.*) was used
970 to root the phylogeny. Heatmap legends for all five taxonomies are colored by their order of
971 appearance in the heatmap, from top to bottom; white heatmap cells denote genomes that could
972 not be assigned to a taxonomic unit within a given taxonomic framework.

973

974

975 **Figure 3**. Maximum likelihood (ML) phylogeny constructed using core SNPs identified among
976 orthologous gene clusters of 1,081 *panC* Group IV *B. cereus s.l.* genomes. The phylogeny was
977 rooted using *panC* Group III *B. anthracis* str. Ames Ancestor as an outgroup (NCBI RefSeq
978 Accession GCF_000008445.1; omitted for readability). Tip label colors and clade labels denote
979 species assigned using GTDB-Tk ("GTDB Species"). The heatmap to the right of the phylogeny
980 denotes (i) whether a strain was sequenced in this study (dark pink) or not (light pink; "Study"),
981 and (ii) multi-locus sequence typing (MLST) sequence types (STs) associated with strains
982 sequenced in this study, where applicable (colored), or not (gray; "ST"). MLST lineages
983 discussed in Table 3 are annotated to the right of the heatmap ("MLST Lineage"). MLST
984 lineages with numerical superscripts contain two or more strains sequenced in this study, which
985 were highly similar on a genomic scale; these lineages are depicted in Figure 4. The tree was
986 rooted and time-scaled using LSD2, with branch lengths reported in years.
987
988 **Figure 4.** *B. cereus s.l.* multi-locus sequence typing (MLST) lineages that contained two or more
989 strains sequenced in this study, which were identical or nearly identical at the whole-genome
990 scale. Lineage names and sequence types (STs), where applicable, are shown in the top right
991 corner of each panel. Geographic and source origins of each strain are displayed in the respective
992 map.
993
994 **Figure 5.** Maximum likelihood phylogeny constructed using core SNPs identified among
995 orthologous gene clusters of 597 *panC* Group II and III *B. cereus s.l.* genomes. The phylogeny
996 was rooted using *panC* Group IV *B. cereus* str. ATCC 14579 as an outgroup (NCBI RefSeq
997 Accession GCF_006094295.1; omitted for readability). Tip label colors and clade labels denote

998 species assigned using GTDB-Tk ("GTDB Species"). The heatmap to the right of the phylogeny
999 denotes (i) whether a strain possesses anthrax toxin-encoding genes *cya, lef,* and *pagA* (dark
1000 orange) or not (light orange; "Anthrax"); (ii) whether a strain possesses cereulide synthetase-
1001 encoding *cesABCD* (dark purple) or not (light purple; "Emetic"); (iii) whether a strain belongs to
1002 *panC* Group II (blue) or III (yellow; *"panC"*); (iv) whether a strain was sequenced in this study
1003 (dark pink) or not (light pink; "Study"); (v) multi-locus sequence typing (MLST) sequence types
1004 (STs) associated with strains sequenced in this study, where applicable (colored), or not (gray;
1005 "ST"). MLST lineages discussed in Table 3 are annotated to the right of the heatmap ("MLST
1006 Lineage"). MLST lineages with numerical superscripts contain two or more strains sequenced in
1007 this study, which were highly similar on a genomic scale; these lineages are depicted in Figure 4.
1008 The tree was rooted and time-scaled using LSD2, with branch lengths reported in years.
1009
1010 **Figure 6.** Maximum likelihood phylogeny constructed using core SNPs identified among
1011 orthologous gene clusters of 219 *panC* Group V *B. toyonensis* genomes. The phylogeny was
1012 rooted using *panC* Group IV *B. cereus* str. ATCC 14579 as an outgroup (NCBI RefSeq
1013 Accession GCF_006094295.1; omitted for readability). The heatmap to the right of the
1014 phylogeny denotes (i) whether a strain was sequenced in this study (dark pink) or not (light pink;
1015 "Study"); (ii) multi-locus sequence typing (MLST) sequence types (STs) associated with strains
1016 sequenced in this study, where applicable (colored), or not (gray; "ST"). MLST lineages
1017 discussed in Table 3 are annotated to the right of the heatmap. The tree was rooted and time-
1018 scaled using LSD2, with branch lengths reported in years.

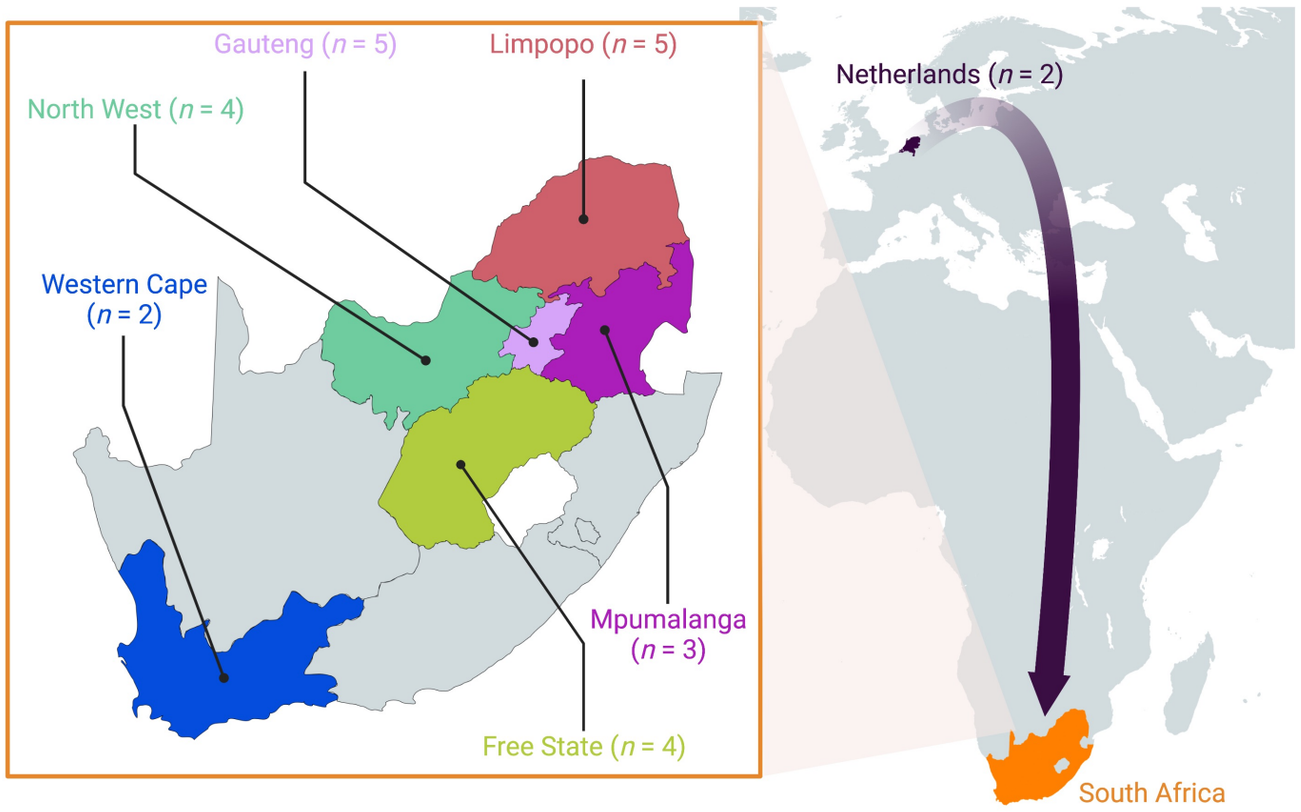

Figure 1. Geographic origins of *B. cereus s.l.* strains sequenced in this study ($n = 25$).

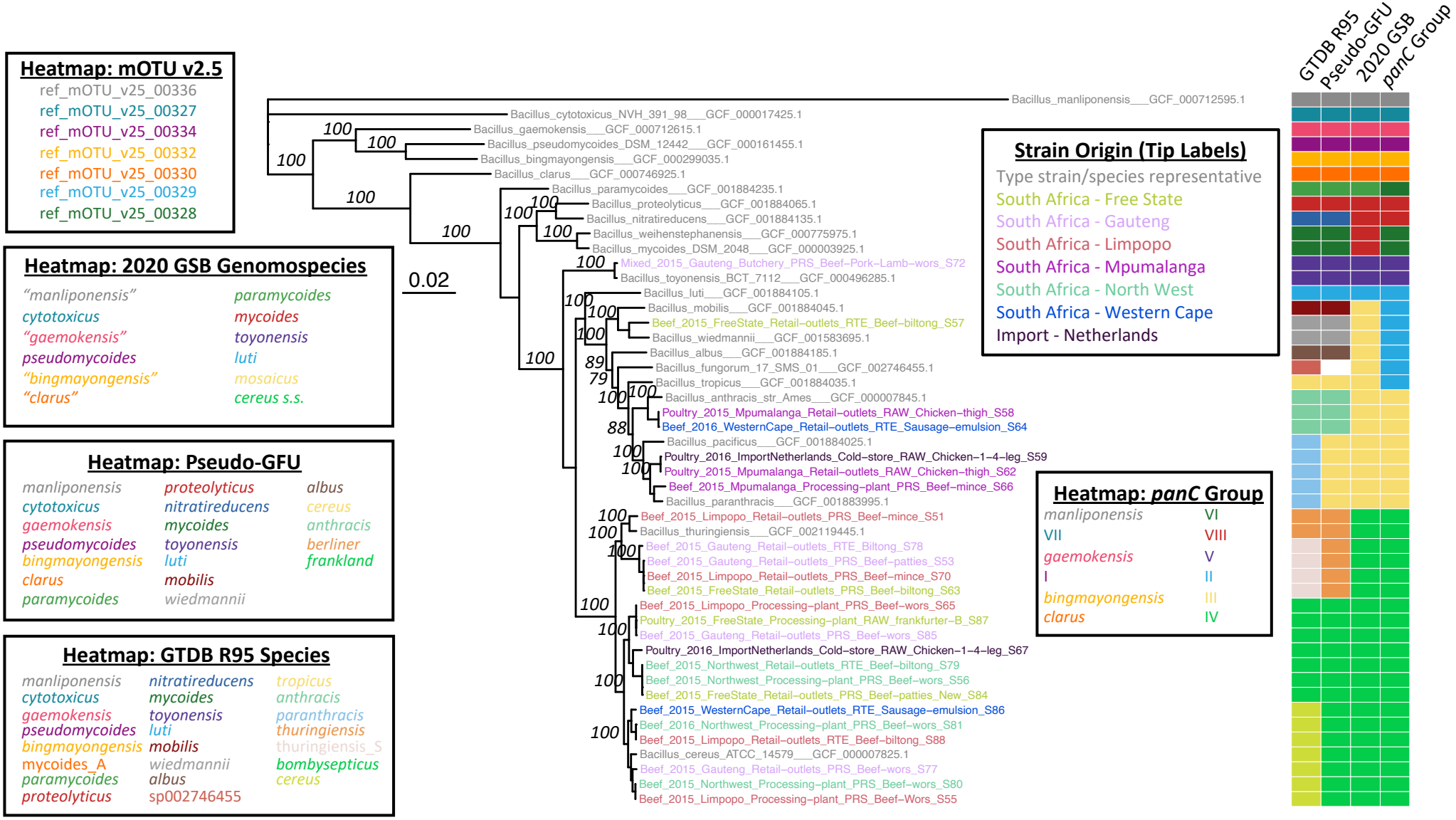

Figure 2. Maximum likelihood (ML) phylogeny constructed using amino acid sequences derived from the 25 *B. cereus s.l.* isolate genomes sequenced in this study (tip labels colored by geographic origin), plus type strain/species representative genomes of 23 published and effective *B. cereus s.l.* species (gray tip labels). The heatmap to the right of the phylogeny showcases species assignments obtained within the following taxonomic frameworks (from left to right): (i) Genome Taxonomy Database (GTDB) Release 05-RS95 and GTDB-Tk (GTDB R95); (ii) pseudo-gene flow units (GFUs) assigned using BTyper3 (Pseudo-GFU); (iii) genomospecies of the 2020 standardized *B. cereus s.l.* genomospecies/subspecies/biovar (GSB) framework and BTyper3 (2020 GSB); (iv) *panC* Group (I-VIII), assigned using BTyper3 (*panC* Group). The phylogeny was constructed using IQ-TREE, using core orthologs identified among all genomes via OrthoFinder as input. Branch lengths are reported in substitutions per site. Branch labels correspond to branch support percentages obtained using 1,000 replicates of the ultrafast bootstrap approximation (selected for readability). The type strain genome of effective *B. cereus s.l.* species "*B. manliponensis*" (the most distant recognized member of *B. cereus s.l.*) was used to root the phylogeny. Heatmap legends for all five taxonomies are colored by their order of appearance in the heatmap, from top to bottom; white heatmap cells denote genomes that could not be assigned to a taxonomic unit within a given taxonomic framework.

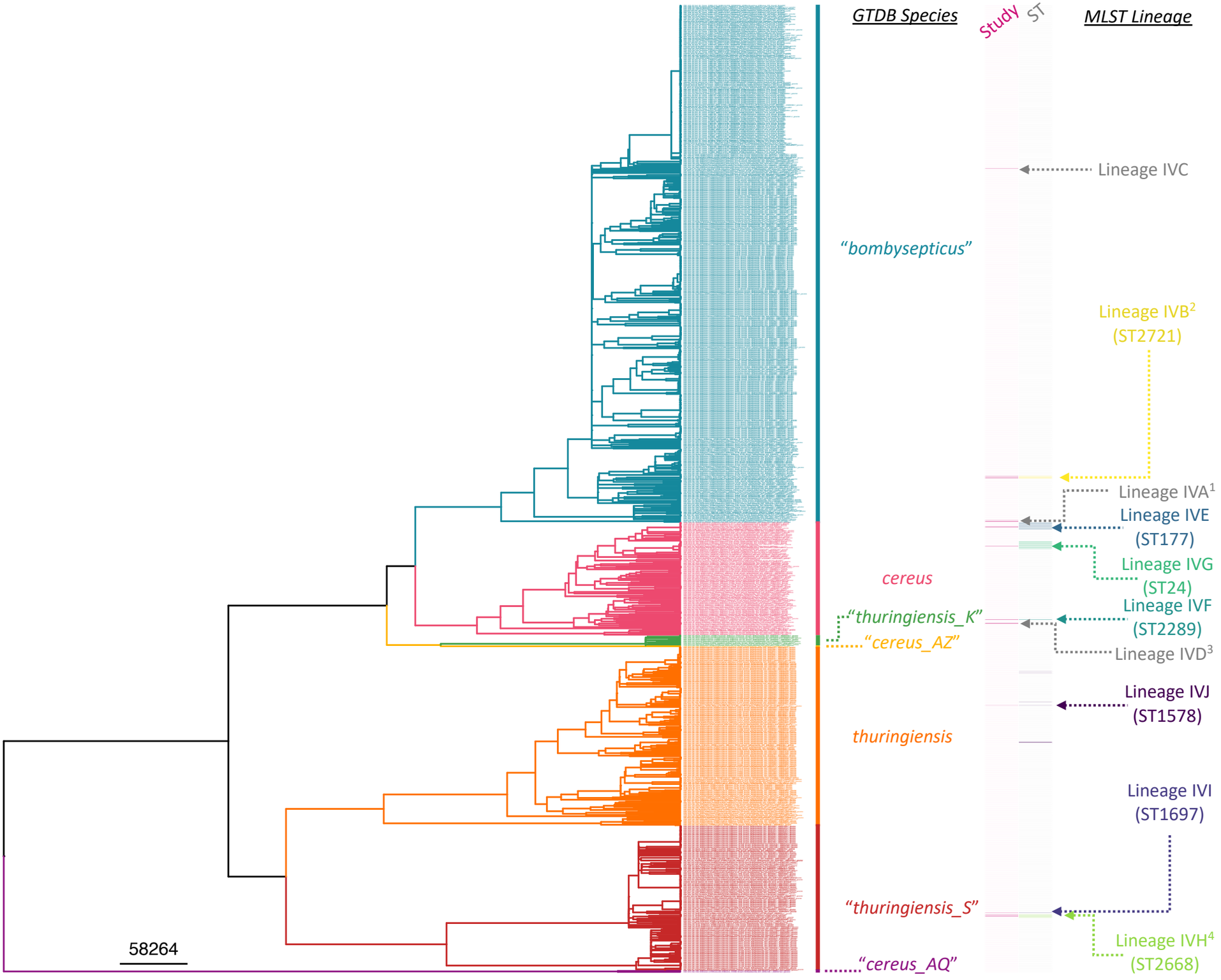

Figure 3. Maximum likelihood (ML) phylogeny constructed using core SNPs identified among orthologous gene clusters of 1,081 *panC* Group IV *B. cereus s.l.* genomes. The phylogeny was rooted using *panC* Group III *B. anthracis* str. Ames Ancestor as an outgroup (NCBI RefSeq Accession GCF_000008445.1; omitted for readability). Tip label colors and clade labels denote species assigned using GTDB-Tk ("GTDB Species"). The heatmap to the right of the phylogeny denotes (i) whether a strain was sequenced in this study (dark pink) or not (light pink; "Study"), and (ii) multi-locus sequence typing (MLST) sequence types (STs) associated with strains sequenced in this study, where applicable (colored), or not (gray; "ST"). MLST lineages discussed in Table 3 are annotated to the right of the heatmap ("MLST Lineage"). MLST lineages with numerical superscripts contain two or more strains sequenced in this study, which were highly similar on a genomic scale; these lineages are depicted in Figure 4. The tree was rooted and time-scaled using LSD2, with branch lengths reported in years.

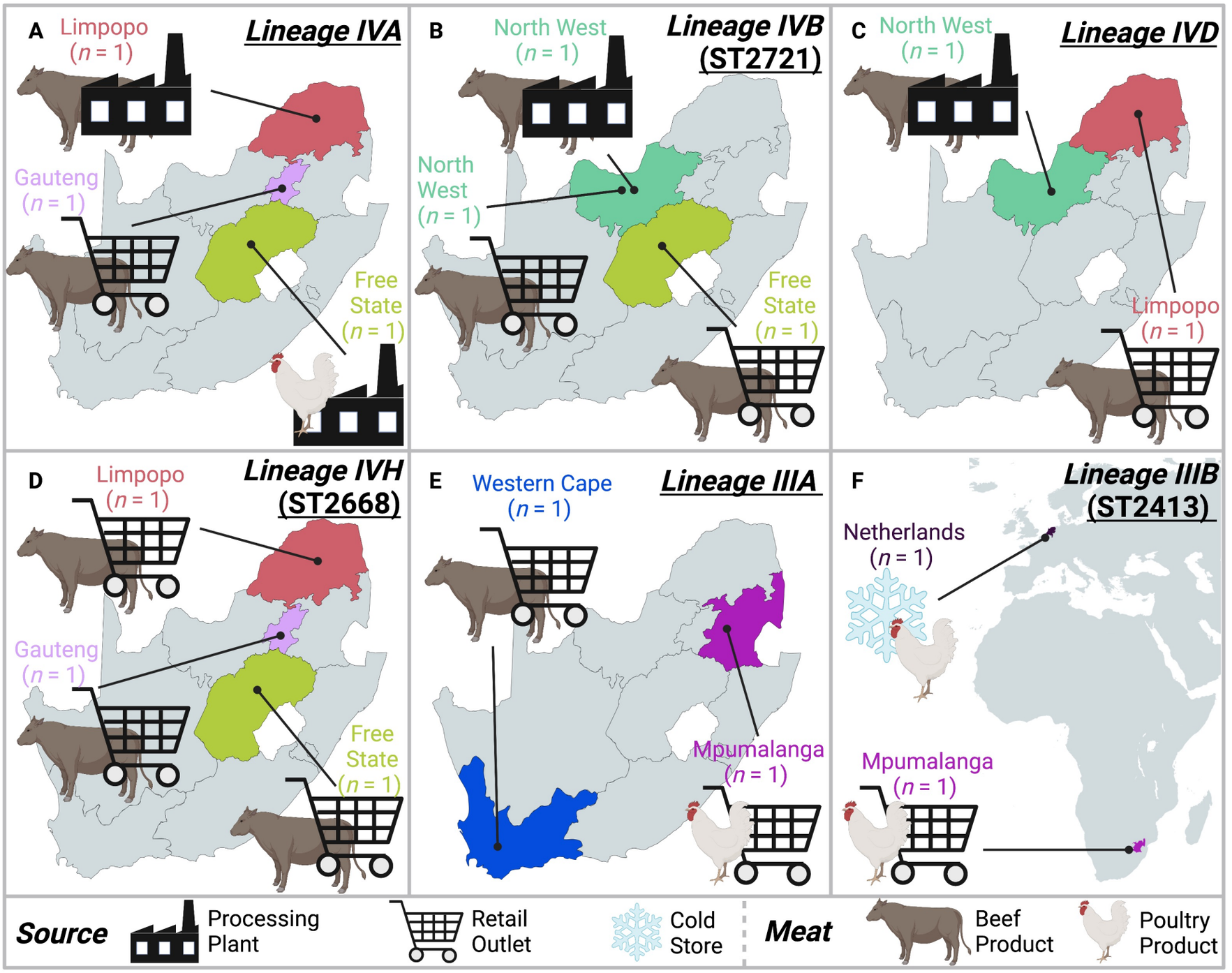

Figure 4. *B. cereus s.l.* multi-locus sequence typing (MLST) lineages that contained two or more strains sequenced in this study, which were identical or nearly identical at the whole-genome scale. Lineage names and sequence types (STs), where applicable, are shown in the top right corner of each panel. Geographic and source origins of each strain are displayed in the respective map.

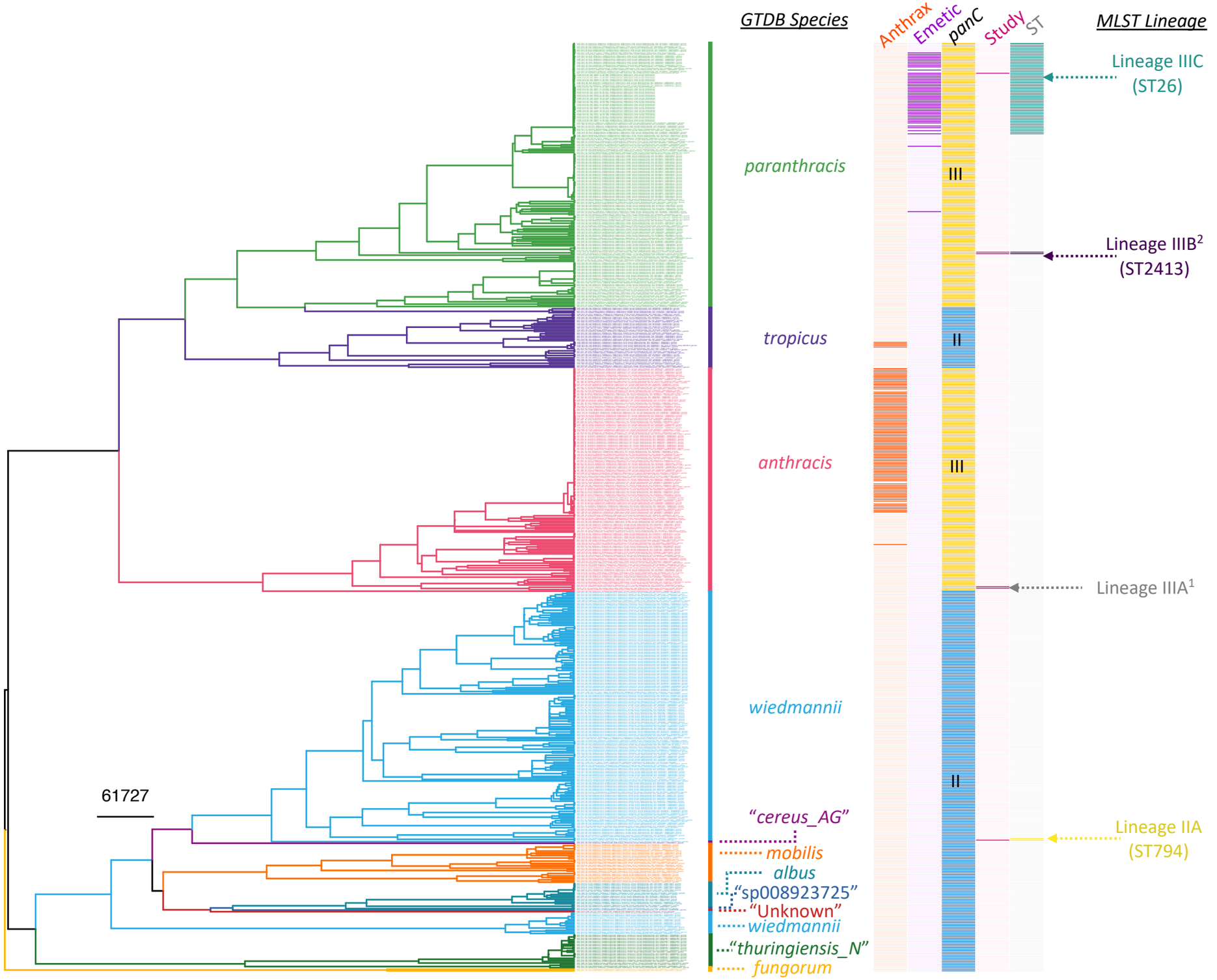

Figure 5. Maximum likelihood phylogeny constructed using core SNPs identified among orthologous gene clusters of 597 *panC* Group II and III *B. cereus s.l.* genomes. The phylogeny was rooted using *panC* Group IV *B. cereus* str. ATCC 14579 as an outgroup (NCBI RefSeq Accession GCF_006094295.1; omitted for readability). Tip label colors and clade labels denote species assigned using GTDB-Tk ("GTDB Species"). The heatmap to the right of the phylogeny denotes (i) whether a strain possesses anthrax toxin-encoding genes *cya*, *lef*, and *pagA* (dark orange) or not (light orange; "Anthrax"); (ii) whether a strain possesses cereulide synthetase-encoding *cesABCD* (dark purple) or not (light purple; "Emetic"); (iii) whether a strain belongs to *panC* Group II (blue) or III (yellow; *"panC"*); (iv) whether a strain was sequenced in this study (dark pink) or not (light pink; "Study"); (v) multi-locus sequence typing (MLST) sequence types (STs) associated with strains sequenced in this study, where applicable (colored), or not (gray; "ST"). MLST lineages discussed in Table 3 are annotated to the right of the heatmap ("MLST Lineage"). MLST lineages with numerical superscripts contain two or more strains sequenced in this study, which were highly similar on a genomic scale; these lineages are depicted in Figure 4. The tree was rooted and time-scaled using LSD2, with branch lengths reported in years.

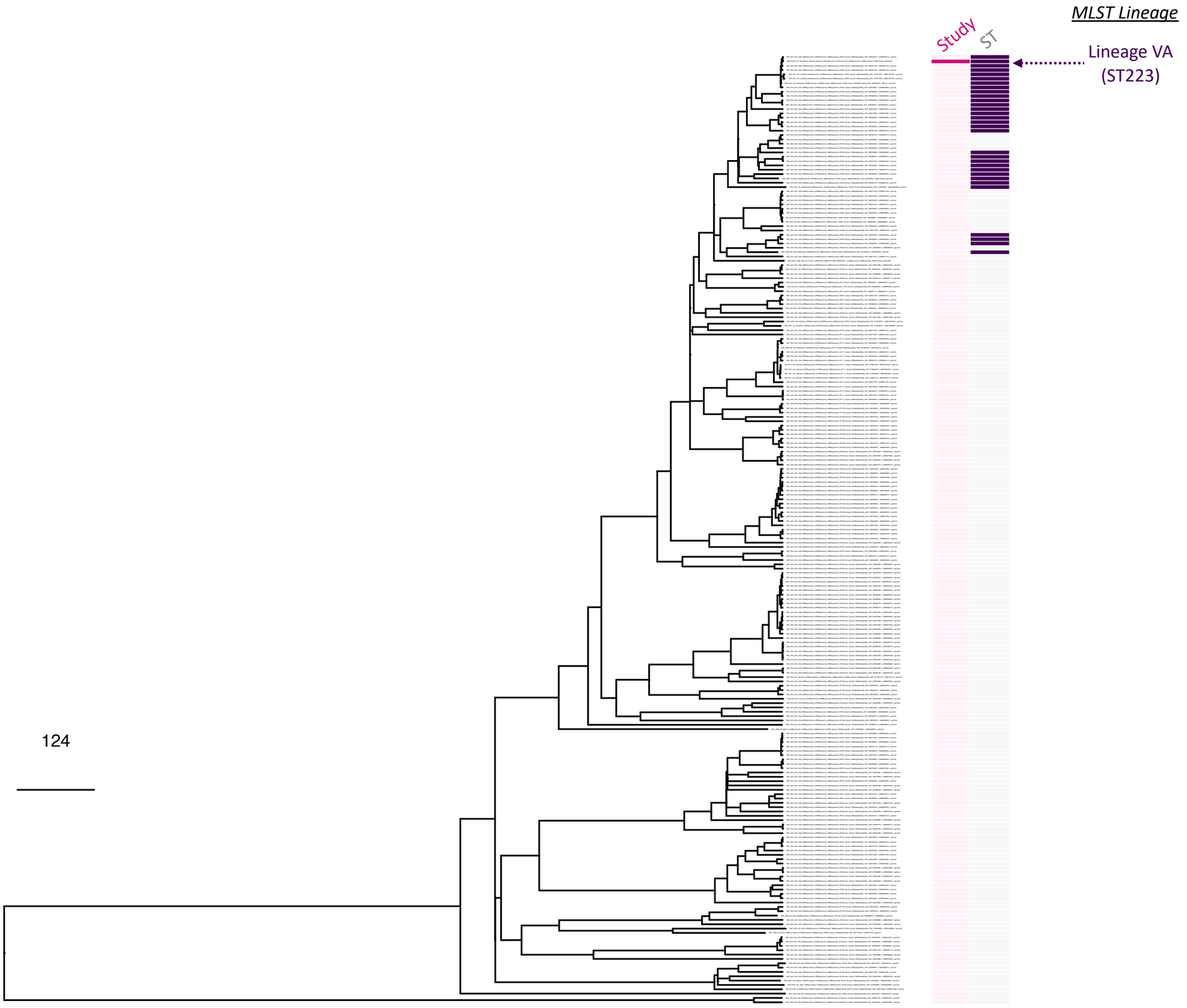

Figure 6. Maximum likelihood phylogeny constructed using core SNPs identified among orthologous gene clusters of 219 *panC* Group V *B. toyonensis* genomes. The phylogeny was rooted using *panC* Group IV *B. cereus* str. ATCC 14579 as an outgroup (NCBI RefSeq Accession GCF_006094295.1; omitted for readability). The heatmap to the right of the phylogeny denotes (i) whether a strain was sequenced in this study (dark pink) or not (light pink; "Study"); (ii) multi-locus sequence typing (MLST) sequence types (STs) associated with strains sequenced in this study, where applicable (colored), or not (gray; "ST"). MLST lineages discussed in Table 3 are annotated to the right of the heatmap. The tree was rooted and time-scaled using LSD2, with branch lengths reported in years.