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: 250 words (Abstract) + 147 words (Importance)

Text Word Count: 4,628 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 strain possessed cereulide (emetic) synthetase-encoding genes. No strains harbored anthrax toxin- or capsule-encoding genes, including two strains assigned to the Genome Taxonomy Database's (GTDB) *B. anthracis* genomospecies. 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 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

47 studies to date have evaluated the potential utility of WGS for *B. cereus s.l.* surveillance and source
48 tracking in foodstuffs. In this proof-of-concept study, we applied WGS to *B. cereus s.l.* strains
49 collected via South Africa's national surveillance program of domestic and imported meat
50 products, and we provide strong evidence that *B. cereus s.l.* can be transmitted intra- and inter-
51 nationally via the agro-food supply chain. Our results showcase that WGS can be used for source
52 tracking of *B. cereus s.l.* in foods, although future WGS and isolate metadata collection efforts are
53 needed to ensure that *B. cereus s.l.* surveillance initiatives are on par with those of other foodborne
54 pathogens.
55
56
57

58 **INTRODUCTION**

59 *Bacillus cereus sensu lato* (*s.l.*), also known as the *B. cereus* group, is a complex of closely
60 related, Gram-positive, spore-forming species, which are widespread throughout the environment
61 (1). While some members of *B. cereus s.l.* have important industrial applications or roles (e.g., as
62 biocontrol agents in agricultural settings, as food spoilage organisms) (2-6), others are capable of
63 causing illnesses or death in humans and/or animals (1, 7-9). Illnesses caused by members of *B.*
64 *cereus s.l.* can range in severity from mild to severe/fatal and include anthrax and anthrax-like
65 illness (8, 10, 11), foodborne emetic intoxication (1, 7, 12-15), foodborne diarrheal toxicoinfection
66 (1, 7, 14, 15), and non-gastrointestinal infections (16, 17). As a foodborne pathogen, "*B. cereus*"
67 is estimated to be responsible for more than 256,000 illnesses globally each year (18), although
68 this is likely an underestimate, due to the relatively mild and self-limiting nature of the symptoms
69 that often accompany foodborne illness caused by *B. cereus s.l.* (1).

70 Food safety surveillance efforts around the world have identified *B. cereus s.l.* strains in a
71 wide variety of foodstuffs (1, 7), including raw intact, processed, and ready-to-eat (RTE) meat and
72 poultry products (19-27). In South Africa specifically, previous surveillance efforts have identified
73 *B. cereus s.l.* in (i) retail meats sold at supermarkets in the Pretoria area (i.e., Vienna sausages,
74 salami, and poultry) (27) and (ii) biltong (a spiced intermediate moisture RTE meat product) sold
75 at supermarkets, stalls, kiosks, and butcheries in Bloemfontein, Free State (28). Most recently, in
76 a study of over two thousand meat product samples collected from butcheries, processing plants,
77 abattoirs, and retail outlets across all nine South African provinces, *B. cereus s.l.* was present in
78 4.5 and 2.7% of domestic and imported meat products, respectively (26).

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

However, it is unclear which *B. cereus s.l.* lineages are present in South African meat and poultry products on a genomic scale. Here, we used whole-genome sequencing (WGS) to characterize 25 *B. cereus s.l.* strains isolated from raw intact, processed, and RTE meat and poultry products collected from processing plants, butcheries, and retail outlets across South Africa, as well as imported meat products. By comparing South African strains sequenced here to all publicly available *B. cereus s.l.* genomes (n = 2,887), we identified multiple *B. cereus s.l.* species present among South African meat and poultry products, and we detected multiple potential inter-national and inter-provincial *B. cereus s.l.* transmission events. Overall, our study serves as the first genome-scale study of South African *B. cereus s.l.* in foodstuffs and showcases the utility of WGS for *B. cereus s.l.* surveillance and source tracking.

RESULTS

***B. cereus s.l.* are present among domestic and imported meat and poultry products in South Africa.** A total of 25 *B. cereus s.l.* strains were isolated from meat and poultry products, which had been collected across South Africa in 2015 and 2016 (Figure 1, Table 1, and Supplemental Table S1). Overall, 19 strains (76.0%) originated from beef products, including: beef wors, a processed South African sausage (n = 7; 28.0%); beef biltong, a South African spiced intermediate moisture RTE meat product (n = 5; 20.0%); processed beef mince (n = 3; 12.0%); RTE beef sausage emulsion (n = 2; 8.0%); and processed beef patties (n = 2; 8.0%, Table 1 and Supplemental Table S1). Five strains (20.0%) were isolated from poultry products, including raw chicken thighs (n = 2; 8.0%), raw chicken quarter-legs (n = 2; 8.0%), and a frankfurter (n = 1; 4.0%, Table 1 and Supplemental Table S1). One strain (4.0%) was isolated from wors, which had been made from a mix of beef, pork, and lamb (Table 1 and Supplemental Table S1).

104 The majority of strains (23 of 25, 92.0%) were obtained from domestic meat and poultry
105 products acquired across six South African provinces (*n* = 5, 5, 4, 4, 3, and 2 strains from Gauteng,
106 Limpopo, Free State, North West, Mpumalanga, and Western Cape, respectively; Figure 1, Table
107 1, and Supplemental Table S1). The remaining two strains (8.0%) were isolated from raw, intact
108 chicken quarter-legs, which had been imported into South Africa from the Netherlands (Figure 1,
109 Table 1, and Supplemental Table S1). Fifteen of the 25 *B. cereus s.l.* strains (60.0%) were isolated
110 from meat or poultry products acquired from retail outlets, while seven (28.0%) were acquired
111 from meat or poultry products obtained from processing plants (Table 1 and Supplemental Table
112 S1). Two strains (from chicken quarter-legs imported from the Netherlands; 8.0%) were derived
113 from poultry in cold stores, while one strain (4.0%) was isolated from the mixed beef-pork-lamb
114 wors acquired from a butchery (Table 1 and Supplemental Table S1).

115 **Multiple species are present among *B. cereus s.l.* from South African meat and poultry**
116 **products.** Species-level taxonomic classification of *B. cereus s.l.* is notoriously challenging (1,
117 29, 30); to avoid taxonomic ambiguities and maximize interpretability, we applied multiple
118 taxonomic assignment and sequence typing methods to the 25 strains sequenced here (Table 1).
119 One such sequence typing framework relies on the pantoate-β-alanine ligase gene (*panC*) to assign
120 *B. cereus s.l.* strains to one of seven or more major phylogenetic groups, which have been proposed
121 to conceptually serve as "species" (31, 32). Using the adjusted eight-group *panC* typing approach
122 implemented in BTyper3 (33), the 25 *B. cereus s.l.* strains sequenced here encompassed four *panC*
123 phylogenetic groups (i.e., "species"; Figure 2 and Table 1). The majority (*n* = 18 of 25, 72.0%) of
124 the strains sequenced here were assigned to *panC* Group IV (Figure 2 and Table 1). The remaining
125 strains were assigned to *panC* Groups III, II, and V (*n* = 5, 1, and 1 strains, representing 20.0%,
126 4.0%, and 4.0% of isolates sequenced here, respectively; Figure 2 and Table 1).

127 Using the Genome Taxonomy Database (GTDB) taxonomy, the 25 *B. cereus s.l.* strains
128 sequenced here encompassed eight genomospecies (Figure 2 and Table 1). The 18 *panC* Group
129 IV strains sequenced here encompassed four GTDB genomospecies: "*B. bombysepticus*" (*n* = 7),
130 *B. cereus* (*n* = 6), "*B. thuringiensis*_S" (*n* = 4), and *B. thuringiensis* (*n* = 1; Figure 2 and Table
131 1). The five *panC* Group III strains sequenced here were assigned to two GTDB genomospecies
132 (*n* = 3 and 2 strains assigned to *B. paranthracis* and *B. anthracis*, respectively), while the *panC*
133 Group II and Group V strains (*n* = 1 each) were assigned to the *B. wiedmannii* and *B. toyonensis*
134 genomospecies, respectively (Figure 2 and Table 1).
135 Using a standardized genomospecies-subspecies-biovar (GSB) nomenclatural framework
136 proposed for *B. cereus s.l.* in 2020 (34) (referred to hereafter as the "2020 GSB" framework), the
137 25 strains sequenced here encompassed three genomospecies (Figure 2 and Tables 1 and 2). All
138 18 *panC* Group IV strains were assigned to the *B. cereus sensu stricto* (*s.s.*) genomospecies (Figure
139 2 and Tables 1 and 2); genes encoding insecticidal toxins (referred to hereafter as "Bt toxin-
140 encoding genes") were detected within all Group IV *B. cereus s.s.* genomes (using
141 BtToxin_scanner2's "old" gene detection approach), meaning that these 18 strains were predicted
142 to belong to the Thuringiensis biovar (i.e., *B. cereus s.s.* bv. Thuringiensis; Table 2). All five *panC*
143 Group III and the single *panC* Group II strain(s) were assigned to genomospecies *B. mosaicus*
144 within the 2020 GSB framework (Figure 2 and Tables 1 and 2). One *panC* Group III *B. mosaicus*
145 strain (i.e., strain S66) was assigned to PubMLST sequence type 26 (ST26) and possessed
146 cereulide (emetic toxin) synthetase-encoding *cesABCD* and was thus assigned to the *cereus*
147 subspecies and biovar Emeticus (i.e., *B. mosaicus* subsp. *cereus* bv. Emeticus; Table 2). The lone
148 *panC* Group V strain sequenced here was assigned to the *B. toyonensis* genomospecies; Bt toxin-
149 encoding genes were detected in this genome (via BtToxin_scanner2's "old" gene detection

150 approach), and thus this strain was predicted to belong to biovar Thuringiensis (i.e., *B. toyonensis*
151 bv. Thuringiensis; Figure 2 and Tables 1 and 2).

152 Using a rapid, average nucleotide identity (ANI)-based pseudo-gene flow unit assignment
153 scheme, which attempts to assign *B. cereus s.l.* genomes to taxonomic units that mimic *B. cereus*
154 *s.l.* "species" previously delineated using recent gene flow (33), the 25 strains sequenced here were
155 assigned to six pseudo-gene flow units (pseudo-GFUs; Figure 2 and Supplemental Table S1). The
156 18 *panC* Group IV strains sequenced here were assigned to two pseudo-GFUs: *frankland* and
157 *berliner* (*n* = 13 and 5, respectively; Figure 2 and Supplemental Table S1). The five *panC* Group
158 III strains sequenced here also encompassed two pseudo-GFUs (*n* = 3 and 2 strains assigned to
159 pseudo-GFUs *cereus* and *anthracis*, respectively; Figure 2 and Supplemental Table S1). The *panC*
160 Group II and Group V strains (*n* = 1 each) were assigned to the *wiedmannii* and *toyonensis* pseudo-
161 GFUs, respectively (Figure 2 and Supplemental Table S1).

162 As mentioned above, considerable phenotypic diversity was predicted among the stains
163 sequenced here, as one strain harbored cereulide synthetase-encoding genes, and 19 strains
164 possessed Bt toxin-encoding genes (detected using BtToxin_scanner2's "old" gene detection
165 approach; Table 2). Notably, no anthrax toxin- or capsule-encoding genes were identified within
166 the genomes of the isolates sequenced here, including among two genomes that most closely
167 resembled the *B. anthracis* species representative genome (via ANI-based methods; Table 2).

168 Overall, regardless of whether the *panC*, GTDB, 2020 GSB, or pseudo-gene flow unit
169 assignment frameworks were used, *B. cereus s.l.* strains isolated from meat and poultry products
170 in South Africa were considerably diverse and represented multiple genomospecies (Figure 2,
171 Tables 1 and 2, and Supplemental Table S1). Additionally, using PubMLST's seven-gene MLST
172 scheme for *B. cereus*, 17 of 25 strains (68.0%) encompassed 11 STs, with eight strains (32.0%)

173 assigned to unknown STs (Table 1 and Supplemental Table S1). Due to the considerable genomic
174 diversity observed among isolates sequenced here, major lineages represented by strains
175 sequenced in this study are discussed individually in detail below, largely within the context of
176 *panC* Groups and/or MLST STs, as these frameworks are well-established (31, 32, 35, 36) and
177 likely the most interpretable to readers.

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

193 Similar results were observed for *panC* Group IV ST2721 (Lineage IVB in Table 3), which
194 was also assigned to GTDB's "*B. bombysepticus*" genomospecies and contained three strains
195 sequenced in this study (S56, S79, and New_S84; Figures 3 and 4 and Table 3). The three ST2721

196 genomes were nearly identical (pairwise core SNP distance = 0, pairwise ANI > 99.99; Table 3),
197 despite the fact that the strains originated from three different meat products (one from each of
198 beef wors, beef biltong, and processed beef patties) obtained from three different establishments
199 (from a processing plant in North West province, a retail outlet in North West province, and a retail
200 outlet in Free State, respectively; Figure 4 and Table 1).

201 A third *panC* Group IV lineage of unknown ST, which was assigned to GTDB's *B. cereus*
202 species (Lineage IVD), contained two isolates sequenced in this study (S81 and S88; Figures 3
203 and 4 and Table 3). One strain (S81) had been isolated from beef wors in a processing plant in the
204 North West province; the other (S88) from RTE beef biltong in a retail outlet in Limpopo (Figure
205 4 and Table 1). Both strains were highly similar on a genomic scale (>99.99 ANI) and differed by
206 three core SNPs (Figure 3 and Table 3). For reference, in a previous point source foodborne
207 outbreak caused by *B. cereus s.l.* (37), outbreak isolates could differ by up to seven core SNPs
208 (using the same SNP calling methodology used here) (38).

209 A fourth *panC* Group IV lineage, ST2668, contained three strains sequenced in this study
210 (S53, S63, and S70), which were assigned to GTDB's "*B. thuringiensis_S*" genomospecies
211 (Lineage IVH; Figures 3 and 4 and Table 3). The three strains sequenced here differed by, at most,
212 a single core SNP (Table 3), even though all had been isolated from different meat products
213 (processed beef patties, beef biltong, and processed beef mince) obtained from different
214 establishments (i.e., from retail outlets in each of Gauteng, Free State, and Limpopo, respectively;
215 Figure 4 and Table 1). One publicly available genome associated with a *B. cereus s.l.* strain isolated
216 from grass in KarieDeshe, Israel (39) was additionally assigned to ST2668 (NCBI RefSeq
217 Accession GCF_005217245.1); however, this strain was not closely related to the three nearly
218 identical South African ST2668 strains sequenced in this study (Table 3).

219 The remaining six MLST-based lineages (i.e., Lineages IVC, IVE, IVF, IVG, IVI, and IVJ,
220 corresponding to an unknown ST, ST177, ST2289, ST24, ST1697, and ST1578, respectively;
221 Table 3) contained isolates sequenced in this study, which were not closely related to any other
222 strains at the whole-genome level (Figure 3 and Table 3). Lineages IVC, IVF, and IVI were
223 singleton lineages, which each contained one genome sequenced in this study (S67, S86, and S78,
224 respectively), shared 98.90-99.35 ANI with their closest publicly available neighbors (Figure 3
225 and Table 3). Two lineages (IVG and IVJ) each contained multiple genomes, but only one genome
226 sequenced in this study (i.e., S77 and S51, respectively; Figure 3 and Table 3); for both lineages,
227 the South African genome sequenced here was not identical to any publicly available genomes.
228 The remaining lineage, Lineage IVE (ST177), contained multiple genomes, as well as multiple
229 genomes sequenced in this study (Figure 3 and Table 3). Notably, the two ST177 strains sequenced
230 in this study (S55 and S80), which were assigned to this lineage, were not identical, and were
231 isolated from beef wors from processing plants in the Limpopo and North West provinces,
232 respectively (Table 1), indicating that WGS can be useful for differentiating closely related *B.*
233 *cereus s.l.* genomes within STs.

234 **A *panC* Group III *B. cereus s.l.* lineage most closely related to *B. anthracis* was detected in**
235 **beef and poultry products from two South African provinces.** Two *panC* Group III *B. cereus*
236 *s.l.* strains of an unknown ST that had been sequenced here were assigned to GTDB's *B. anthracis*
237 genomospecies (S58 and S64; Figure 5 and Table 1); however, these genomes did not possess
238 anthrax toxin- or capsule-encoding genes and were not members of the closely related *B. anthracis*
239 lineage most commonly associated with anthrax illness (i.e., the "*B. anthracis*" subspecies within
240 the 2020 GSB framework; Figure 5 and Tables 1 and 2). Despite having been isolated from
241 different products (S58 from a raw chicken thigh, S64 from RTE sausage emulsion) from different

242 establishments (retail outlets in Mpumalanga and Western Cape, respectively), the strains were
243 nearly identical (pairwise core SNP distance = 0, pairwise ANI > 99.84; Figures 4 and 5 and Table
244 3), indicating that these two strains share a common source.

245 **A *panC* Group III *B. cereus s.l.* strain with cereulide synthetase-encoding genes belongs to**
246 **the "high-risk" ST26 lineage.** One *panC* Group III *B. cereus s.l.* strain (S66) was assigned to
247 GTDB's *B. paranthracis* genomospecies and possessed cereulide synthetase-encoding genes
248 (Figure 5 and Table 2). This strain, which had been isolated from processed beef mince from a
249 processing plant in Mpumalanga, was a member of ST26 (Figure 5 and Table 2), the ST to which
250 most cereulide-producing *B. cereus s.l.* strains belong (although it should be noted that ST26
251 strains may be capable of producing enterotoxins and causing diarrheal illness, regardless of
252 whether they produce cereulide or not) (38, 40, 41). While members of ST26 are comparatively
253 closely related (>99.52 pairwise ANI), WGS was able to distinguish the South African strain
254 sequenced here from closely related ST26 strains (pairwise core SNP distance >913 relative to
255 publicly available genomes; Figure 5 and Table 3).

256 **A *panC* Group III *B. cereus s.l.* lineage assigned to ST2413 shows evidence of intercontinental**
257 **transmission.** Two *panC* Group III *B. cereus s.l.* strains sequenced in this study (S59 and S62)
258 were assigned to ST2413 within GTDB's *B. paranthracis* species (Figure 5 and Table 3). Unlike
259 the ST26 strain, which was also assigned to GTDB's *B. paranthracis* genomospecies, neither
260 ST2413 strain possessed cereulide synthetase-encoding genes (Figure 5 and Table 2). Both strains
261 were isolated from raw chicken; however, S59 was isolated from a chicken quarter-leg that had
262 been imported from the Netherlands, and S62 from a chicken thigh sold at a retail outlet in
263 Mpumalanga (Figure 4 and Table 1). Notably, these strains were nearly identical on a genomic

264 scale (pairwise core SNP distance = 1, pairwise ANI = 100.0; Figure 5 and Table 3), despite
265 originating from different continents (i.e., Europe and Africa; Figure 4 and Table 1).

266 **A *panC* Group II *B. cereus s.l.* strain assigned to ST794 is most closely related to a food-**
267 **associated strain responsible for diarrheal illness in Norway.** One *panC* Group II *B. cereus s.l.*
268 strain was sequenced in this study (S57) and was assigned to ST794 within GTDB's *B. wiedmannii*
269 species (Figure 5 and Table 1). Strain S57 had been isolated from RTE beef biltong sold at a retail
270 outlet in Free State (Table 1). When compared to the two publicly available ST794 genomes, S57
271 shared > 99.9 ANI with both publicly available genomes and differed by 31 and 100 core SNPs
272 (relative to genomes with NCBI RefSeq Assembly Accessions GCF_900094845.1 and
273 GCF_007671965.1, respectively; Figure 5 and Table 3). Notably, beef biltong-associated strain
274 S57 sequenced here was most closely related to strain NVH 0674-98, a psychrotolerant strain that
275 had been isolated in Norway from mashed swedes, which were reportedly responsible for diarrheal
276 foodborne illness (42). The other ST794 strain, DE0555, was an environmental isolate collected
277 in 2018 from Durham, North Carolina in the United States (NCBI BioSample Accession
278 SAMN11792715).

279 **A *panC* Group V *B. cereus s.l.* strain from South African mixed-meat wors most closely**
280 **resembles a plant-associated strain from the United States.** One *panC* Group V *B. cereus s.l.*
281 strain was sequenced in this study (S72) and was assigned to ST223 within GTDB's *B. toyonensis*
282 species (Figure 6 and Tables 1-3). S72 had been isolated in a butchery in Gauteng, from a
283 processed wors composed of a mix of beef, pork, and lamb (Table 1). Strain S72 was most closely
284 related to a publicly available genome, strain AFS092321 (NCBI RefSeq Accession
285 GCF_002552615.1), which had been isolated in 2014 from a tree leaf in North Carolina, United

286 States (NCBI BioSample Accession SAMN07598612) (43); the two strains shared >99.9 ANI and
287 differed by 33 core SNPs (Figure 6 and Table 3).
288
289 **DISCUSSION**
290 ***B. cereus s.l.* lineages can be disseminated inter- and intra-nationally via the food supply**
291 **chain.** The movement of commodities (e.g., foods, animals, animal products, agricultural products,
292 consumer products) through inter- and intra-national trade can contribute to the global, regional,
293 and local dissemination of microorganisms, including pathogens (44-48). The international agro-
294 food trade specifically plays an increasingly pivotal role in providing food supplies to communities
295 around the globe but can contribute to the dissemination of foodborne pathogens (47, 48).
296 Consequently, high-resolution technologies such as WGS are being employed increasingly to
297 monitor the spread of pathogens along the food supply chain (49, 50).
298 Using WGS, we identified six South African *B. cereus s.l.* lineages, which showcased
299 evidence of inter-regional dissemination via the trade and transport of meat products (Figure 4 and
300 Table 3). Notably, one *B. cereus s.l.* lineage showed evidence of intercontinental spread between
301 Europe and Africa: a *B. cereus s.l.* ST2413 strain isolated from raw chicken sold in retail outlets
302 in Mpumalanga, South Africa was identical to a ST2413 strain isolated from chicken imported
303 from the Netherlands. We may hypothesize that the raw chicken sold in Mpumalanga originated
304 from the Netherlands, as the Netherlands was the second-largest exporter of chicken meat products
305 to South Africa in 2014-2016 (i.e., the timeframe in which the strains sequenced here were
306 collected) (51); however, this is merely a hypothesis, as we were unable to confirm this with the
307 retail outlet.

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

316 **Nomenclatural frameworks that incorporate both genomic and phenotypic data can improve**
317 **strain-level *B. cereus s.l.* risk assessment.** *B. cereus s.l.* species delineation is notoriously
318 challenging, and numerous *B. cereus s.l.* taxonomic frameworks have been proposed (29).
319 Phenotypic traits historically used for *B. cereus s.l.* species assignment (e.g., motility, colony
320 morphology, ability to cause illness) have long been known to be inconsistent with genome
321 evolution (29, 31, 32). Taxonomies that rely solely on genomic data, however, may lead to
322 incorrect assumptions of a strain's pathogenic potential, particularly when taxonomic labels have
323 deep roots in medicine or industry (29, 34).

324 Within ANI-based taxonomic frameworks, such as GTDB, the *B. anthracis* genomospecies
325 encompasses *B. cereus s.l.* strains that resemble "*B. cereus*" phenotypically (e.g., they are motile,
326 hemolytic), but are incapable of causing anthrax (34). These non-anthrax-causing "*B. cereus*"
327 strains have been isolated from diverse environments, including food (e.g., milk, egg, whites,
328 spices), consumer products (e.g., baby wipes), and soil, indicating that these organisms are not
329 uncommon in environmental and industrial settings (34). Thus, as WGS becomes more popular in
330 clinical and industrial settings, it is possible that professionals who rely solely on increasingly

331 popular genomic methods for taxonomic delineation (e.g., GTDB, ANI-based comparisons to
332 species type strains) may incorrectly assume that these strains can cause anthrax, due to their "*B.*
333 *anthracis*" species labels (29, 34). Here, we showed that this is not just a hypothetical scenario:
334 during routine surveillance of meat products in South Africa, we identified two *B. cereus s.l.* strains
335 from beef and poultry products sold at retail outlets, which belong to the "*B. anthracis*"
336 genomospecies using GTDB and similar ANI-based methods (Table 1). However, these "*B.*
337 *anthracis*" strains do not possess anthrax toxin- or capsule-encoding genes, nor do they belong to
338 the "*B. anthracis*" lineage most commonly linked to anthrax disease (Table 2). Thus, referring to
339 these isolates as "*B. anthracis*" would be misleading and could potentially have disastrous public
340 health and industrial consequences.

341 Similarly, we isolated three *B. cereus s.l.* strains from beef and poultry products, which
342 were assigned to the "*B. paranthracis*" genomospecies via GTDB and similar ANI-based methods
343 (Table 1). As noted previously, "*B. paranthracis*" was proposed as a "novel" species in 2017 (52)
344 but was later found to encompass the well-known foodborne pathogen "emetic *B. cereus*" within
345 its genomospecies boundary (29, 37, 38). One of the "*B. paranthracis*" strains isolated here indeed
346 possessed cereulide synthetase-encoding genes and belonged to ST26 (Table 2), the ST that
347 encompasses most *B. cereus s.l.* strains capable of producing emetic toxin (38). This strain thus
348 likely poses a food safety threat and would most likely be referred to as "emetic *B. cereus*" in
349 clinical or industrial settings. Referring to this strain as "*B. paranthracis*" could be misleading to
350 researchers, clinicians, and other professionals who are not well-versed and up to date in *B. cereus*
351 *s.l.* taxonomy (29, 34, 38). However, not all "*B. paranthracis*" strains are capable of producing
352 emetic toxin: here, two ST2413 strains isolated from poultry were assigned to the "*B.*
353 *paranthracis*" genomospecies but did not possess cereulide synthetase-encoding genes (Table 2),

354 indicating that these strains cannot cause emetic intoxication. Thus, differentiating potentially
355 emetic from non-emetic strains is critical for informing public health and food safety efforts.

356 Recently, we proposed a standardized nomenclatural framework for *B. cereus s.l.* (i.e., the
357 2020 GSB framework), which can utilize genomic, genetic, and/or phenotypic information for
358 taxonomic classification (33, 34). Importantly, the 2020 GSB framework relies on a standardized
359 collection of biovars (i.e., biovar Anthracis, Emeticus, and Thuringiensis), which can be applied
360 to individual *B. cereus s.l.* strains to convey phenotypes of clinical and/or industrial importance
361 (i.e., ability to produce anthrax, emetic, and insecticidal toxins, respectively) (33, 34). Within this
362 framework, the absence of the Anthracis biovar terms denotes that the two "*B. anthracis*" strains
363 sequenced here cannot produce anthrax toxin, while the presence/absence of the Emeticus biovar
364 term differentiates cereulide-producing "*B. paranthracis*" from non-cereulide-producing strains
365 (Table 2). While the 2020 GSB framework provides a standardized set of *B. cereus s.l.*
366 genomospecies (Tables 1 and 2) (33, 34), researchers and other professionals may prefer to use
367 more established names for lineages (e.g., obtained via MLST, *panC* group assignment); biovar
368 terms can thus be appended to *B. cereus s.l.* lineage names (e.g., the ST26 strain sequenced here,
369 which possesses cereulide synthetase, can be referred to as "*B. cereus s.l.* ST26 biovar Emeticus").
370 Overall, standardized taxonomic frameworks that can incorporate both genomic/genetic and
371 phenotypic information may improve strain-level risk evaluation of *B. cereus s.l.*

372 **WGS may improve *B. cereus s.l.* genomic surveillance, traceback investigations, and source**
373 **tracking.** WGS has been shown to improve surveillance and source tracking efforts for numerous
374 foodborne pathogens, including *Escherichia coli, Salmonella enterica*, and *Listeria*
375 *monocytogenes* (50, 53). While the amount of publicly available WGS data derived from members
376 of *B. cereus s.l.* is increasing (34), efforts to sequence the genomes of food-associated *B. cereus*

377 *s.l.* strains are lagging relative to other foodborne pathogens. Here, we showed that WGS can
378 conceptually be used for *B. cereus s.l.* surveillance and source tracking; however, our study is
379 limited by a lack of publicly available (i) genomic data and (ii) corresponding metadata associated
380 with *B. cereus s.l.* strains. For example, we identified two identical *B. cereus s.l.* ST2413 strains,
381 which were present in both Dutch and South African raw poultry. However, due to a lack of
382 additional publicly available ST2413 *B. cereus s.l.* genomes, we were unable to gain additional
383 insights into exactly where this lineage originated. Thus, future *B. cereus s.l.* surveillance and
384 WGS initiatives in clinical, industrial, and environmental settings are needed to improve *B. cereus*
385 *s.l.* source tracking and traceback efforts. Furthermore, it is essential that data and metadata
386 obtained in such future initiatives are made publicly available, as international sharing of WGS
387 data can decrease both the amount of time required to solve foodborne outbreaks and the public
388 health burden caused by foodborne pathogens (54). Overall, the proof-of-concept study detailed
389 here highlights the benefits of WGS for *B. cereus s.l.* surveillance and source tracking, even among
390 closely related lineages, and future studies will benefit from increasingly available publicly
391 available WGS data and metadata.
392

393 **MATERIALS AND METHODS**

394 **Isolate selection and whole-genome sequencing.** A subset of 34 isolates were selected from a
395 total of 79 *B. cereus s.l.* isolates from our previous study (26) using simple random sampling
396 without replacement (55) via random numbers generated in Microsoft Excel. Culturing and
397 genomic DNA extraction was performed as described previously (26), using the High Pure PCR
398 Template preparation kit (Roche, Germany). WGS of selected isolates was performed at the
399 Biotechnology Platform, Agricultural Research Council, Onderstepoort, South Africa. DNA

400 libraries were prepared using TruSeq and Nextera DNA library preparation kits (Illumina, San
401 Diego, CA, USA), followed by sequencing on HiSeq and MiSeq instruments (Illumina, San Diego,
402 CA, USA).

403 **Data pre-processing and quality control.** Paired-end reads associated with each of the 34 isolates
404 were supplied as input to Trimmomatic v0.38 (56), which was used to remove Illumina adapters
405 and leading and trailing low-quality/ambiguous bases (LEADING:3 and TRAILING:3,
406 respectively); reads with average per-base quality scores <15 within a 4 bp sliding window
407 (SLIDINGWINDOW:4:15) were additionally cut, and reads with length < 36 bp were removed.
408 FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to assess
409 the quality of the resulting trimmed paired-end reads.

410 SKESA v2.4.0 (57) was used to assemble each genome, using the trimmed paired-end
411 reads as input and default settings. SPAdes v3.13.1 (58) was additionally used to assemble each
412 genome in "careful" mode, using trimmed paired-end reads as input. QUAST v4.5 (59) was used
413 to assess the quality of each resulting assembly, and CheckM v1.1.3 (60) was used to evaluate
414 genome contamination/completeness. MultiQC v1.8 (61) was used to assess genome quality in
415 aggregate. Assemblies produced using SKESA were used in all subsequent steps, as they were of
416 higher quality based on metrics produced by QUAST (e.g., N50, number of contigs). Genomes
417 with (i) < 95% completeness (via CheckM), (ii) > 5% contamination (via CheckM), and/or (iii)
418 N50 < 20 Kbp were considered to be of low quality and were excluded (n = 4), yielding a
419 preliminary set of 30 genomes used in subsequent analyses.

420 ***In silico* typing and taxonomic characterization.** BTyper3 v3.1.1 (33) was used to characterize
421 each assembled genome (see section "Data pre-processing and quality control" above) using: (i)
422 ANI-based genomospecies, (ii) ANI-based subspecies, and (iii) biovar assignment, using a

423 standardized nomenclatural framework for *B. cereus s.l.* (34) and dependencies FastANI v1.31
424 (62) and BLAST v2.9.0 (63); (iv) ANI-based pseudo-gene flow unit assignment (33) (also via
425 FastANI); (v) *in silico* seven-gene MLST using the PubMLST *B. cereus* database (accessed 25
426 October 2020); (vi) *panC* phylogenetic group assignment, using an adjusted eight-group (Group
427 I-VIII) framework (33). All aforementioned analyses were performed using default settings, as
428 well as with virulence factor minimum coverage thresholds lowered to 0% (--virulence_coverage
429 0) to confirm virulence factor absence. Because BTyper3 uses a conservative approach for Bt toxin
430 gene detection, the command-line implementation of BtToxin_scanner v1.0
431 (BtToxin_scanner2.pl) was used to identify Bt toxin genes in each genome using default settings
432 (64).

433 Taxonomic classification of assembled genomes was additionally performed using GTDB
434 Release 05-RS95 (17 July 2020) and GTDB-Tk v. 1.3.0 (i.e., "GTDB R95"), using GTDB-Tk's
435 "classify_wf" workflow (65-67). Notably, five genomes were assigned to species outside of *B.*
436 *cereus s.l.* (i.e., three genomes classified as *Escherichia flexneri,* one as *Escherichia dysenteriae*,
437 and one as *Staphylococcus saprophyticus* via GTDB-Tk) and were thus excluded, yielding a final
438 set of 25 *B. cereus s.l.* genomes used in subsequent analyses (Table 1 and Supplemental Table S1).

439 **Phylogenomic comparison of South African *B. cereus s.l.* genomes to *B. cereus s.l.* species**
440 **type strains.** Prokka v1.14.6 (68) was used to annotate each of the 25 *B. cereus s.l.* genomes
441 sequenced in this study. Protein coding sequences derived from the type strain genomes of each
442 of the 23 validly published and effective *B. cereus s.l.* species (accessed 28 August 2021) were
443 downloaded from NCBI's RefSeq Assembly database (see Table 1 of Méndez Acevedo, et al. for
444 all type strain accessions) (69). OrthoFinder v2.5.2 (70, 71) was used to identify orthologues
445 among protein coding sequences associated with all 47 genomes (25 *B. cereus s.l.* genomes

446 sequenced in this study, plus 23 *B. cereus s.l.* species type strain genomes), using MAFFT v7.475
447 (72, 73) for sequence alignment and RAxML-NG v1.0.2 (74) for phylogeny construction.
448 The resulting amino acid sequence alignment was supplied as input to IQ-TREE v1.5.4
449 (75), which was used to construct a maximum likelihood (ML) phylogeny, using 1,000 replicates
450 of the ultrafast bootstrap approximation (76), plus the optimal amino acid substitution model
451 selected using ModelFinder (i.e., the general matrix model with empirical amino acid frequencies
452 and the FreeRate model with six categories; JTT+F+R6) (77-80). The resulting phylogeny was
453 rooted using effective species "*B. manliponensis*" (i.e., the most distant recognized member of *B.*
454 *cereus s.l.*) (81) and annotated using the bactaxR package (34) in R v4.1.2 (82).
455 **Acquisition of publicly available *B. cereus s.l.* genomes.** All assembled genomes submitted to
456 the National Center for Biotechnology Information (NCBI) RefSeq database (83, 84) as one of 23
457 validly published or effective *B. cereus s.l.* species (i.e., *albus, anthracis,* "*bingmayongensis*",
458 *cereus,* "*clarus*", *cytotoxicus, fungorum,* "*gaemokensis*", *luti,* "*manliponensis*", *mobilis, mycoides,*
459 *nitratireducens, pacificus, paramycoides, paranthracis, proteolyticus, pseudomycoides,*
460 *thuringiensis, toyonensis, tropicus, weihenstephanensis, wiedmannii*) (1, 29, 52, 69, 81, 85-97)
461 were downloaded (*n* = 2,733; accessed 20 March 2021). QUAST and CheckM were used to assess
462 the quality of each assembled genome (see section "Data pre-processing and quality control"
463 above), and BTyper3 (using default settings) and GTDB-Tk were used for typing and/or taxonomic
464 assignment as described above (see section "*In silico* typing and taxonomic characterization"). The
465 rentrez package (v1.2.3) was used to download metadata associated with each genome's
466 BioSample in R v3.6.1 (82, 98, 99). Publicly available genomes meeting all of the following
467 quality thresholds were used in subsequent analyses (*n* = 2,664; Supplemental Table S2): (i) > 95

468 % completeness (via CheckM), (ii) < 5% contamination (via CheckM), (iii) N50 > 20 Kbp (via
469 QUAST), (iv) composed of < 1,000 contigs (via QUAST).

470 **Acquisition of genomes from a study of *B. thuringiensis* outbreaks.** All sequencing reads
471 associated with isolates from a previous study of outbreaks caused by *B. thuringiensis* in France
472 (100) were downloaded from NCBI's Sequence Read Archive (SRA) database using the SRA
473 Toolkit v 2.8.2 (101, 102). Genomic data for all 171 isolates were pre-processed, assembled, and
474 taxonomically classified as described above, with genomes assembled using SKESA used in
475 subsequent steps (see sections "Data pre-processing and quality control" and "*In silico* typing and
476 taxonomic characterization" above). Four genomes did not meet the quality thresholds employed
477 in this study (see section "Acquisition of publicly available *B. cereus s.l.* genomes" above) and
478 were thus excluded, yielding 167 genomes from the study, which were used in subsequent analyses
479 (Supplemental Table S3).

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

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

498 For each of the three major lineages, Prokka was used to annotate each genome; the
499 resulting GFF files associated with each genome were supplied as input to Panaroo v1.2.8 (103),
500 which was used to partition genes into core- and pan-genome orthologous gene clusters, using the
501 following parameters (all other parameters were set to their default values): (i) "strict" mode (--
502 clean-mode strict), (ii) core genome alignment using MAFFT (-a core --aligner mafft), (iii) a core
503 genome sample threshold of 95% (--core_threshold 0.95). The resulting core genome (nucleotide)
504 alignment was queried using snp-sites v2.5.1 (104), which was used to identify (i) core SNPs and
505 (ii) constant sites among all genomes in the major lineage. The resulting core SNP alignment was
506 supplied as input to IQ-TREE v1.5.4, which was used to construct a ML phylogeny using the
507 General Time-Reversible (GTR) nucleotide substitution model (105), one thousand replicates of
508 the ultrafast bootstrap approximation (76), and an ascertainment bias correction obtained using
509 constant sites output by snp-sites.

510 For each of the three major lineages, all aforementioned steps were repeated, with the
511 addition of an outgroup genome: for the *panC* Group IV phylogeny, *panC* Group III *B. anthracis*
512 str. Ames Ancestor was used as an outgroup (NCBI RefSeq Accession GCF_000008445.1); for

513 the *panC* Groups II/III and *panC* Group V phylogenies, *panC* Group IV *B. cereus* str. ATCC 14579
514 was used as an outgroup (NCBI RefSeq Accession GCF_006094295.1). Additionally, only
515 genomes with detailed metadata (i.e., a reported year of isolation, isolation source, and geographic
516 location) were included in this analysis (Supplemental Tables S1-S4). Each of the three resulting
517 phylogenies were rooted and scaled using LSD2 v1.4.2.2 (106) and the following parameters: (i)
518 tip dates corresponding to the year of isolation associated with each genome; (ii) constrained mode
519 (-c), with the root estimated using constraints on all branches (-r as); (iii) variances calculated
520 using input branch lengths (-v 1); (iv) 1,000 samples for calculating confidence intervals for
521 estimated dates (-f 1000); (v) a sequence length of 5,500,000 (-s 5500000); (vi) rooting along the
522 outgroup genome (-g -k). The resulting phylogenies were annotated using the bactaxR package in
523 R.

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

531 For each MLST lineage, Snippy v4.6.0 (https://github.com/tseemann/snippy) was used to
532 identify core SNPs among all genomes assigned to the respective MLST lineage, using (i) a
533 genome sequenced in this study as a reference genome (Table 3); (ii) paired-end reads associated
534 with each genome as input (for the genomes sequenced in this study, as well as the genomes from
535 the Bonis, et al. study and the New York State outbreak study) (37, 100) and/or assembled genomes

536 as input (for NCBI genomes); (iii) default settings. For MLST lineages with more than four
537 genomes (i.e., ST24, ST26, ST177, ST223, ST1578), Gubbins v3.1.3 (107) was used to remove
538 recombination, and core SNPs were identified within the resulting filtered alignment using snp-
539 sites. For all MLST lineages, pairwise core SNP distances were calculated within the MLST
540 lineage (i) among all genomes, (ii) among genomes sequenced in this study, and (iii) between
541 genomes sequenced in this study and publicly available genomes (Table 3), using the dist.gene
542 function in the ape package (108, 109) in R.

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

548

549 **ACKNOWLEDGMENTS**

570 The following organisations are acknowledged for their contributions:

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

577

578

579

580

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.

605 **REFERENCES**

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

608 2. Jouzani GS, Valijanian E, Sharafi R. 2017. *Bacillus thuringiensis*: a successful
609 insecticide with new environmental features and tidings. Appl Microbiol Biotechnol
610 101:2691-2711.

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

614 4. Azizoglu U. 2019. *Bacillus thuringiensis* as a Biofertilizer and Biostimulator: a Mini-
615 Review of the Little-Known Plant Growth-Promoting Properties of Bt. Curr Microbiol
616 76:1379-1385.

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

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

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

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

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

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

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

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

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

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

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

641 16. Bottone EJ. 2010. *Bacillus cereus*, a volatile human pathogen. Clin Microbiol Rev
642 23:382-98.

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

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

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

655 20. Tewari A, Singh SP, Singh R. 2015. Incidence and enterotoxigenic profile of *Bacillus*
656 *cereus* in meat and meat products of Uttarakhand, India. J Food Sci Technol 52:1796-
657 801.

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

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

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

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

671 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,
672 Pang R, Zhang Y, Wei X, Zhang J, Wu Q, Ding Y. 2020. A Study on Prevalence and
673 Characterization of *Bacillus cereus* in Ready-to-Eat Foods in China. Frontiers in
674 Microbiology 10.

675 26. Madoroba E, Magwedere K, Chaora NS, Matle I, Muchadeyi F, Mathole MA, Pierneef R.
676 2021. Microbial Communities of Meat and Meat Products: An Exploratory Analysis of
677 the Product Quality and Safety at Selected Enterprises in South Africa. Microorganisms
678 9.

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

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

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

687 30. Liu Y, Lai Q, Goker M, Meier-Kolthoff JP, Wang M, Sun Y, Wang L, Shao Z. 2015.
688 Genomic insights into the taxonomic status of the *Bacillus cereus* group. Sci Rep
689 5:14082.

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

693 32. Guinebretiere MH, Velge P, Couvert O, Carlin F, Debuyser ML, Nguyen-The C. 2010.
694 Ability of *Bacillus cereus* group strains to cause food poisoning varies according to
695 phylogenetic affiliation (groups I to VII) rather than species affiliation. J Clin Microbiol
696 48:3388-91.

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

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

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

705 36. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics:
706 BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res
707 3:124.

708 37. Carroll LM, Wiedmann M, Mukherjee M, Nicholas DC, Mingle LA, Dumas NB, Cole
709 JA, Kovac J. 2019. Characterization of Emetic and Diarrheal *Bacillus cereus* Strains
710 From a 2016 Foodborne Outbreak Using Whole-Genome Sequencing: Addressing the
711 Microbiological, Epidemiological, and Bioinformatic Challenges. Front Microbiol
712 10:144.

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

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

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

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

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

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

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

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

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

740 46. Rohr JR, Barrett CB, Civitello DJ, Craft ME, Delius B, DeLeo GA, Hudson PJ, Jouanard
741 N, Nguyen KH, Ostfeld RS, Remais JV, Riveau G, Sokolow SH, Tilman D. 2019.
742 Emerging human infectious diseases and the links to global food production. Nat Sustain
743 2:445-456.

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

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

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

753 50. Brown B, Allard M, Bazaco MC, Blankenship J, Minor T. 2021. An economic evaluation
754 of the Whole Genome Sequencing source tracking program in the U.S. PLoS One
755 16:e0258262.

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

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

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

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

768 55. Rose S. 2015. Management research : applying the principles, First Edition. ed.
769 Routledge, Taylor & Francis Group, New York.

770 56. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
771 sequence data. Bioinformatics 30:2114-20.

772 57. Souvorov A, Agarwala R, Lipman DJ. 2018. SKESA: strategic k-mer extension for
773 scrupulous assemblies. Genome Biology 19:153.

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

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

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

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

785 62. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput
786 ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun
787 9:5114.

788 63. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.
789 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

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

793 65. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify
794 genomes with the Genome Taxonomy Database. Bioinformatics
795 doi:10.1093/bioinformatics/btz848.

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

799 67. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. 2020. A
800 complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol
801 38:1079-1086.

802 68. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics
803 30:2068-9.

804 69. Mendez Acevedo M, Carroll LM, Mukherjee M, Mills E, Xiaoli L, Dudley EG, Kovac J.
805 2020. Novel Effective *Bacillus cereus* Group Species "*Bacillus clarus*" Is Represented by
806 Antibiotic-Producing Strain ATCC 21929 Isolated from Soil. mSphere 5.

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

809 71. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for
810 comparative genomics. Genome Biol 20:238.

811 72. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid
812 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res
813 30:3059-66.

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

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

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

822 76. Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic
823 bootstrap. Mol Biol Evol 30:1188-95.

824 77. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
825 ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods
826 14:587-589.

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

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

832 80. Yang Z. 1995. A space-time process model for the evolution of DNA sequences.
833 Genetics 139:993-1005.

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

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

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

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

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

848 86. Sternbach G. 2003. The history of anthrax. The Journal of Emergency Medicine 24:463-
849 467.

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

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

856 89. Guinebretiere MH, Auger S, Galleron N, Contzen M, De Sarrau B, De Buyser ML,
857 Lamberet G, Fagerlund A, Granum PE, Lereclus D, De Vos P, Nguyen-The C, Sorokin
858 A. 2013. *Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus*
859 *cereus* Group occasionally associated with food poisoning. Int J Syst Evol Microbiol
860 63:31-40.

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

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

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

868 93. Nakamura LK. 1998. *Bacillus pseudomycoides* sp. nov. Int J Syst Bacteriol 48 Pt 3:1031-
869 5.

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

872 95. Jimenez G, Urdiain M, Cifuentes A, Lopez-Lopez A, Blanch AR, Tamames J, Kampfer
873 P, Kolsto AB, Ramon D, Martinez JF, Codoner FM, Rossello-Mora R. 2013. Description
874 of *Bacillus toyonensis* sp. nov., a novel species of the *Bacillus cereus* group, and pairwise
875 genome comparisons of the species of the group by means of ANI calculations. Syst Appl
876 Microbiol 36:383-91.

877 96. Miller RA, Beno SM, Kent DJ, Carroll LM, Martin NH, Boor KJ, Kovac J. 2016.
878 *Bacillus wiedmannii* sp. nov., a psychrotolerant and cytotoxic *Bacillus cereus* group
879 species isolated from dairy foods and dairy environments. Int J Syst Evol Microbiol
880 66:4744-4753.

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

884 98. Winter DJ. 2017. rentrez: An R package for the NCBI eUtils API. The R Journal 9:520-
885 526.

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

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

894 101. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database C.
895 2011. The sequence read archive. Nucleic Acids Res 39:D19-21.

896 102. Kodama Y, Shumway M, Leinonen R, International Nucleotide Sequence Database C.
897 2012. The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids
898 Res 40:D54-6.

899 103. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA,
900 Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J. 2020.
901 Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol
902 21:180.

903 104. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-
904 sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom
905 2:e000056.

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

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

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

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

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

917

918

919

920

921 **TABLES**

922 **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*	*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*	*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*

923 [a]Year of isolation
924 [b]RTE, Ready-to-Eat
925 [c]Two strains were isolated from meat products imported from the Netherlands (NL)
926 [d]*panC* phylogenetic group assigned using BTyper3
927 [e]Sequence type (ST) assigned using the PubMLST seven-gene multi-locus sequence typing (MLST) scheme for *B. cereus* and
928 BTyper3

929 [f]Genome Taxonomy Database (GTDB) species assigned using GTDB-Tk
930 [g]Species and subspecies (where applicable), assigned using the 2020 Genomospecies-Subspecies-Biovar (GSB) nomenclatural
931 framework for *B. cereus s.l.* and BTyper3 (see Table 2 for predicted biovar/phenotypic information)
932

933 **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*	-	-	+	-	+	-/-/+	
S59	III	2413	*paranthracis*	-	-	+	-	-	-/-/-	*B. mosaicus*
S62	III	2413	*paranthracis*	-	-	+	-	-	-/-/+	
S64	III	NA	*anthracis*	-	-	+	-	+	-/-/+	
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

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

937 [c]Genome Taxonomy Database (GTDB) species assigned using GTDB-Tk
938 [d]Viurlence factors detected in each genome using BTyper3 and default thresholds (70% amino acid identity and 80% coverage);
939 presence and absence of virulence factors are denoted by "+" and "-", respectively
940 [e]All of *cya, lef, pagA* (toxin), *capABCDE*, *hasABC*, and/or *bpsXABCDEFGH* (capsules)
941 [f]*B. thuringiensis* (Bt) insecticidal toxin-encoding genes detected using (i) BTyper3 (which uses a conservative BLAST-based approach
942 and minimum default amino acid and coverage thresholds of 50 and 70%, respectively), and BtToxin_scanner2's (ii) "old" and (iii)
943 "new" gene detection approaches (detected using default thresholds), each separated by a slash ("/")
944 [g]Species, subspecies (where applicable), and biovar (where applicable) assigned using the 2020 Genomospecies-Subspecies-Biovar
945 (GSB) nomenclatural framework for *B. cereus s.l.* and BTyper3; multiple taxonomic labels are listed for strains that can be referenced
946 using shorted subspecies and/or biovar notation (separated by a semi-colon)
947 [h]Biovar Thuringiensis was assigned to genomes in which Bt insecticidal toxin-encoding genes were detected using
948 BtToxin_scanner2's "old" gene detection approach and default settings (which is less conservative than BTyper3's Bt toxin gene
949 detection approach)
950
951

952 **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*	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)

953 [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
954 asterisk contain two or more strains sequenced in this study, which were highly similar on a genomic scale; NA, not
955 available/applicable
956 [b]Genome Taxonomy Database (GTDB) species assigned using GTDB-Tk
957 [c]Sequence type (ST) assigned using the PubMLST seven-gene MLST scheme for *B. cereus* and BTyper3
958 [d]FastANI average nucleotide identity (ANI) range and mean values between all genomes in the lineage (excludes self-comparisons)
959 [e]Strain sequenced in this study, which was used as a reference genome for single-nucleotide polymorphism (SNP) calling among all
960 genomes within the lineage via Snippy
961 [f]Pairwise SNP distances calculated among all genomes within the lineage (excludes self-comparisons)
962 [g]Pairwise SNP distances calculated among all genomes within the lineage that were sequenced in this study (excludes self-
963 comparisons)
964 [h]Pairwise SNP distances calculated between genomes sequenced in this study and public genomes within the same lineage (excludes
965 self-comparisons)
966 [i]For polyphyletic ST24, one outlier ST24 genome was excluded (NCBI RefSeq Accession GCF_010580595.1)
967 [j]ST was polyphyletic

968 **FIGURE LEGENDS**

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

970

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

988

989

990 **Figure 3**. Maximum likelihood (ML) phylogeny constructed using core SNPs identified among
991 orthologous gene clusters of 1,081 *panC* Group IV *B. cereus s.l.* genomes. The phylogeny was
992 rooted using *panC* Group III *B. anthracis* str. Ames Ancestor as an outgroup (NCBI RefSeq
993 Accession GCF_000008445.1; omitted for readability). Tip label colors and clade labels denote
994 species assigned using GTDB-Tk ("GTDB Species"). The heatmap to the right of the phylogeny
995 denotes (i) whether a strain was sequenced in this study (dark pink) or not (light pink; "Study"),
996 and (ii) multi-locus sequence typing (MLST) sequence types (STs) associated with strains
997 sequenced in this study, where applicable (colored), or not (gray; "ST"). MLST lineages
998 discussed in Table 3 are annotated to the right of the heatmap ("MLST Lineage"). MLST
999 lineages with numerical superscripts contain two or more strains sequenced in this study, which
1000 were highly similar on a genomic scale; these lineages are depicted in Figure 4. The tree was
1001 rooted and time-scaled using LSD2, with branch lengths reported in years.
1002
1003 **Figure 4.** *B. cereus s.l.* multi-locus sequence typing (MLST) lineages that contained two or more
1004 strains sequenced in this study, which were identical or nearly identical at the whole-genome
1005 scale. Lineage names and sequence types (STs), where applicable, are shown in the top right
1006 corner of each panel. Geographic and source origins of each strain are displayed in the respective
1007 map.
1008
1009 **Figure 5.** Maximum likelihood phylogeny constructed using core SNPs identified among
1010 orthologous gene clusters of 597 *panC* Group II and III *B. cereus s.l.* genomes. The phylogeny
1011 was rooted using *panC* Group IV *B. cereus* str. ATCC 14579 as an outgroup (NCBI RefSeq
1012 Accession GCF_006094295.1; omitted for readability). Tip label colors and clade labels denote

1013 species assigned using GTDB-Tk ("GTDB Species"). The heatmap to the right of the phylogeny
1014 denotes (i) whether a strain possesses anthrax toxin-encoding genes *cya, lef,* and *pagA* (dark
1015 orange) or not (light orange; "Anthrax"); (ii) whether a strain possesses cereulide synthetase-
1016 encoding *cesABCD* (dark purple) or not (light purple; "Emetic"); (iii) whether a strain belongs to
1017 *panC* Group II (blue) or III (yellow; *"panC*"*);* (iv) whether a strain was sequenced in this study
1018 (dark pink) or not (light pink; "Study"); (v) multi-locus sequence typing (MLST) sequence types
1019 (STs) associated with strains sequenced in this study, where applicable (colored), or not (gray;
1020 "ST"). MLST lineages discussed in Table 3 are annotated to the right of the heatmap ("MLST
1021 Lineage"). MLST lineages with numerical superscripts contain two or more strains sequenced in
1022 this study, which were highly similar on a genomic scale; these lineages are depicted in Figure 4.
1023 The tree was rooted and time-scaled using LSD2, with branch lengths reported in years.
1024
1025 **Figure 6.** Maximum likelihood phylogeny constructed using core SNPs identified among
1026 orthologous gene clusters of 219 *panC* Group V *B. toyonensis* genomes. The phylogeny was
1027 rooted using *panC* Group IV *B. cereus* str. ATCC 14579 as an outgroup (NCBI RefSeq
1028 Accession GCF_006094295.1; omitted for readability). The heatmap to the right of the
1029 phylogeny denotes (i) whether a strain was sequenced in this study (dark pink) or not (light pink;
1030 "Study"); (ii) multi-locus sequence typing (MLST) sequence types (STs) associated with strains
1031 sequenced in this study, where applicable (colored), or not (gray; "ST"). MLST lineages
1032 discussed in Table 3 are annotated to the right of the heatmap. The tree was rooted and time-
1033 scaled using LSD2, with branch lengths reported in years.

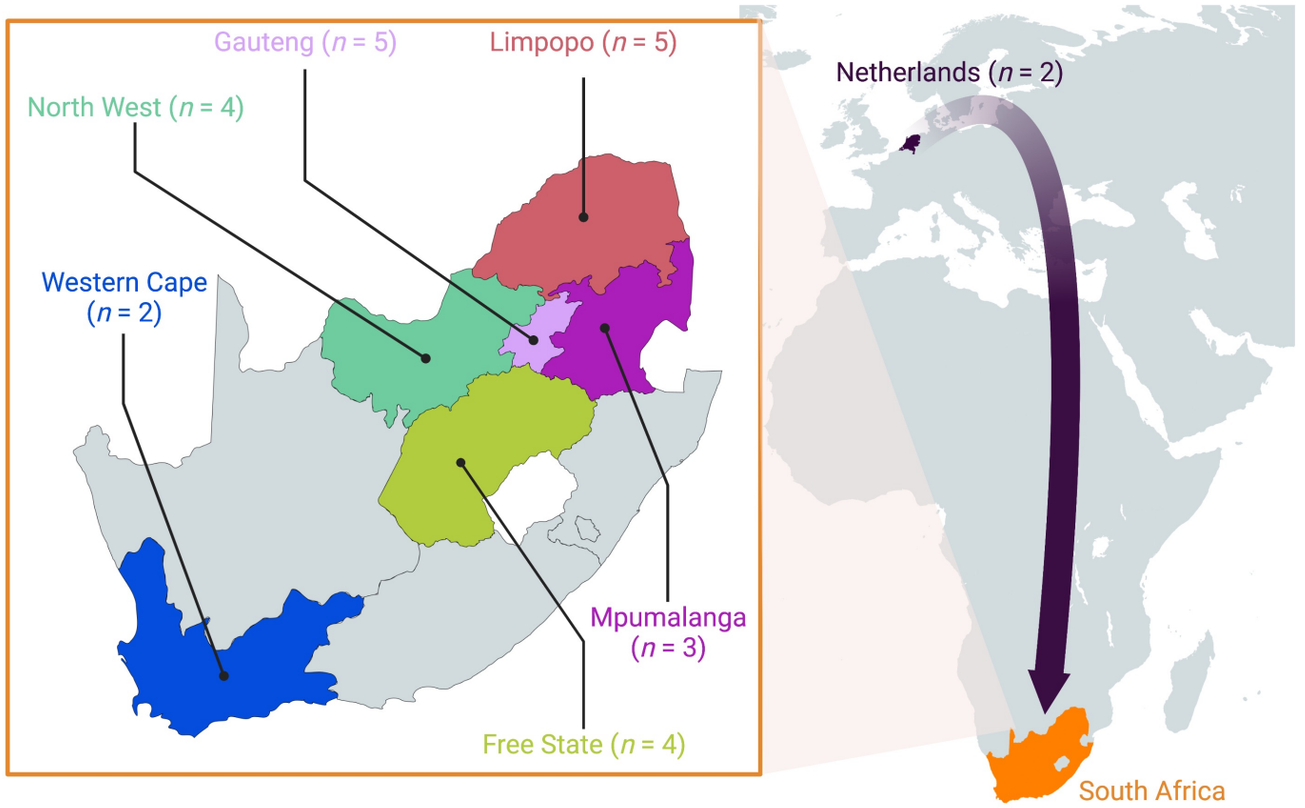

Figure 1. Geographic origins of *B. cereus s.l.* strains sequenced in this study ($n = 25$).

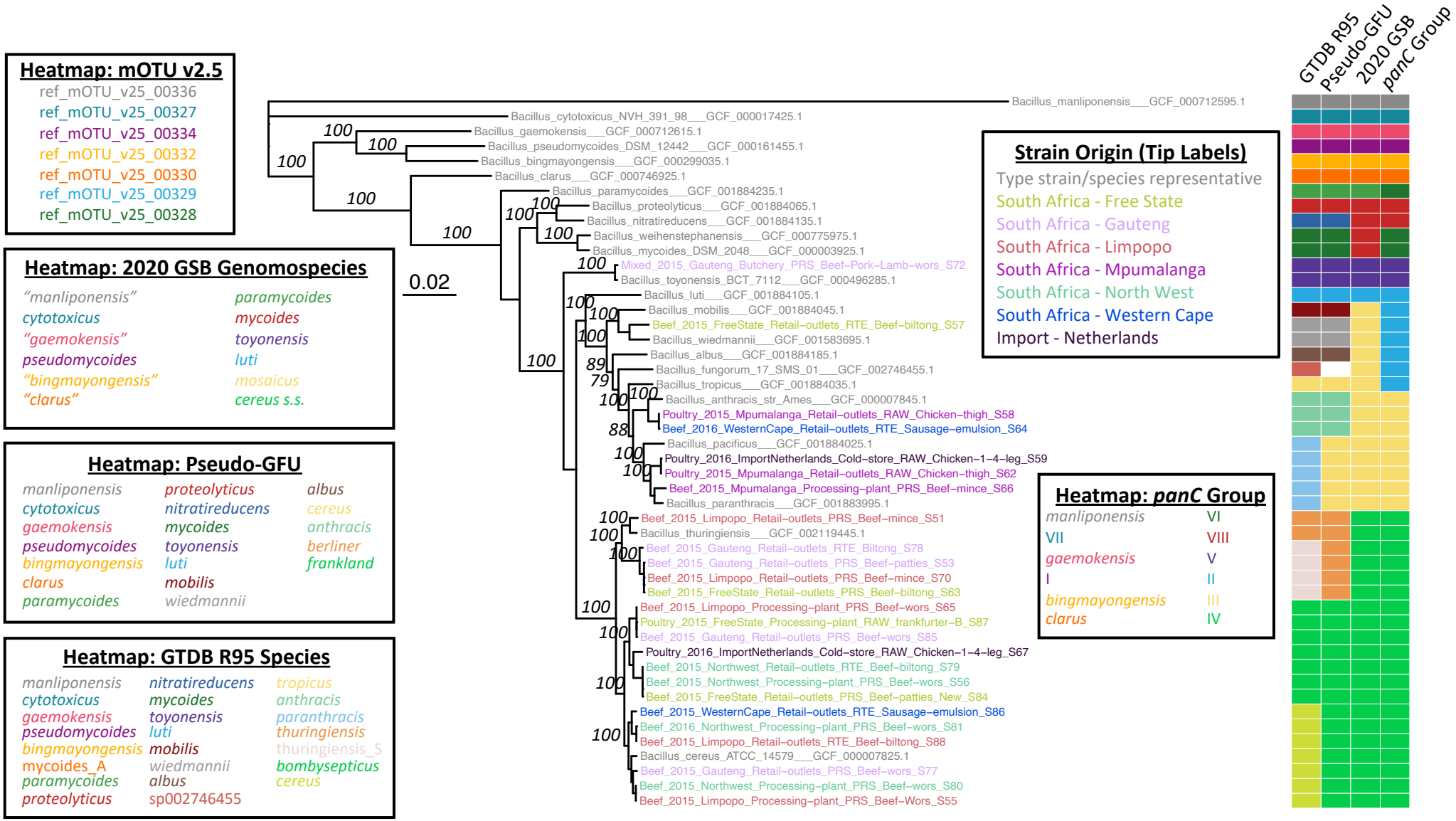

Figure 2. Maximum likelihood (ML) phylogeny constructed using amino acid sequences derived from the 25 *B. cereus s.l.* isolate genomes sequenced in this study (tip labels colored by geographic origin), plus type strain/species representative genomes of 23 published and effective *B. cereus s.l.* species (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.

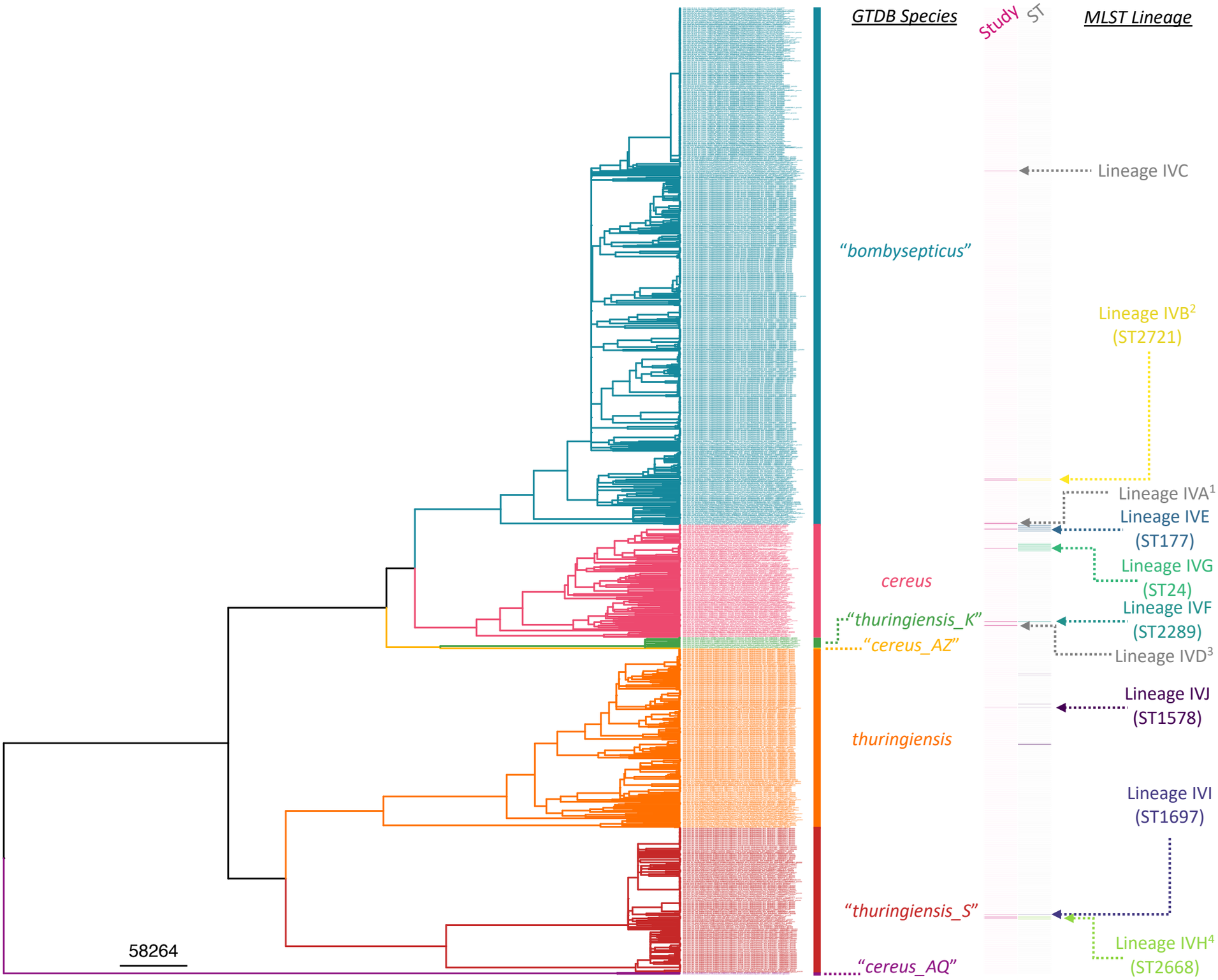

Figure 3. Maximum likelihood (ML) phylogeny constructed using core SNPs identified among orthologous gene clusters of 1,081 *panC* Group IV *B. cereus s.l.* genomes. The phylogeny was rooted using *panC* Group III *B. anthracis* str. Ames Ancestor as an outgroup (NCBI RefSeq Accession GCF_000008445.1; 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.

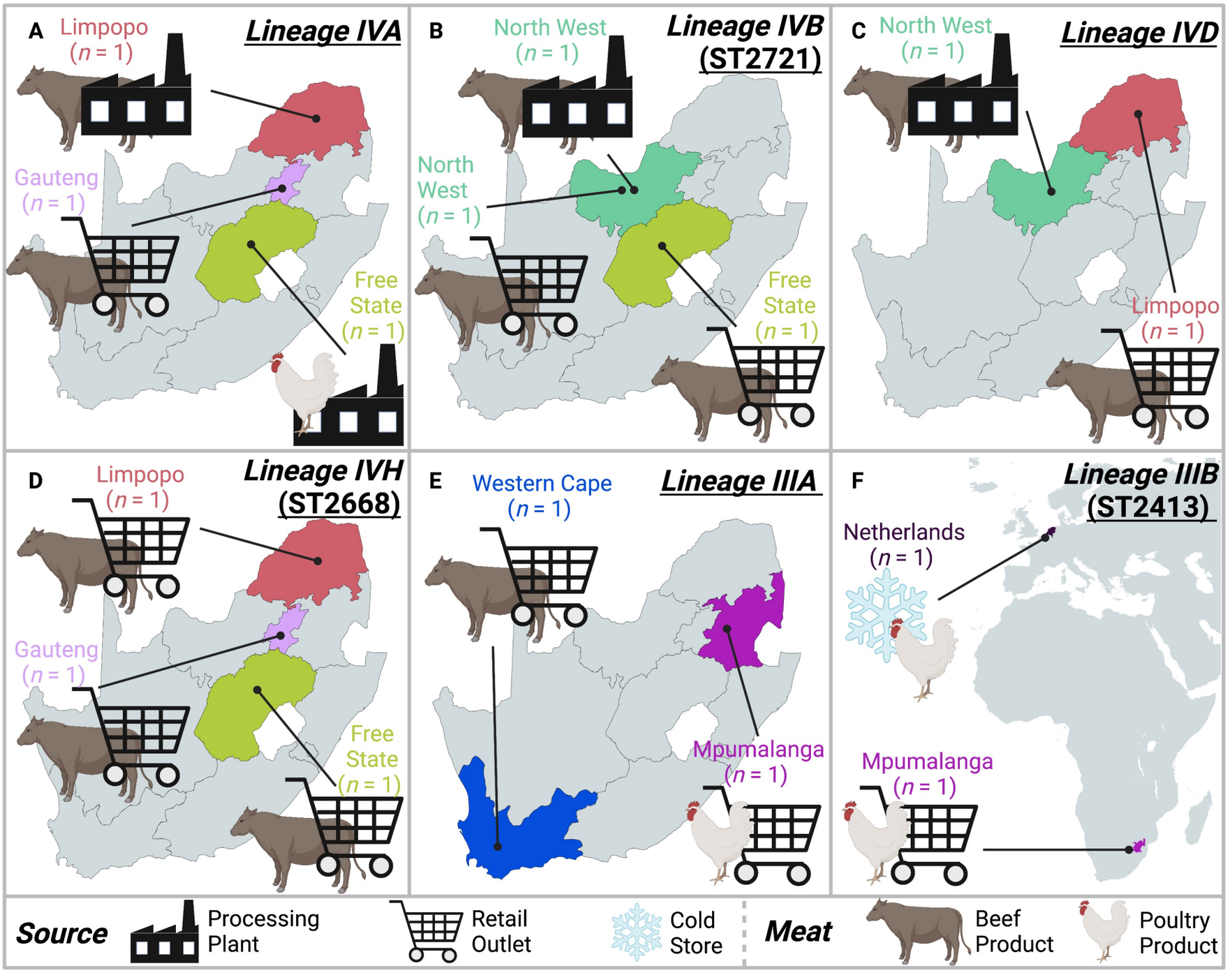

Figure 4. *B. cereus s.l.* multi-locus sequence typing (MLST) lineages that contained two or more strains sequenced in this study, which were identical or nearly identical at the whole-genome scale. Lineage names and sequence types (STs), where applicable, are shown in the top right corner of each panel.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

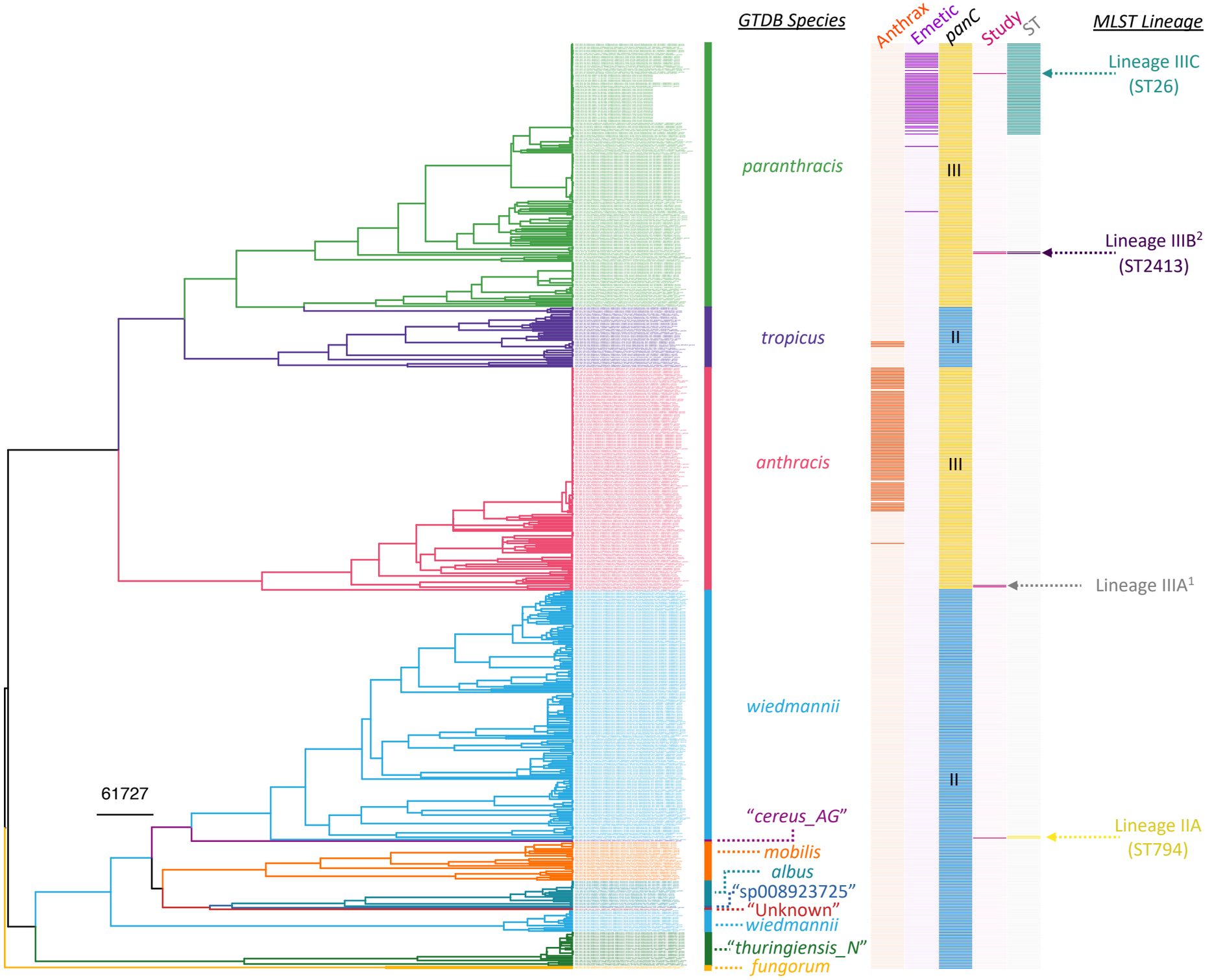

Figure 5. Maximum likelihood phylogeny constructed using core SNPs identified among orthologous gene clusters of 597 *panC* Group II and III *B. cereus s.l.* genomes. The phylogeny was rooted using *panC* Group IV *B. cereus* str. ATCC 14579 as an outgroup (NCBI RefSeq Accession GCF_006094295.1; omitted 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.

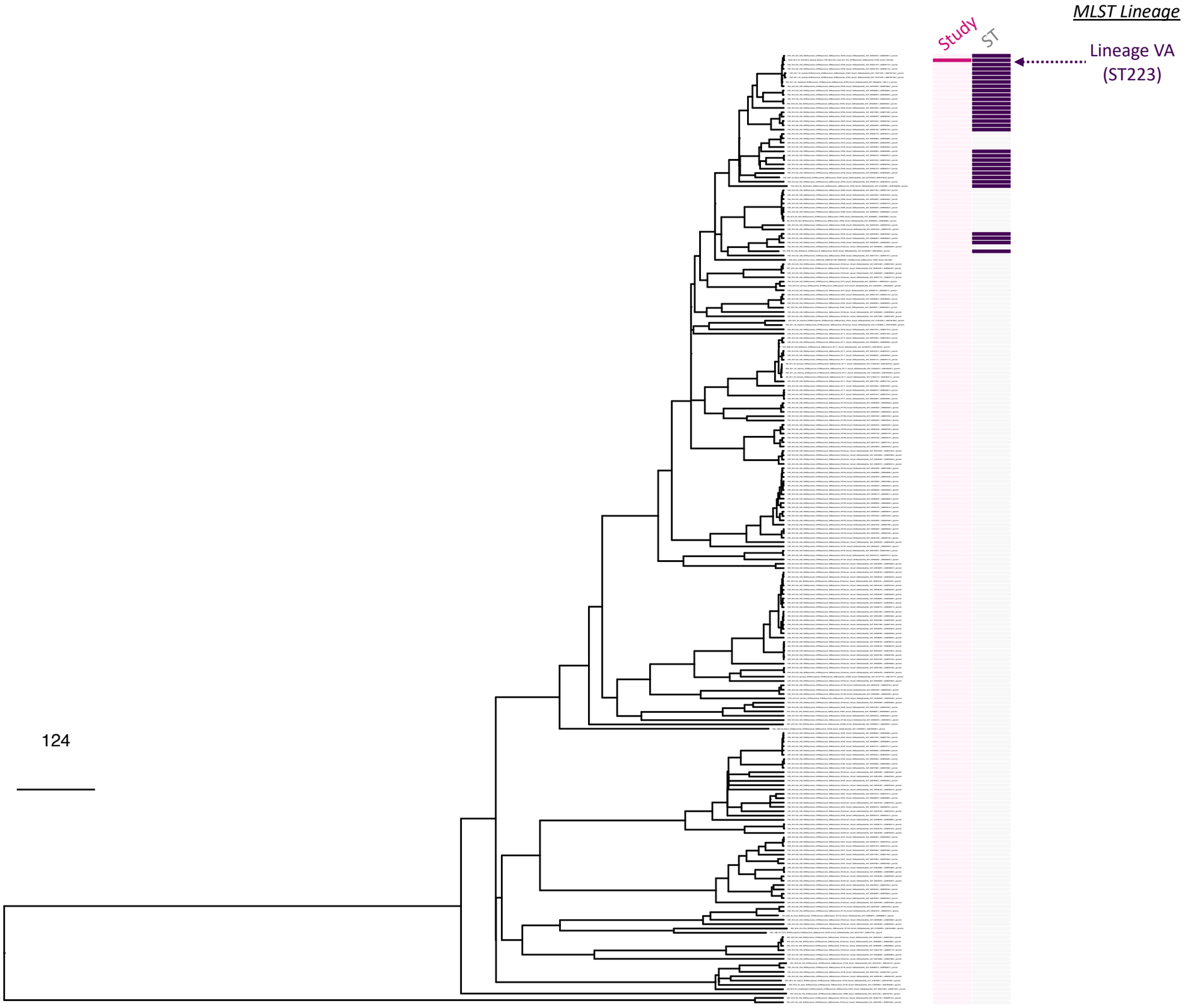

Figure 6. Maximum likelihood phylogeny constructed using core SNPs identified among orthologous gene clusters of 219 *panC* Group V *B. toyonensis* genomes. The phylogeny was rooted using *panC* Group IV *B. cereus* str. ATCC 14579 as an outgroup (NCBI RefSeq Accession GCF_006094295.1; omitted for readability). 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.