International Journal of Food Microbiology 439 (2025) 111266

Contents lists available at ScienceDirect

International Journal of Food Microbiology

journal homepage: www.elsevier.com/locate/ijfoodmicro

Whole-genome sequencing of soil- and foodborne *Bacillus cereus* sensu lato indicates no clear association between their virulence repertoire, genomic diversity and food matrix

Bram Jacobs [a,b,c,*], Bert Bogaerts [d], Marie Verhaegen [c], Kevin Vanneste [d], Sigrid C.J. De Keersmaecker [d], Nancy H.C. Roosens [d], Andreja Rajkovic [b], Jacques Mahillon [c], Tom Van Nieuwenhuysen [a], Koenraad Van Hoorde [a]

[a] *Foodborne Pathogens, Sciensano, Juliette Wytsmanstraat 14, Brussels, Belgium*
[b] *Laboratory of Food Microbiology and Food Preservation, Department of Food Technology, Safety and Health, Faculty of Bioscience Engineering, Ghent University, Coupure Links 635, Ghent, Belgium*
[c] *Laboratory of Food and Environmental Microbiology, Earth and Life Institute, Catholic University of Louvain, Croix du Sud 2, Louvain-la-Neuve, Belgium*
[d] *Transversal activities in Applied Genomics, Sciensano, Juliette Wytsmanstraat 14, Brussels, Belgium*

ARTICLE INFO

Keywords:
Bacillus cereus
WGS
cgMLST
SNP
Virulence genes
Risk assessment

ABSTRACT

Bacillus cereus sensu lato is frequently involved in foodborne toxico-infections and is found in various foodstuff. It is unclear whether certain strains have a higher affinity for specific food matrices, which can be of interest for risk assessment. This study reports the characterization by whole-genome sequencing of 169 *B. cereus* isolates, isolated from 12 food types and soil over two decades. Any potential links between the food matrix of isolation, the isolate's genetic lineage and/or their (putative) virulence gene reservoir were investigated. More than 20 % of the strains contained the genes for the main potential enterotoxins (*nheABC*, *hblCDA* and *cytK_2*). Cereulide biosynthesis genes and genes encoding hemolysins and phospholipases, were detected in multiple isolates. Strain typing revealed a high diversity, as illustrated by 84 distinct sequence types, including 26 not previously described. This diversity was also reflected in the detection of all seven *panC* types and 71 unique virulence gene profiles. Core-genome MLST was used for phylogenomic investigation of the entire collection and SNP-based clustering was performed on the four most abundant sequence types, which did not reveal a clear affinity for specific *B. cereus* lineages or (putative) virulence genes for certain food matrices. Additionally, minimal genetic overlap was observed between soil and foodborne isolates. Clusters of closely-related isolates with common epidemiological metadata were detected. However, some isolates from different food matrices or collected several years apart were found to be genetically identical. This study provides elements that can be used for risk assessment of *B. cereus* in food.

1. Introduction

B. cereus is often denoted as *Bacillus cereus senso lato* (*s.l.*) or the *Bacillus cereus* group as an umbrella term to cover lineages of multiple *Bacillus* species (Carroll et al., 2020a; Torres Manno et al., 2020). *B. cereus s.l.* varies from commensal to probiotic, bioinsecticidal and food pathogenic agents. Its food poisoning activity is mediated by the production of two types of toxins leading to intoxication often distinguished by two distinct syndromes, emesis and diarrhea.

The emetic syndrome is caused by the ingestion of cereulide, a small circular oligopeptide secreted in food prior to ingestion. The biosynthesis machinery of this emetic toxin is encoded by a non-ribosomal peptide synthesis (NRPS) gene cluster (*cesABCDPTH*) located on the 270 kb megaplasmid pCER270 (Ehling-Schulz et al., 2006; Hoton et al., 2005). The proteins encoded by *cesABCD* are involved in the production and translocation of cereulide, whilst the proteins encoded by *cesPT* are involved in NRPS activation and post-translational regulation. Additionally, the protein encoded by the *cesH* gene, possessing its own promotor sequence, is involved in controlling cereulide biosynthesis at the transcriptional level (Lücking et al., 2015). Typically, strains producing

* Corresponding author at: Foodborne Pathogens, Sciensano, Juliette Wytsmanstraat 14, Brussels, Belgium.
E-mail address: Bram.Jacobs@sciensano.be (B. Jacobs).

https://doi.org/10.1016/j.ijfoodmicro.2025.111266
Received 9 January 2025; Received in revised form 29 April 2025; Accepted 10 May 2025
Available online 14 May 2025

cereulide are detected in relatively low numbers in foods (Jovanovic et al., 2021). The resistance of cereulide to extreme temperatures, pH and (digestive) enzymes makes it very difficult to eliminate during food processing (Decleer et al., 2018; Rajkovic et al., 2008).

In addition to the emetic syndrome, *B. cereus s.l.* strains can also cause a diarrheic syndrome by the production of different enterotoxins. These enterotoxins are produced by the bacterium in the gastrointestinal tract of the host after ingestion and subsequent germination of endospores. The three main potential enterotoxins are the non-hemolytic enterotoxin (Nhe), the hemolytic enterotoxin (Hbl) and cytotoxin K2 (CytK_2). Nhe and Hbl are two tripartite toxins encoded by the *nheABC* and *hblCDA* operons, respectively. On the contrary, CytK_2 is a monopartite toxin encoded by the *cytK_2* gene**.** CytK_2 is widely distributed over *B. cereus s.l.* but its contribution to the development of disease remains a matter of debate (Castiaux et al., 2015). CytK_2 has a homologue enterotoxin, CytK_1, to which it shares 89 % protein identity. The occurrence of CytK_1 is uniquely related to the thermotolerant member of the group: *Bacillus cytotoxicus* (Guinebretière et al., 2013). Several strains of *B. cereus s.l.* are also capable of producing other (putative) virulence factors such as hemolysins (e.g. Clo, HlyII and HlyIII), phospholipases (PlcA, PlcB), immune inhibitor proteases (*inhA1* and *inhA2*) (Jovanovic et al., 2021). All these enterotoxins and (putative) virulence factors, except for *ces*, are regulated by the phospholipase C regulator PlcR (Gohar et al., 2008). These (putative) virulence genes can be found throughout the *B. cereus* group, whilst the genes involved in the etiology of anthrax disease (*lef*, *cya* and *pagA*) and the anthrax toxin gene regulator *atx* are specific to *Bacillus anthracis*, the most notorious member of the *B. cereus s.l.* group (Ehling-Schulz et al., 2019; Stenfors Arnesen et al., 2008).

The European Food Safety Authority (EFSA) reported a significant increase in the number of foodborne outbreaks caused by *B. cereus s.l.* in 2022 compared to 2021 (EFSA and ECDC, 2023). Therefore, it is important to understand its potential hazard for an adequate risk assessment. One of the elements is the detection of (putative) virulence factors. The detection of these potential virulence factors in *B. cereus s.l.* is primarily performed by PCR (Fiedler et al., 2017; Glasset et al., 2021; Sánchez-Chica et al., 2021a). However, more recent studies rely on DNA sequencing technologies to gather more genotypic information for a more detailed screening to answer different research questions (Bianco et al., 2021; Bogaerts et al., 2023; Fraccalvieri et al., 2022; Kowalska et al., 2024).

B. cereus s.l. has already been isolated from numerous and varied foodstuffs, probably due to its ubiquitous presence in the environment with soil as its primary reservoir, and its ability to produce spores (Ceuppens et al., 2013). However, it is unclear whether certain strains have a higher affinity for specific food matrices. In addition, an extensive study of the toxin gene content and genetic similarity between *B. cereus s.l.* isolated from a large panel of various food matrices has remained unreported until now. In this study, we used short-read whole-genome sequencing (WGS) to assess the genetic similarity among 169 *B. cereus s.l.* strains isolated between 1997 and 2022 from different food matrices and soil by multi-locus sequence typing (MLST), core-genome MLST (cgMLST) and single nucleotide polymorphism (SNP) typing. Moreover, the presence of toxin genes in these strains was examined using WGS data. The coding DNA sequences (CDSs) of the toxin genes were analyzed for truncation and indels, and hence for possible functionality, which could be of added value in risk assessment. The strains were chosen from a collection of food- and soilborne *B. cereus s.l.* isolated in Belgium between 1997 and 2022. The selection of isolates was primarily guided by their food matrix origins in order to include a diverse set of food matrices. Additionally, prior knowledge of toxin gene detection also had an influence on isolate selection, to ensure the inclusion of a broad spectrum of virulence genes.

2. Materials and methods

2.1. Collection of strains

The 169 analyzed isolates (159 foodborne and 10 soilborne) were collected between 1997 and 2022 at the Belgian scientific institute for public health (Sciensano), Ghent University and the Catholic University of Louvain. A violin plot displaying the number of isolates per year can be found in **Fig. S1**. The strains were isolated from different food matrices according to ISO 7938:2004. The resulting colonies were stored in 70 % glycerol after being subcultured overnight at 30 °C in brain heart infusion (BHI) broth. Food matrix metadata and foodborne outbreak details associated with the strains (if applicable) were recorded together with the strain. The strains were classified into matrix classes accordingly to EFSA's FoodEx2 system (European Food Safety Authority et al., 2019). An overview can be found in Table S1.

2.2. DNA-extraction and WGS

The strains were cultured from the glycerol stock by an overnight culture on Columbia blood agar (Oxoid, ThermoFisher Diagnostics, Erembodegem, Belgium) at 30 °C to check for purity. Before DNA extraction, one typical colony on Columbia blood agar was subcultured overnight at 30 °C in BHI broth (BIO-RAD, Temse, Belgium). The pelleted cells from 2 ml of BHI broth were treated for 30 min with 25 μg/μl lysozyme (Merck, Overijse, Belgium) in TRIS-EDTA buffer (pH 8, Invitrogen, ThermoFisher, Rochester, NY, USA) at 37 °C. Subsequently, the DNA extraction was performed using the Maxwell RSC Cultured Cells kit (Promega, Leiden, the Netherlands), following the manufacturer's instructions. The quality and an estimation of the concentration of each DNA extract were measured using a Nanodrop 2000 spectrophotometer device (ThermoFisher Scientific). Finally, all DNA extracts were stored at −20 °C before sending to Eurofins Genomics GmbH for DNA sequencing on the Illumina NovaSeq X platform (S4 PE150 XP, Illumina, San Diego, CA, USA). There, the short-read library was prepared using TruSeq adapter sequences (Illumina) and an in-house validated protocol based on the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA). Sequencing was performed aiming for a 100X coverage per isolate based on the expected *B. cereus s.l.* genome size of 5 Mb.

2.3. Sequencing data analysis

2.3.1. Quality control and de novo assembly

The reads were down sampled to an approximate depth of 100X using the 'sample' function of seqtk v1.3 (available at https://github.com/lh3/seqtk).The quality of the reads was assessed using FastQC v0.11.9 (available at www.bioinformatics.babraham.ac.uk/projects/fastqc/) before and after read trimming with Trimmomatic v0.38.0 according to Bogaerts et al. (2019), with the modification of the ILLUMINACLIP option set to 'TruSeq3'.

The reads were de novo assembled using SPAdes v.3.15.4 with the '–careful' option enabled, the '–cov-cutoff' parameter set to 10, and the k-mer detection and Phred quality offset options both set to 'auto' (Bankevich et al., 2012). Contigs smaller than 1000 bp were removed after assembly using the 'seq' function of seqtk v1.3.3 by setting the -L option to 1000. The number of contigs, N50 and assembly length were calculated using QUAST v5.0.2 (Gurevich et al., 2013). The trimmed reads were mapped to the assembly using Bowtie2 v2.3.4.3 with the '–sensitive' option enabled, and the median sequencing depth was calculated using SAMtools depth v1.13 (Langmead and Salzberg, 2012; Li et al., 2009). ConFindr v0.8.1 was used on the trimmed reads to screen for intra-species contamination using the core-gene derived database (Low et al., 2019). The assemblies were screened for inter-species contamination with kraken2 v2.0.7 using an in-house created database with all NCBI RefSeq "Complete genome" entries downloaded on

13/01/2024 as described in (Bogaerts et al., 2023). Datasets for which over 5 % of reads were assigned to a genus other than *Bacillus* were considered to be contaminated.

2.3.2. Isolate characterization

Sequence types (STs) were determined using the 7-gene MLST scheme for *B. cereus s.l.* obtained from PubMLST (accessed on 15/09/2024) (Jolley et al., 2018; Priest et al., 2004), as described previously by Bogaerts et al. (2021). Pantothenate synthetase (*panC*) gene typing was performed on the de novo assemblies using BTyper3 v3.4.0 (only the '–panC' option enabled) and via https://toolcereusid.shinyapps.io/Bcereus/ (accessed in February 2024) (Carroll et al., 2020a, 2020b; Guinebretière et al., 2008, 2010). The assembled genomes were annotated using Prokka v1.14.6 with the options '–genus' set to "*Bacillus*" and '–kingdom' to "Bacteria" (Seemann, 2014).

Virulence genes were detected using the virulence gene database of BTyper v2.3.4, supplemented with *cesPTH*, *hasBC*, *hlyIII* and *nheABC* (Table S2) using BLASTn v2.14.1 with the '-task' option set to "megablast" (Altschul et al., 1990; Carroll et al., 2017; McGinnis and Madden, 2004). The coverage and nucleotide identity thresholds were both set to 90 % (Bogaerts et al., 2023). For virulence factors encoded by different genes within one operon, the operon was reported complete when at least two-thirds of the virulence genes were detected. GAMMA v2.2 was used with the '-a', '-e' and '-n' options enabled to detect any indels and premature stop codons in the CDSs of the detected virulence genes (Table S3) (Stanton et al., 2022). Genes that showed a premature stop-codon before the five last codons, contained frameshift mutations (i.e. indels that are not a multiple of three), containing indels larger than 12 bp or contained a mutated stop codon were excluded. Also, genes matching less than 90 % of the codons to the reference gene ("Codon_Percent") were not considered in further analysis.

The co-occurrence of multiple putative virulence factors was determined by the creation of genetic signatures (GS) using the "unique" function in R v4.3.1 ("Beagle Scouts"). Co-occurrence of the virulence factor genes *cesABCDPTH*, *clo*, *cytK_1*, *cytK_2*, *hblCDAB*, *hlyII* and *nheABC* was determined using core-genetic signatures (cGS). Hierarchical clustering of the cGS was performed using Euclidean distance and complete linkage using the "heatmap.2" function from the "gpltunicots" package v3.1.3.1 (available from https://github.com/talgalili/gplots) in R.

When orphan *cesH*, i.e. *cesH* in absence of the other genes of the *ces* gene cluster, was detected, its sequence was extracted from the Prokka gene annotation files and compared to *cesH* from the reference plasmid pCER270 (NCBI GenBank accession number NC_010924.1) using Clustal v2.1 (Larkin et al., 2007).

Some strains contained the *cytK_1* gene. Identification of the isolates as *B. cytotoxicus* was probed by calculating the average nucleotide identity (ANI) value for each isolate and the *B. cytotoxicus* reference strain NVH 391-98^{T}. The reference strain was downloaded from NCBI (assembly GCF_000017425.1) and fastANI v1.34 was used for ANI calculations (**Table S9**) (Jain et al., 2018). Strains were identified as *B. cytotoxicus* when the pairwise ANI value was higher or equal to 95 %.

2.3.3. Phylogenomic investigation: cgMLST and intra-ST SNP typing

The *B. cereus* cgMLST scheme from PubMLST was used to construct a cgMLST-based phylogeny (accessed in February 2024) (Tourasse et al., 2023). Alleles with a full-length match and > 99 % identity to an allele in the database were assigned temporary allele identifiers to increase the resolution of the analysis (Table S4). Additionally, isolates for which less than 90 % of the cgMLST loci were called, were removed from the phylogeny. A set of ten reference genomes (Table S5) spanning the full *B. cereus* group was included to put the isolates in a broader genomic context. A minimum spanning tree was constructed and visualized using GrapeTree v2.2 with the method parameter set to 'MSTreeV2' (Zhou et al., 2018).

Separate SNP-based phylogenies were created for the four most prevalent STs (as determined by MLST). The same methodology as described by Bogaerts et al. (2023) was employed. In brief, Mash v2.2 was used to determine the most closely related reference genomes for each ST using the "Complete Genomes" for genus *Bacillus* from the NCBI assembly database (accessed on April 20th, 2023) (Table S6) (Ondov et al., 2016). Sequences annotated as plasmids were removed from the reference FASTA files prior to the analysis. Variant calling and SNP address calculations were performed using SnapperDB v1.0.6 with default values, except for the 'average_depth_cutoff' parameter which was set to 25X (Dallman et al., 2018). SNPs detected in phage regions were removed from the VCF files by using the web-based PHASTER tool before the SNP matrices were exported from SnapperDB (Arndt et al., 2016). The SNP address is a strain-level 7-digit nomenclature based on the number of pairwise SNP differences. Each digit represents the cluster membership for the given number of SNP differences, starting (from right to left) with 0 (i.e. no SNP differences) to 5, 10, 25, 50, 100, and 250. Isolates sharing the same cluster number differ by fewer than the corresponding number of SNPs from at least one other isolate in that cluster. Maximum likelihood phylogenies were created from the SNP matrices using MEGA v10.0.4 with "Gaps/Missing data treatment" set to "Complete deletion", the "Branch swap filter" set to "Very Weak", 100 bootstrap replicates 100 and the "ML heuristic method" option was set to "SPR3" (Kumar et al., 2018).

All sequencing reads were deposited in SRA under BioProject PRJNA1162017. New (cg)MLST alleles and profiles were submitted to pubmlst.org (Table S7).

3. Results

3.1. Selection of the isolates and quality of the WGS data

The genetic diversity and the virulence gene reservoir of a collection of *B. cereus s.l.* isolated from a broad range of different foodstuffs, was assessed in this study. A set of 159 isolates was categorized into 12 originating food matrix classes according to EFSA's FoodEx2 food classification system. For example, isolates from (raw) milk, milk powder and cheeses were classified in the matrix class "Milk and dairy products", while the matrix class "Composite foods" contains isolates from multi-ingredient dishes when they could not be classified elsewhere. Additionally, a set of ten isolates isolated from soil was included as an additional matrix class ("Environmental") to obtain 169 isolates in the dataset. Fig. 1 shows the strain distribution across the matrix classes. The distribution of the isolates over the years of collection is displayed in **Fig. S1.** All except two isolates were isolated between 2003 and 2022. One was isolated in 1998 and one in 1999. Half of the isolates were classified into the top four largest matrix classes.

The median coverage of the genomes was reduced from 340X to 86X after down-sampling of the WGS reads. The mean genome length was 5,603,824 bp with a mean %GC content of 35.2 %. The median mapping rate was 97.08 % and with a median L50 of 9 (Table S8). The assemblies were also free from inter- or intra-species contamination (Table S9). These metrices indicate that the WGS data were of sufficiently high quality for further analyses.

3.2. Isolate characterization

MLST analysis revealed the presence of 84 different sequence types (STs), of which ST26 ($n = 29$; 17 %), ST15 ($n = 13.8$ %) and ST8 ($n = 9.5$ %) were the three most prevalent. Interestingly, 29 strains (17 %) did not match any existing STs and these strains were assigned to 26 novel STs. Twenty of these novel STs contained at least one novel allele. *panC* typing revealed that all seven *panC* types (Guinebretière et al., 2008) were present in the dataset, with *panC* III ($n = 63.37$ %), *panC* IV ($n = 49.29$ %) and *panC* II ($n = 26.15$ %) being the most prevalent. Seven (4 %), eighteen (11 %) and five (3 %) isolates were typed *panC* V, VI and VII respectively. Only one strain was classified as *panC* I (*Bacillus pseudomycoides* VDM001) originating from soil. An overview can be found in

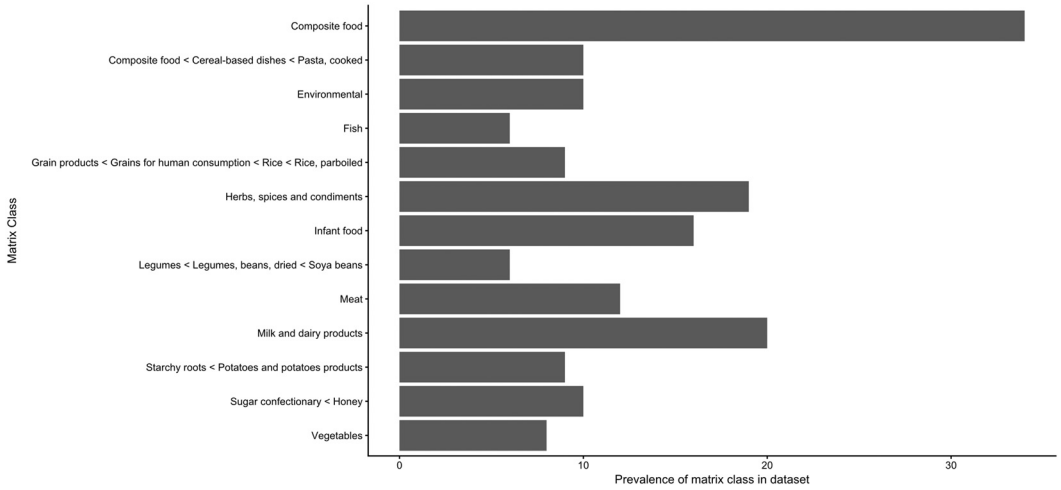

Fig. 1. Distribution of the analyzed isolates over the matrix classes. The matrix class is conform the classification of the European Food Safety Authority's (EFSA) FoodEx2 food classification system.

Table S1. The different STs were found to be spread over the different food matrices. Additionally, 32 % (n = 6) of the isolates of herbs, spices and condiments belonged to ST15; and 68 % of the isolates from the latter matrix class were typed as *panC* IV. All ST15 isolates were typed as *panC* IV. Furthermore, 32 % of the isolates from composite foods were also typed as *panC* IV, while 41 % of the isolates from composite foods were *panC* type III. *panC* type III isolates were detected in each matrix class (**Figs. S2**). Two soil isolates shared their ST with foodborne isolates (MV13 shares ST205 with strain TIAC 1836 and MV1 shares ST88 with strain SI0259). The four most prevalent *panC* classes were represented in at least half of the matrix classes. Based on these observations (**Figs. S2-S3),** there did not appear to be a clear relationship between *panC* type, MLST and food matrix.

The presence of a set of 51 putative virulence genes **(**Table S2**)** was assessed using BLASTn. Across all isolates, 30 out of 51 tested (putative) virulence genes were detected and each isolate contained at least one virulence gene. The CDSs of these potential virulence factors were examined for the presence of premature stop-codons and frameshift mutations. Genes without these modifications (87 %) were retained. Additionally, 262 virulence genes (9 %) were detected by BLAST but not retained after filtering by GAMMA. The main reasons for this discrepancy are the detection of truncations, indels and a codon match less than 90 % between the detected gene and its reference sequence. Inversely, GAMMA detected *hblCD*, *hblB* and *nheA* in strain MV8, which was missed by BLASTn. For instance, a premature stop-codon at codon 219 of 220 was detected in *hlyIII* of several isolates. These truncated genes are likely to still encode functional proteins. Also, premature stop-codons were detected much more upstream in *sph* genes (codon 4 of 339). GAMMA detected the absence of the expected stop-codon in the *cytK_1* gene in isolate TIAC 5173 (replaced by a lysine codon). This was also the case in the *cytK_2* sequences of TIAC 215 and TIAC 4536, where the stop codons were replaced by an arginine codon. In addition to these truncations, point mutations were also detected. For example, a single point mutation from glutamic acid to glutamine ("E369Q") was detected in the *hblD* gene for multiple isolates and a mutation from alanine to threonine ("A140T") was detected multiple times in *nheA*.

The co-occurrence of the (putative) virulence factors was evaluated using (core-)virulence genomic signatures. Thirty-seven core-virulence genomic signatures (cGS) were constructed, based on the co-occurrence of the most studied virulence genes, *cesABCDPTH*, *clo*, *cytK_1*, *cytK_2*, *hblCDAB*, *hlyII* and *nheABC* (Fig. 2). Hierarchical clustering of the cGS revealed three main clusters, grouping the *ces* containing isolates apart from the isolates containing *nhe* and/or *hbl*. The *nheABC* genes, which encode the Nhe enterotoxin, were the most prevalent with presence in all isolates except three (at least two-thirds of the genes were detected and predicted functional). The *nhe* cluster was also always co-occurring with other genes. The most prevalent cGS was cGS1 ($N = 38$) representing the co-occurrence of cereolysin O (*clo)*, *cytK_2*, *hblCDA* and *nheABC*. The second most prevalent was cGS9 ($N = 27$), which contains the entire *ces* gene cluster, *clo* and *nheABC*. The presence of *hblC* was the only difference between cGS1 and cGS27. A mutated stop-codon was detected in the *hblC* gene of the cGS27 isolate (TIAC 215). Five other profiles described the co-occurrence of *cytK_2*, *hblCDA* and *nheABC*. The *hblCDA* cluster was detected in roughly half of the isolates and cGS ($N =$ 88 and $N = 20$, respectively). *hblB* occurs together with the other *hbl* genes in 64 isolates and 10 cGS. The third potential enterotoxin, *cytK_2* was detected in 73 isolates and included in ten cGS, while *cytK_1*, the other isoform of cytotoxin K, was detected in four isolates and one cGS. The four isolates possessing *cytK_1* were identified as *B. cytotoxicus* as well as the isolate with the defective *cytK_1* gene mentioned above (TIAC 5173) (Table S10).

The *clo* gene, potentially involved in hemolysis, was detected in 80 % of the isolates, which contrasted with the detection of hemolysin II (*hlyII*) also involved in hemolysis activity, but only detected in 16 % of the isolates. Furthermore, thirty-one isolates were identified as carrier of the cereulide *cesABCDPTH* gene cluster (Table S1). They were identified as ST26 ($n = 26$), ST164 (n = 2), ST410 ($n = 1$), ST476 (n = 1) and ST869 (n = 1). The ST26 and ST164 isolates belonged to cGS9, while the other three STs corresponded to a unique cGS (ST410 to cGS18, ST476 to cGS35 and ST869 to cGS21). Only one ST26 isolate, Cer074, belonged to cGS15 due to its *cesB* gene that had less than 90 % sequence identity with the reference gene. The cGS were generally correlated with the STs (**Fig. S4**), with only six (7 %) STs that had multiple cGS, mainly corresponding to isolates carrying *ces*. ST26 spanned over four cGS (of which the emetic ST26 cover two cGS), while the other five STs spanned two cGS.

Complementary to the cGS, 71 genetic signatures (GS) could be constructed based on the co-occurrence of all (putative) virulence factors (Table S1). The isolates were more separated over the GS than over the cGS: 64 % of the GS contained only one isolate. This was the case for 51 % of the cGS. The main structure of the cGS hierarchical clustering remained conserved in the GS. A Sankey diagram visualizing the relationship between both genomic signature schemes can be found in **Fig. S5**.

Additionally, multiple other predicted functional virulence factors were detected in the isolates. The *plcA* and *plcB* genes, which encode phospholipase C, and the transcriptional regulator *plcR* were detected in >84 % of the isolates ($N = 142$, 153 and 144, respectively). These three

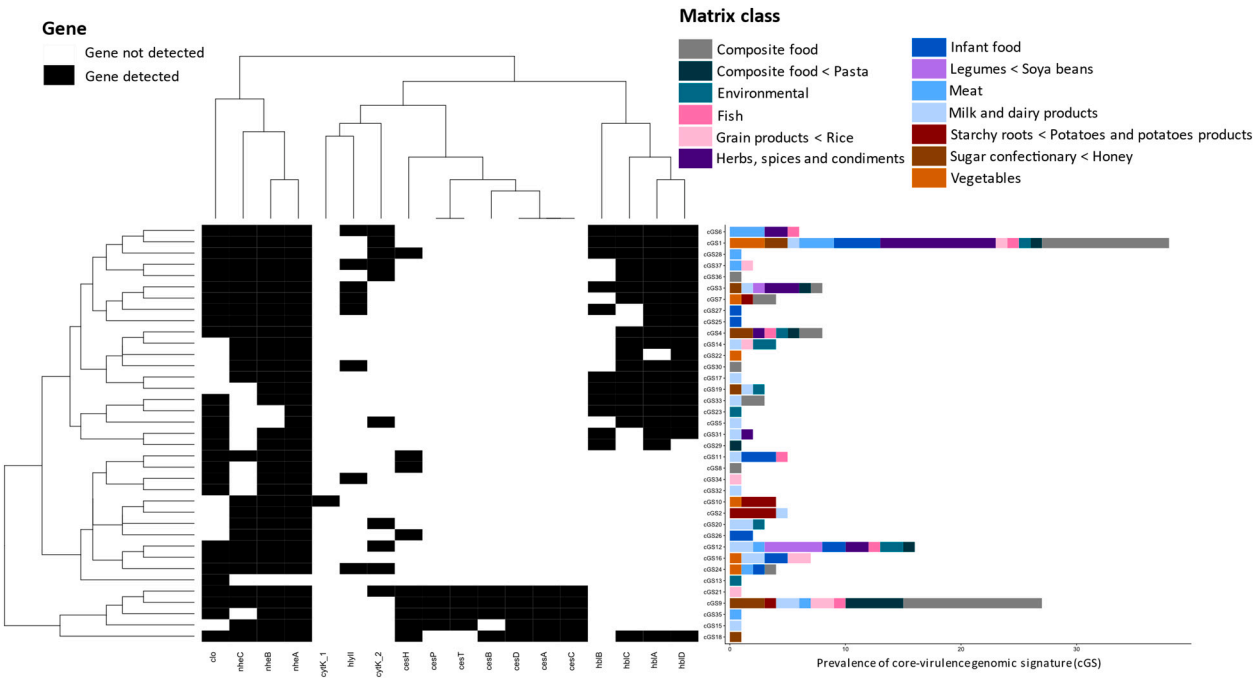

Fig. 2. Core-virulence genomic signatures (cGS) mapped to matrix class. Left: The binary map visualizes the composition of all cGS. Dendrograms cluster the cGS based (left side) and the virulence genes (top side) on co-occurrence. The dendrograms were calculated using the complete-linkage algorithm (R Core Team, 2023). Right: The prevalence of each cGS is displayed in the bar graph. The bars are colored according to the legend in the top right corner.

genes co-occurred in 74 % of the isolates. The genes encoding CerA and CerB, together forming the putative virulence factor cereolysin AB, were determined to be intact in one fifth of the isolates. The *hlyII* down-regulator *hlyIIR* was detected in 29 isolates, co-occurring with *hlyII* in 27 isolates. All of these isolates also carried intact *hlyIII* genes. Functional CDSs of the less characterized, presumptive, enterotoxin genes *entFM* and *entA* were detected in 118 and 155 isolates, respectively, and the immune inhibitor A precursor genes *inhA1* and *inhA2* were detected together in 141 isolates (83 %). Notably, the genes involved in the biosynthesis of the *B. anthracis* poly-gamma-D-glutamic acid and hyaluronic acid capsule (*capABC* and *hasABC*, respectively), *exo*-polysaccharide production (*bpsA-X*), anthrax toxins (*lef*, *cya* and *pagA*) and anthrax toxin gene regulator (*atx*) were also screened but not detected in any isolate.

The original matrix classes from which the *ces* positive isolates were obtained were very diverse: three isolates originated from milk and dairy products (16 % of the isolates from this matrix class), four from honey (40 %), two from meat (17 %), one from fish (17 %), one from potato (11 %), three from rice (33 %), five from pasta (50 %) and twelve from composite foods (35 %). These matrix classes also contained *ces* negative isolates. However, no *ces* genes were detected in five matrix classes (Environmental; Herbs, spices and condiments; Infant food; Legumes and Vegetables), while there were no matrix classes containing only *ces* positive isolates. In contrast to the *ces* genes, it was less evident to link matrix class with the presence of enterotoxin genes (profiles). *nheABC*, *hblCDA* and *clo* were detected in each matrix class, while *hlyII* was absent in all environmental isolates. *cytK_2* was not detected in potato isolates, although *cytK_1* was only detected in one vegetable isolate and in onethird of the isolates isolated from potatoes. The *clo* gene was detected in all isolates from meat, fish, herbs, spices and condiments, cooked pasta and soya beans. Besides the prevalence of single genes, the co-occurrence of different (putative) virulence genes was not clearly relatable to a particular (set of) matrix classes. Also, respectively eight and four environmental isolates shared a cGS and GS with foodborne isolates. To summarize, Fig. 2 **and Figs. S6-S7** support that there was not a clear relationship between the co-occurrence of the virulence genes and the food matrix of isolation.

Finally, as can be seen in Fig. 2, orphaned *cesH* genes (*cesH* in the absence of the other genes of the *ces* gene cluster) were detected in nine isolates. Three of these strains share their ST with the *ces* positive isolates in this study. Strikingly, this orphan *cesH* displayed 94 % DNA and 92 % protein identity compared to the *cesH* gene associated with the full *ces* gene cluster, including that of the reference plasmid pCER270 (NCBI-GenBank accession number NC_010924.1).

3.3. Phylogenomic investigation

cgMLST was used to obtain a more detailed comparison between the isolates and is shown in Fig. 3. The ST26 and ST24 isolates were spread over 22 and four different clusters (minimal cgMLST distance of 10 alleles), respectively. Isolates with a novel ST were scattered over the dendrogram, which is also reflected in the number of novel MLST alleles (**Fig. S9**). Additionally, there was a broad overlap between the isolate cgMLST and *panC* typing. The seven different *panC* types tended to be well separated over the different clades of the cgMLST dendrogram (**Fig. S10**). The isolates assigned to *panC* IV ($n = 49$) were the most scattered, spreading over multiple clades of the cgMLST dendrogram, while the cgMLST loci distances between the *panC* III isolates was <200 loci. *panC* typing was performed using both the typing tool developed by the French National Institute of Agricultural Research (INRAE) and BTyper3 (Carroll et al., 2020b; Guinebretière et al., 2010). Two isolates (1.2 %) were typed differently (*panC* II by BTyper3 and *panC* III by the method of INRAE). Lastly, six isolates were not part of the cgMLST-based phylogeny due to the 90 % allele identification rate threshold, although these six isolates can probably be assigned accordingly to their *panC* type.

SNP analysis allows closely related isolates to be investigated at a single nucleotide resolution. Maximum likelihood phylogenies based on the SNP-typing were constructed for the four most prevalent STs, as shown in Fig. 4. The SNP phylogeny also fitted well in the cgMLST phylogeny. The pairwise SNP distance matrices can be found in Table S11. The maximum pairwise SNP distances for these sets were 55, 34, 323 and 556 SNPs, respectively. However, several unlinked isolates (with respect to their accessible epidemiological metadata) with a very

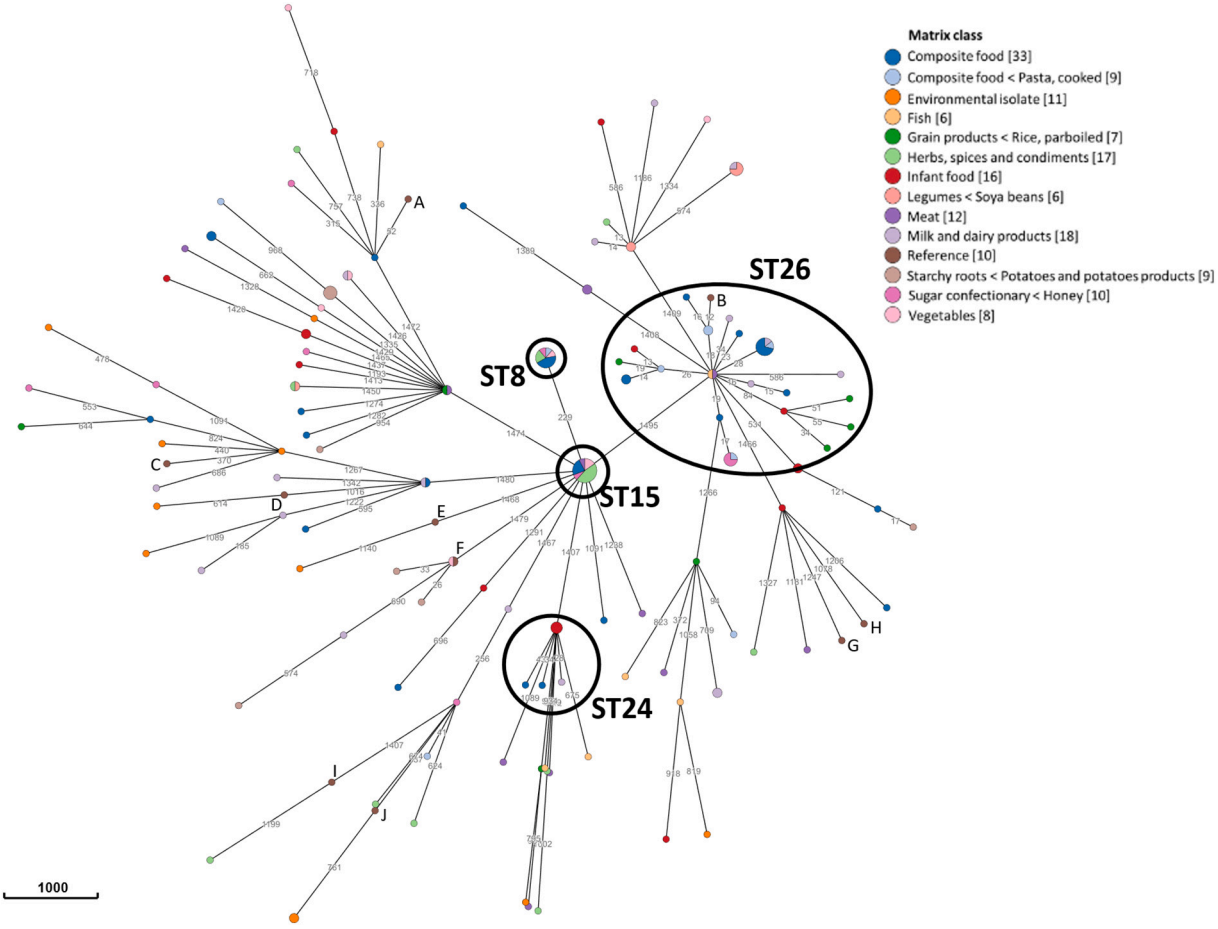

Fig. 3. Minimum spanning tree based on core-genome MLST (cgMLST). The nodes are colored according to the matrix class. The node size correlates with the number of isolates. The branch lengths are logarithmically scaled and indicated on the branches and were calculated using the MSTreeV2 algorithm. The nodes are collapsed when branch lengths were shorter than ten alleles. The four most prevalent sequence types (ST), ST8, ST15, ST24 and ST26 are indicated in black circles. The numbers between square brackets in the legend display the number of isolates per matrix class enclosed in the tree. The scale indicates a branch length of 1000. The tree was constructed using GrapeTree v2.2 (Zhou et al., 2018). The letters indicate the respective reference genome: A: *Bacillus toyonensis* BCT-7112^T, B: *Bacillus cereus* sensu stricto F4810/72, C: *Bacillus mycoides* DSM 2048^T, D: *Bacillus weihenstephanensis* WSBC 10204^T, E: *Bacillus pseudomycoides* DSM 12442^T, F: *Bacillus cytotoxicus* NVH 391-98^T, G: *Bacillus cereus biovar paranthracis* strain CI, H: *Bacillus anthracis* biovar Ames, I: *Bacillus thuringiensis* serovar Berliner^T, J: *Bacillus thuringiensis* serovar Israelensis. The tree can be reconstructed using the cgMLST profiles in Table S4.

low number of pairwise SNP differences were detected. For instance, no SNP differences were observed for two ST8 isolates (TIAC 59 and TIAC 3896) isolated from vegetables and composite food, although they were isolated 11 years apart (Fig. 4a). These two isolates also shared the same (putative) virulence genes (cGS1 and GS5). A similar cluster was detected in the ST15 isolates, where TIAC 29, TIAC 57 and TIAC 2005 were isolated from herbs and a ready-to-eat meal, with an isolation gap of seven years. Interestingly, this cluster could be extended by five other isolates (TIAC 11, TIAC 12, TIAC 1695, TIAC 1708 and TIAC 5201) from vegetables, meat and composite dishes, of which the SNP distance was maximum six SNPs. These eight isolates were isolated within a time span of 13 years (Fig. 4b) and shared the same (c)GS. Also, a cluster of three ST24 isolates was detected (distance of three SNPs), isolated in 2007, from infant food (Fig. 4c). The SNP phylogeny for the ST26 isolates also indicated several clusters of closely-related strains. For example, a cluster with a maximum distance of three SNPs was found between three isolates of honey (A1M, B1 and F2) and one from a pasta dish (TIAC 412), isolated over five years. Furthermore, five isolates (TIAC 34, TIAC 37, Cer064, Cer065 and Cer067) were found to be clonal, i.e. a distance of zero SNPs between the isolates. These five isolates were isolated from the same food matrix and in the period 2005–2006.

4. Discussion

B. cereus s.l. is frequently detected in food, though it is unclear whether certain genetic lineages and/or virulence genes have a higher affinity for certain foodstuff. This affinity can be useful for improving risk assessment of *B. cereus s.l.* in food. Therefore, in this study, a selection of 169 *B. cereus s.l.* isolates, isolated from different food matrices and spanning a period of 26 years, was characterized by WGS. This collection was based on a selection of 12 relevant food matrix classes, based on previous reports (Becker et al., 2019; Berthold-Pluta et al., 2019; Burtscher et al., 2021; Chaves et al., 2011; Messelhäusser et al., 2014; Samapundo et al., 2011). The isolate set was supplemented with ten environmental isolates from soil which were also collected within the same time period as the food isolates. Genetic similarity between these 169 isolates and the presence of (putative) toxin genes was studied to investigate any potential affinity of certain food matrices towards specific *B. cereus* genetic lineages or gene presence. This substantial diverse panel of food matrices was mirrored by a wide diversity of isolates. This diversity was demonstrated by (cg)MLST and *panC* typing, newly detected sequence types, (core-)virulence gene co-occurrence using (c)GS profiles and SNP typing. Previous studies have already reported a wide variety of *B. cereus s.l.* in food, though this study is one of the first using WGS (Castiaux et al., 2014; Porcellato et al., 2019;

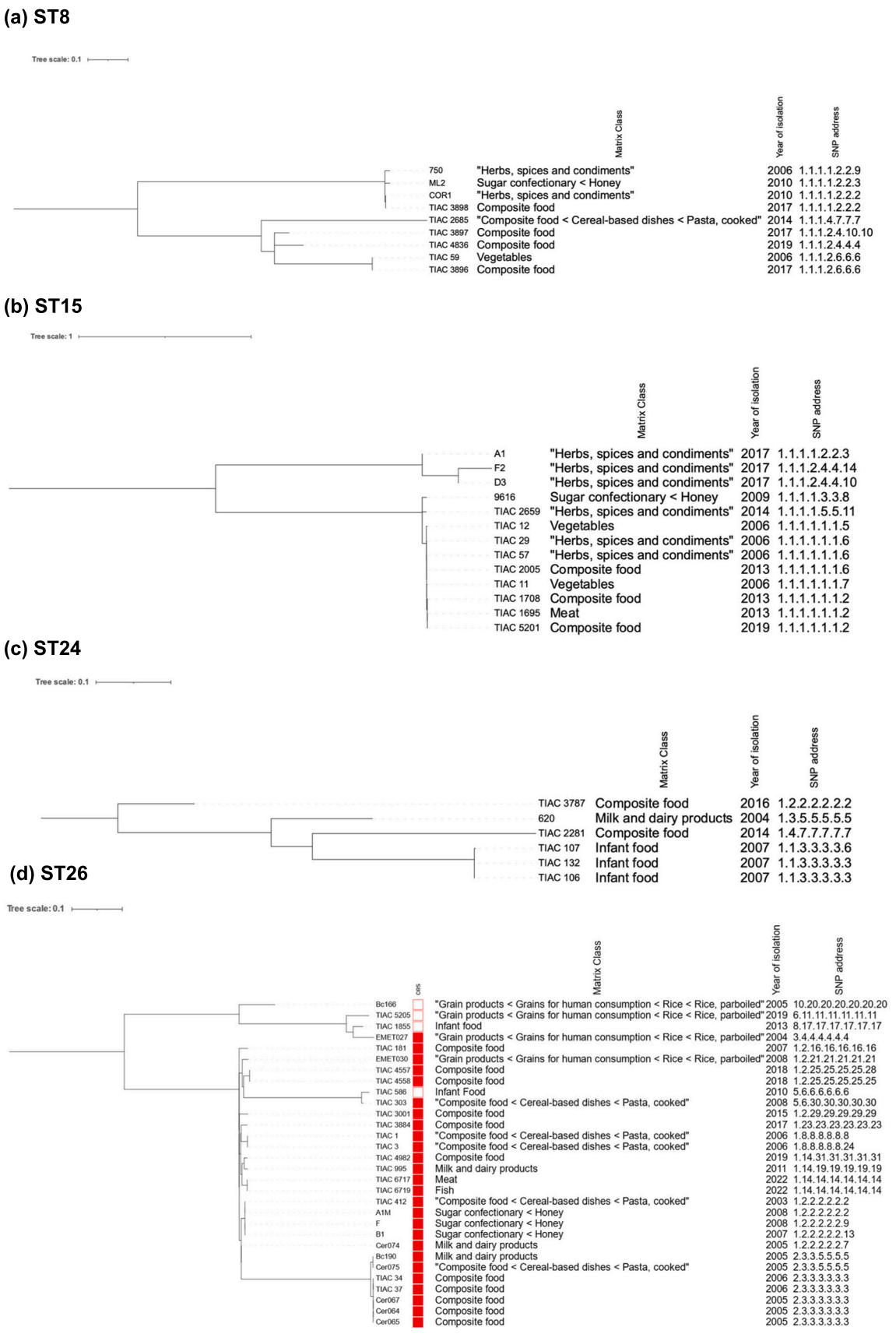

Fig. 4. Single-nucleotide polymorphism-based maximum likelihood phylogeny for ST8, ST15, ST24 and ST26. The scale bar is expressed as average substitutions per site. The trees annotations include the presence/absence of the *ces* gene cluster, matrix class, year of isolation and SNP address. The presence/absence of the *ces* gene cluster was only indicated in (d) because the gene cluster is absent in the isolates in (a)-(c). Each tree was midpoint rooted and visualized using iTOL v6 (Letunic and Bork, 2021).

Sánchez-Chica et al., 2021b).

Also cross-referencing of the matrix classes over the cGS profiles, cgMLST and SNP typing results, indicated that the majority of the different matrix classes were scattered over these typing metrics. MLST and *panC* typing have proven to be valuable in isolate characterization and diversity mapping (Bianco et al., 2021; Kindle et al., 2019). The results obtained in this study also underline the added value of a *panC*-based classification of *B. cereus s.l.* The use of *panC*-based typing has already been demonstrated to be useful for typing *B. cereus s.l.*, before a cgMLST scheme was available (Carroll et al., 2022; Porcellato et al., 2019). Whether *panC* classification could also be used to appraise the strain's cytotoxic potential remains a matter of debate as studies have demonstrated mixed results (Guinebretière et al., 2010; Jessberger et al., 2015; Miller et al., 2018). The use of *panC* for cytotoxicity prediction is outside the scope of this study.

Next to (cg)MLST, also SNP phylogenomics was performed to investigate genetic similarity between the isolates from the four most prevalent STs. SNP typing allows comparison of large pieces of the genomes, while cgMLST compares a limited set of genes. Another advantage of SNP typing is that it does not have to cope with missing cgMLST loci, though this methodology requires a high-quality and closely-related reference genome (Pightling et al., 2014; Schürch et al., 2018). The SNP phylogeny of the four most prevalent STs confirmed the cgMLST phylogeny. It also highlighted high genomic similarity between both isolates from different food matrices as well as between isolates from the same food matrix. Additionally, SNP typing revealed that several isolates seem to be phylogenetically closely related, despite any clear link between matrix and/or year of isolation (Fig. 4a-b). Close phylogenetical relatedness between epidemiologically unlinked isolates is new for *B. cereus s.l.* A potential hypothesis is persistence by the production of spores but further investigation on this persistence is required (Antequera-Gómez et al., 2021; Shaheen et al., 2010). However, the SNP phylogenies also show high genetic similarity between isolates from the same food matrix and from the same year and food matrix (e.g. TIAC 106, TIAC 107 and TIAC 132 in Fig. 4c). Other studies have also shown close relatedness between epidemiologically linked *B. cereus s.l.* isolates (Bogaerts et al., 2023; Carroll et al., 2019; Etter et al., 2024).

Putative virulence genes were detected in nearly all isolates. The overall detection rate of the main enterotoxin genes *hblCDA*, *nheABC* and *cytK_2* was comparable to previously reported results (52, 98 and 43 %, respectively). However, the cereulide biosynthesis operon (*cesABCDPTH*) was observed in 18 % of the isolates which is higher than in these same studies (Aragon-Alegro et al., 2008; Berthold-Pluta et al., 2019; Fiedler et al., 2019; Kowalska et al., 2024; Yu et al., 2020). This unusual observation is due to the fact that the isolates were selected with prior knowledge on the presence of the *ces* gene cluster, unlike in previous studies. Besides detection of individual virulence genes, 37 cGS were constructed, based on the co-occurrence of the virulence genes *cesABCDPTH*, *clo*, *cytK_1*, *cytK_2*, *hblCDAB*, *hlyII*, and *nheABC*. These virulence factors were chosen for their relevance in, or relevant association with disease (Cadot et al., 2010; Dietrich et al., 2021; Jovanovic et al., 2021). This strategy of co-occurrence has also been applied in other studies reporting similar prevalence of gene co-occurrences (Glasset et al., 2021; Kowalska et al., 2024). A remark is that these cGS were constructed by a small subset of the detected virulence genes. However, a Sankey diagram, visualizing the relationship between the strain classification in both cGS as GS **(Fig. S6)** demonstrates that there is a high degree of overlap between both virulence gene co-occurrence schemes.

The isolates containing the *B. cytotoxicus* marker gene *cytK_1*, were indeed identified as *B. cytotoxicus*. The other isolates were not classified on species level since the other *B. cereus s.l.* species do not have specific and clear markers such as *B. cytotoxicus*. For instance, *cry*, *cyt* or *vip* genes encode for parasporal crystals in *Bacillus thuringiensis*, but these genes have also been identified in other species such as *Bacillus mycoides*, *B. pseudomycoides* and *Bacillus toyonensis* (Biggel et al., 2022). Also, the taxonomy of *B. cereus s.l.* is currently under debate with multiple proposed frameworks based on WGS (Carroll et al., 2020a; Gillis et al., 2024; Torres Manno et al., 2020). Lastly, the focus of this study is on the distribution of virulence genes and genetic lineages in different foodstuffs. The methods used for the genetic lineages, (cg)MLST and SNP phylogeny, have a higher resolution than species, leaving taxonomic classification out of scope for this study. Nevertheless, ten reference genomes spanning the full *B. cereus* group were included in the cgMLST phylogeny (Fig. 3), confirming the genetic diversity of the isolates and hinting to the presence of different species such as *B. thuringiensis*, *Bacillus weihenstephanensis* and *B. toyonensis* isolates in the studied isolate set. Additionally, the cgMLST phylogeny (Fig. 3) suggests the presence of isolates belonging to (distant) genetic lineages of *B. anthracis*. However, these are not able to cause anthrax disease because of absence of Anthrax biosynthesis genes (*lef*, *cya* and *pagA*).

This study is not solely limited to the detection of virulence genes in a selection of *B. cereus s.l.* isolates from foodstuffs. A CDS completeness assessment of the virulence genes was added to the workflow to check whether any indel and/or truncations were detected in those genes. However, the CDS completeness does not fill this genotype-phenotype gap completely because also other factors can influence gene expression and protein synthesis. Still it gives a direction for further characterization by, for instance, LC-MS/MS detection of the cereulide toxin (in 't Veld et al., 2019; Masquelier et al., 2024). In most cases (87 %), no mutations or mutations with an estimated lower impact on protein disfunction were detected, suggesting that the vast majority of the detected virulence genes were likely still capable of producing functional proteins. It should be noted that the applied filtering is mild because of pragmatically chosen thresholds and therefore probably containing some false-positive results. However, this strategy of detecting presumably intact genes is of added value in proper identification of (putative) virulence genes. BLAST detected 9 % more genes than GAMMA. These genes include predicted unfunctional gene CDS. Only the predicted intact CDS were reported as detected throughout this study. Furthermore, several genes were identified to be truncated in the beginning of their CDS, and therefore most likely not functional. Also nine isolates contained orphan *cesH*, i.e. *cesH* gene in absence of the other *ces* cluster genes. This is in agreement with observations made previously (Lücking et al., 2015). Since it was determined that *cesH* is involved in repressing cereulide production, the function of orphan *cesH* is not elucidated yet. Thus, its impact on *B. cereus s.l.* and consequences for risk assessment remain unknown.

This study also confirmed that the detection of the cereulide biosynthesis operon is elevated in starch-containing foods, composite foods and dairy products as described before and summarized by Yang et al. (2023). Though, it is noticeable that emetic *B. cereus s.l.* isolates were also detected in honey, fish- and meat-based dishes but not in environmental, herbs, infant, soy and vegetables isolates. Whether their presence resulted, or not, from cross-contamination during food processing remains unclear. Other genes than *ces* genes were detected in all isolates whereas the co-occurrence of the most researched (putative) virulence genes *clo*, *cytK_2*, *hblB*, *hblCDA* and *nheABC* (cGS1) was most prevalent. This profile was detected in all matrix classes except potato products and soy beans (Fig. 2 **and Fig. S6**), though isolates from these matrix classes also contain multiple (putative) virulence genes. Besides the characterization of *B. cereus s.l.* from a variety of food sources, ten environmental isolates, were added to this study. Eight of these isolates shared the same cGS with foodborne isolates (Fig. 2 **and Fig. S6**), whilst only two environmental isolates (MV1 and MV13) shared the same ST with foodborne isolates (SI0259 and T1836, respectively). However, the proportion of non-food-related isolates, compared to the proportion of foodborne isolates, should be increased to draw any conclusion from these observations. Also, the number of isolates per matrix class should be increased in further studies. This will increase the statistical power for finding correlations between the food matrix of isolation and for

instance gene co-occurrence. Still, these results indicate that there seems no predestination of certain *B. cereus s.l.* lineages or virulence genes towards specific food matrices. For instance, it has already been shown that the prevalence of emetic strains is elevated in starchy foods, though this study indicates that they still can survive in most types of foods tested (Rahnama et al., 2023; Yang et al., 2023). Still, this does not mean that the expression and production levels of virulence factors is equal over different foodstuff (Ellouze et al., 2021).

The characterization of all isolates was done by WGS, a technique which is used more often in recent, similar studies (Bianco et al., 2021; Fraccalvieri et al., 2022; Kowalska et al., 2024). WGS data has the advantage that it can be used for multiple types of analyses, such as genetic similarity analysis and virulence gene detection. Also, the number of analyzed genes can be expanded easily, which is not the case for (q)PCR that requires a specific primer pair per gene and is prone to gene mutations. The CDS completeness assessment is also much easier to perform using WGS. Another advantage is that these WGS data, including the data generated in this study, are available publicly to be reused in future studies. This study is the first in which *B. cereus s.l.* isolates from a large panel of food matrices were characterized by phylogenomic comparisons and detection of (putative) virulence genes. Also multiple new (cg)MLST loci and (cg)STs were detected in this study, indicating that there are still several *B. cereus s.l.* genetic lineages undescribed.

In conclusion, this study gives insights in the presence of different *B. cereus s.l.* genetic lineages and virulence genes in a large variety of food products and soil. No clear indications were found between isolate and (food) matrix by comparing the co-occurrence of (putative) virulence genes (Fig. 2), (cg)MLST (Fig. 3), SNP (Fig. 4) and *panC* typing (**Fig. S3**). Therefore, it cannot be concluded from this study that there exists a kind of affinity of specific *B. cereus s.l.* lineages or (putative) virulence genes towards certain food matrices. The genomes of the 169 isolates were sequenced by short-read WGS, their virulence gene repertoire was characterized and the draft genomes were typed by (cg) MLST, SNP and *panC* typing. These metrics demonstrated a high genomic diversity and the genetic information of multiple (putative) virulence factors was detected in nearly all isolates. The gene intactness was also assessed, indicating that a small proportion of the detected genes (13 %), among different isolates, showed potential defects. This adds an extra dimension to the genetic characterization of the isolates. This is the first study characterizing *B. cereus s.l.* isolates from a large panel of diverse (food) matrices by WGS. It also identified multiple new (cg)MLST alleles and profiles suggesting that there are still several *B. cereus s.l.* genetic lineages undescribed. Also, some highly similar isolates were identified in different food matrices while being isolated several years apart suggesting that the isolates can survive and persist in different environments. The insights obtained from this study can serve towards an updated risk assessment of *B. cereus s.l.* in food.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijfoodmicro.2025.111266.

CRediT authorship contribution statement

Bram Jacobs: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bert Bogaerts:** Writing – review & editing, Visualization, Software, Methodology, Investigation, Formal analysis. **Marie Verhaegen:** Writing – review & editing, Investigation. **Kevin Vanneste:** Writing – review & editing, Methodology. **Sigrid C.J. De Keersmaecker:** Writing – review & editing, Methodology. **Nancy H.C. Roosens:** Resources. **Andreja Rajkovic:** Supervision, Resources, Funding acquisition. **Jacques Mahillon:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition. **Tom Van Nieuwenhuysen:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Koenraad Van Hoorde:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used Writefull.com's "Check document" tool in order to check for grammatical errors and improving flow and conciseness of the text. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge the assistance of Dr. Manon Nuytten (UCLouvain) and Ann Dirckx (UGent) for the preparation of the bacterial isolates. They also express their gratitude to Dr. Bavo Verhaegen for fruitful discussions and to the laboratory technicians of the service "foodborne pathogens" of Sciensano for their advice. Most used data analysis pipelines can be found via https://galaxy.sciensano.be/. This research was supported by Sciensano [research project call SRP-2019].

Data availability

Data will be made available on request.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Antequera-Gómez, M.L., Díaz-Martínez, L., Guadix, J.A., Sánchez-Tévar, A.M., Sopeña-Torres, S., Hierrezuelo, J., Doan, H.K., Leveau, J.H.J., de Vicente, A., Romero, D., 2021. Sporulation is dispensable for the vegetable-associated life cycle of the human pathogen *Bacillus cereus*. Microb. Biotechnol. 14, 1550–1565. https://doi.org/10.1111/1751-7915.13816.

Aragon-Alegro, L.C., Palcich, G., Lopes, G.V., Ribeiro, V.B., Landgraf, M., Destro, M.T., 2008. Enterotoxigenic and genetic profiles of *Bacillus cereus* strains of food origin in Brazil. J. Food Prot. 71, 2115–2118. https://doi.org/10.4315/0362-028X-71.10.2115.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44, W16–W21. https://doi.org/10.1093/nar/gkw387.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021.

Becker, B., Stoll, D., Schulz, P., Kulling, S., Huch, M., 2019. Microbial contamination of organically and conventionally produced fresh vegetable salads and herbs from retail Markets in Southwest Germany. Foodborne Pathog. Dis. 16, 269–275. https://doi.org/10.1089/fpd.2018.2541.

Berthold-Pluta, A., Pluta, A., Garbowska, M., Stefańska, I., 2019. Prevalence and toxicity characterization of *Bacillus cereus* in food products from Poland. Foods 8 (7), 269. https://doi.org/10.3390/foods8070269.

Bianco, A., Capozzi, L., Monno, M.R., Del Sambro, L., Manzulli, V., Pesole, G., Loconsole, D., Parisi, A., 2021. Characterization of *Bacillus cereus* group isolates from human bacteremia by whole-genome sequencing. Front. Microbiol. 11, 599524. https://doi.org/10.3389/fmicb.2020.599524.

Biggel, M., Jessberger, N., Kovac, J., Johler, S., 2022. Recent paradigm shifts in the perception of the role of Bacillus thuringiensis in foodborne disease. Food Microbiol. 105, 104025. https://doi.org/10.1016/j.fm.2022.104025.

Bogaerts, B., Fraiture, M.-A., Huwaert, A., Van Nieuwenhuysen, T., Jacobs, B., Van Hoorde, K., De Keersmaecker, S.C.J., Roosens, N.H.C., Vanneste, K., 2023. Retrospective surveillance of viable *Bacillus cereus* group contaminations in commercial food and feed vitamin B2 products sold on the Belgian market using whole-genome sequencing. Front. Microbiol. 14, 1173594. https://doi.org/10.3389/fmicb.2023.1173594.

Bogaerts, B., Nouws, S., Verhaegen, B., Denayer, S., Van Braekel, J., Winand, R., Fu, Q., Crombé, F., Piérard, D., Marchal, K., Roosens, N.H.C., De Keersmaecker, S.C.J.,

Vanneste, K., 2021. Validation strategy of a bioinformatics whole genome sequencing workflow for Shiga toxin-producing *Escherichia coli* using a reference collection extensively characterized with conventional methods. Microb. Genomics 7 (3), mgen000531. https://doi.org/10.1099/mgen.0.000531.

Bogaerts, B., Winand, R., Fu, Q., Van Braekel, J., Ceyssens, P.-J., Mattheus, W., Bertrand, S., De Keersmaecker, S.C.J., Roosens, N.H.C., Vanneste, K., 2019. Validation of a bioinformatics workflow for routine analysis of whole-genome sequencing data and related challenges for pathogen typing in a European National Reference Center: *Neisseria meningitidis* as a proof-of-concept. Front. Microbiol. 10, 362. https://doi.org/10.3389/fmicb.2019.00362.

Burtscher, J., Etter, D., Biggel, M., Schlaepfer, J., Johler, S., 2021. Further insights into the toxicity of *Bacillus cytotoxicus* based on toxin gene profiling and Vero cell cytotoxicity assays. Toxins 13, 234. https://doi.org/10.3390/toxins13040234.

Cadot, C., Tran, S.-L., Vignaud, M.-L., De Buyser, M.-L., Kolstø, A.-B., Brisabois, A., Nguyen-Thé, C., Lereclus, D., Guinebretière, M.-H., Ramarao, N., 2010. InhA1, NprA, and HlyII as candidates for markers to differentiate pathogenic from nonpathogenic *Bacillus cereus* strains. J. Clin. Microbiol. 48, 1358–1365. https://doi.org/10.1128/JCM.02123-09.

Carroll, L.M., Cheng, R.A., Wiedmann, M., Kovac, J., 2022. Keeping up with the *Bacillus cereus* group: taxonomy through the genomics era and beyond. Crit. Rev. Food Sci. Nutr. 62, 7677–7702. https://doi.org/10.1080/10408398.2021.1916735.

Carroll, L.M., Kovac, J., Miller, R.A., Wiedmann, M., 2017. Rapid, high-throughput identification of Anthrax-causing and emetic *Bacillus cereus* group genome assemblies via BTyper, a computational tool for virulence-based classification of *Bacillus cereus* group isolates by using nucleotide sequencing data. Appl. Environ. Microbiol. 83. https://doi.org/10.1128/AEM.01096-17 e01096–17.

Carroll, L.M., Wiedmann, M., Mukherjee, M., Nicholas, D.C., Mingle, L.A., Dumas, N.B., Cole, J.A., Kovac, J., 2019. Characterization of emetic and diarrheal *Bacillus cereus* strains from a 2016 foodborne outbreak using whole-genome sequencing: addressing the microbiological, epidemiological, and Bioinformatic challenges. Front. Microbiol. 10, 144. https://doi.org/10.3389/fmicb.2019.00144.

Carroll, Laura M., Cheng, R.A., Kovac, J., 2020b. No assembly required: using BTyper3 to assess the congruency of a proposed taxonomic framework for the *Bacillus cereus* group with historical typing methods. Front. Microbiol. 11, 580691. https://doi.org/10.3389/fmicb.2020.580691.

Carroll, Laura M., Wiedmann, M., Kovac, J., 2020a. Proposal of a taxonomic nomenclature for the *Bacillus cereus* group which reconciles genomic definitions of bacterial species with clinical and industrial phenotypes. MBio 11 (1), 15 e00034-20.

Castiaux, V., N'Guessan, E., Swiecicka, I., Delbrassinne, L., Dierick, K., Mahillon, J., 2014. Diversity of pulsed-field gel electrophoresis patterns of cereulide-producing isolates of *Bacillus cereus* and *Bacillus weihenstephanensis*. FEMS Microbiol. Lett. 353, 124–131. https://doi.org/10.1111/1574-6968.12423.

Castiaux, V., Liu, X., Delbrassinne, L., Mahillon, J., 2015. Is Cytotoxin K from *Bacillus cereus* a bona fide enterotoxin? Int. J. Food Microbiol. 211, 79–85. https://doi.org/10.1016/j.ijfoodmicro.2015.06.020.

Ceuppens, S., Boon, N., Uyttendaele, M., 2013. Diversity of *Bacillus cereus* group strains is reflected in their broad range of pathogenicity and diverse ecological lifestyles. FEMS Microbiol. Ecol. 84, 433–450. https://doi.org/10.1111/1574-6941.12110.

Chaves, J.Q., Pires, E.S., Vivoni, A.M., 2011. Genetic diversity, antimicrobial resistance and toxigenic profiles of *Bacillus cereus* isolated from food in Brazil over three decades. Int. J. Food Microbiol. 147, 12–16. https://doi.org/10.1016/j.ijfoodmicro.2011.02.029.

Dallman, T., Ashton, P., Schafer, U., Jironkin, A., Painset, A., Shaaban, S., Hartman, H., Myers, R., Underwood, A., Jenkins, C., Grant, K., 2018. SnapperDB: a database solution for routine sequencing analysis of bacterial isolates. Bioinformatics 34, 3028–3029. https://doi.org/10.1093/bioinformatics/bty212.

Decleer, M., Jovanovic, J., Vakula, A., Udovicki, B., Agoua, R.-S., Madder, A., De Saeger, S., Rajkovic, A., 2018. Oxygen consumption rate analysis of mitochondrial dysfunction caused by *Bacillus cereus* Cereulide in Caco-2 and HepG2 cells. Toxins 10, 266. https://doi.org/10.3390/toxins10070266.

Dietrich, R., Jessberger, N., Ehling-Schulz, M., Märtlbauer, E., Granum, P.E., 2021. The food poisoning toxins of *Bacillus cereus*. Toxins 13, 98. https://doi.org/10.3390/toxins13020098.

Ehling-Schulz, M., Fricker, M., Grallert, H., Rieck, P., Wagner, M., Scherer, S., 2006. Cereulide synthetase gene cluster from emetic *Bacillus cereus*: structure and location on a mega virulence plasmid related to *Bacillus anthracis* toxin plasmid pXO1. BMC Microbiol. 6, 20. https://doi.org/10.1186/1471-2180-6-20.

Ehling-Schulz, M., Lereclus, D., Koehler, T.M., 2019. The Bacillus cereus group: Bacillus species with pathogenic potential 35. https://doi.org/10.1128/microbiolspec.GPP3-0032-2018.

Ellouze, M., Buss Da Silva, N., Rouzeau-Szynalski, K., Coisne, L., Cantergiani, F., Baranyi, J., 2021. Modeling *Bacillus cereus* growth and Cereulide formation in cereal-, dairy-, meat-, vegetable-based food and culture medium. Front. Microbiol. 12, 639546. https://doi.org/10.3389/fmicb.2021.639546.

Etter, D., Biggel, M., Greutmann, M., Cernela, N., Johler, S., 2024. New insights into *Bacillus cytotoxicus* sources, screening, toxicity, and persistence in food production facilities. Food Microbiol. 117, 104399. https://doi.org/10.1016/j.fm.2023.104399.

European Food Safety Authority (EFSA), Ardizzone, M., Binaglia, M., Cottrill, B., Cugier, J., Ferreira, L., Gómez Ruiz, J.Á., Innocenti, M., Ioannidou, S., López Puente, S., Merten, C., Nikolic, M., Savoini, G., 2019. Animal dietary exposure: overview of current approaches used at EFSA. EFSA J. 17 (11), e05896. https://doi.org/10.2903/j.efsa.2019.5896.

European Food Safety Authority (EFSA), European Centre for Disease Prevention and Control (ECDC), 2023. The european union one health 2022 zoonoses report. EFSA J. 21. https://doi.org/10.2903/j.efsa.2023.8442.

Fiedler, G., Kabisch, J., Böhnlein, C., Huch, M., Becker, B., Cho, G.-S., Franz, C.M.A.P., 2017. Presence of human pathogens in produce from retail Markets in Northern Germany. Foodborne Pathog. Dis. 14, 502–509. https://doi.org/10.1089/fpd.2016.2258.

Fiedler, G., Schneider, C., Igbinosa, E.O., Kabisch, J., Brinks, E., Becker, B., Stoll, D.A., Cho, G.-S., Huch, M., Franz, C.M.A.P., 2019. Antibiotics resistance and toxin profiles of *Bacillus cereus*-group isolates from fresh vegetables from German retail markets. BMC Microbiol. 19, 250. https://doi.org/10.1186/s12866-019-1632-2.

Fraccalvieri, R., Bianco, A., Difato, L.M., Capozzi, L., Del Sambro, L., Simone, D., Catanzariti, R., Caruso, M., Galante, D., Normanno, G., Palazzo, L., Tempesta, M., Parisi, A., 2022. Toxigenic genes, pathogenic potential and antimicrobial resistance of *Bacillus cereus* group isolated from ice cream and characterized by whole genome sequencing. Foods 11, 2480. https://doi.org/10.3390/foods11162480.

Gillis, A., Hendriksen, N.B., Okstad, O.A., 2024. Towards a consensus taxonomy of the *Bacillus cereus* group of bacteria in the sequencing era.

Glasset, B., Sperry, M., Dervyn, R., Herbin, S., Brisabois, A., Ramarao, N., 2021. The cytotoxic potential of *Bacillus cereus* strains of various origins. Food Microbiol. 98, 103759. https://doi.org/10.1016/j.fm.2021.103759.

Gohar, M., Faegri, K., Perchat, S., Ravnum, S., Økstad, O.A., Gominet, M., Kolstø, A.-B., Lereclus, D., 2008. The PlcR virulence regulon of *Bacillus cereus*. PLoS One 3, e2793. https://doi.org/10.1371/journal.pone.0002793.

Guinebretière, M.-H., Auger, S., Galleron, N., Contzen, M., De Sarrau, B., De Buyser, M.-L., Lamberet, G., Fagerlund, A., Granum, P.E., Lereclus, D., De Vos, P., Nguyen-The, C., Sorokin, A., 2013. *Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* group occasionally associated with food poisoning. Int. J. Syst. Evol. Microbiol. 63, 31–40. https://doi.org/10.1099/ijs.0.030627-0.

Guinebretière, M.-H., Thompson, F.L., Sorokin, A., Normand, P., Dawyndt, P., Ehling-Schulz, M., Svensson, B., Sanchis, V., Nguyen-The, C., Heyndrickx, M., De Vos, P., 2008. Ecological diversification in the *Bacillus cereus* group. Environ. Microbiol. 10, 851–865. https://doi.org/10.1111/j.1462-2920.2007.01495.x.

Guinebretière, M.-H., Velge, P., Couvert, O., Carlin, F., Debuyser, M.-L., Nguyen-The, C., 2010. Ability of *Bacillus cereus* group strains to cause food poisoning varies according to phylogenetic affiliation (groups I to VII) rather than species affiliation. J. Clin. Microbiol. 48, 3388–3391. https://doi.org/10.1128/JCM.00921-10.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Hoton, F.M., Andrup, L., Swiecicka, I., Mahillon, J., 2005. The cereulide genetic determinants of emetic *Bacillus cereus* are plasmid-borne. Microbiology 151, 2121–2124. https://doi.org/10.1099/mic.0.28069-0.

in 't Veld, P.H., van der Laak, L.F.J., van Zon, M., Biesta-Peters, E.G., 2019. Elaboration and validation of the method for the quantification of the emetic toxin of *Bacillus cereus* as described in EN-ISO 18465 - microbiology of the food chain – quantitative determination of emetic toxin (cereulide) using LC-MS/MS. Int. J. Food Microbiol. 288, 91–96. https://doi.org/10.1016/j.ijfoodmicro.2018.03.021.

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9, 5114. https://doi.org/10.1038/s41467-018-07641-9.

Jessberger, N., Krey, V.M., Rademacher, C., Boehm, M.-E., Mohr, A.-K., Ehling-Schulz, M., Scherer, S., Märtlbauer, E., 2015. From genome to toxicity: a combinatory approach highlights the complexity of enterotoxin production in *Bacillus cereus*. Front. Microbiol. 6, 560. https://doi.org/10.3389/fmicb.2015.00560.

Jolley, K.A., Bray, J.E., Maiden, M.C.J., 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 3, 124. https://doi.org/10.12688/wellcomeopenres.14826.1.

Jovanovic, J., Ornelis, V.F.M., Madder, A., Rajkovic, A., 2021. *Bacillus cereus* food intoxication and toxicoinfection. Compr. Rev. Food Sci. Food Saf. 20, 3719–3761. https://doi.org/10.1111/1541-4337.12785.

Kindle, P., Etter, D., Stephan, R., Johler, S., 2019. Population structure and toxin gene profiles of *Bacillus cereus sensu lato* isolated from flour products. FEMS Microbiol. Lett. 366, fnz240. https://doi.org/10.1093/femsle/fnz240.

Kowalska, J., Maćkiw, E., Korsak, D., Postupolski, J., 2024. Characterization of the *Bacillus cereus* group isolated from ready-to-eat foods in Poland by whole-genome sequencing. Foods 13, 3266. https://doi.org/10.3390/foods13203266.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404.

Letunic, I., Bork, P., 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49, W293–W296. https://doi.org/10.1093/nar/gkab301.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Low, A.J., Koziol, A.G., Manninger, P.A., Blais, B., Carrillo, C.D., 2019. ConFindr: rapid detection of intraspecies and cross-species contamination in bacterial whole-genome sequence data. PeerJ 7, e6995. https://doi.org/10.7717/peerj.6995.

Lücking, G., Frenzel, E., Rütschle, A., Marxen, S., Stark, T.D., Hofmann, T., Scherer, S., Ehling-Schulz, M., 2015. Ces locus embedded proteins control the non-ribosomal

synthesis of the cereulide toxin in emetic *Bacillus cereus* on multiple levels. Front. Microbiol. 6. https://doi.org/10.3389/fmicb.2015.01101.

Masquelier, J., Segers, C., Jacobs, B., Van Nieuwenhuysen, T., Delbrassinne, L., Van Hoeck, E., 2024. Validation of a targeted LC–MS/MS method for Cereulide and application in food and Faeces. Toxins 16, 13. https://doi.org/10.3390/toxins16010013.

McGinnis, S., Madden, T.L., 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 32, W20–W25. https://doi.org/10.1093/nar/gkh435.

Messelhäusser, U., Frenzel, E., Blöchinger, C., Zucker, R., Kämpf, P., Ehling-Schulz, M., 2014. Emetic *Bacillus cereus* are more volatile than thought: recent foodborne outbreaks and prevalence studies in Bavaria (2007–2013). Biomed. Res. Int. 2014, 1–9. https://doi.org/10.1155/2014/465603.

Miller, R.A., Jian, J., Beno, S.M., Wiedmann, M., Kovac, J., 2018. Intraclade variability in toxin production and cytotoxicity of *Bacillus cereus* group type strains and dairy-associated isolates. Appl. Environ. Microbiol. 84. https://doi.org/10.1128/AEM.02479-17.

Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., Phillippy, A.M., 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 17, 132. https://doi.org/10.1186/s13059-016-0997-x.

Pightling, A.W., Petronella, N., Pagotto, F., 2014. Choice of reference sequence and assembler for alignment of *listeria monocytogenes* short-read sequence data greatly influences rates of error in SNP analyses. PLoS One 9, e104579. https://doi.org/10.1371/journal.pone.0104579.

Porcellato, D., Aspholm, M., Skeie, S.B., Mellegård, H., 2019. Application of a novel amplicon-based sequencing approach reveals the diversity of the *Bacillus cereus* group in stored raw and pasteurized milk. Food Microbiol. 81, 32–39. https://doi.org/10.1016/j.fm.2018.01.014.

Priest, F.G., Barker, M., Baillie, L.W.J., Holmes, E.C., Maiden, M.C.J., 2004. Population structure and evolution of the *Bacillus cereus* group. J. Bacteriol. 186, 7959–7970. https://doi.org/10.1128/JB.186.23.7959-7970.2004.

R Core Team, 2023. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rahnama, H., Azari, R., Yousefi, M.H., Berizi, E., Mazloomi, S.M., Hosseinzadeh, S., Derakhshan, Z., Ferrante, M., Conti, G.O., 2023. A systematic review and meta-analysis of the prevalence of *Bacillus cereus* in foods. Food Control 143, 109250. https://doi.org/10.1016/j.foodcont.2022.109250.

Rajkovic, A., Uyttendaele, M., Vermeulen, A., Andjelkovic, M., Fitz-James, I., in 't Veld, P., Denon, Q., Vérhe, R., Debevere, J., 2008. Heat resistance of *Bacillus cereus* emetic toxin, cereulide. Lett. Appl. Microbiol. 46, 536–541. https://doi.org/10.1111/j.1472-765X.2008.02350.x.

Samapundo, S., Heyndrickx, M., Xhaferi, R., Devlieghere, F., 2011. Incidence, diversity and toxin gene characteristics of *Bacillus cereus* group strains isolated from food products marketed in Belgium. Int. J. Food Microbiol. 150, 34–41. https://doi.org/10.1016/j.ijfoodmicro.2011.07.013.

Sánchez-Chica, J., Correa, M.M., Aceves-Diez, A.E., Castañeda-Sandoval, L.M., 2021a. Enterotoxin gene distribution and genotypes of *Bacillus cereus* sensu lato isolated from cassava starch. Toxins 13, 131. https://doi.org/10.3390/toxins13020131.

Sánchez-Chica, J., Correa, M.M., Aceves-Diez, A.E., Castañeda-Sandoval, L.M., 2021b. Genetic and toxigenic diversity of *Bacillus cereus* group isolated from powdered foods. J. Food Sci. Technol. 58, 1892–1899. https://doi.org/10.1007/s13197-020-04700-2.

Schürch, A.C., Arredondo-Alonso, S., Willems, R.J.L., Goering, R.V., 2018. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. Clin. Microbiol. Infect. 24, 350–354. https://doi.org/10.1016/j.cmi.2017.12.016.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Shaheen, R., Svensson, B., Andersson, M.A., Christiansson, A., Salkinoja-Salonen, M., 2010. Persistence strategies of *Bacillus cereus* spores isolated from dairy silo tanks. Food Microbiol. 27, 347–355. https://doi.org/10.1016/j.fm.2009.11.004.

Stanton, R.A., Vlachos, N., Halpin, A.L., 2022. GAMMA: a tool for the rapid identification, classification and annotation of translated gene matches from sequencing data. Bioinformatics 38, 546–548. https://doi.org/10.1093/bioinformatics/btab607.

Stenfors Arnesen, L.P., Fagerlund, A., Granum, P.E., 2008. From soil to gut: *Bacillus cereus* and its food poisoning toxins. FEMS Microbiol. Rev. 32, 579–606. https://doi.org/10.1111/j.1574-6976.2008.00112.x.

Torres Manno, M.A., Repizo, G.D., Magni, C., Dunlap, C.A., Espariz, M., 2020. The assessment of leading traits in the taxonomy of the *Bacillus cereus* group. Antonie Van Leeuwenhoek 113, 2223–2242. https://doi.org/10.1007/s10482-020-01494-3.

Tourasse, N.J., Jolley, K.A., Kolstø, A.-B., Økstad, O.A., 2023. Core genome multilocus sequence typing scheme for *Bacillus cereus* group bacteria. Res. Microbiol. 174 (6), 104050. https://doi.org/10.1016/j.resmic.2023.104050.

Yang, S., Wang, Y., Liu, Y., Jia, K., Zhang, Z., Dong, Q., 2023. Cereulide and emetic *Bacillus cereus*: characterizations. Impacts and Public Precautions. Foods 12, 833. https://doi.org/10.3390/foods12040833.

Yu, S., Yu, P., Wang, J., Li, C., Guo, H., Liu, C., Kong, L., Yu, L., Wu, S., Lei, T., Chen, M., Zeng, H., Pang, R., Zhang, Y., Wei, X., Zhang, J., Wu, Q., Ding, Y., 2020. A study on prevalence and characterization of *Bacillus cereus* in ready-to-eat foods in China. Front. Microbiol. 10, 3043. https://doi.org/10.3389/fmicb.2019.03043.

Zhou, Z., Alikhan, N.-F., Sergeant, M.J., Luhmann, N., Vaz, C., Francisco, A.P., Carriço, J. A., Achtman, M., 2018. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Res. 28, 1395–1404. https://doi.org/10.1101/gr.232397.117.