Genetic and genomic diversity of NheABC locus from *Bacillus* strains

Yan Cai² · Tingxuan Huang³ · Yuekang Xu¹ · Guoping Zhou³ · Ping Zou² · Guifeng Zeng² · Xiaojin Liu¹

Received: 5 December 2016 / Revised: 13 January 2017 / Accepted: 27 January 2017 / Published online: 10 March 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract Non-hemolytic enterotoxin (NHE), a tri-partite, proteinaceous toxin encoded by contiguous *nheA, nheB* and *nheC* genes of *Bacillus cereus sensu lato* (*B. cereus* s.l.), is considered to be associated with the foodborne diarrheic syndrome. However, *B. cereus* s.l. includes a number of closely related strains, and the occurrence of NHE among them, and other members of *Bacillus* is unclear. Consequently, we aimed to determine the distribution and evolution of NHE within *Bacillus* by confirming the presence of the *nheA, B* and *C* sequences and variation within them using published data, and to analyze the genomic and genetic diversity. The phylogenetic tree of NHE proteins (*NheA, NheB, and NheC*) from 81 different *B. cereus* s.l. strains was constructed. And on the genetic determinants of the NHE toxin did not bring any obvious link between the *nheABC* genes sequence of a strain and its virulence in the diarrhoeal pathogenesis. Analysis of the genomic diversity of the *nheA, B* and *C* loci revealed that their upstream regions were more conserved than the downstream sequences. Multilocus sequence typing schemes (MLST) based on seven concatenated housekeeping genes and *nheA, B* and *C* genes of the 75 strains were developed. The neighbor joining phylogenetic tree based on seven housekeeping genes together with *nheA, B* and *C* genes was similar to published *Bacillus* phylogenetic trees. And on the genetic determinants of the NHE toxin did not bring any obvious link between the *nheABC* genes sequence of a strain and its virulence in the diarrhoeal pathogenesis. The results indicate that *nheA, B* and *C* genes do not affect the diversity of housekeeping genes, and this specific NHE protein does not participate in the diarrheic syndrome.

Keywords *Bacillus* · *NheABC* · Genetic diversity · MLST

Introduction

*Bacillus cereus* s.l. is widely distributed in food, soil and plants (Okinaka and Keim 2016). This Gram-positive, spore-forming bacterium may behave as an opportunistic human pathogen. Long known to be responsible for two forms of food poisoning, characterized by either diarrhea or nausea and vomiting. The diarrheal symptom includes the following symptoms, which usually last generally less than 48 h: abdominal pain, profuse watery diarrhea, sometimes nausea and vomiting within 8–16 h. Although most cases are generally mild, more serious and even lethal cases have been reported in Europe.

Concerning the diarrhoeal syndrome, no definitive hypothesis that correlates the symptoms to a unique component exists to date. Indeed, several putative enterotoxins have been reported to be potentially responsible, alone or in combination, of the diarrhoeic pathotypes. These include...
the tripartite enterotoxins as Haemolysin BL (HBL) and Nonhaemolytic enterotoxin (NHE), but also the single-component toxin, Cytotoxin K (CytK) also sometimes named Haemolysin IV (HlyIV). In addition to these three major candidates, other molecules are also cited as potential enterotoxins involved in the diarrhoeic syndrome, such as enterotoxin FM (EntFM) (Boonchai et al. 2008), enterotoxin S (entS), enterotoxin-T (bceT) and pore-forming haemolysins like the Cereolysin O (CerO), Haemolysin II (HemII) and Haemolysin III (HlyIII). Besides, other virulence factors seem to contribute to the B. cereus foodborne diseases, such as phosphoholipases or the sphingomyelinase (SMase).

The non-hemolytic enterotoxin NHE is encoded by nheA, nheB and nheC (Kim et al. 2015). NHE is a tripartite pore-forming toxin that requires the combination of three proteins: NheA, NheB and NheC. NHE was first isolated from the supernatant of a B. cereus s.l. culture that caused a large food poisoning outbreak in Norway in 1995 (Lund and Granum 1996). NHE proteins are secreted independently and maximal toxic activity on Vero cells requires all the three parts in a molar ratio 10:10:1 of NheA, NheB and NheC, respectively. NheB is the binding component of the enterotoxin complex and an increase in the concentration of NheC results in a decrease in Nhe toxic activity (Lindbäck et al. 2004).

However, the genes coding for these putative enterotoxins are, for the most part, largely distributed among the B. cereus group isolates, irrespective of their diarrheic activities (McIntyre et al. 2008; Swiecicka et al. 2006). Moreover, with the exception of the rabbit ileum assay, no animal model can be used to specifically test for the diarrheal properties of the strains (or for purified proteins) (Beecher et al. 1995). Only classical assays on animal cell lines are readily available, but they only give information on the generic cytotoxicity of these putative enterotoxins (Jeßberger et al. 2014).

B. cereus s.l. includes a number of closely related species, viz., Bacillus cereus sensu stricto, Bacillus mycoides, Bacillus pseudomycoides, Bacillus thuringiensis, Bacillus weihenstephanensis, Bacillus anthracis and Bacillus cytoxicus, and many strains thereof (Lindbäck et al. 2004). The occurrence of NHE within the group, and other members of Bacillus is unclear, e.g., we have found NHE in some non-B. cereus s.l. members. Analysis methods such as MLST indexes the sequence variation present in a small number (usually seven) of housekeeping gene fragments located around the bacterial genome, is to provide a highly discriminating typing system that can be particularly helpful for the typing of bacterial pathogens (Keith and Martin 2014). And PHYLOViZ software make the data easy to be visualized and export the results in graphic formats (Alexandre et al. 2012). In this study, we were interested in determining the particular strains that produce NHE and to trace the molecular evolution and variation of the nheABC genes and the phylogenetic relationship of the various Bacillus cereus s.l. strains to others within Bacillus. Our approach involved sequence-based typing analysis, interrogation of online databases of allelic profiles and associated epidemiological data collected.

With the aim of assessing the potential implication of NHE in the diarrheic syndrome, the genomic and genetic diversity, as well as the occurrence and the evolutionary ecology of nheABC were studied in detail on a collection of Bacillus cereus strains.

Materials and methods

Bioinformatics and strains information

The presence of nheABC genes and their corresponding putative sequences was screened in the NCBI database among the 174 genomes from the B. cereus group as available (Aug. 15, 2015). The NheA, B and C protein sequences from B. cereus ATCC14579 were applied to find all the homologous proteins among the 174 strains by BLAST program in http://blast.ncbi.nlm.nih.gov/Blast.cgi.

A set of strains (Table 1) coming from food poisonings (FP), food products (F), the environment (E), clinical cases (C) and unknown origin (U) was collected to investigate the presence of nheABC genes. Then, a panel of positive strains was selected for MLST analysis (Table 2 and below). This selection was based on the origin, the year and the country of strains and the species to obtain the most diversified panel of nheABC positive strains.

MLST analysis

Seven loci encoding housekeeping genes were chosen for MLST analysis: glp (glycerol kinase), gmk (guanylate kinase), ilv (isoleucine-valine), pta (phosphate acetyltransferase), pur (purine synthesis), pyc (pyruvate carboxylase), tpi (triosephosphate isomerase) (Lampe and English 2016). 75 strains for which the nheABC genes are available in the genome databases, were selected for the MLST analysis (Table 2). The MLST result were shown in minimum-spanning tree (http://pubmlst.org/analysysis/). The nucleotide sequence diversity of glp, gmk, ilv, pta, pur, pyc, tpi and nheABC genes was analyzed at two levels: by constructing a phylogenetic tree for each loci with the CLC Main Workbench 7 software (CLCbio, a Qiagen Company) using the neighbor joining (NJ) algorithm with Jukes Cantor as substitution rate model, and by building sequence types (ST) of the various strains using the non-redundant database (NRDB) for allele comparison (http://pubmlst.org/).
Table 1  The occurrence of nheABC in all the 92 B. cereus s.l. group strains whose whole genome sequence available online

| Source | nheABC+ | nheABC- | Positive ratio |
|--------|---------|---------|----------------|
| C      | Ba A012, Ba Ames, Ba A0193, Ba Sterne, Ba A0389, Ba A0174, Ba A0442, Ba A0465, Ba A0488, Ba A1055, Ba Australia 94, Ba CNEVA-9066, Ba Kruger B, Ba Volum, Ba WNA USA6153, Ba ‘Ames Ancestor’, Ba 03BB108, Ba 03BB102, Ba G9842, Ba AH1272, Ba AH1273, Ba R309803, Ba AH820, Ba 84264, Ba 172560W, Ba 95/8201, Ba F65185, Ba AH1134, Ba G9241, Ba IBL 200 | Ba A0248 | 96.8% (30/31) |
| F      | Bc AH603, Bc MM3, Bc m1293, Bc NVH0597-99, Bc m1550 | | |
| FP     | Bc ATCC 10,987, Bc AH187, Bc W, B. cytotoxicus NVH 391-98 | | |

E environment, F food, FP food poisoning, C clinical isolations, U unknown source

To cluster the strains according to their ST, BURST analysis was performed which defines a group when at least 5/7 loci were identical (http://pubmlst.org/perl/mlstanalyse/). Sequences of the seven chromosomal loci were also concatenated with or without adding the nheABC genes to construct and compare the respective phylogeny trees generated by CLC Main Workbench 7 software using UPGMA algorithm and Kimura 80 mathematical model using. The correctness of the results was evaluated using a 100-step bootstrap test (Virginie et al. 2015).

Results and discussion

nheABC occurrence among the B. cereus s.l. strains

While the mechanisms involved in the B. cereus diarrheal pathogenesis are still largely unknown, a large panel of enterotoxins have been designated as potential causative agents led to this syndrome. And NHE is regularly cited as a candidate because of its cytotoxic, necrotic and haemolytic activities on human intestinal cell lines (Lindback et al. 2004; Zhu et al. 2016). Some B. cereus genomes harbor the nhe operon which codes for the cytolytic protein NheA and the binding components NheB and NheC (Wehrle et al. 2009). In this study, the nheABC sequence from B. cereus ATCC14579 were applied to find all the homologous genes among the 174 strains by BLAST program in http://blast.ncbi.nlm.nih.gov/Blast.cgi. Since the first bacterial genome sequence was completed in 1995, 174 sequences of Bacillus strains genomes have been published in NCBI database until Aug. 2015 (the number of assembled and annotated B. cereus genome is 294 until Dec. 2016). The nheABC loci were found in 81 strains, including all the 31 B. anthracis strains, 24 B. cereus strains, 21 B. thuringiensis strains, 2 B. welshenstephanensis strains, 1 Bacillus bombaysepticus strain and 1 Bacillus toyonensis strain. 2 B. mycoides strains were sequenced but only one strain contains nheABC loci.

However, Bacillus anthracis, Bacillus clausii, Bacillus coagulans, Bacillus infantis, Bacillus lehensis, Bacillus licheniformis, Bacillus megaterium, Bacillus methanolicus, Bacillus methylotrophicus, Bacillus pseudefirmus, Bacillus pumilus, Bacillus sp., Bacillus subtilis, Bacillus halodurans, Geobacillus kaustophilus and Bacillus cellulosilyticus do not contain nheABC loci (Table 1). The nheABC operon occurrence in NCBI database is 46.6% (81/174). And the mean value frequency of nheA, nheB, nheC found in the literature is 82, 81 and 78%, respectively (Swiecicka et al. 2006; Gaviria et al. 2000; Hansen and Hendriksen 2001; Banerjee et al. 2011; Moravek et al. 2004; De Jonghe et al. 2010; Krause et al. 2010; Zhou et al. 2010; Samapundo et al. 2011; Chon et al. 2015). These frequencies are slightly higher than the ratio we measured from database. Maybe because more and more Bacillus genomes sequencing are completed, and nheABC operon is absent from the recently released genomes, so the ratio measured from database is low.

Genomic diversity of nheABC

81 genetic regions (30 kb in size) from the nheABC+ strains mentioned above centered on the nheABC operon
Table 2 Origin, typing and BURST-grouping data of *B. cereus*s. *strains*

| Group | Strain                        | Country | Type of sample | ST | glp | gmk | ilv | pta | pur | pyc | tpi |
|-------|-------------------------------|---------|----------------|----|-----|-----|-----|-----|-----|-----|-----|
| 1     | *B. cereus* ATCC4342          | USA     | F              | 38 | 24  | 12  | 50  | 21  | 23  | 31  | 19  |
|       | *B. anthracis* A0248          | Unknown | C              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. cereus* 03BB87            | USA     | C              | 78 | 24  | 22  | 33  | 37  | 34  | 34  | 5   |
|       | *B. cereus* 03BB108           | USA     | U              | 62 | 38  | 1   | 32  | 1   | 18  | 33  | 24  |
|       | *B. thuringiensis* 97-27      | Sarajevo| C              | 113| 62  | 1   | 57  | 52  | 55  | 37  | 43  |
|       | *B. cereus* 03BB102           | USA     | C              | 11 | 34  | 1   | 32  | 1   | 33  | 37  | 24  |
|       | *B. cereus* g9241            | Unknown | C              | 78 | 24  | 22  | 33  | 37  | 34  | 34  | 5   |
|       | *B. cereus* C1                | Ivory Coast | E    | 935| 34  | 1   | 83  | 1   | 18  | 29  | 5   |
|       | *B. anthracis* C4ac02         | China   | U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* PAK-1          | Pakistan| U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* Vollum         | Unknown | U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* 2000031021     | USA     | U              | 933| 65  | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* HYU01          | South Korea | U    | 3  | 2   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* SVA11          | Unknown | C              | 3  | 2   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* A16            | Unknown | C              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* H9401          | South Korea | C    | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* Sterne chromosome | Unknown | U    | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* RA3            | France  | U              | 3  | 2   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* V770-NP-1R     | USA     | U              | 2  | 1   | 1   | 2   | 1   | 1   | 1   | 2   |
|       | *B. anthracis* BA1035         | South Africa | U    | 3  | 2   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* BA1015         | USA     | U              | 2  | 1   | 1   | 2   | 1   | 1   | 2   | 1   |
|       | *B. anthracis* Sterne         | Unknown | E              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* Pasteur        | Unknown | U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* SK-102         | USA     | U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* Ohio ACB       | USA     | U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* K3             | Unknown | U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* Vollum 1B      | USA     | U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* CDC 684        | Unknown | C              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* 2002013094     | USA     | U              | 933| 65  | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* Canadian bison | Canada  | U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* Turkey32       | Unknown | C              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. thuringiensis* HD571      | Unknown | U              | 109| 34  | 1   | 32  | 1   | 51  | 37  | 24  |
|       | *B. cereus* 3a                | Unknown | FP             | 145| 65  | 1   | 52  | 1   | 1   | 37  | 24  |
|       | *B. cereus* S2-8              | USA     | E              | 145| 65  | 1   | 52  | 1   | 1   | 37  | 24  |
|       | *B. cereus* AH820             | Unknown | C              | 460| 65  | 1   | 56  | 1   | 1   | 53  | 24  |
|       | *B. cereus* F837/76           | Unknown | C              | 75 | 44  | 1   | 32  | 1   | 18  | 33  | 24  |
|       | *B. cereus* FT9               | Brazil  | E              | 1262| 24 | 12 | 50 | 21 | 23 | 58 | 19 |
|       | *B. cereus* D17               | Unknown | F              | 1263| 34 | 1 | 124 | 16 | 18 | 33 | 89 |
|       | *B. anthracis* Ames A0462     | USA     | U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* Ames Ancestor  | Unknown | E              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* Ames chromosome | Unknown | U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|       | *B. anthracis* Ames BA1004    | USA     | U              | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| 2     | *B. mycoides* ATCC6462        | Unknown | E              | 116| 25  | 10 | 22 | 53 | 57 | 23 | 44 |
|       | *B. thuringiensis* HD-1       | USA     | E              | 10 | 15  | 6  | 10 | 8  | 3  | 7  | 14 |
|       | *B. cereus* NC7401            | Japan   | FP             | 26 | 3   | 2 | 31 | 5 | 16 | 3 | 4 |
|       | *B. cytotoxicus* 391-98       | Unknown | FP             | 930| 211 | 127 | 221 | 221 | 57 | 10 | 171 |
|       | *B. weihenstephanensis* KBAB4 | Unknown | E              | 958| 18 | 10 | 79 | 36 | 77 | 70 | 18 |
|       | *B. weihenstephanensis* WSBC10204 | Germany | F    | 196| 64 | 10 | 79 | 36 | 56 | 22 | 11 |
|       | *B. thuringiensis* HD1002     | Israel  | U              | 16 | 15  | 7  | 7  | 2  | 6  | 8  | 13 |

© Springer
were collected and aligned for genomic diversity analysis. The putative ORFs (more than 200aa) from all fragments were annotated (Table 3), and compared to illustrate their genetic features. By using the \textit{nheABC} locus of \textit{B. cereus} ATCC14579 as a typical reference, good conservation was observed in the upstream region of this gene cluster among the other 76 strains. Interestingly, the rest four strains showed different pattern of conservation, therefore, we studied the genomic diversity in two parts.

In the first part, 30 kb DNA sequences centered on the \textit{nheABC} loci from 76 strains were aligned and labelled in the sketch map. The alignment result showed a higher degree of conservation in the upstream region in terms of gene content, relative to the downstream region (Fig. 1). The downstream of \textit{nheABC} locus contained six distinct conserved regions, which was different from their gene contents and organizations (indicated as A to F, and their relative frequencies of 5/76, 24/76, 2/76, 12/76, 31/76 and 2/76, respectively). Specifically, Branch A contained \textit{B. thuringiensis} YBT-1518, \textit{B. cereus} AH187, etc. strains. Branch B contained all the 31 \textit{B. anthracis} strains. Branch F contained \textit{B. weihenstephanensis} KBAB4 and WSBC10204.

In the \textit{nheABC} locus upstream regions (left part of \textit{nheABC} loci in Fig. 1), seven genes (two-component response regulator vanR, C terminus of sensor histidine kinase, M24/M37 family peptidase, manganase transport protein MntH, sulfur transferase, hypothetical protein...
and 2-dehydropantoate 2-reductase) were present in all the 76 fragments. Interestingly, in 49 of the 76 strains, a 505aa ORF (amino acid permease) was found immediately upstream of the NheA, while it was broken up into two smaller ORFs (also annotated as amino acid permease) in the 32 remaining strains, including which included 31 B. anthracis strains plus B. thuringiensis 97-27 strain.

In the nheABC locus downstream regions (right part of nheABC loci in Fig. 1), there was obvious genes constituent change and rearrangement in different branch. The gene rhtB (homoserine/threonine efflux protein) was conserved in Branch A–D, but disappeared in Branch E and F. The four genes cluster deoR (deoxyribonucleoside regulator), deoC (deoxyribose-phosphate aldolase), nupC (nucleoside transporter), pdp (pyrimidine-nucleoside phosphorylase) was well kept in the same direction in Branch B–F. And the gene yndJ encoded a putative membrane protein also maintained in Branch D and E. The other genes showed different variety. In the downstream regions, the composition and arrangement of genes diversified in different Branch strains.

But the rest five nheABC+ strains, B. cereus ATCC4342, D17, FT9, G9241 and B. cytotoxicus

| Identifier | Gene function | Identifier | Gene function |
|------------|---------------|------------|---------------|
| 247        | Two-component response regulator vanR | 551        | multicopper oxidase family protein |
| 226        | sensor histidine kinase, C terminus  | 216        | CAAX amino protease |
| 262        | M24/M37 family peptidase              | 235        | No published record |
| 431        | manganese transport protein MntH      | 291        | No published record |
| 306        | hypothetical protein BA_1881          | 316        | inosine/uridine-prefering nucleoside hydrolase |
| 491        | No published record                   | 318        | Virginiamycin B lyase |
| 314        | 2-dehydropantoate 2-reductase         | 591        | S-layer domain protein |
| 262        | conserved hypothetical protein         | 439        | Aminopeptidase |
| 218        | amino acid permease                   | 456        | carboxylic ester hydrolase |
| 334        | non-hemolytic enterotoxin A           | 206        | Transcriptional regulator, TetR |
| 402        | non-hemolytic enterotoxin B           | 207        | acetyltransferase |
| 305        | non-hemolytic enterotoxin C           | 209        | CAAX amino protease |
| 208        | homoserine/threonine efflux protein   | 364        | No published record |
| 315        | deoxyribonucleoside regulator DeoR    | 211        | No published record |
| 223        | deoxyribose-phosphate aldolase        | 569        | DNA topoisomerase III |
| 393        | nucleoside transporter NupC           | 1152       | molybdate metabolism regulator |
| 437        | pyrimidine-nucleoside phosphorylase   | 368        | No published record |
| 316        | membrane protein, putative            | 768        | No published record |
| 226        | No published record                   | 368        | VWA domain containing CoxE-like protein |
| 294        | BNR repeat-containing protein         | 524        | zinc finger, swim domain protein |
| 427        | xaa-pro aminopeptidase                |            |               |
NVH391-98, sequences showed high genomic diversity, could not be included into any branch.

To illustrate genes distribution of the whole 30 kb fragments in different branches, six representative strains were selected. Based on the 5.7 kb fragment from \textit{B. cereus} NVH883-00, genes arrangement surrounding the \textit{nheABC} locus from \textit{B. anthracis} Ames with the counterparts from \textit{B. mycoides} Rock1-4, 3-17, \textit{B. pseudomycoides} DSM12442, \textit{B. cereus} NVH391-98 were shown in Fig. 2. All orthologous genes are indicated by arrows of the same size and matches are indicated in gray shadow.
and *B. mycoides* Rock 3-17 had almost the same genes distribution. It is easy to recognize gene cluster rearrangement and gene insertion in *B. anthracis* Ames and *B. cereus* NVH391-98, though these two strains shared high similarity in most genes rearrangement.

Genetic diversity of NheA, B and C

To further explore the genetic diversity based on the 81 collected NheA, B and C protein sequences, each protein was analyzed, respectively. The size of 73 NheA proteins was 386aa, and the size of the rest eight strains, such as *B. cytotoxicus* NVH391-98, *B. cereus* G9241, FT9, Ba HYU01, SVA11, RA3, *B. anthracis* BA1035, 2002013094 was 387aa, 312aa, 279aa, 388aa, 388aa, 388aa, 388aa and 318aa, respectively. Based on sequences diversity, all the 73 NheA proteins were divided into four groups (Fig. S1). In group A, the NheA from different strains shared 97–100% identities with *B. cereus* ATCC14579-NheA. In group B, the NheA from strains *B. cereus* 94–100% identities with *B. cereus* NVH391-98, though these two strains shared high similar-

Seven housekeeping genes MLST analysis from 75 *nheABC* strains

To further characterize the sequence variation among the 75 *nheABC* positive strains of *B. cereus*, MLST was performed using the seven housekeeping genes: *glp, gmk, ilv, pta, pur, pyc* and *tpi*. The sequence variability of each locus was also studied in details. Based on the allelic profiles of the seven loci, ST could be defined for all the isolates. The high number of ST is another illustration of the high diversity existing among the *nheABC* positive *B. cereus* strains. Of note, Table S1 displays the correspondence between the allelic profiles defined in this work and those reported in previous MLST schemes (Helgason et al. 2004; Sorokin et al. 2006). The 75 isolates (Table S1) were then subjected to a BURST analysis to group the strains according to the similarity of their allelic profile. The isolates were grouped together when five out of the seven analyzed loci were identical. Based on this criterion, seven clusters were formed, as shown in Table 2. 75 strains were clustered in seven groups. Strains coming from foodborne outbreaks spread in different groups. This result showed that there was no obvious association between similar allelic profiles and geographic or source origins.

To visualize the influence of *nheABC* gene on the relatedness of the strains, 75 strains displaying *nheABC* positive (originating from different countries) were selected for the MLST analysis of the *Bacillus* strains. A tree was built with the sequences of the seven loci concatenated with *nheABC*. PHYLOViZ was built upon the goeBURST implementation (available at http://goeburst.phyloviz.net) and it allows to integrate and display multiple sources of information. The result was showed in Fig. 3. The whole tree was built on all the selected species about 2033 *Bacillus cereus* strains from different sources (data not shown), and the 75 selected strains assigned black color were located in this split tree. Though the observed frequency of each strain was different, all the selected strains spread randomly in all the part of MLST split tree. The number of locus differences between each pair strains could be observed from the line distances of the tree. This explained that the *nheABC* sequences variation had no favor of any type of strains, and also showed

 Springer
Fig. 3 MLST split tree of *Bacillus cereus*. The location of each strain illustrated the number of locus differences with other strains in MLST split tree. The distances represented the number of locus differences between every pair of samples. The *size of the circles* indicated the observed frequencies. The *black circles* represented the strains from Table S1, the *blue* and *green circles* were 2033 *Bacillus cereus* strains (data not shown). The *numbers in the circles* were ST of each strains.  
*a* the whole tree.  
*b* the partial enlargement of A.
again that there was no obvious association between similar allelic profiles and geographic or source origins.

To verify that nheABC operon was not prone to lateral transfer, 30 kb genome fragments centered on the nheABC operon locus were investigated. Although the genomic neighborhood of nheABC showed some variability, no element potentially involved in horizontal transfer was found (data not shown). The comparison of the 76 Bacillus fragments has provided valuable insights into defining the key genetic complement of the organisms, which forms the basic genetic support to define the organism’s pathogenicity ability (Lisdawati et al. 2015).

To investigate the genetic diversity of nheABC and seven housekeeping genes (glp, gmk, ilv, pta, pur, pyc and tpi), the concatenated sequences of the seven housekeeping genes (I), nheA, B and C genes (II), the concatenated sequences of the seven housekeeping genes plus the nheA, B and C genes (III) were collected from 35 B. cereus s.l. group strains for which the seven genes sequences were all available in NCBI database. The sequences were then analyzed by online software clustalw (http://www.genome.jp/tools/clustalw/). The neighbor joining trees based on seven housekeeping genes together with nheA, B and C genes were built. In tree I (Fig. 4), B. cereus NVH391-98 was far away from all the other 35 strains. Genes from different strains clustered together, according to the genetic relationship five main clusters are noted as A–E. In tree II and III, though the order of five main clusters changed, the constituent of each cluster stayed the same. The structure of these trees were similar to published Bacillus phylogenetic trees (Virginie et al. 2015). This result indicated that no striking differences was observed between the various trees, and meant that the influence of nheABC on strain clustering is limited. This result also showed that the genetic determinants of the NHE had no any obvious relationship with the nheABC genes sequence of a strain and its virulence in the diarrhoeal pathogenesis.

In this study, we found that the genetic determinants of the NHE toxin did not bring any obvious link between the nheABC genes sequence of a strain and its virulence in the diarrhoeal pathogenesis. To assess whether NHE is a significant factor in this disease, a transcriptomic study should be considered to take the genes expression of the toxin into account. And the potential action of NHE also should be investigated in concert with other possible enterotoxins (e.g., HBL, CytK or HlyII) and other virulence factors to evaluate the diarrheic potential of B. cereus strains. To elucidate the actual involvement of these molecules in the diarrhoeal syndrome, it is necessary to find an adequate animal model. In fact, due to their proteinaceous nature, these putative enterotoxins may be prone to a rapid inactivation in the intestinal tract, unless they would be released by the B. cereus cells in the immediate vicinity of host’s intestinal epithelium being protected by the mucus layer. However, this hypothesis has still to be verified.

Conclusions

The nheABC genes do not affect the diversity displayed by housekeeping genes, and this specific protein is probably not implicated in the diarrheal syndrome. Our data provide

![Fig. 4 The NJ trees built based on the concatenated sequences of the six housekeeping genes and nheABC genes. I (based on six housekeeping genes), II (based on nheABC genes) and III (based on the concatenation of six housekeeping genes and nheABC genes)](image-url)
a scientific basis for us to know more about nheABC loci in food poisoning \textit{B. cereus}.

**Acknowledgements** This study was supported by grants from Key project of Anhui Science and Technology Department (1408085QC59), Key project of Anhui Education Department (KJ2016A849), Anhui Provincial Key Lab of the Conservation and Exploitation of Biological Resources, Innovation Team of Scientific Research Platform in Anhui Universities, Chinese National Natural Science Foundation of China (No.31500418), Jiangsu Province Natural Science Foundation of China (BK200140261).

**Compliance with ethical standards**

**Conflict of interest** The authors declare no conflict of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

**References**

Alexandre P, Francisco, Catia V et al (2012) PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics, 13:87–92.

Banerjee M, Nair GB et al (2011) Phenotypic & genetic characterization of \textit{Bacillus cereus} isolated from the acute diarrhoeal patients. Indian J Med Res 133:88–95

Beecher DJ, Schoeni JL, Wong AC (1995) Enterotoxic activity of hemolysin BL from \textit{Bacillus cereus}. Infect Immun 63(11):4423–4428

Böhm ME, Huptas C, Krey VM et al (2015) Massive horizontal gene transfer, strictly vertical inheritance and ancient duplications differentially shape the evolution of \textit{Bacillus cereus} enterotoxin operons hbl, cytK and nhe. BMC Evol Biol 15:246

Boonchai N, Asano SI, Bando H et al (2008) Study on cytotoxicity and nucleotide sequences of enterotoxin FM of \textit{Bacillus cereus} isolated from various food sources. J Med Assoc Thai 91(9):1425–1432

Chon JW, Kim HS et al (2015) Quantitative prevalence and toxin gene profile of \textit{Bacillus cereus} from ready-to-eat vegetables in South Korea. Foodborne Pathog Dis 12(9):795–799

De Jonghe V, Coorevits A, De Block J et al (2010) Toxigenic and spoilage potential of aerobic spore-formers isolated from raw milk. Int J Food Microbiol 136(3):318–325

Gaviria Rivera AM, Granum PE et al (2000) Common occurrence of enterotoxin genes and enterotoxicity in \textit{Bacillus thuringiensis}. FEMS Microbiol Lett 190(1):151–155

Hansen BM, Hendriksen NB (2001) Detection of enterotoxic \textit{Bacillus cereus} and \textit{Bacillus thuringiensis} strains by PCR analysis. Appl Environ Microbiol 67(1):185–189

Helgason E, Tourasse NJ, Meisal R et al (2004) Multilocus sequence typing scheme for bacteria of the \textit{Bacillus cereus} group. Appl Environ Microbiol 70:191–201

Jeßberger N, Dietrich R, Bock S (2014) \textit{Bacillus cereus} enterotoxins act as major virulence factors and exhibit distinct cytotoxicity to different human cell lines. Toxicicon 77:49–57

Kim MJ, Han JK, Park JS et al (2015) Various enterotoxin and other virulence factor genes widespread among \textit{Bacillus cereus} and \textit{Bacillus thuringiensis} strains. J Microbiol Biotechnol 25(6):872–879

Krause N, Moravek M et al (2010) Performance characteristics of the Duopath(R) cereus enterotoxins assay for rapid detection of enterotoxinogenic \textit{Bacillus cereus} strains. Int J Food Microbiol 144(2):322–326

Kremer J, Stahlberg K, Morgenstern H et al (2008) Identification of \textit{Bacillus cereus} group species isolated from food poisoning outbreaks in British Columbia, Canada. Appl Environ Microbiol 74(23):7451–7453

Lampe BJ, English JC (2016) Toxicological assessment of nattokinase derived from \textit{Bacillus subtilis var. natto}. Food Chem Toxicol 88:87–99

Lindhäck T, Fagerlund A, Rödlund MS et al (2004) Characterization of the \textit{Bacillus cereus} Nhe enterotoxin. Microbiology 150:3959–3967

Lisdawati V, Puspandari N, Rif’atî L et al (2015) Molecular epidemiology study of \textit{Mycobacterium tuberculosis} and its susceptibility to anti-tuberculosis drugs in Indonesia. BMC Infect Dis 15:366

Lund T, Granum PE (1996) Characterisation of a non-haemolytic enterotoxin complex from \textit{Bacillus cereus} isolated after a food-borne outbreak. FEMS Microbiol Lett 141:151–156

McIntyre L, Bernard K, Beniac D et al (2008) Identification of \textit{Bacillus cereus} group species associated with food poisoning outbreaks in British Columbia, Canada. Appl Environ Microbiol 74(23):7451–7453

Moravek M, Wegscheider M et al (2004) Colony immunoblot assay for the detection of hemolysin BL enterotoxin producing \textit{Bacillus cereus}. FEMS Microbiol Lett 238(1):107–113

Okinaka RT, Keim P (2014) The phylogeny of \textit{Bacillus cereus sensu lato}. Microbiol Syst 4(1): TBS-0012-2012

Samaipundo S, Heyndrickx M et al (2011) Incidence, diversity and toxin gene characteristics of \textit{Bacillus cereus} group strains isolated from food products marketed in Belgium. Int J Food Microbiol 150(1):34–41

Sorokin A, Candelon B, Guilloux K et al (2006) Multiple-locus sequence typing analysis of \textit{Bacillus cereus} and \textit{Bacillus thuringiensis} reveals separate clustering and a distinct population structure of psychrotrophic strains. Appl Environ Microbiol 72:1569–1578

Swiecicka I, Van der Auwera GA, Mahillon J (2006) Hemolytic and nonhemolytic enterotoxin genes are broadly distributed among \textit{Bacillus thuringiensis} isolated from wild mammals. Microb Ecol 52(3):544–551

Virginie C, Xiaoqin L, Laurence D et al (2015) Is Cytotoxin K from \textit{Bacillus cereus} a bona fide enterotoxin? Int J Food Microbiol 211:79–85

Wehrle E, Moravek M, Dietrich R et al (2009) Comparison of multiplex PCR, enzyme immunoassay and cell culture methods for the detection of enterotoxigenic \textit{Bacillus cereus}. J Microbiol Methods 78(3):265–270

Zhou G, Zheng D, Dou L et al (2010) Occurrence of psychrotolerant \textit{Bacillus cereus} group strains in ice creams. Int J Food Microbiol 137:143–146

Zhu Y, Chen P, Bao Y et al (2016) Complete genome sequence and transcriptomic analysis of a novel marine strain \textit{Bacillus weihaiensis} reveals the mechanism of brown algae degradation. Sci Rep 30(6):38248