PLOS ONE

RESEARCH ARTICLE

Phylogenetic and protein prediction analysis reveals the taxonomically diverse distribution of virulence factors in *Bacillus cereus* strains

Ming Zhang[1], **Jun Liu**[1], **Zhenzhen Yin**[1]*, **Li Zhang**[2]*

1 School of Yunkang Medicine and Health, Nanfang College, Guangzhou, Guangdong, China, **2** School of Life Science, Liaoning University, Shenyang, Liaoning, China

* 837949196@qq.com (ZY); fmtzhangli@qq.com (LZ)

Citation: Zhang M, Liu J, Yin Z, Zhang L (2022) Phylogenetic and protein prediction analysis reveals the taxonomically diverse distribution of virulence factors in *Bacillus cereus* strains. PLoS ONE 17(5): e0262974. https://doi.org/10.1371/journal.pone.0262974

Editor: Fatih Oz, Ataturk Universitesi, TURKEY

Received: January 7, 2022

Accepted: April 21, 2022

Published: May 19, 2022

Data Availability Statement: All files are available from the NCBI database: https://www.ncbi.nlm.nih.gov/genome/?term=Bacillus+cereus.

Funding: The authors received "Doctor foundation project of nanfang college guangzhou (2020BQ23)" and "Science and technology projects in guangzhou (2022-1980)" fundings for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Bacillus cereus is a food contaminant with widely varying enterotoxic potential due to its virulence proteins. In this article, phylogenetic analysis of the amino acid sequences from the whole-genomes of 41 strains, evolutionary distance calculation of the amino acid sequences of the virulence genes, and functional and structural predictions of the virulence proteins were performed to reveal the taxonomically diverse distribution of virulence factors. The genome evolution of the strains showed a clustering trend based on the protein-coding virulence genes. The strains of *B. cereus* have evolved into non-toxic risk and toxic risk clusters with medium-high- and medium-low-risk subclusters. The evolutionary transfer distances of incomplete virulence genes relative to housekeeping genes were greater than those of complete virulence genes, and the distance values of HblACD were higher than those of nheABC and CytK among the complete virulence genes. Cytoplasmic localization was impossible for all the virulence proteins, and NheB, NheC, Hbl-B, and Hbl-L_1 were predicted to be extracellular. Nhe and Hbl proteins except CytK had similar spatial structures. The predicted structures of Nhe and Hbl mainly showed 'head' and 'tail' domains. The 'head' of NheA and Hbl-B, including two α-helices separated by β-tongue strands, might play a special role in the formation of Nhe trimers and Hbl trimers, respectively. The 'cap' of CytK, which includes two 'latches' with many β-sheets, formed a β-barrel structure with pores, and a 'rim' balanced the structure. The evolution of *B. cereus* strains showed a clustering tendency based on the protein-coding virulence genes, and the complete virulence-gene operon combination had higher relative genetic stability. The beta-tongue or latch associated with β-sheet folding might play an important role in the binding of virulence structures and pore-forming toxins in *B. cereus*.

Introduction

Bacillus cereus (*B. cereus*), which is one of twelve closely related species in the *Bacillus cereus* group [1], is a Gram-positive bacterium occurring ubiquitously in nature with widely varying pathogenic potential [2]. Cells are rod-shaped, with some in chains or occasionally long

Abbreviations: ANI, average nucleotide identity; MLSA, multilocus sequence analysis; Nhe, nonhemolytic enterotoxin; Hbl, hemolysin BL; CytK, cytotoxin K; DV, distance value; CV, composition vector; GMQE, global model quality estimate; NCBI, National Center for Biotechnology Information; *adk*, adenylate kinase; *ccpA*, catabolite control protein A; *glpF*, glycerol uptake facilitator protein; *glpT*, glycerol-3-phosphate transporter; *panC*, pantoate-beta-alanine ligase; *pta*, phosphate acetyltransferase; *pyc*, pyruvate carboxylase.

filaments, and are aerobic or facultative anaerobic. Most species will grow on common media such as nutrient agar and blood agar. They are characteristically large (2–7 mm in diameter) and vary in shape from circular to irregular, with matt or granular textures, while smooth and moist colonies are also common. *B. cereus*, the spores of which can survive at high temperatures and germinated vegetative cells of which can multiply and produce toxins under favorable conditions, is recognized as the most frequent cause of food-borne disease [3]. Its toxins cause two distinct forms of food poisoning, the emetic type (uncommon) and the diarrheal type (common). Diarrheal strains produce three enterotoxins, which belong to the family of pore-forming toxins: nonhemolytic enterotoxin (Nhe), hemolysin BL (Hbl), and cytotoxin K (CytK). Nhe comprises three proteins, NheA, NheB, and NheC, encoded by one operon containing one of three genes, namely, *nheA*, *nheB*, and *nheC*, respectively. Hbl consists of a single B-component (encoded by *hblA*) and two L-components, L_1 (*hblC*) and L_2 (*hblD*), all of which are essential for activity, with no individual or pairwise activity [4]. CytK (*cytK*) is a single-component toxin [5]. The genes that encode NheABC can be detected in nearly all enteropathogenic *B. cereus* strains, *hblACD* can be detected in approximately 45% to 65% of such strains, and *cytK* is less prevalent [6,7]. *B. cereus* species, which were compared on the basis of 16S rRNA (identity values >98%), were closely homologous to each other [8]. Nevertheless, the suitability of this marker for the classification of *B. cereus* might be limited, as it is unable to effectively distinguish between the closely related species [9]. Some papers have reported that the species affiliation of *B. cereus* group, which could lead to an exchange of virulence plasmids between species, often does not match patterns of phylogenetic relatedness [10,11]. While the enterotoxins of *B. cereus* are chromosome-coded, the unique characteristics are observed for plasmids and are thus present throughout the *B. cereus* group [12]. Lapidus et al. reported a large plasmid with an operon encoding all three Nhe components in a *B. cereus* strain [13]. There is evidence that extensive gene exchange occurs between plasmids and the chromosome during the evolution of the *B. cereus* group [14]. Therefore, some genes encoded on plasmids can spread via horizontal gene transfer among *B. cereus* and the transfer of a single plasmid from one species to another [15]. Didelot et al. detected three phylogenetic groups (clades) in a study on the evolution of pathogenicity in the *B. cereus* group [16]. *B. cereus*, as a genomospecies, could be mainly found in clade two based MLSA and multiple comparative analysis of ANI values [17]. Later, seven major phylogenetic groups with ecological differences were identified in the *B. cereus* group [10]. A recent study suggested that nine phylogenetic clades of isolates may be better for assessing the risk of diarrheal foodborne disease caused by *B. cereus* group isolates [18]. These studies of virulence factors of *B. cereus* concern the evolutionary classification of virulence genes, and there have been few comparative analyses of the relative evolutionary distance of virulence genes and the prediction of virulence protein function and structure. In this study, the genome, virulence gene sequences and predicted virulence proteins of 41 *B. cereus* strains were comparatively analyzed. This work aims to examine the species diversity of *B. cereus* strains and the phylogenetic relationships among virulence factors, to systematically evaluate the distribution of virulence genes, and to comparatively analyze the structures and functions of virulence proteins.

Materials and methods

Characterization of *B. cereus* strains

Forty-one strains of *B. cereus* with complete- and chromosome-level assemblies in the National Center for Biotechnology Information (NCBI) database were selected for comparative analysis. Datasets that passed the completeness test (acceptable level is >85%) and contamination test (acceptable level is <5%) were composed of sequences submitted on deadline March 5, 2021.

Table 1. The forty-one *B. cereus* strains and one *S. terrae* control strain used in this study.

Strain	Length (Mb)	G+C (%)	Accession number	Origin	Strain	Length (Mb)	G+C (%)	Accession number	Origin
B4264	5.42	35.30	GCA_000021205.1	Pneumonia	172560W	5.70	34.80	GCA_000160935.1	Human tissue
ATCC14579	5.43	35.31	GCA_000007825.1	Soil	Rock3-29	5.88	34.90	GCA_000161215.1	Soil
FORC-013	5.68	35.20	GCA_001518875.1	Foodborne	Rock1-3	5.86	34.90	GCA_000161155.1	Soil
03BB102	5.45	35.29	GCA_000022505.1	Pneumonia	Rock4-18	5.92	35.00	GCA_000161295.1	Soil
ATCC10987	5.43	35.52	GCA_000008005.1	Soil	Rock1-15	5.77	34.90	GCA_000161175.1	Soil
G9842	5.74	35.05	GCA_000021305.1	Stool	BDRD-ST24	5.44	35.10	GCA_000161055.1	Nd
AH820	5.59	35.31	GCA_000021785.1	Periodontitis	ATCC10876	5.94	34.80	GCA_000160895.1	Nd
F837/76	5.29	35.40	GCA_000239195.1	Enteritis	Rock4-2	5.77	34.90	GCA_000161275.1	Soil
Q1	5.51	35.50	GCA_000013065.1	Nd	BGSC6E1	5.73	35.00	GCA_000160915.1	Nd
CI	5.49	35.27	GCA_000143605.1	Anthrax	F65185	6.13	34.70	GCA_000161315.1	Wound
E33L	5.84	35.17	GCA_000833045.1	Carcass	Rock3-28	6.04	35.75	GCA_000161195.1	Soil
AH187	5.60	35.52	GCA_000021225.1	Nd	Rock3-42	5.20	35.20	GCA_000161235.1	Soil
NC7401	5.55	35.54	GCA_000283675.1	Soil	m1293	5.27	35.35	GCA_000003645.1	Food
SGAir0263	6.47	34.91	GCA_010223795.1	Air	BDRD-ST196	5.58	35.20	GCA_000161095.1	Nd
SGAir0260	6.30	34.97	GCA_010232525.2	Air	AH676	5.59	35.00	GCA_000161395.1	Soil
BHU1	5.20	35.10	GCA_002504105.1	Soil	AH1271	5.66	35.30	GCA_000161375.1	Lamp
Co1-1	6.38	35.10	GCA_007923085.1	Wastewater	AH1272	5.79	35.20	GCA_000161395.1	Amniotic fluid
ATCC4342	5.23	35.20	GCA_000161015.1	Food	AH1273	5.79	35.25	GCA_000003955.1	Blood
BDRD-Cer4	5.40	35.10	GCA_000161115.1	Food	R309803	5.59	35.30	GCA_0000160995.1	Clinic
m1550	5.25	35.10	GCA_000161035.1	Food	BDRD-ST26	5.57	35.25	GCA_000161075.1	Nd
95/8201	5.58	35.10	GCA_000161135.1	Endocarditis	Control 70–3	3.31	45.30	GCA_009176625.1	Nd

Nd: Not determined.

https://doi.org/10.1371/journal.pone.0262974.t001

Thirty-one pathogenic and ten nonpathogenic strains that were isolated from food, patients, the environment, and unknown sources, were eligible, along with a control strain, *Sporolactobacillus terrae* 70–3 (*S. terrae* 70–3), that belonged to a different genus. In terms of evolution, *S. terrae* which has a defined taxonomic and phylogenetic status, is closely related to *B cereus* in Bacillales. The sequences and annotation information for the stains were downloaded from the NCBI (details in Table 1).

Quality assessment of genomic sequences

The contamination and completeness of the metagenomic sequences were evaluated by CheckM software version v1.1.3 [19].

Phylogenetic and average nucleotide identity (ANI) analysis

The first phylogenetic tree, based on whole-genome amino acid sequences of each strain, was constructed by using the CVTree4 webserver (http://cvtree.online/v4/prok/index.html), which constructs whole-genome-based phylogenetic trees without sequence alignment by using a composition vector (CV) approach, and the K-tuple length was 6 [20]. Every genome sequence was represented by a composition vector, which was calculated as the difference between the frequencies of k-strings and the prediction frequencies by the Markov model [21]. The shape and text content of the phylogenetic tree were modified by Molecular Evolutionary Genetics

Analysis (MEGA-X version 10.2.2) [22]. In this study, the same genome sequence data were subjected to ANI analysis to verify the significance of the first phylogenetic tree. ANI analysis was performed using JSpeciesWS Online Service (http://jspecies.ribohost.com/jspeciesws/) as described by Richter et al. [23]. The distance matrix, which was calculated by the distance value (DV) using the formula DV = 1-[ANIb value], was used to construct the second phylogenetic tree, which was generated from the resulting Newick format file using Njplot [24]. The formula was balanced using the mean value method and was subjected to calculation using DrawGram in the PHYLIP package version 3.695 [25]. To determine the associations between each protein-coding gene and the different clusters, statistical enrichment analyses were conducted with PhyloGLM V2.6 [17].

Multilocus sequence analysis (MLSA)

A total of forty-one strains containing gene sequences, which were downloaded from the NCBI, were found and further analyzed for the presence of seven housekeeping and three enterotoxin genes. The housekeeping genes adenylate kinase (*adk*), catabolite control protein A (*ccpA*), glycerol uptake facilitator protein (*glpF*), glycerol-3-phosphate transporter (*glpT*), pantoate-beta-alanine ligase (*panC*), phosphate acetyltransferase (*pta*), and pyruvate carboxylase (*pyc*) were chosen to calculate the basic evolutionary distances of the species. These housekeeping genes, scattered across the entire chromosome, are suitable for MLSA [26]. The types of enterotoxin genes (*nhe*, *hbl*, and *cytK*) were divided into different groups, and the base sequences were concatenated for further MLSA. Thus, rearrangement of genes was unnecessary because the order of the genes within the operons was conserved in all strains. The distances of concatenated genes were calculated in MEGA X using the maximum likelihood (ML) algorithms, which are based on the Tamura-Nei model with a discrete gamma distribution [22]. The model applied for MLSA of DVs was the ideal substitution model according to the 'find best DNA/Protein models' function [27]. The housekeeping genes were of the same length in all strains, as were the different virulence genes. The same settings for the calculation of all phylogenetic DVs were used to ensure comparability of the results. We calculated the relative changes in genetic DVs between the virulence genes, which were concatenated housekeeping genes minus the simple housekeeping gene, representing the change in virulence gene transfer.

Prediction of virulence protein function and structure

SMART software (http://smart.embl-heidelberg.de/), which is a simple modular architecture research tool, was used to predict the domain architecture of the virulence proteins in this study [28]. PSORT (http://www.psort.org/psortb2) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM) software were employed to predict the subcellular location and transmembrane helices of virulence proteins, respectively [29,30]. The SIGNALP-5.0 (http://www.cbs.dtu.dk/services/SignalP/), SWISS-MODEL (http://swissmodel.expasy.org) and AlphaFold v2.1.1 (https://github.com/deepmind/alphafold) servers were used to predict the signal peptide cleavage and three-dimensional (3-D) structures of the enterotoxin proteins, respectively [31–34]. The amino acid sequences of the virulence proteins analyzed were submitted in FASTA format. To predict structure, we performed homology modeling to generate 3-D virulence protein structures.

Results

General genome characteristics and quality assessment of sequences

A summary of the features of the forty-one genomes of *B*. *cereus* and the control genome of the closely related species *S*. *terrae* is provided in Table 1. The genome sizes of *B*. *cereus* strains

varied from 5.20 to 6.47 MB. The G+C contents of the forty-one genomes ranged from 34.70% to 35.75%. Compared with the control genome from *S. terrae* 70–3 (3.31 MB and 45.30%), the genomes of *B. cereus* were much larger and had lower G+C contents. The contamination and completeness of the sequences were 0–2.02% and 89.81%-98.99%, respectively (shown in Table 2). These results suggested that these sequences are of high quality, have low contamination (values <2.02%, acceptable level is <5%), and have high completeness (values >89.81%, acceptable level is >85%); thus, they were appropriate for analysis. In this study, the strains originated from food (5/41), the clinic (12/41), the environment (17/41), and undetermined sources (7/41). The enterotoxic risk potential based on the virulence genes of forty-one *B. cereus* strains is listed in Table 2. Enterotoxicity, which was reflected by virulence gene numbers, was categorized into levels of three types (10/41), two types (14/41), one type (7/41), and no types (10/41) levels. The genes detected as enterotoxic were *nheABC* (29/41), *hblACD* (19/41), *cytk* (17/41), *nheAB* (10/41), *hblCD* (6/41), *hblAD* (2/41), and *hblD* (2/41) (shown in Table 2).

Phylogenetic analysis based on whole amino acid sequences

Two whole-genome-based methods were used to construct phylogenetic trees. The first phylogenetic tree was constructed with the CV method using the whole amino acid sequences of forty-one *B. cereus* strains and the outgroup species *S. terrae* 70–3 [35]. To ensure the accuracy of the results, we added the inbuilt sequence AH1273 and the sequence of *S. terrae* 70–3 (control) from the webserver database (Fig 1). According to enterotoxic risk potential, the forty-one strains of *B. cereus* had evolved into five distinct clusters, which were likely risk regions I, IV and V and nonrisk regions II and III. However, there were individual nonconformities, such as nontoxicity of BDRD ST196 in region V. Region I was dominated by medium- and high-risk strains (15/17) but also included two low-risk strains (BHU1 and CO1-1), region II and III included only nonrisk strains (9/9), and regions IV and V were dominated by medium- and low-risk strains (13/15) but also included AH820 (high-risk strain) and BDRD ST196 (nonrisk strain), respectively.

To verify the above results and obtain more accurate molecular evolutionary relationships, we established a second phylogenetic tree based on ANI analysis (Fig 2). According to enterotoxic risk potential, the forty-one strains of *B. cereus* had evolved into six distinct regions: likely risk regions A, C, E, and D_2 and nonrisk regions B and D_1. The two phylogenetic trees were similar in terms of the regions where enterotoxic risk was likely. The only difference between the two phylogenetic trees was a change in the evolutionary cluster of two strains. The ATCC 10987 and ATCC 4342 strains, which belonged region IV in the first tree, were assigned to region D in the second tree. Region A, which was dominated by medium-high-risk strains (15/17) but also included two low-risk strains (BHU1 and CO1-1), was the same as region I. Regions B and D_1, which included only nonrisk strains (9/9), were the same as regions II and III. Regions C, D_2 and E, which were also dominated by medium- and low-risk strains (13/15) but included AH820 (high-risk strain in C) and BDRD ST196 (nonrisk strain in E), were the same as regions IV and V. By taking advantage of the updated enterotoxic risk regions found in the current *B. cereus* strains, we decided to use a statistical approach to evaluate whether the occurrence of virulence factor-encoding genes (detailed in Table 3) correlate with a particular region. We observed that *nheABC* was significantly present in region C ($p < 0.05$), and *hblA* and *nheC* were significantly present in region A and C ($p < 0.05$), respectively. The results showed that *nheABC* and *nheC* were significantly enriched in the medium-low-risk region, and *hblA* was significantly enriched in the medium-high-risk region.

Table 2. The results of the completeness, contamination, enterotoxic genes and risk potential of the strains.

Strain	Completeness (%)	Contamination (%)	Enterotoxic risk potential	Analysis of enterotoxic genes			Strain	Completeness (%)	Contamination (%)	Enterotoxic risk potential	Analysis of enterotoxic genes		
				nheABC	*hblACD*	*cytK*					*nheABC*	*hblACD*	*cytK*
B4264	98.99	0.00	H	+	+	+	172560W	98.99	0.00	M	+	*CD*	+
ATCC14579	98.99	1.35	H	+	+	+	Rock3-29	98.99	0.00	-	*AB*	*CD*	-
FORC_013	98.99	0.00	H	+	+	+	Rock1-3	98.99	0.00	-	*AB*	*CD*	-
03BB102	98.99	0.00	L	+	-	-	Rock4-18	98.99	0.00	-	*AB*	*CD*	-
ATCC10987	98.99	0.00	M	+	-	+	Rock1-15	98.99	1.01	H	+	+	+
G9842	98.99	2.02	M	+	+	-	BDRD-ST24	98.99	0.00	H	+	+	+
AH820	98.48	0.00	H	+	+	+	ATCC10876	98.99	0.00	H	+	+	+
F837/76	98.99	0.00	M	+	+	-	Rock4-2	89.81	0.34	H	+	+	+
Q1	98.99	0.34	-	*AB*	-	-	BGSC6E1	98.99	0.00	M	+	+	-
CI	98.99	1.01	L	+	-	-	F65185	97.98	0.34	M	+	*AD*	+
E33L	98.99	0.00	M	+	-	+	Rock3-28	96.46	0.34	-	*AB*	*CD*	-
AH187	98.99	0.00	-	*AB*	-	-	Rock3-42	98.99	0.00	M	+	-	+
NC7401	98.99	0.00	-	*AB*	-	-	m1293	98.99	0.34	-	*AB*	-	-
SGAir0263	98.99	0.34	M	+	+	-	BDRD-ST196	98.99	0.34	-	*AB*	*CD*	-
SGAir0260	98.99	0.34	M	+	+	-	AH676	98.99	1.01	M	+	*AD*	+
BHU1	98.99	0.00	L	-	+	-	AH1271	98.99	0.00	M	+	+	-
Co1-1	98.99	0.00	L	-	+	-	AH1272	97.47	1.35	L	+	*D*	-
ATCC4342	98.99	0.34	M	+	+	-	AH1273	96.93	1.35	L	+	*D*	-
BDRD-Cer4	97.98	0.34	H	+	+	+	R309803	98.99	0.00	L	+	-	-
m1550	98.99	0.00	H	+	+	+	BDRD-ST26	98.99	0.00	-	*AB*	-	-
95/8201	98.99	1.01	M	+	-	+	Control 70–3	98.99	0.00	-	-	-	-

Enterotoxic risk potential: Toxicity via the expression of the *nheABC*, *hblACD*, and *cytK* genes; + or -: All or none, respectively; H, M, or L: Three, two, or one type(s) of toxic genes, respectively.

https://doi.org/10.1371/journal.pone.0262974.t002

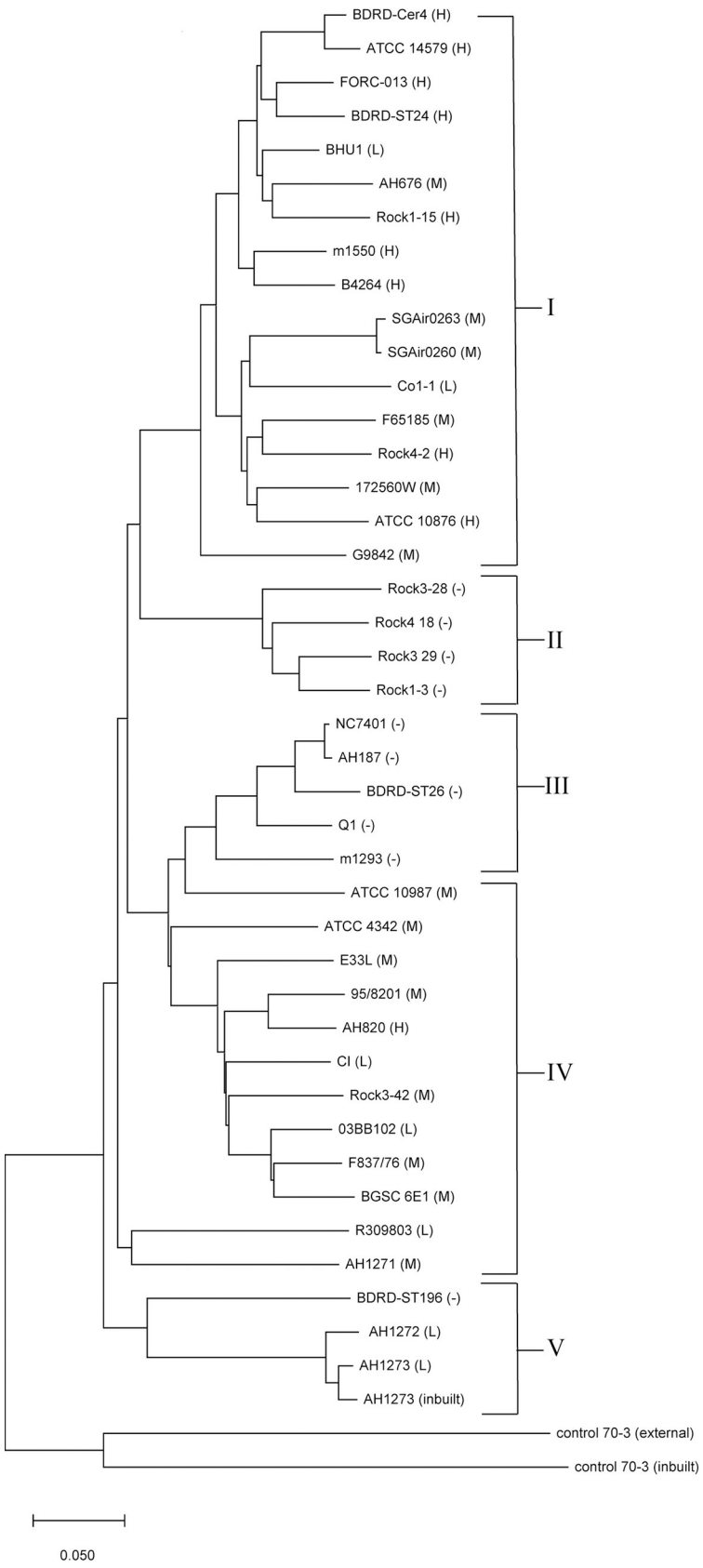
BDRD-Cer4 (H)
ATCC 14579 (H)
FORC-013 (H)
BDRD-ST24 (H)
BHU1 (L)
AH676 (M)
Rock1-15 (H)
m1550 (H)
B4264 (H)
SGAir0263 (M)
SGAir0260 (M)
Co1-1 (L)
F65185 (M)
Rock4-2 (H)
172560W (M)
ATCC 10876 (H)
G9842 (M)
I
Rock3-28 (-)
Rock4 18 (-)
Rock3 29 (-)
Rock1-3 (-)
II
NC7401 (-)
AH187 (-)
BDRD-ST26 (-)
Q1 (-)
m1293 (-)
III
ATCC 10987 (M)
ATCC 4342 (M)
E33L (M)
95/8201 (M)
AH820 (H)
CI (L)
Rock3-42 (M)
03BB102 (L)
F837/76 (M)
BGSC 6E1 (M)
R309803 (L)
AH1271 (M)
IV
BDRD-ST196 (-)
AH1272 (L)
AH1273 (L)
AH1273 (inbuilt)
V
control 70-3 (external)
control 70-3 (inbuilt)
0.050

Fig 1. Phylogenetic relationships of the amino acid sequences of forty-one *B. cereus* strains and one external *S. terrae* control strain used in this study. Two additional inbuilt strains from the software were used as internal controls. H, M, L, and "-" symbols indicate high (three types of virulence genes), middle (two types of virulence genes), low (one type of virulence gene), and no enterotoxic risk potential.

https://doi.org/10.1371/journal.pone.0262974.g001

Phylogenetic distance analysis based on concatenated housekeeping and virulence genes

To analyze the evolution and phylogenetic relationships of virulence gene transfer in relation to DVs, the *nheABC*, *hblACD*, and *cytK* genes of the forty-one strains, which need to be compared to the housekeeping genes of the strains, were studied. To this end, we concatenated the sequences of virulence proteins from the strains and seven housekeeping proteins (Adk-CcpA-GlpF-GlpT-PanC-Pta-Pyc) from the *B. cereus* core genome. The genetic DV of virulence gene transfer was evaluated by calculating the average difference in the phylogenetic DV of the ATCC14579 strain compared with forty other strains. The hblD and hblAD virulence proteins, which were observed in only two strains, were excluded. As shown in Table 2, the relative genetic DVs were calculated for nheAB (10/41), hblCD (6/41), nheABC (29/41), hblACD (19/41), and CytK (17/41). As shown in Fig 3, the average evolutionary DVs of virulence gene transfer from high to low were 0.015 (*nheAB*), 0.012 (*hblCD*), 0.005 (*hblACD*), 0.003 (*nheABC*) and 0.001 (*cytK*). The DVs of incomplete virulence genes (*nheAB* and *hblCD*) were higher than those of complete virulence genes (*nheABC*, *hblACD*, and *CytK*). The average evolutionary DV of *nheAB* was higher than the DV of *hblCD* among the incomplete virulence genes; the DV of *hblACD* was the highest and that of *cytK* was the lowest among the complete virulence genes.

Comparative prediction analysis of the function and structure of virulence proteins

As shown in Table 4 and Fig 4, we obtained the scores of the seven virulence proteins for subcellular localization prediction. The scores of NheB, NheC, Hbl-B, and Hbl-L_1 were all 9.73, and that of CytK was 9.98, all consistent with extracellular localization. The localization of NheA and Hbl-L_2 was unknown because the scores were all lower in the cytoplasmic membrane (3.33/4.6), cell wall (3.33/2.48), and extracellular space (3.33/2.92), making it impossible for the virulence proteins to appear in the cytoplasm. NheB and Hbl-L_1 had two helices, which were transmembrane region sequences 235-257/267-286 and 239-261/268-290, and NheC had only one helix, of which the transmembrane region was 228–250, but the others had none. The virulence protein cleavage sites of all strains were in the sequence 30–32 with 0.93–0.99 likelihood levels, except NheA, for which the site was in the 26–27 sequence (0.81 likelihood level). The amino acid sequences of Nhe, Hbl and CytK contained N-terminal signal peptides for secretion ($<$ 31 amino acids). The signal peptide start-end was between 1 and 31 sequences but not found for Hbl-B and Hbl-L_1 were not found, and the domain start-end was between 35 and 329 sequences.

Examination of the phylogenetic tree constructed using the Hbl, Nhe, and CytK sequences of ATCC 14579 (Fig 5) showed that NheA and Hbl-L_2, as well as NheBC and Hbl-L_1, were more closely related to one another than to the other components, and CytK was the least evolutionarily related. This result was also reflected in the evaluation parameters of the 3-D enterotoxin protein structures. As shown in Table 5, the closest template for NheB and NheC was Hbl-L_1 (sequence identity of 40.82% and 36.83%, respectively), and that for Hbl-L_2 was NheA (24.85%). The closest template for CytK was alpha-hemolysis (30.39%), as expected, with considerable amino acid sequence homology to *S. aureus* leukocidin [36]. The templates


E
D
D2
D1
C
B
A
AH1272 (L)
AH1273 (L)
BDRD-ST196 (-)
ATCC 4342 (M)
ATCC10987 (M)
AH187 (-)
BDRD-ST26 (-)
NC7401 (-)
Q1 (-)
m1293 (-)
03BB102 (L)
F837/76 (M)
BGSC 6E1 (M)
CI (L)
ROCK3-42 (M)
AH820 (H)
95/8201 (M)
E33L (M)
AH1271 (M)
R309803 (L)
ROCK1-3 (-)
ROCK3-29 (-)
ROCK3-28 (-)
ROCK4-18 (-)
G9842 (M)
172560W (M)
Co1-1 (L)
Rock4-2 (H)
F65185 (M)
SGAir0260 (M)
SGAir0263 (M)
ATCC 10876 (H)
ATCC 14579 (H)
BDRD-Cer4 (H)
BDRD-ST24 (H)
FRC-013 (H)
Rock1-15 (H)
BHU1 (L)
AH676 (M)
m1550 (H)
B4264 (H)
Cotrol 70-3
0.050

Fig 2. ANI analysis of the phylogenetic relationships of the forty-one *B. cereus* strains and one *S.terrae* control strain used in this study. H, M, L, and "-" have the same meanings as in Table 2.

https://doi.org/10.1371/journal.pone.0262974.g002

of NheA, and Hbl-L_1 were included in the SWISS-MODEL server with high sequence identity (97.22% and 99.73%, respectively), and Hbl-B had acceptable sequence identity (71.99%). The sequence coverage and range of all structures were 0.71–0.93 and 33–439, respectively, with GMQE evaluation values (0.54–0.88) above 0.5, which indicated reliable model construction. Each residue is allotted a reliability score between 0 and 1, indicating the expected resemblance to the native structure. Higher numbers represent higher reliability of the residues [37]. In general, a sequence identity of >30% for each template was acceptable based on the SWISS-MODEL server. To verify the above results (especially the sequence identity of Hbl-L_2, which was 24.85%) and obtain the predicted Nhe-trimer and Hbl-trimer structures, we used AlphaFold software for secondary structural prediction. The results were acceptable, the predicted local-distance difference test (plDDt) values of monomers were 81.90–94.14, and the scores of Nhe trimers and Hbl trimers were 0.68 and 0.36 (ipTM+pTM), respectively [33,34].

Due to the sequence similarity of NheB and NheC with Hbl-B, homology models based on the Hbl-B structure were established. As shown in Fig 6, NheA and Hbl-B had highly similar structures (Fig 6A), and the NheA, NheB, NheC, HBl-B, HBl-L_1, and HBl-L_2 structures showed that there were two main domains, a 'head' and 'tail' (Figs 6B–6D, 6F–6H, and 7A–7F). The main body of the structure was formed by the 'tail' domain, which consisted of five major helices, and the 'head' domain of NheA included two long α-helices separated by β-tongue strands (Figs 6B and 7A). Multiple β-tongue strands were detected in Hbl-B (Fig 7D) but are not shown in Fig 6F because of the prediction method. Another difference was the 'head' of Hbl-L_2, possibly related to the low sequence identity (Figs 6H and 7F). The 'latch' with many β-sheets of CytK folded the 'cap' domain, which was the toxic area (Figs 6E and 7G). The amino 'latch', which included a short helix in all known pore structures, was observed on the top of the conformation, which extended into the pore to form a β-barrel and was folded into a stranded antiparallel β-sheet in the monomer. Although the amino 'latch' protrudes and interacts with the adjacent protomer in the pore, it is located at the edge of the β-sheet of the 'cap' region [39]. The 'rim' domain, which was composed of three strands of short β-sheets, formed the main body of the balanced structure. The trimers of Nhe and Hbl were horizontally arranged. The Hbl trimer (arranged in the sequence B, -L_1, -L_2) was more similar than the Nhe trimer (arranged in the sequence A, B, C) based on the structural features. The β-tongue strands of Hbl-B and NheA might play an important structural and functional role in the formation of trimers.

Discussion

In this study, forty-one strains of *B. cereus* were subjected to phylogenetic analyses based on whole amino acid sequences. Enterotoxicity, which was evaluated on the basis of *nheABC*, *hblACD*, and *cytK* gene expression, was classified into levels of three types, two types, one type, and no types. In terms of evolutionary relationships, clusters of virulence and nonvirulence gene strains were evident, and the regional distribution of the number of types of virulence genes was also presented, further confirmed by ANI-based phylogenetic analyses. We found that the two phylogenetic trees were similar. All non-toxic-risk strains were concentrated in two clusters, and all but two of the medium-high- and medium-low-toxic-risk strains formed clusters. The results suggest the possibility of virulence gene transfer, which may be related to frequent exchange of pathogenicity factors during *B. cereus* virulence evolution, including so-called probiotic or nonpathogenic species [15]. Previous taxonomic results for the *B. cereus*

Table 3. Analysis of the presence of toxic genes enriched by the PhyloGLM tool.

Clade	nheA		nheB		nheC		hblA		hblC		hblD		cytK		nheABC		hblACD	
	E	pvalue	E	pvalue	E	pvalue	E	pvalue	E	pvalue	E	pvalue	E	pvalue	E	pvalue	E	pvalue
A	-0.0324	0.9259	-0.0324	0.9259	0.3325	0.5176	1.7076	0.0265	0.0696	0.7994	0.1587	0.7026	0.2329	0.5356	0.3326	0.5176	0.4710	0.2952
B	2.6099	0.4195	2.6327	0.4202	-0.1009	0.7644	-0.3859	0.4192	0.4412	0.3372	0.4722	0.3986	-0.0551	0.8126	-0.1009	0.7644	-0.5876	0.3012
C	0.0003	0.9992	-0.0002	0.9995	2.7819	0.0306	0.0142	0.9645	-0.0477	0.8810	-0.1765	0.7184	0.0474	0.8796	2.7819	0.0306	0.0157	0.9517
D	-0.0082	0.9832	-0.0082	0.9832	-0.9999	0.2557	-0.0182	0.9555	-0.0605	0.8506	-1.4942	0.1320	-0.0519	0.8722	-0.9999	0.2557	-0.0086	0.9728
E	2.2231	0.3705	2.2245	0.3704	-0.0364	0.8931	-0.3678	0.4250	-1.8756	0.1716	0.4070	0.4224	-0.0499	0.8161	-0.0364	0.8931	-0.0642	0.7707

E: Estimated value for the Generalized Linear Model.

https://doi.org/10.1371/journal.pone.0262974.t003

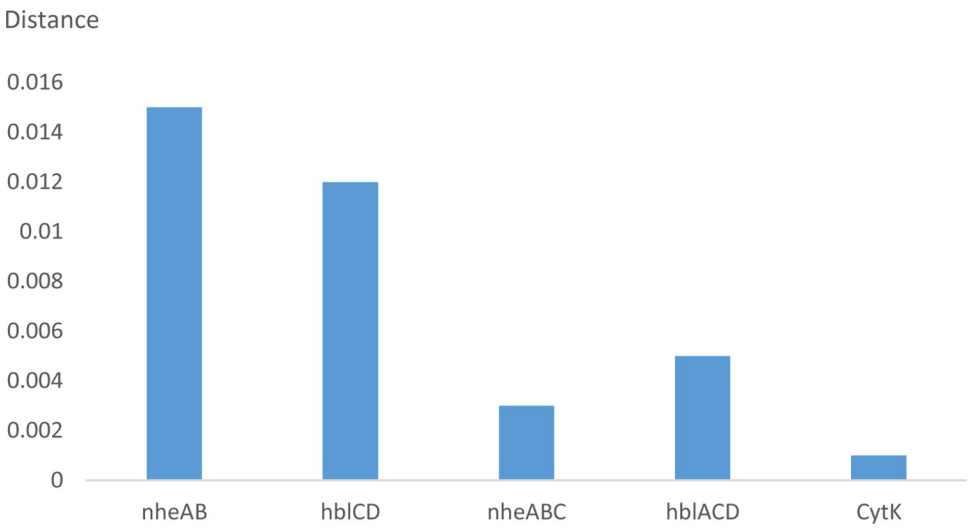

Fig 3. The average difference in the phylogenetic distance values of the ATCC14579 strain compared with forty other strains for virulence genes plus housekeeping genes and housekeeping genes examined by MLSA.

https://doi.org/10.1371/journal.pone.0262974.g003

group are largely based on inadequate criteria such as virulence characteristics, which residing on virulence plasmids [9]. Due to rampant horizontal gene transfer in bacterial ecosystems, increasing numbers of "core" genes should be found and defined based on refined species classification [41]. Recently, phylogeny-aware methods based on linear regression models were applied at the whole-genome scale to study the genomes of bacteria [17]. By using this statistical approach, we hereby observed that the virulence genes *nheABC* and *nheC* positively correlated with enrichment in the medium-low-risk cluster, and *hblA* was found in the medium-high-risk cluster. The inconsistent evolutionary distribution of individual virulence genes may be due to other factors, which needs further study.

The *Bacillus* hemolytic and nonhemolytic enterotoxin family of proteins consists of several *Bacillus* enterotoxins, which can cause food poisoning in humans [42]. Hemolytic BL and

Table 4. Subcellular localization, transmembrane helix and region, signal peptide, and domain prediction results of the virulence proteins of *B. cereus* strains.

Name	Final localization	Final score	Scores for different localizations				Helix number	Transmembrane region	Cleavage site	Signal peptide likelihood	Signal peptide	Domain
			Cytoplasmic membrane	Cell wall	Extracellular	Cytoplasmic					start-end	start-end
NheA	Unknown	3.33/4.60	3.33/4.60	3.33/2.48	3.33/2.92	0.00	0	-	26 and 27	0.81±0.07	1–26	41–216
NheB	Extracellular	9.73	0.09	0.18	9.73	0.00	2	235–257;267–286	30 and 31	0.99±0.01	1–30	54–232
NheC	Extracellular	9.73	0.09	0.18	9.73	0.00	1	228–250	30 and 31	0.98±0.02	1–23	46–224
Hbl-B	Extracellular	9.73	0.09	0.18	9.73/9.72	0.00	0	-	31 and 32	0.96±0.04	-	45–224
Hbl-L_1	Extracellular	9.73	0.09	0.18	9.73	0.00	2	239–261;268–290	30 and 31	0.98±0.01	1–20	44–230
Hbl-L_2	Unknown	3.33/4.60	3.33/4.60	3.33/2.48	3.33/2.92	0.00	0	-	32 and 33	0.93±0.05	-	35–207
CytK	Extracellular	9.98	0.00	0.02	9.98	0.00	0	-	31 and 32	0.99±0.01	1–31	64–329

https://doi.org/10.1371/journal.pone.0262974.t004

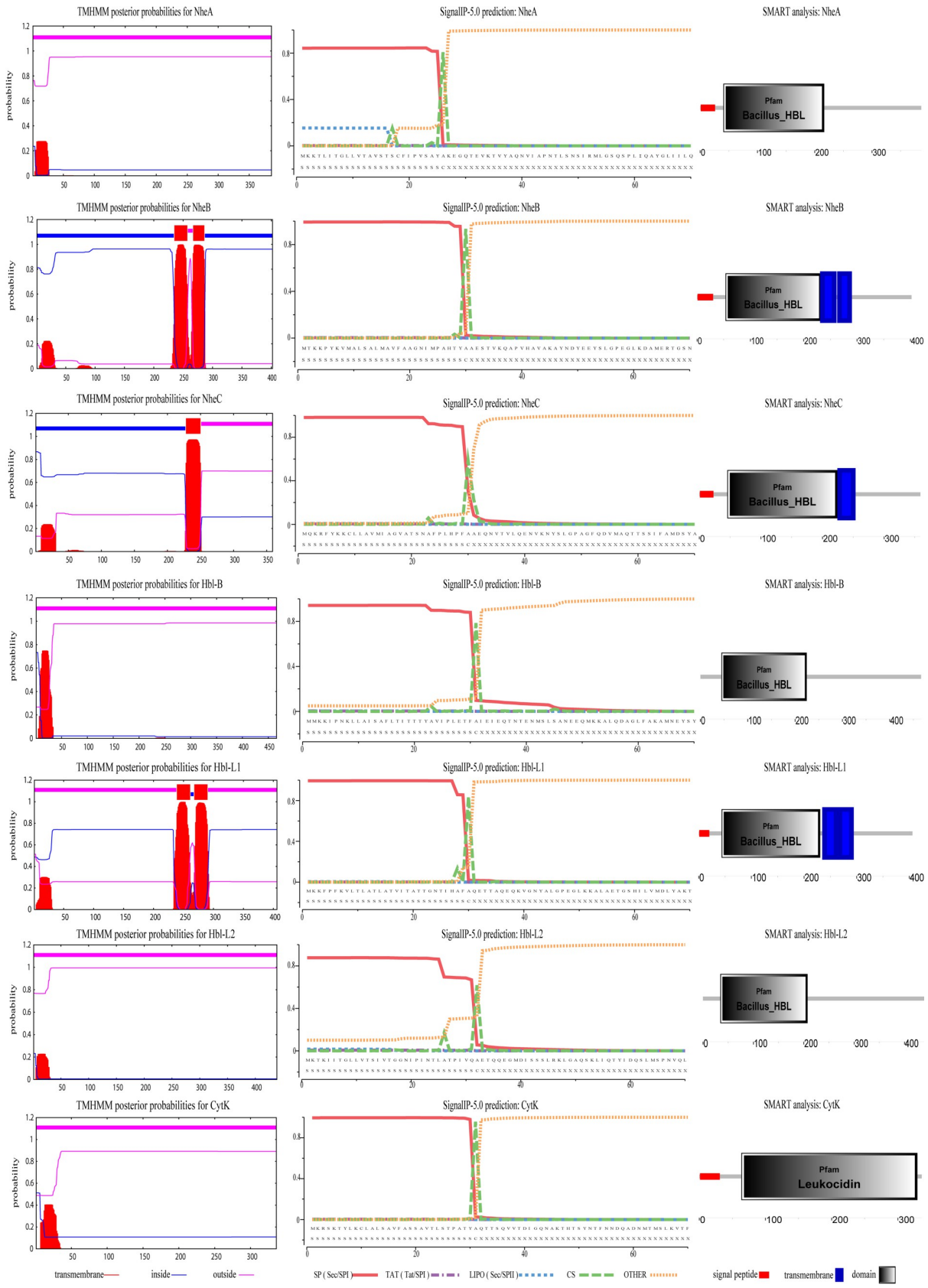
TMHMM posterior probabilities for NheA
SignalIP-5.0 prediction: NheA
SMART analysis: NheA
Pfam
Bacillus_HBL
TMHMM posterior probabilities for NheB
SignalIP-5.0 prediction: NheB
SMART analysis: NheB
Pfam
Bacillus_HBL
TMHMM posterior probabilities for NheC
SignalIP-5.0 prediction: NheC
SMART analysis: NheC
Pfam
Bacillus_HBL
TMHMM posterior probabilities for Hbl-B
SignalIP-5.0 prediction: Hbl-B
SMART analysis: Hbl-B
Pfam
Bacillus_HBL
TMHMM posterior probabilities for Hbl-L1
SignalIP-5.0 prediction: Hbl-L1
SMART analysis: Hbl-L1
Pfam
Bacillus_HBL
TMHMM posterior probabilities for Hbl-L2
SignalIP-5.0 prediction: Hbl-L2
SMART analysis: Hbl-L2
Pfam
Bacillus_HBL
TMHMM posterior probabilities for CytK
SignalIP-5.0 prediction: CytK
SMART analysis: CytK
Pfam
Leukocidin
probability
transmembrane
inside
outside
SP (Sec/SPI)
TAT (Tat/SPI)
LIPO (Sec/SPII)
CS
OTHER
signal peptide
transmembrane
domain

Fig 4. The locations of transmembrane helices, cleavage sites, signal peptides, and domain start-ends were predicted by TMHMM, SignalP, and SMART software with the ATCC14579 strain.

https://doi.org/10.1371/journal.pone.0262974.g004

cytotoxin K (encoded by *hblACD* and *cytK*) and nonhemolytic enterotoxin (encoded by *nheABC*) represent the significant enterotoxins produced by *B*. *cereus*. Cardazzo et al. detected horizontal gene transfer in the evolution of enterotoxins within *B*. *cereus* strains [43]. Our MLSA results showed that in the process of toxin molecular evolution, there were differences between the results for complete and incomplete virulence proteins, and two toxic-type genes had a more significant effect in relation to DVs than three toxic-type genes. The results suggested that the complete virulence-gene operon combination has higher relative genetic

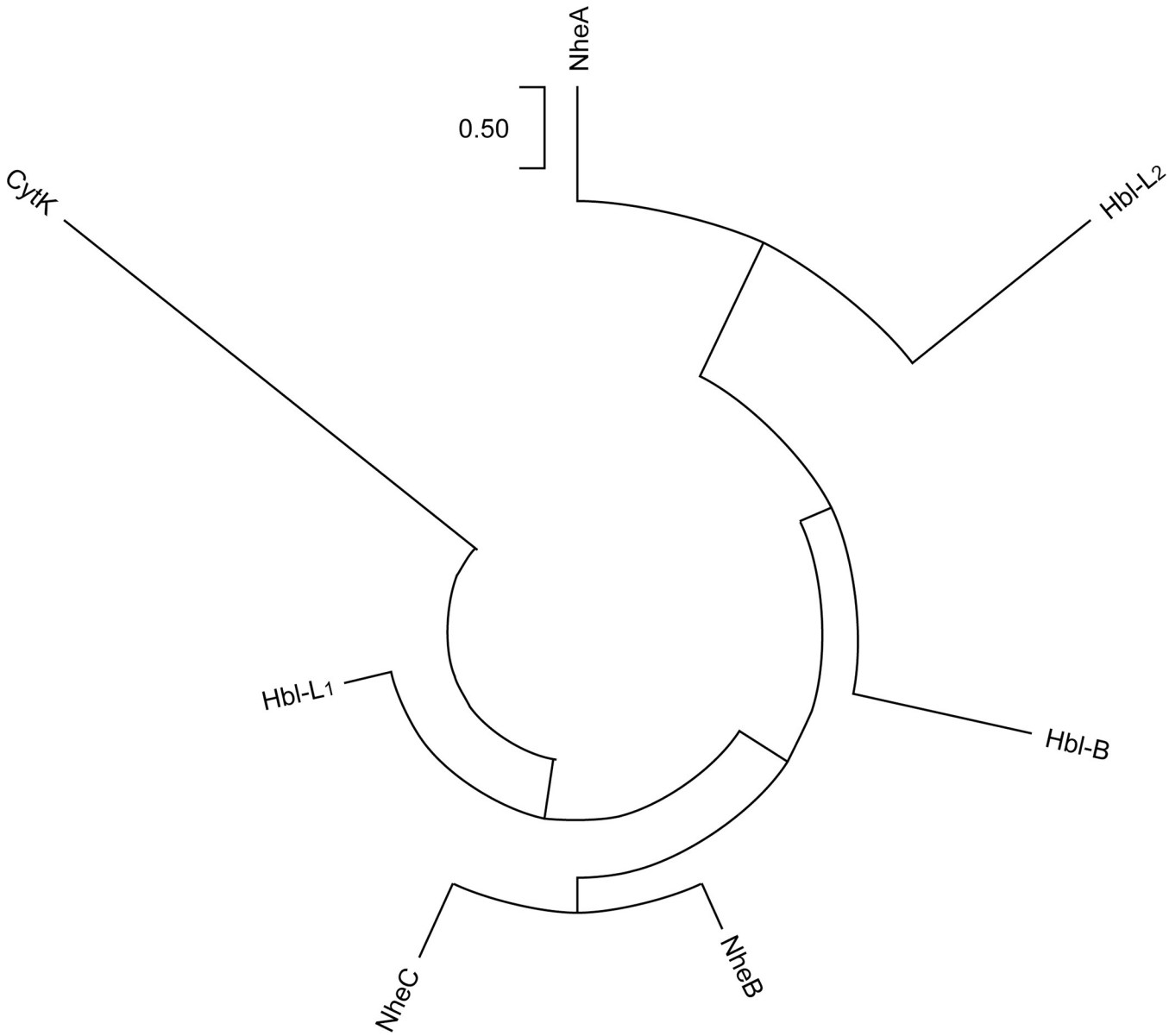

Fig 5. The phylogenetic tree of Hbl-B, Hbl-L_1, Hbl-L_2, NheA, NheB, NheC, and CytK component sequences with the ATCC14579 strain.

https://doi.org/10.1371/journal.pone.0262974.g005

Table 5. The evaluation parameters of the 3-D enterotoxin protein structures predicted by the SWISS-MODEL server and AlphaFold software with the ATCC14589 strain.

Name	Template number	Template description	Sequence identity	Sequence coverage	Sequence range	GMQE	AlphaFold plDDt	AlphaFold ipTM+pTM
NheA	4k1p.1.A	NheA	97.22%	0.93	42–386	0.88	93.86	-
NheB	7nmq.1.A	Hbl-L_1	40.82%	0.85	52–397	0.69	83.42	-
NheC	7nmq.1.A	Hbl-L_1	36.83%	0.88	44–358	0.68	90.25	-
Hbl-B	2nrj.1.A	Hbl-B	71.99%	0.71	33–364	0.61	91.73	-
Hbl-L_1	7nmq.1.A	Hbl-L_1	99.73%	0.90	41–405	0.88	81.90	-
Hbl-L_2	4k1p.1.A	NheA	24.85%	0.74	38–439	0.54	94.14	-
CytK	3yhd.1.A	Alpha-hemolysin	30.39%	0.84	35–334	0.61	89.75	-
Nhe-trimer	-	-	-	-	-	-	-	0.68
Hbl-trimer	-	-	-	-	-	-	-	0.36

The global model quality estimate (GMQE) is a quality estimate that combines properties from the target-template alignment and the template structure [38]. plDDt: Predicted local-distance difference test [29]; pTM: Predicted TM score; ipTM: Interface pTM; ipTM+pTM: Evaluation result of multimer prediction [34].

https://doi.org/10.1371/journal.pone.0262974.t005

stability. The DV of hemolysin Bl was greater than that of nonhemolytic cytotoxin K. *nheABC*, which was responsible for most of the cytotoxic activity of *B. cereus* isolates, showed stable, strictly vertical inheritance [44]. In contrast to *hbl*, duplication or deletion of *nhe*, which was

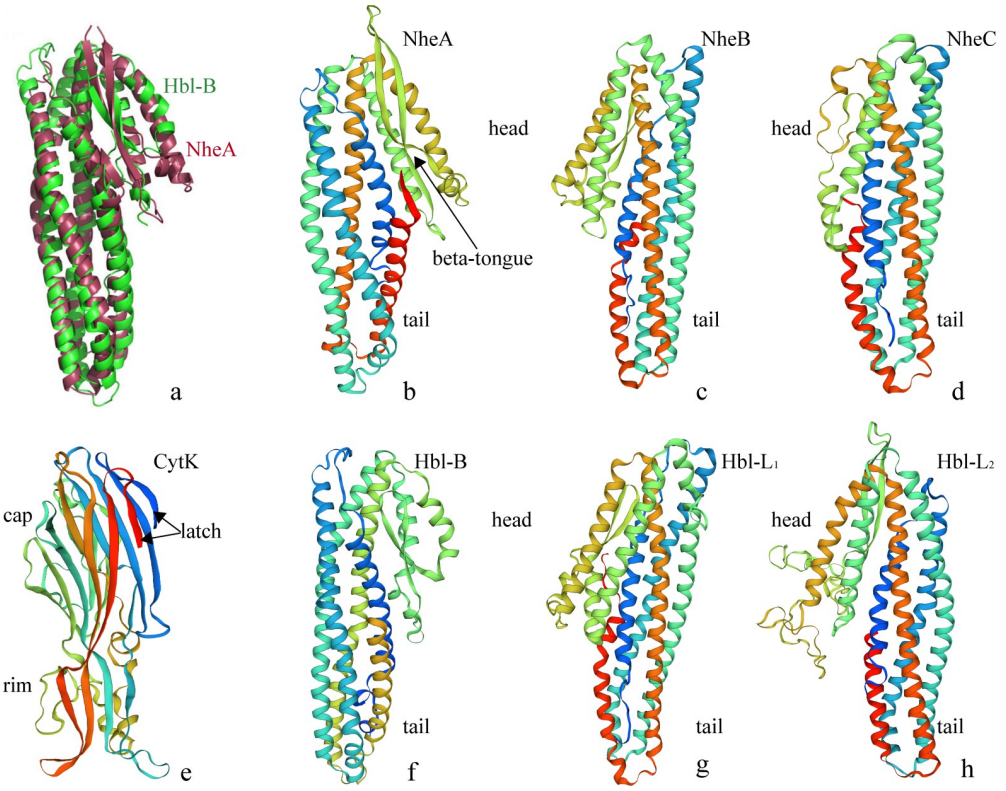

Fig 6. Overview of the structure predicted by the SWISS-MODEL server. a shows the superposition of the structures of Hbl-B (green) and NheA (burgundy) [40]. b, c, d, f, g, and h show the structures of NheA, NheB, NheC, Hbl-B, Hbl-L1 and Hbl-L_2, which are annotated with the 'head' and 'tail', respectively. b shows a beta-tongue in the 'head' region. e shows the structure of CytK, which is annotated with 'latch', 'cap' and 'rim'.

https://doi.org/10.1371/journal.pone.0262974.g006

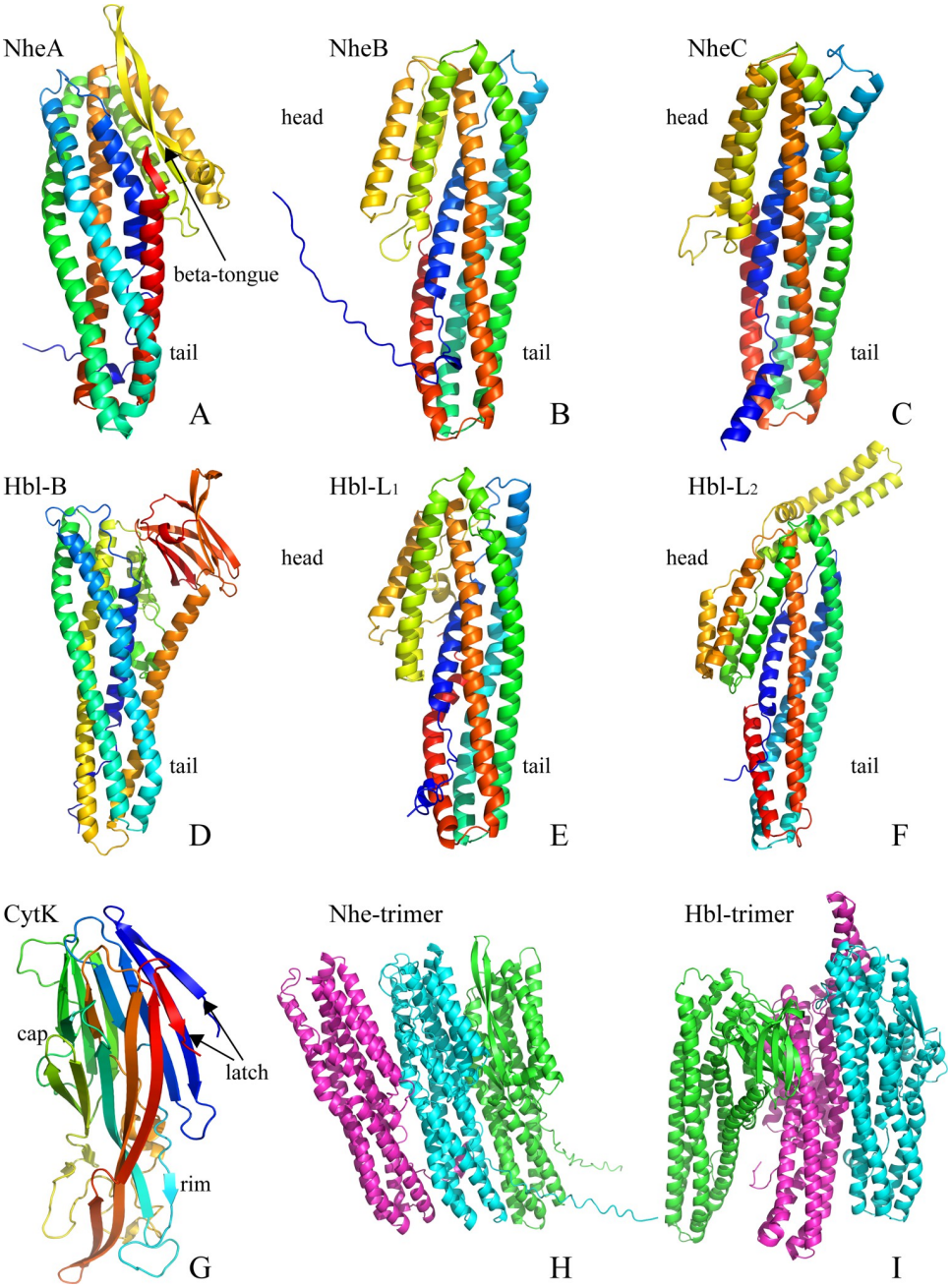

Fig 7. Overview of the structure predicted by AlphaFold software. A, B, C, D, E, and F are the structures of NheA, NheB, NheC, Hbl-B, Hbl-L_1, and Hbl-L_2, which are annotated with the 'head' and 'tail', respectively. A and D show beta-tongues in the 'head' region. G is the structure of CytK, which is annotated with 'latch', 'cap' and 'rim'. H and I show the trimers of the structures of NheA and Hbl-B (green), NheB and Hbl-L_2 (wathet), and NheC and Hbl-L_1 (pink).

https://doi.org/10.1371/journal.pone.0262974.g007

almost exclusively transmitted vertically, was rarely observed, and *cytK*, a one-type gene, had the highest relative genetic stability [15].

Bazinet revealed significant associations of particular genes with phenotypic traits shared by groups of taxa [41]. Currently, it is commonly accepted that the toxicity potential of *B.*

cereus is not driven by enterotoxin gene types because the expression of enterotoxin genes is highly complex and probably strain-specifically affected by transcription, posttranscriptional and posttranslational modification [45–47]. Carroll et al. suggested that further classification and descriptions of phenotypes should be added on the basis of genotype classification in *B. cereus* [2]. In this study, both Nhe and Hbl are three-component cytotoxins composed of binding components A and B and two lytic components B, C and -L_1, -L_2, with all three subunits acting synergically to cause illness. The amino acid sequences of all Nhe, Hbl and Cytk components containing N-terminal signal peptides indicated toxin secretion via the secretory translocation pathway. The final positions of Hbl-B, Hbl-L_1, NheB, NheC, and CytK were all extracellular and did not appear in the cytoplasm, and the two transmembrane regions of NheB and Hbl-L_1 might be responsible for transporting the assembled three-component cytotoxins across the membrane to complete the toxic effect. Dietrich et al. found that the factor triggering enterotoxin production under simulated intestinal conditions by various cell lines from different organisms and compartments was independent of cell differentiation [48]. Similarities were found when predicted transmembrane helices were compared. NheA and Hbl-B had no such helices, and NheB and Hbl-L_1 had two that may play an important role in molecular docking and transmembrane activities. Furthermore, the Nhe components seem to be additionally processed in the extracellular space after separation from the signal peptide for secretion [48]. The difference is that NheC had one such component and Hbl B had none, which may strengthen the secretion of the Nhe protein.

The Nhe and Hbl proteins share sequence similarities, both between the three components of each complex and between the two enterotoxin complexes [12]. The structural and functional properties were consistent with those of the superfamily of pore-forming cytotoxins of Hbl and Nhe [49–51]. The NheA, NheB, NheC, HBl-B, HBl-L_1, and HBl-L_2 structures showed two main domains, a 'head' and 'tail'. The 'heads' of NheA and Hbl-B, including two α-helices separated by β-tongue strands, play a special role in Nhe trimers and Hbl trimers, respectively. Upon contact with lipids, cell membranes or detergents, the protein oligomerizes and forms ring-shaped structures acting as transmembrane pores [52,53], and the hydrophobic β-tongue is assumed to be inserted into the membrane first [54]. It is worth noting that NheB, NheC, Hbl-L_1, and Hbl-L_2 had few or no β-strands in the 'head', which were either responsible for conformational changes of NheA and Hbl-B or for the stabilization of the 'head' domain [50] or might lead to reduced toxicity or only ligand function [41]. The difference was reflected in the triplet prediction results, i.e., a significant difference in structural arrangement compactness between Nhe trimers (noncompact type) and Hbl trimers (compact type). Ganash et al. speculated that the Nhe trimer requires interaction with unknown proteins of an additional function [41]. A specific binding order of the three Nhe and Hbl components is also necessary for pore formation [55,56]. We found that NheB and Hbl-L_1 were close to NheA and Hbl-B in the predicted trimer structure. Didier et al. found that NheA is important for attaching to cell-bound NheB and NheC and that NheB is the main interaction partner of NheA [57], and a further correlation was found for the amounts of Hbl B and Hbl L_1 [46]. Cytotoxin K is a single protein with β-barrel pore-forming toxin in contrast to the tripartite toxin complexes Hbl and Nhe. The CytK structure, which exhibits two 'latches' with many β-sheets folded beside the 'cap' domain forming a β-barrel, was the pore structure on top of the conformation. The 'rim' region, which was folded into a three-stranded antiparallel β-sheet, balanced the structure in the monomer. The predicted structure revealed that CytK was likely to belong to the leukocyte toxin family. These monomers diffuse to target cells and are attached to them by specific receivers [58], which are lipids and proteins that cause lysis of red blood cells by destroying their cell membrane [59].

Conclusion

In this study, we describe the molecular evolution, function and structural diversity of virulence factors in *B*. *cereus* strains. The evolution of *B*. *cereus* strains showed a clustering trend based on the coding virulence genes. The complete virulence gene operon combination had higher relative genetic stability than the incomplete operon. The two α-helices in the 'head' of the NheA and Hbl-B structures, which are separated by β-tongue strands, and two 'latches' with many β-sheets folded beside the 'cap' of the CytK structure might play a special role in the binding of virulence structures and pore-forming toxins in *B*. *cereus*. Overall, the exact mechanism by which *B*. *cereus* causes diarrhea remains unknown, but our results provide helpful information for better understanding the taxonomically diverse distribution of virulence factors in *B*. *cereus* strains.

Author Contributions

Conceptualization: Ming Zhang.

Formal analysis: Jun Liu.

Funding acquisition: Ming Zhang.

Methodology: Zhenzhen Yin.

Resources: Ming Zhang.

Writing – original draft: Ming Zhang.

Writing – review & editing: Li Zhang.

References

1. Liu Y, Du J, Lai Q, Zeng RY, Ye DZ, Xu J, et al. Proposal of nine novel species of the *Bacillus cereus* group. Int J syst Evol Microbiol. 2017; 67: 2499–2508. https://doi.org/10.1099/ijsem.0.001821 PMID: 28792367
2. Carroll LM, Wiedmann M, Kovac J. Proposal of a taxonomic nomenclature for the *Bacillus cereus* group which reconciles genomic definitions of bacterial species with clinical and industrial phenotypes. mBio. 2020; 11(1): e00034–20. https://doi.org/10.1128/mBio.00034-20 PMID: 32098810
3. Webb MD, Barker GC, Goodburn KE, Peck MW. Risk presented to minimally processed chilled foods by psychrotrophic *Bacillus cereus*. Trends Food Sci Technol. 2019; 93: 94–105. https://doi.org/10.1016/j.tifs.2019.08.024 PMID: 31764911
4. Bhunia AK. Foodborne microbial pathogens: Mechanisms and pathogenesis. Springer, New York; 2008.
5. Rajkovic A, Jovanovic J, Monteiro S, Decleer M, Andjelkovic M, Foubert A, et al. Detection of toxins involved in foodborne diseases caused by Gram-positive bacteria. Compr Rev Food Sci Food Saf. 2020; 19(4):1605–1657. https://doi.org/10.1111/1541-4337.12571 PMID: 33337102
6. Jessberger N, Dietrich R, Schwemmer S, Tausch F, Schwenk V, Didier A, et al. Binding to the target cell surface is the crucial step in pore formation of hemolysin BL from *Bacillus cereus*. Toxins. 2019; 11 (5): 281–297. https://doi.org/10.3390/toxins11050281 PMID: 31137585
7. Koné KM, Douamba Z, Halleux MD, Bougoudogo F, Mahillon J. Prevalence and diversity of the thermotolerant bacterium *Bacillus* cytotoxicus among dried food products. J Food Protec. 2019; 82(7): 1210–1216. https://doi.org/10.4315/0362-028X.JFP-19-006 PMID: 31233363
8. Thompson CC, Chimetto L, Edwards RA, Swings J, Stackebrandt E, Thompson FL. Microbial genomic taxonomy. BMC Genomics. 2013; 14: 913. https://doi.org/10.1186/1471-2164-14-913 PMID: 24365132
9. Liu Y, Lai QL, Göker M, Meier-Kolthoff JP, Wang M, Sun YM, et al. Genomic insights into the taxonomic status of *Bacillus cereus* group. Sci Rep 2015; 5: 14082. https://doi.org/10.1038/srep14082 PMID: 26373441

10. Guinebretière MH, Thompson FL, Sorokin A, Normand P, Dawyndt P, Ehling-Schulz M, et al. Ecological diversification in the *Bacillus cereus* Group. Environ Microbiol. 2008; 10(4): 851–865. https://doi.org/10.1111/j.1462-2920.2007.01495.x PMID: 18036180

11. Helgason E, Økstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, et al. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*-one species on the basis of genetic evidence. Appl Environ Microbiol. 2000; 66(6): 2627–2630. https://doi.org/10.1128/AEM.66.6.2627-2630.2000 PMID: 10831447

12. Dietrich R, Jessberger N, Ehling-Schulz M, Märtlbauer E, Granum PE. The food poisoning toxins of *Bacillus cereus*. Toxins. 2021; 13(2): 98. https://doi.org/10.3390/toxins13020098 PMID: 33525722

13. Lapidus A, Goltsman E, Auger S, Galleron N, Ségurens B, Dossat C, et al. Extending the *Bacillus cereus* group genomics to putative food-borne pathogens of different toxicity. Chem. Interact. 2008; 171(2): 236–249. https://doi.org/10.1016/j.cbi.2007.03.003 PMID: 17434157

14. Zheng J, Guan Z, Cao S, Peng D, Ruan L, Jiang D, Sun M. Plasmids are vectors for redundant chromosomal genes in the *Bacillus cereus* group. BMC Genom. 2015; 16(1): 1–10. https://doi.org/10.1186/s12864-014-1206-5 PMID: 25608745

15. Böhm ME, Huptas C, Krey VM, Scherer S. Massive horizontal gene transfer, strictly vertical inheritance and ancient duplications differentially shape the evolution of *Bacillus cereus* enterotoxin operons hbl, cytK and nhe. BMC Evol Biol. 2015; 15(1): 246. https://doi.org/10.1186/s12862-015-0529-4 PMID: 26555390

16. Didelot X, Barker M, Falush D, Priest FG. Evolution of pathogenicity in the *Bacillus cereus* group. Syst Appl Microbiol. 2009; 32(2): 81–90. https://doi.org/10.1016/j.syapm.2009.01.001 PMID: 19200684

17. Torres Manno MA, Repizo GD, Magni C. Dunlap CA, Espariz M. The assessment of leading traits in the taxonomy of the *Bacillus cereus* group. Antonie van Leeuwenhoek. 2020; 113: 2223–2242. https://doi.org/10.1007/s10482-020-01494-3 PMID: 33179199

18. Kovac J, Miller RA, Carroll LM, Kent DJ, Jian JH, Beno SM, et al. Production of hemolysin BL by *Bacillus cereus* group isolates of dairy origin is associated with whole-genome phylogenetic clade. BMC genomics. 2016; 17: 581. https://doi.org/10.1186/s12864-016-2883-z PMID: 27507015

19. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015; 25 (7):1043–1055. https://doi.org/10.1101/gr.186072.114 PMID: 25977477

20. Zuo G, Hao B. CVTree3 web server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy. Genom Proteom Bioinf. 2015; 13: 321–331. https://doi.org/10.1016/j.gpb.2015.08.004 PMID: 26563468

21. Zuo G. CVTree: a parallel alignment-free phylogeny and taxonomy tool based on composition vectors of genomes. Genom Proteom Bioinf. 2021; 19: 1–6. https://doi.org/10.1016/j.gpb.2021.03.006 PMID: 34119695

22. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018; 35(6): 1547–1549. https://doi.org/10.1093/molbev/msy096 PMID: 29722887

23. Richter M, Rosselló-Móra R, Glöckner FO, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics. 2016; 32(6): 929–931. https://doi.org/10.1093/bioinformatics/btv681 PMID: 26576653

24. Perrière G, Gouy M. WWW-query: an on-line retrieval system for biological sequence banks. Biochimie. 1996; 78(5): 364–369. https://doi.org/10.1016/0300-9084(96)84768-7 PMID: 8905155

25. Baum BR. PHYLIP: phylogeny inference package. version 3.2. Joel Felsenstein. Biology. 1989; 64(4): 539–541. https://doi.org/10.1086/416571

26. Tourasse NJ, Helgason E, Økstad OA, Hegna IK, Kolst AB. The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. J Appl Microbiol. 2006; 101(3): 579–593. https://doi.org/10.1111/j.1365-2672.2006.03087.x PMID: 16907808

27. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013; 30(12): 2725–2729. https://doi.org/10.1093/molbev/mst197 PMID: 24132122

28. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. 2021; 49(D1): D458–D460. https://doi.org/10.1093/nar/gkaa937 PMID: 33104802

29. Peabody MA, Laird MR, Vlasschaert C, Lo R, Brinkman FSL. PSORTdb: expanding the bacteria and archaea protein subcellular localization database to better reflect diversity in cell envelope structures. Nucleic Acids Res. 2016; 44(D1): D663–D668. https://doi.org/10.1093/nar/gkv1271 PMID: 26602691

30. Möller S, Croning MDR, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics. 2001; 17(7): 646–653. https://doi.org/10.1093/bioinformatics/17.7.646 PMID: 11448883

31. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019; 37(4): 420–423. https://doi.org/10.1038/s41587-019-0036-z PMID: 30778233

32. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014; 42: 252–258. https://doi.org/10.1093/nar/gku340 PMID: 24782522

33. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596(7873): 583–589. https://doi.org/10.1038/s41586-021-03819-2 PMID: 34265844

34. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv preprint. 2021 October 4. https://doi.org/10.1101/2021.10.04.463034

35. Thitiprasert S, Piluk J, Tolieng V, Tanaka N, Shiwa Y, Fujita N, et al. Draft genome sequencing of *Sporolactobacillus terrae* SBT-1, an efficient bacterium to ferment concentrated sugar to d-lactic acid. Arch Microbiol. 2021; 203: 3577–3590. https://doi.org/10.1007/s00203-021-02352-0 PMID: 33961074

36. Sugawara T, Yamashita D, Kato K, Zhao P, Ueda J, Kaneko J, et al. Structural basis for pore-forming mechanism of staphylococcal α-hemolysin. Toxicon. 2015; 108: 226–231. https://doi.org/10.1016/j.toxicon.2015.09.033 PMID: 26428390

37. Sharma A, Ponmariappan S, Sarita R, Alam SI, kamboj DV, Shukla S. Identification of cross reactive antigens of *C. botulinum* types A, B, E & F by immunoproteomic approach. Curr Microbiol. 2018; 75(5): 531–540. https://doi.org/10.1007/s00284-017-1413-9 PMID: 29332140

38. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics. 2013; 29(21): 2722–2728. https://doi.org/10.1093/bioinformatics/btt473 PMID: 23986568

39. Takaki S, Daichi Y, Koji K, Zhao P, Junki U, Jun K, et al. Structural basis for pore-forming mechanism of staphylococcal α-hemolysin. Toxicon. 2015; 108: 226–231. https://doi.org/10.1016/j.toxicon.2015.09.033 PMID: 26428390

40. Ganash M, Phung D, Sedelnikova SE, Lindbäck T, Granum PE, Artymiuk PJ. Structure of the NheA component of the Nhe toxin from *Bacillus cereus*: Implications for function. PLoS ONE. 2013; 8(9): e74748. https://doi.org/10.1371/journal.pone.0074748 PMID: 24040335

41. Bazinet AL. Pan-genome and phylogeny of *Bacillus cereus sensu lato*. BMC Evol Biol. 2017; 17:176. https://doi.org/10.1186/s12862-017-1020-1 PMID: 28768476

42. Phelps RJ, McKillip JL. Enterotoxin production in natural isolates of *Bacillaceae* outside the *Bacillus cereus* group. Appl Environ Microbiol. 2002; 68(6): 3147–3151. https://doi.org/10.1128/AEM.68.6.3147-3151.2002 PMID: 12039781

43. Cardazzo B, Negrisolo E, Carraro L, Alberghini L, Patarnello T, Giaccone V. Multiple-locus sequence typing and analysis oftoxin genes in *Bacillus cereus* food-borne isolates. Appl Environ Microbiol. 2008; 74(3): 850–860. https://doi.org/10.1128/AEM.01495-07 PMID: 18083872

44. Moravek M, Dietrich R, Buerk C, Broussolle V, Guinebretière MH, Granum PE, et al. Determination of the toxic potential of *Bacillus cereus* isolates by quantitative enterotoxin analyses. FEMS Microbiol Lett. 2006; 257(2): 293–298. https://doi.org/10.1111/j.1574-6968.2006.00185.x PMID: 16553866

45. Dietrich R, Moravek M, Bürk C, Granum PE, Märtlbauer E. Production and characterization of antibodies against each of the three subunits of the *Bacillus cereus* nonhemolytic enterotoxin complex. Appl Environ Microbiol. 2005; 71(12): 8214–8220. https://doi.org/10.1128/AEM.71.12.8214-8220.2005 PMID: 16332805

46. Jeßberger N, Dietrich R, Bock S, Didier A, Märtlbauer E. *Bacillus cereus* enterotoxins act as major virulence factors and exhibit distinct cytotoxicity to different human cell lines. Toxicon. 2014; 77(1): 49–57. https://doi.org/10.1016/j.toxicon.2013.10.028 PMID: 24211313

47. Jeßberger N, Krey VM, Rademacher C, Böhm ME, Mohr AK, Ehling-Schulz M, et al. From genometo toxicity: a combinatory approach highlights the complexity of enterotoxin production in *Bacillus cereus*. Front. Microbiol. 2015; 6(6): 560. https://doi.org/10.3389/fmicb.2015.00560 PMID: 26113843

48. Dietrich R, Jeßberger N, Ehling-Schulz M, Märtlbauer E, Granum PE. The food poisoning toxins of *Bacillus cereus*. Toxins. 2021; 13(2): 98. https://doi.org/10.3390/toxins13020098 PMID: 33525722

49. Fagerlund A, Lindbäck T, Storset AK, Granum PE, Hardy SP. *Bacillus cereus* Nhe is a pore-forming toxin with structural and functional properties similar to the ClyA (HlyE, SheA) family of haemolysins, able to induce osmotic lysis in epithelia. Microbiol. 2008; 154 (Pt 3): 693–704. https://doi.org/10.1099/mic.0.2007/014134-0 PMID: 18310016

50. Madegowda M, Eswaramoorthy S, Burley SK, Swaminathan S. X-ray crystal structure of the B component of Hemolysin BL from *Bacillus cereus*. Proteins Struct Funct & Bioinform. 2008: 71(2): 534–540. https://doi.org/10.1002/prot.21888 PMID: 18175317

51. Phung D, Ganash M, Sedelnikova SE, Lindbäck T, Granum PE, Artymiuk PJ. Crystallization and preliminary crystallographic analysis of the NheA component of the Nhe toxin from *Bacillus cereus*. Acta Cryst. 2012; 68 (Pt 9): 1073–1076. https://doi.org/10.1107/S1744309112030813 PMID: 22949198

52. Wallace AJ, Stillman TJ, Atkins A, Jamieson SJ, Bullough PA, Green J, et al. *E. coli* Hemolysin E (HlyE, ClyA, SheA). Cell. 2000; 100 (2): 265–276. https://doi.org/10.1016/s0092-8674(00)81564-0 PMID: 10660049

53. Eifler N, Vetsch M, Gregorini M, Ringler P, Chami M, Philippsen A, et al. Cytotoxin ClyA from *Escherichia coli* assembles to a 13-meric pore independent of its redox-state. EMBO J. 2006; 25 (11): 2652–2661. https://doi.org/10.1038/sj.emboj.7601130 PMID: 16688219

54. Mueller M, Grauschopf U, Maier T, Glockshuber R, Ban N. The structure of a cytolytic α-helical toxin pore reveals its assembly mechanism. Nature. 2009: 459: 726–730. https://doi.org/10.1038/nature08026 PMID: 19421192

55. Worthy HL, Williamson LJ, Auhim HS, Leppla SH, Sastalla I, Jones DD, et al. The crystal structure of *Bacillus cereus* $HblL_1$. Toxins. 2021; 13(4): 253–266. https://doi.org/10.3390/toxins13040253 PMID: 33807365

56. Lindbäck T, Fagerlund A, Rødland MS, Granum PE. Characterization of the *Bacillus cereus* Nhe enterotoxin. Microbi. 2004; 150(Pt 12): 3959–3967. https://doi.org/10.1099/mic.0.27359-0 PMID: 15583149

57. Didier A, Dietrich R, Märtlbauer E. Antibody binding studies reveal conformational flexibility of the *Bacillus cereus* nonhemolytic enterotoxin (Nhe) A-component. PLoS One. 2016; 11(10): e0165135. https://doi.org/10.1371/journal.pone.0165135 PMID: 27768742

58. Thompson JR, Cronin B, Bayley H, Wallace MI. Rapid assembly of a multimeric membrane protein pore. Biophys J. 2011; 101(11): 2679–2683. https://doi.org/10.1016/j.bpj.2011.09.054 PMID: 22261056

59. Stipcevic T, Piljac T, Isseroff RR. Di-rhamnolipid from *Pseudomonas aeruginosa* displays differential effects on human keratinocyte and fibroblast cultures. J Dermatol Sci. 2005; 40(2): 141–143. https://doi.org/10.1016/j.jdermsci.2005.08.005 PMID: 16199139