TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Mikrobielle Ökologie

Comparative phylogenetic and transcriptional analysis of the *Bacillus cereus* sensu lato enterotoxin genes *nhe, hbl* and *cytK*

Maria-Elisabeth Böhm

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. R. F. Vogel
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. S. Scherer
	2.	Univ.-Prof. Dr. W. Liebl

Die Dissertation wurde am 29. 02. 2016 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 12. 05. 2016 angenommen.

Table of Contents

Abstract

Bacillus cereus sensu lato comprises eight closely related species including the human pathogens *Bacillus anthracis* and *Bacillus cereus*. *Bacillus cereus* is a food-contaminant with greatly varying enteropathogenic potential. All known strains possess the genes for at least one of the three enterotoxins Nhe, Hbl and CytK. However, some strains show no cytotoxicity, while others have caused lethal outbreaks of food-poisoning.

Due to historical and medical reasons, the taxonomy of *B. cereus* sensu lato is still based on phenotypic characteristics (presence or absence of virulence factors) and morphologic conspicuities, even though these characteristics are known to be insufficient for species demarcation since they are often encoded on mobile genetic elements. Within *B. cereus* sensu lato, chromosomally and plasmid-encoded toxins exist. While plasmid-mediated horizontal gene transfer of the emetic toxin, anthrax and insecticidal toxins is known, evolution of enterotoxin genes within the group has not been studied.

Within this study 30 *B. cereus* sensu lato genomes were sequenced and assembled *de novo*. These sequences together with genomic sequences deposited in public databases were analyzed with respect to species affiliation, evolution of the enterotoxins and the regulatory role of promoter regions and 5' untranslated sequences of the enterotoxin operons *nhe* and *hbl*.

Species affiliation was reviewed by multilocus sequence analysis (MLSA) of 142 strains and validated by a phylogenetic network based on average nucleotide identity (ANI) derived from genome sequences and a phylogeny based on whole-genome single nucleotide polymorphism (SNP) analysis. The data clearly support subdivision of *B. cereus* sensu lato into seven phylogenetic groups. While group I, V and VII represent *B. pseudomycoides*, *B. toyonensis* and *B. cytotoxicus*, which are distinguishable at the species level (ANI border ≥ 96 %), strains ascribed to the other five species do not match phylogenic groups. The chromosomal enterotoxin operons *nheABC* and *hblCDAB* are abundant within *B. cereus* strains both isolated from infections and from the environment (100 % and 63 % of the analyzed strains, respectively). While the duplicated *hbl* variant hbl_a is present in 22 % of all strains investigated, duplication of *nheABC* is extremely rare (2.2 %) and appears to be phylogenetically unstable. The distribution of toxin genes was matched to a MLSA master tree based on seven concatenated housekeeping genes, which depicts the phylogenetic species relationships in *B. cereus* sensu lato as accurately as whole-genome comparisons.

Comparison of the master tree to the phylogeny of individual enterotoxin operons uncovered ample evidence for horizontal transfer of *hbl, cytK* and *plcR,* as well as frequent deletion of both toxins. Frequent exchange of the pathogenicity factors *hbl, cytK* and *plcR* in *B. cereus* sensu lato appears to be an important mechanism of *B. cereus* virulence evolution, including so-called probiotic or non-pathogenic species, which might have consequences for risk assessment procedures. In contrast, exclusively vertical inheritance of *nhe* was observed, and since *nhe*-negative strains appear to be extremely rare, fitness loss may be associated with deletion or horizontal transfer of the *nhe* operon.

Comparison of 142 genomic sequences revealed partially conserved but always exceptionally long 5' untranslated regions (5'UTRs) upstream of the translational start of the *nhe* and *hbl* operons. These 5'UTRs accumulated recognition sites for an entire orchestra of transcriptional regulators, including the virulence regulator PlcR, redox regulators ResD and Fnr as well as the master regulator for biofilm formation SinR.

Bioluminescent promotor fusions were constructed lacking various (internal) parts of the *hbl* and *nhe* 5'UTR to demarcate uncharacterized functional promoter regions. It was shown that the entire 331 bp *nhe* 5'UTR is necessary for full promoter activity, while presence of the 606 bp *hbl* 5' UTR represses promoter activity. The repressing sequence could be narrowed down to a 268 bp sequence directly downstream of the transcription start. Thus, the highly complex and differential regulation of enterotoxin transcription in *B. cereus* is mediated by unusually long promoter regions of *nhe* and *hbl*. A specific interaction between the nutrient-sensitive regulator CodY and both *nhe* and *hbl* promoter regions was shown. Carbon (catabolite repression) and nitrogen sources have long been known to greatly impact expression of *nhe* and *hbl*. Absence of free amino acids, general nutrient deficiency and low oxygen content – as present under gut-simulating conditions – increase enterotoxin promoter activity. This study shows that the *B. cereus* enteropathogenic potential is strain-specific as well as essentially influenced by growth conditions, which further complicates reliable risk assessment processes in addition to possible horizontal transfer of virulence genes.

Zusammenfassung

Bacillus cereus sensu lato umfasst acht sehr nah verwandte Spezies, unter anderen die humanpathogenen Arten *B. cereus* und *B. anthracis*. *B. cereus* tritt häufig als Verunreinigung in Lebensmitteln in Erscheinung und verfügt über ein stark variierendes enteropathogenes Potential. Alle bekannten Stämme besitzen die Gene für mindestens eines der drei Enterotoxine Nhe, Hbl und CytK. Manche Stämme zeigen jedoch keine Cytotoxizität, während andere für tödliche Fälle von Lebensmittelvergiftungen verantwortlich sind.

Obwohl bekannt ist, dass phänotypische Charakteristika (An- oder Abwesenheit von Virulenzfaktoren) und morphologische Auffälligkeiten oft auf mobilen genetischen Elementen codiert und damit nicht ausreichend für eine Spezieszuordnung sind, beruht die Taxonomie der *B. cereus* Gruppe aus historischen und medizinischen Gründen nach wie vor auf der althergebrachten Einteilung. Innerhalb von *B. cereus* sensu lato existieren Toxingene sowohl im Chromosom als auch plasmid-codiert. Plasmid-abhängiger horizontaler Gentransfer des emetischen Toxins, der Anthrax- und insektiziden Toxine ist seit langem bekannt, die Evolution der Enterotoxingene dieser Gruppe ist jedoch noch nicht im Detail untersucht worden.

30 *B. cereus* sensu lato Genome wurden im Rahmen dieser Arbeit *de novo* sequenziert und assembliert. Diese und weitere Genomsequenzen aus öffentlichen Datenbanken wurden analysiert, um einen besseren Einblick in die Spezieszuordnung, die Evolution der Enterotoxine und die regulatorische Funktion der 5‘ untranslatierten Regionen der Enterotoxinoperons *nhe* und *hbl* zu gewinnen.

Die Spezieszugehörigkeit wurde durch Multilocus Sequenz Analyse (MLSA) von 142 Stämmen begutachtet. Diese wurde validiert durch ein phylogenetisches Netzwerk auf der Basis der genomweiten durchschnittlichen Identität der Nukleotide (ANI) und durch eine phylogenetische Einteilung basierend auf einer genomweiten Einzelnukleotid Polymorphismus (SNP) Analyse. Die MLSA Spezies-Phylogenie beruht auf der Sequenz von sieben konkatenierten Housekeeping Genen und spiegelt die Spezies-Verwandtschaft der *B. cereus* Gruppe ebenso präzise wider wie Genomvergleiche. Die Ergebnisse unterstützen eindeutig eine Unterteilung von *B. cereus* sensu lato in sieben phylogenetische Gruppen. Während die Gruppen I, V und VII *B. pseudomycoides*, *B. toyonensis* und *B. cytotoxicus* repräsentieren, die auf Speziesebene unterscheidbar sind (ANI Grenzwert ≥ 96 %), können die restlichen fünf Spezies keinen phylogenetischen Gruppen zugeordnet werden. Die chromosomalen Enterotoxin-Operons *nheABC* (100 %) und *hblCDAB* (63 %) sind häufig in *B. cereus* Isolaten aus Infektionen und aus Umweltproben zu finden. Das Duplikat von *hbl*, hbl_a, wurde in 22 % aller untersuchten Stämme entdeckt, während das *nhe* Duplikat extrem selten (nhe_a, 2.2 %) gefunden wurde und phylogenetisch instabil zu sein scheint. Die Verteilung der Toxine wurde in die Spezies-Phylogenie integriert.

Vergleiche der Spezies-Phylogenie mit den phylogenetischen Bäumen einzelner Enterotoxin-Operons zeigten zahlreiche Hinweise auf horizontalen Transfer von *hbl*, *cytK* und *plcR* ebenso wie häufige

Deletionen beider Toxine. Der häufige Austausch der Pathogenitätsfaktoren *hbl*, *cytK* und *plcR* scheint ein wichtiger Mechanismus der Evolution von *B. cereus* Virulenz zu sein, einschließlich der sogenannten probiotischen oder nicht-pathogenen Spezies, und könnte weitreichende Konsequenzen für die Risikoeinschätzung und Diagnostikprozesse haben. Im Gegensatz dazu wurde aus den Daten auf eine rein vertikale Vererbung des *nhe* Operons geschlossen. Da *nhe*-negative Stämme extrem selten sind, kann ein Fitnessverlust mit der Deletion oder dem horizontalen Transfer von *nhe* verbunden sein.

Der Vergleich von 142 genomischen Sequenzen ergab partiell konservierte aber in allen Fällen außergewöhnlich lange 5‘ untranslatierte Regionen (5’UTRs) stromaufwärts der *nhe* und *hbl* Translationsstarts. Diese 5’UTRs akkumulierten Erkennungsstellen für eine ganze Reihe von Transkriptionsregulatoren, unter anderem für den Virulenzregulator PlcR, die Redoxregulatoren ResD und Fnr, und den Regulator der Biofilmbildung SinR.
Biolumineszente Promotorfusionen mit Deletion verschiedener (interner) Teile der *hbl* und *nhe* 5’UTR wurden konstruiert, um bisher nicht charakterisierte funktionale Promotorregionen zu erkennen. Es konnte gezeigt werden, dass die gesamte 331 bp lange *nhe* 5’UTR für eine volle Promotoraktivität nötig ist, während die 606 bp lange *hbl* 5’UTR die Promotoraktivität hemmt. Die Region mit reprimierender Funktion konnte auf einen 268 bp langen Bereich direkt stromabwärts des Transkriptionsstarts eingegrenzt werden.
Damit konnte gezeigt werden, dass die hochkomplexe und differentielle Regulation der Enterotoxintranskription in *B. cereus* von den ungewöhnlich langen *nhe* und *hbl* Promotorregionen beeinflusst wird. Eine spezifische Interaktion zwischen dem Nährstoff-sensitiven Regulator CodY und den *nhe* und *hbl* Promotorregionen konnte nachgewiesen werden. Es ist seit langem bekannt, dass Kohlenstoff- (Katabolitrepression) und Stickstoffquellen einen großen Einfluss auf die Expression von *nhe* und *hbl* haben. Mangel an freien Aminosäuren, allgemeine Nährstoffknappheit und ein geringer Sauerstoffgehalt – wie unter darmsimulierenden Bedingungen – stimulieren die Promotoraktivität der Enterotoxingene. In dieser Arbeit konnte gezeigt werden, dass das enteropathogene Potential von *B. cereus* Stämmen variabel und stark von Wachstumsbedingungen abhängig ist, was eine verlässliche Risikoabschätzung zusätzlich zu der Möglichkeit eines horizontalen Transfers von Virulenzgenen erschwert.

Publications

Böhm, M.-E., Huptas, C., Krey, V. M. & 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.* **15,** 246 (2015). doi: 10.1186/s12862-015-0529-4

Böhm, M.-E., Huptas, C., Krey, V. M. & Scherer, S. Draft Genome Sequence of *Bacillus cytotoxicus* CVUAS 2833, a Very Close Relative to Type Strain NVH 391-98 Isolated from a Different Location. *Genome Announc* **3,** (2015). doi: 10.1128/genomeA.00901-15

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

Böhm, M.-E., Krey, V. M., Jeßberger, N., Frenzel, E. & Scherer, S. Comparative Bioinformatics and Experimental Analysis of the Intergenic Regulatory Regions of *Bacillus cereus hbl* and *nhe* Enterotoxin Operons and the Impact of CodY on Virulence Heterogeneity. *Front Microbiol* **7,** 768 (2016). doi: 10.3389/fmicb.2016.00768

Abstracts and Conferences:

Böhm, M.-E. Comparative phylogenetic and transcriptional analysis of *Bacillus cereus* sensu lato enterotoxin genes *nhe, hbl* and *cytK. 16. Fachsymposium Lebensmittelmikrobiologie der VAAM und DGHM*, Stuttgart-Hohenheim, Germany, 30.03.-01.04.2016 (Presentation)

Böhm, M.-E., Huptas, C., Krey, V., Scherer, S. Genomic insights into species affiliation and propagation of virulence determinants in *B. cereus* sensu lato. *Annual Conference 2015 of the Association for General and Applied Microbiology (VAAM)*, Marburg/Lahn, Germany, 01.-04.03.2015 (Poster)

Böhm, M.-E., Krey, V., Jeßberger, N., Dietrich, R., Scherer, S. Genomic insights into potential horizontal gene transfer in enteropathogenic *Bacillus cereus*. *Microbiology after the genomics revolution: Genomes 2014*, Paris, France, 24.-27.06.2014 (Poster)

Böhm, M.-E., Doll, V., Jeßberger, N., Dietrich, R., Rademacher, C., Scherer, S. Genome sequencing of enterotoxin genes of *Bacillus cereus*. *14. Fachsymposium Lebensmittelmikrobiologie der VAAM und DGHM*, Tutzing, Germany, 22.-24.04.2013 (Poster)
Awarded first prize by the Association for General and Applied Microbiology (VAAM) and the German Society for Hygiene and Microbiology (DGHM).

Abbreviations

AA: Amino Acid

AAI: Average Amino Acid Identity

Adk: Gene encoding adenylate kinase

AF: Alignment Fraction

ANI: Average Nucleotide Identity

ANIb: Calculation of ANI based on the BLAST algorithm

AFLP: Amplified Fragment Length Polymorphism

APS: Ammonium persulfate

ATCC: American Type Culture Collection

BCAA: Branched-chain Amino Acid

BSA: Bovine Serum Albumin

Cap: Operon encoding *B. anthracis* poly-γ-D-glutamic acid capsule

CcpA: Gene encoding catabolite control protein A

Ces: Seven gene operon encoding among others both cereulide synthetase subunits A and B

CFU: Colony Forming Units

ClyA: Gene encoding the pore-forming hemolysin cytolysin A

COG: Cluster of Orthologous Groups

Cry: Genes encoding *B. thuringiensis* insecticidal crystal endotoxins

CspA: Gene encoding cold shock protein A

Cya: Gene encoding *B. anthracis* edema factor

Cyt: Genes encoding *B. thuringiensis* cytolysins

CytK: Gene encoding cytotoxin K, exists in the variants CytK-1 and CytK-2

DDH: DNA-DNA Hybridization

DTT: (*2S,3S*)-1,4-Bis-sulfanylbutane-2,3-diol (Dithiothreitol)

EDTA: 2,2',2'',2'''-(1,2-Ethanediyldinitrilo)tetraacetic acid

EMSA: Electro Mobility Shift Assay

FCK: Fraction of Core k-mers

FTIR: Fourier Transform Infrared Spectroscopy

gANI: genome-wide ANI

GBDP: Genome BLAST Distance Phylogeny

GlpF: Gene encoding glycerol uptake facilitator protein

GlpT: Gene encoding glycerol-3-phosphate transporter

GTP: Guanosine-5'-triphosphate

Hbl: Operon encoding the tripartite enterotoxin hemolysin BL, exists in the variants Hbl and Hbl_a

HGT: Horizontal Gene Transfer

His_6: Hexahistidine-tag

HlyII: Hemolysin II

ILV: L-isoleucine/L-leucine/L-valine

LB: Luria Bertani

Lef: Gene encoding *B. anthracis* lethal factor

Lux: Operon encoding luciferase subunits and fatty acid reductase complex

ML: Maximum Likelihood

MLEE: Multilocus Enzyme Electrophoresis

MLSA: Multilocus Sequence Analysis

MLST: Multilocus Sequence Typing

MYP: Mannitol Egg Yolk Polymyxin

Nhe: Operon encoding the tripartite non-hemolytic enterotoxin, exists in the variants Nhe and Nhe_a

OD: Optical Density

PAA: Polyacrylamide

PAGE: Polyacrylamide Gel Electrophoresis

PagA: Gene encoding *B. anthracis* protective antigen A

PanC: Gene encoding pantoate-β-alanine ligase

PapR: Gene encoding the quorum sensing peptide PapR necessary for activation of PlcR

PC: Plate Count

PCR: Polymerase Chain Reaction

PEMBA: Polymyxin Egg Yolk Mannitol Bromothymol blue Agar

PFT: Pore-forming Toxin

PlcR: Gene encoding the Phospholipase C virulence regulator

PMSF: Phenylmethylsulfonyl fluoride

Pta: Gene encoding phosphotransacetylase

pXO1: *B. anthracis* virulence mega-plasmid containing the pathogenicity island encoding among others PagA, Lef and Cya

pXO2: *B. anthracis* virulence mega-plasmid containing among others the *cap* operon

PycA: Gene encoding pyruvate carboxylase subunit A

RAPD-PCR: Random Amplification of Polymorphic DNA – PCR

REP-PCR: Repetitive Extragenic Palindromic – PCR

RPMI: Rowell Park Memorial Institute (cell culture medium)

Rrn: 16S rDNA

SDS: Sodium Dodecyl Sulfate

SM: Sphingomyelin

SMase: Sphingomyelinase

SNP: Single Nucleotide Polymorphism

sp.: Species (sing.)

spp.: Species (pl.)

TAE: Tris/Acetate/EDTA

TBE: Tris/Borate/EDTA

TEMED: *N*,*N*,*N*′,*N*′-Tetramethylethane-1,2-diamine

T_m: Melting Temperature

v/v: Volume per Volume

w/v: Weight per Volume

List of Figures

List of Tables

1. Introduction

1.1 Characteristics of *B. cereus* and relationship of *B. cereus* sensu lato members

Bacillus cereus sensu lato comprises eight species of gram-positive, low–GC, facultative anaerobe and endospore-forming bacteria with greatly varying pathogenic potential. *B. cereus* sensu stricto was discovered in 1887 [1] as a bacterium occurring ubiquitously in nature, e.g. in soil and plant material as well as in the intestinal flora of animals [2]. The species contains probiotic as well as pathogenic strains. Probiotic *B. cereus* strains are used as feed additive and for human nutrition (e.g. Toyocerin®, Bactisubtil®, Subtyl). However, their toxic potential is controversially discussed and some are not licensed in the EU, e.g. AICare™, Esporafeed Plus® and Paciflor® due to the proven production of enterotoxins by the contained *B. cereus* strains [3, 4].

B. cereus is also used in industrial biotechnological processes. Chinese bioengineers are taking advantage of the beneficial effect of *B. cereus* on vitamin C production by *Ketogulonicigenium vulgare* in co-culture and could already show an increase of the vitamin C precursor 2-ketogulonic acid by 16 % after serial subcultivation of the co-cultures [5, 6].

B. cereus spores end up in a remarkable range of different foods (rice, pasta, vegetables, potato- and dairy products, spices, etc., see also isolation sources of the strains listed in Table S1) from soil, dust and plants [2, 7, 8]. They are transferred as aerosols as well as by cross-contamination from food and food-processing equipment. As an opportunistic pathogen, *B. cereus* sensu stricto are mostly associated with food-borne illness characterized by diarrhea or vomiting [9-11], but occasionally *B. cereus* is responsible for severe infections, e.g. endophthalmitis, periodontitis, meningitis or pneumonia [12-16]. Two different forms of food poisoning are recognized: the emetic type, caused by the small, cyclic and heat-stable dodecadepsipeptide cereulide with a short incubation period of 0.5 – 6 h [17], which ends rarely fatal [9, 18, 19], and the diarrheal type, which can be caused by one or several heat-labile enterotoxins after a longer incubation period of 8 – 16 h [20, 21]. The former shows symptoms similar to *Staphylococcus aureus* enterotoxins [22], while the latter resembles *Clostridium perfringens* food poisoning, characterized by diarrhea and abdominal pain [20]. The three most important and well-known enterotoxins are the non-hemolytic enterotoxin (Nhe), hemolysin BL (Hbl) and cytotoxin K (CytK). In 2011 the European Food Safety Authority (EFSA) reported an increase of 122.2 % in *B. cereus* food-borne intoxications and toxicoinfections in comparison to 2010 [23]. In 2012 a 117.7 % and in 2013 a 107.3 % increase compared to the previous year was noticed. The majority of outbreaks (between 70.0 and 84.9 %) including rare lethal cases were reported in France [24, 25]. Due to the usually transient and mild symptoms, its broad range of pathogenicity, symptomatic similarity to other food poisoning bacteria and difficulties in strain differentiation, *B. cereus* is still underestimated as a food-borne pathogen [22, 26].

The relationship between *B. cereus* sensu lato species is shown in Fig. 1. *B. cereus, B. thuringiensis* and *B. anthracis* are suspected to be one species on the basis of very close genomic relatedness (≥ 99.7 % 16S rRNA sequence identity) [27]. Several *B. mycoides* strains were shown to be very similar to

B. weihenstephanensis and their reclassification was suggested [28]. The 16S rRNA similarity between *B. mycoides* and *B. pseudomycoides* was determined to be 98 % [29] and *B. cytotoxicus* showed 97 – 98 % 16S rRNA sequence identity [30], confirming the close bonds between even the most distant *B. cereus* sensu lato species.

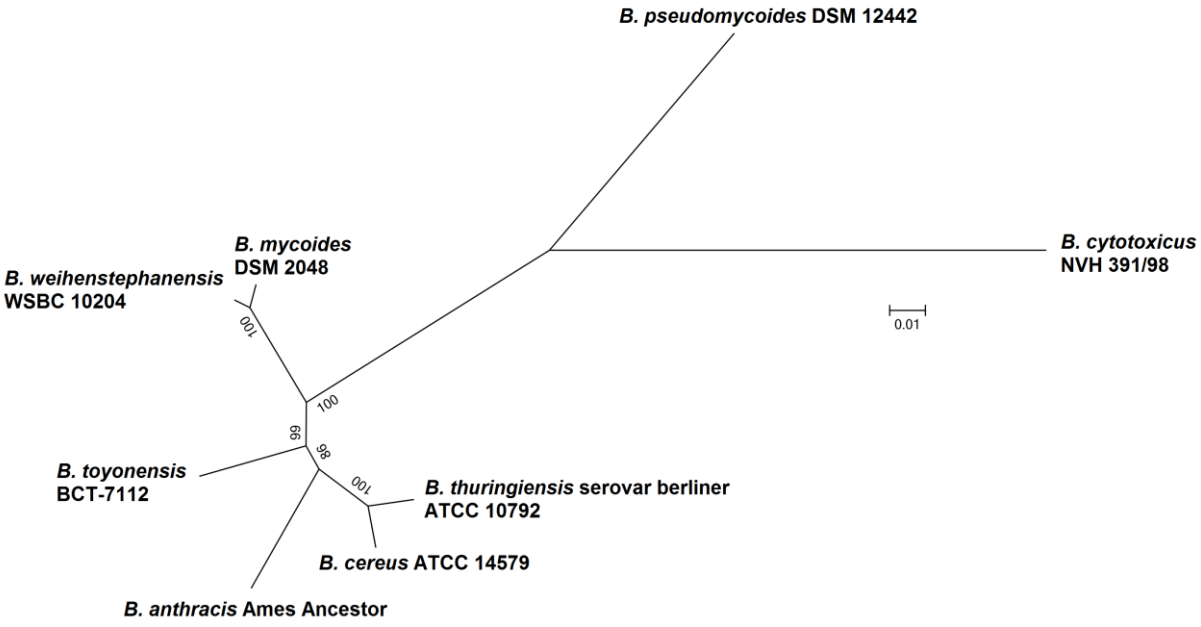


Fig. 1: MLSA-based species relationship within *B. cereus* sensu lato.
The phylogenetic tree (Maximum Likelihood Method) was calculated using the concatenated sequence of seven housekeeping genes (*adk, ccpA, glpF, glpT, panC, pta,* and *pycA)* from the eight *B. cereus* sensu lato type strains. For a detailed description of the method see section 2.5.5.

B. thuringiensis produces insecticidal parasporal protein crystals of Cry (crystal) and/or Cyt (cytolytic) proteins that are mostly encoded on plasmids. Occasionally *B. thuringiensis* have been found responsible for human infections very similar to *B. cereus* [31, 32]. *B. anthracis* is the best known human and animal pathogen of the *B. cereus* group and was demonstrated to be the causative agent of anthrax by Robert Koch in 1876 [33]. The anthrax-associated plasmids pXO1 (encoding anthrax toxin genes *pag, lef* and *cya*) and pXO2 (encoding the poly-γ-D-glutamic acid capsule genes *cap*) have been found in a few *B. cereus* strains such as *B. cereus* G9241 and *B. cereus* biovar *anthracis* CA [34, 35] with a similar pathogenic potential as *B. anthracis*. *B. weihenstephanensis* is psychrotolerant and able to grow below 7 °C [36]. Occasionally, this species houses the emetic toxin cereulide [37, 38]. The psychrotolerant *B. mycoides* is closely related to the other *B. cereus* sensu lato species (16S rRNA sequences showed > 99 % identity [27]), but it can easily be distinguished by its rhizoidal colonial growth [39]. No infections by *B. mycoides* have been reported yet, although it carries both *nhe* and *hbl* and its cytotoxicity was shown [40]. Within *B. mycoides* a group of bacteria with a clearly

distinguishable fatty acid profile was recognized and described as *B. pseudomycoides* [29]. In 1998, a highly enterotoxic and rare variant of cytotoxin K, CytK-1, was discovered in *B. cereus* NVH 391-98, a strain responsible for severe food poisoning. This strain was published in 2013 as the type strain of the new species *B. cytotoxicus* on the basis of presence of the *cytK-1* gene, its thermotolerance (growth at up to 50 °C), a distinctive fatty acid profile, DNA-DNA hybridization and multilocus sequence typing (MLST) [30]. Only a few *B. cytotoxicus* strains have been described yet, but the majority of them have been found in potato products [30, 41, 42]. The eighth member of *B. cereus* sensu lato was isolated in Japan in 1966 but has been described as a separate species *B. toyonensis* only recently [43]. It is commercially available as a probiotic (Toyocerin®). *B. toyonensis* was distinguished from other *B. cereus* sensu lato type species by pairwise calculations of the average nucleotide identity (ANI).

1.2 Species demarcation and horizontal gene transfer among *B. cereus* sensu lato

1.2.1 Prokaryotic species demarcation

The question 'What is a species?' has occupied numerous biologists and taxonomists for a long time. The biological species concept, which is valid for sexually reproducing animals, defines a species as a group of organisms that can interbreed [44]. This definition cannot be applied to asexual prokaryotes. Therefore, microbial taxonomy relies on a combination of phenotypic, genotypic and biochemical characteristics and methods to determine the degree of similarity between organisms. During the 18th century classification of microorganisms was solely based on morphological observations and initially prokaryotes were viewed as one species with a plethora of different shapes [45]. Later, morphology was combined with knowledge about the environmental conditions and pathogenicity. Since the late 18th century a systematic arrangement of microorganism was attempted and developed in parallel with methods for cultivation and (bio-)chemical analyses [45]. These methods include now morphological characterization, biochemical/enzymatical characterization, DNA-DNA hybridization (DDH), 16S rDNA sequencing, comparison of G+C contents and identification of cell wall lipid and polysaccharide composition. Unfortunately, the mentioned methods are limited due to error-prone, expensive and labor-intensive experiments. As a result of the increasing number of available draft and completed genomes the trend in prokaryotic species distinction is moving towards comparison of entire genomes [46-51]. Genome BLAST distance phylogeny (GBDP) [51] and average nucleotide identity (ANI) [52] are based on the pairwise comparison of genome sequences. The average amino acid identity (AAI) [53] compares coding sequences on a whole-genome scale. ANI is now considered the new gold standard of species delineation (ANI species border 95 – 96 % corresponding to 70 % DDH similarity), although reinforcement by other techniques might be necessary [48, 49, 54].

The basis for the demarcation of species is the definition of the term 'species' as the unit of taxonomy and evolution. At least 22 species concepts have been in use and most of them are inappropriate to describe the biological diversity [55]. The debate on the species concept has lasted for decades [56, 57]

and continues on [45, 49]. The modern polyphasic species distinction states that a group of strains belonging to one species must have > 70 % DDH similarity, < 5 °C ΔT_m, < 5 % mol G+C difference of total genomic DNA and > 98 % 16S rRNA identity [46]. This species concept was critized as too conservative, responsible for an underestimation of diversity and lacking congruency with the eukaryotic species definition [45, 58, 59]. Thus, the species definition remains highly controversial.

1.2.2 Species demarcation among *B. cereus* sensu lato

Multilocus sequence typing (MLST), amplified fragment length polymorphism (AFLP) and multilocus enzyme electrophoresis (MLEE) are capable of species distinction based on genomic relatedness and have been used to extensively analyze *B. cereus* sensu lato [60-63]. The database HyperCAT was developed to combine phylogenetic information from the available five *B. cereus* MLST schemes, AFLP and MLEE data [64]. MLST, which provides a standardized approach by examination of fragments of multiple housekeeping gene nucleotide sequences, was considered to become gold standard for bacterial phylogeny due to its unambiguity and portability between laboratories [64, 65]. Now it is likely to be replaced by methods that exploit the wealth of information provided by the increasing number of entire genomes, such as the genome-wide average nucleotide identity.

MLST, AFLP, MLEE and whole-genome based approaches are time-consuming and costly methods. Some of them require ample background knowledge and experience for correct interpretation. Thus, they are inconvenient for routine diagnostics. Since it is not possible to distinguish all species within *B. cereus* sensu lato via the commonly used routine diagnostics methods (selective culture media, such as PEMBA and MYP), several other approaches for fast differentiation have been developed or adapted [66], including Fourier transform infrared spectroscopy (FTIR), RAPD- (random amplification of polymorphic DNA) and REP- (repetitive extragenic palindromic) PCR [67, 68]. However, species distinction is not necessary if the objective is to discern pathogenic from innocuous strains. Several multiplex PCR approaches are used to detect *B. cereus* sensu lato virulence plasmids [69, 70] and enterotoxins [71-74].

1.2.3 Horizontal gene transfer

‘*Horizontal gene transfer is the collective name for processes that permit the exchange of DNA among organisms of different species*` [75].

The transfer of genes between (unrelated) species allows the rapid distribution of newly evolved genes, increases genetic diversity and accelerates genome evolution [76]. The significance of horizontal gene transfer (HGT) was recognized when the spread of virulence factors, e.g. antibiotic resistances and toxin genes, became apparent [77-80]. Prerequisition of the transfer is the ability of (micro-) organims to exchange genes and the necessity to do so due to environmental factors. While the transfer is often

limited to a few linages within a similar environment, it influences nevertheless the molecular evolution of microorganisms and speciation [76, 81]. Among prokaryotes, horizontal transfer can be classified into the (i) acquisition of new genes, (ii) acquisition of paralogs of existing genes, and (iii) xenologous (from a distant linage) gene displacement [82]. Genes that evolved via duplication are called paralogs, as opposed to orthologs, which originate from a single ancestral gene in the last common ancestor of the compared genomes [83].

Several processes and mechanisms are known that allow the transfer of DNA from a donor to a recipient cell [79]: Transformation, the uptake of naked DNA from the environment, is potentially possible between distantly related organisms, but some species show sequence specificity that enhances transformation efficiency between closely related organisms. In contrast, transfer of genetic material via bacteriophages (transduction) is limited by the target range of the phage. Conjugation is a process that mediates the transfer of genetic material (chromosomal DNA for integration or plasmid) from one cell to another via physical contact. Transposons encode the proteins essential for excision and integration of the genetic material into the recipient genome. The transferred DNA integrates into the recipient genome by homologous recombination between closely related species or by non-homologous recombination between distantly related organisms without sequence homology [84]. Recombination interrupts linkage disequilibrium – the non-random pattern of association between alleles at different loci within a population – and is an important characteristic in gene mapping and association studies [85, 86]. Phylogenetic studies therefore should consider reticulate events such as hybridization, horizontal gene transfer, recombination and duplication or loss of genes in order to display true relationships [87].

Horizontal gene transfer (HGT) among distantly related bacteria can be inferred by several approaches such as determination and comparison of atypical base composition, codon usage, oligonucleotide frequencies and number of open reading frames [79, 88], but these are difficult to apply when bacteria as closely related as *B. cereus* sensu lato are investigated [89].

1.2.4 Horizontal gene transfer among *B. cereus* sensu lato

Differentiation of the closely related *B. cereus* species is still based on the presence or absence of phenotypic characteristics for historical and medical reasons. It has been reported that species affiliation of *B. cereus* group strains often does not match the phylogenetic relatedness [27, 90, 91]. One reason for such discrepancies may be the exchange of virulence plasmids between species. Plasmids pXO1 and pXO2 encoding the anthrax toxin complex and the poly-γ-D-glutamic acid capsule are found not only in *B. anthracis*, but also in some *B. cereus* strains [34, 92, 93]. The *B. cereus* cereulide synthetase gene cluster is also located on a large pXO1-like plasmid [17] and is not restricted to a single lineage within *B. cereus* sensu lato [38, 94]. Some of the *B. thuringiensis* insecticidal *cry* genes are encoded on plasmids [95] and can spread via horizontal gene transfer (HGT) among *B. cereus* sensu lato. Therefore,

the transfer of a single plasmid from one species to another species may result in a change of species affiliation.

Plasmids are exchanged or lost frequently among bacteria depending on environmental conditions, but a stable horizontal transfer of chromosomal material additionally requires integration into the bacterial genome via recombination. Several studies have addressed the general possibility of HGT among *B. cereus* sensu lato, obtaining sometimes controversial results. *B. cereus* and *B. thuringiensis* isolates have been studied by MLEE resulting in a high variety of closely related electrophoretic types, evidence of extensive recombination between species and a low degree of clonality [96]. MLST analyses revealed that *B. anthracis* is a homogeneous and clonal cluster within *B. cereus* sensu lato, while *B. cereus* and *B. thuringiensis* are of higher diversity [97]. Similar studies have been conducted with a different set of housekeeping genes for MLST, amplified fragment length polymorphism (AFLP) and genomic comparisons, concluding that HGT involves chromosomal genes and is probably mediated by transposable elements [96, 98, 99]. Other studies conclude from MLST, AFLP, MLEE and genomic data that chromosomal recombination events are generally rare, but appear more often among *B. cereus/B. thuringiensis*, while only plasmids are transmitted by HGT [63, 100, 101]. Horizontal spread of virulence plasmids is not unusual, but little is known about the lateral transfer of chromosomal *B. cereus* virulence factors [97, 99].

1.3 *B. cereus* enterotoxins

The *nhe*, *hbl* and *cytK* genes encode the most important enterotoxins (non-hemolytic enterotoxin, hemolysin BL and cytotoxin K) that are exclusively found in *B. cereus* sensu lato. They are present in *B. cereus* sensu stricto as well as in all other members of the *B. cereus* group. Between 10^5 and 10^8 *B. cereus* CFU (colony forming units) g^{-1} food seem to be necessary to cause disease, but sometimes even 10^3 viable cells or spores cannot be considered safe for consumption [102]. Since the spores are far better equipped to survive passage through the gastric acid, lower numbers can probably cause diarrhea [103]. It was shown that Nhe, Hbl and CytK are secreted via the Sec translocation pathway rather than by the flagellar export apparatus [104].

All three are pore-forming toxins (PFTs) that undergo large conformational changes of the soluble monomers to achieve pore-formation [105, 106]. The β-PFTs, which include CytK, assemble at the membrane surface to a pre-pore followed by insertion into the membrane, while α-PFTs perform of assembly and insertion concomitantly [106]. Both Nhe and Hbl are tripartite α-PFTs that form large conductance pores (>4 nm) in planar lipid bilayers [107]. While all three components are necessary to develop full cytotoxic activity, it was shown that complexes of NheB and NheC alone are capable of formation of stable transmembrane channels at a diameter of ~2 nm [108]. Nhe and Hbl are highly similar among each other, but show no sequence similarity to known protein families. Furthermore,

remarkable tertiary structure similarities to the pore-forming cytolysin A (ClyA) from *Escherichia coli* was discovered [2, 107, 109].

CytK and hemolysin II (HlyII) belong to the same family of oligomeric β-barrel pore-forming toxins (~30 % amino acid identity) that includes among others the α-hemolysin from *Staphylococcus aureus* [110, 111]. The latter forms as a heptamer a pore of 14 – 46 Å in diameter [112], while for the CytK oligomer a minimum pore diameter of 7 Å was predicted [113].

1.3.1 Non-hemolytic enterotoxin

Nhe was first identified as a supposedly non-hemolytic enterotoxin in the *hbl* negative strain *B. cereus* NVH 0075-95, which was involved in a large food poisoning outbreak in Norway in 1995 [114]. Nhe is a tripartite toxin encoded by the operon *nheABC* (Fig. 2A) and is cytolytic against epithelia and erythrocytes due to its ability to cause osmotic lysis by forming pores in the plasma membrane of the host cell [107]. The two components NheB and NheC were found to attach to Vero cells and after subsequent binding of NheA the tripartite complex showed cytotoxic activity [115]. All three components NheA, NheB and NheC are required for full toxic activity although NheC is only expressed in small amounts due to translational repression. A NheA : NheB : NheC ratio of 10 : 10 : 1 leads to maximum cytotoxic activity [116]. The formation of a complex between NheB and NheC is necessary to induce Nhe toxicity, but an insufficient amount of free NheB limits toxic activity [117]. Thus, the correct functioning of Nhe requires not only a specific binding order, but also a defined ratio of the protein components and the initial NheB-NheC complex, proving the pore-formation to be highly complex.

All known *B. cereus* strains possess the *nhe* operon, but some strains exhibit considerable polymorphism, such as *B. cereus* NVH 391-98 (now: *B. cytotoxicus* NVH 391-98) with only 70 – 80 % DNA sequence identity to other *nhe* operons [118-120]. This divergent *nhe* sequence could not been detected by PCR or the available monoclonal antibodies [120]. The cytotoxicity seems to correlate with the expression of *nheABC*, because *B. cereus* strains producing both enterotoxin complexes (Nhe and Hbl) are not more cytotoxic than sole Nhe producers and toxic activity showed good correlation with the concentration of NheB in bacterial supernatant [121]. The production of high-affinity antibodies against all three components of Nhe for the specific quantification of the biological active toxin complex was reported in 2005 [122]. Together with the monoclonal antibodies against the Hbl complex a quantitative evaluation of the two tripartite enterotoxins produced by *B. cereus* is possible, e.g. the cytotoxicity of strain DSM 4384 (= F4430/73) is attributed to Nhe for 60 % and to Hbl for 40 %.

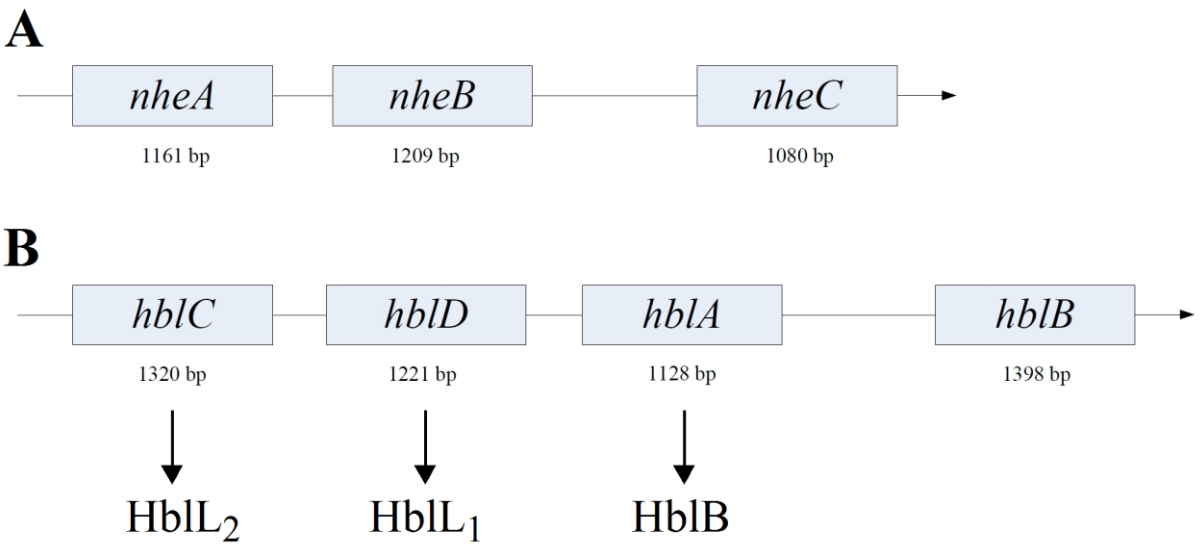


Fig. 2: Genetic organization of the enterotoxin operons in *B. cereus* sensu lato.
The scheme is not drawn to scale. **A:** *Nhe* operon of Nhe reference streain *B. cereus* NVH 0075-95 [123]. **B:** *Hbl* operon of Hbl reference strain *B. cereus* F837/76 [124].

1.3.2 Hemolysin BL

Hbl consists of a binding component B and the two lytic components L_1 and L_2, which were first characterized in 1991 [125]. These proteins are encoded in one operon *hblCDAB* (Fig. 2B) by the genes *hblA*, *hblD* and *hblC*, respectively [118]. It has been shown that *hblB* is not essential for enterotoxic activity, because *hblCDAB* mRNA synthesis appears to terminate within the *hblB* gene [126], which might be a pseudogene [127]. *B. cereus* F837/76, which was isolated from a surgical wound infection, is the Hbl reference strain and a high-level producer of enterotoxin [128]. On their own, none of the three proteins showed toxic activity, but the combination of all three caused fluid accumulation in ligated rabbit ileal loops, necrosis of rabbit skin, cytotoxicity to cultured eucaryotic cells [129] and also hemolysis on blood agar [130]. The production of monoclonal antibodies against the components of the Hbl enterotoxin complex, which enables the specific quantification of the toxin protein(s) in culture supernatant was reported in 1999 [131]. To deploy its hemolytic activity the initial binding of HblB to the membrane of target cells (erythrocytes, epithelial cells) is necessary. Subsequently, the addition of the components L_1 and L_2 lead to maximal activity [125]. HblB represents the limiting part of complex formation since it is binding to the cell surface much slower than the other components [132]. As soon as the concentration of B and/or L_1 exceeded 1.3 nM hemolysis was inhibited, causing the paradoxical zone phenomenon in blood agar: lysis begins several millimeters away from the source of diffusion and cells at intersections of lysis zones remain intact longer [132]. The Hbl- and Nhe-complexes show remarkable structural similarities leading to the assumption that they form together with the ClyA families from e. g. *Escherichia coli*, *Shigella flexneri* or *Salmonella enterica* a superfamily of pore-forming cytotoxins [107]. The same study revealed significant amino acid sequence identity between

Nhe and Hbl components exhibiting similar functions: NheA and $HblL_2$ (23 %), NheB and $HblL_1$ (40 %) and NheC and HblB (25 %). In addition, cross-reactivity of a monoclonal antibody against $HblL_1$ with NheB suggests a common epitope [131]. This led to the assumption that the *hbl* and *nhe* operons resulted from gene duplication of ancestor genes and spread by horizontal gene transfer [2]. In accordance with this hypothesis, a second homologuous Hbl set with similar toxic activity was discovered in 2000 and named Hbl_a [133].

1.3.3 Cytotoxin K

The third diarrhea causing agent is the single-component CytK (34 kDa), which is a hemolytic, dermonecrotic and β-barrel pore-forming enterotoxin [111]. The strain *B. cereus* NVH 391-98, from which CytK was first isolated, was responsible for a severe outbreak of diarrheal syndrome food poisoning in France in 1998 [111]. It has been denoted *B. cytotoxicus* sp. nov. after genotypical and phenotypical differentiation from the other five species of the *B. cereus* group [30]. CytK shows homology with other β-barrel pore-forming toxins, such as the staphylococcal α–hemolysin or the *Clostridum perfringens* β-toxin [111]. The transmembrane channel formed by CytK is weakly selective for monovalent anions and cations, but with an estimated minimum pore diameter of 7 Å at a length of 100 Å smaller than that of the *Staphylococcus aureus* α-toxin [113]. Human intestinal cells (Caco-2) are five times more susceptibel for CytK than Vero cells (African green monkey kidney epithelial cells) [111, 113]. A CytK polymorphism was discovered, when after the original (now: CytK-1) another CytK variant (CytK-2) was isolated. 89 % of the amino acid sequence of CytK-2 is identical to CytK-1. CytK-2 is also able to form pores in planar lipid bilayers, but it shows only 20 % of CytK-1 toxicity [134]. The *cytK-1* gene of *B. cytotoxicus* is a quite rare variant of the cytotoxin K in comparison to the more abundant *cytK-2* [135].

1.3.4 Detection of enterotoxins and additional virulence factors

Since PCR approaches cannot provide information about the actual toxin production, enterotoxins (Nhe and Hbl) are detected in routine diagnostics by immunological methods. Both the 3M TECRA™ BDE-VIA (*Bacillus cereus* diarrheal enterotoxin visual immunoassay) kit and the Oxoid™ BCET-RPLA (*Bacillus cereus* enterotoxin reversed passive latex agglutination) kit use polyclonal antibodies for the detection of NheA and $HblL_2$ enterotoxin subunits, respectively. Both kits are widely used, but their specificity is debatable [136, 137]. The Duopath® Cereus Enterotoxins kit (Merck Millipore) uses monoclonal antibodies for the simultaneous detection of NheB and $HblL_2$ giving more reliable results [138]. Quantitative detection of functional enterotoxin is furthermore complicated since correlation between cytotoxicity and toxin production was only shown for the components NheB, Hbl L_1 and HblB

[121, 139], while commercial kits target NheA and $HblL_2$. So far, no detection kit for CytK or any of the other enterotoxins is commercially available.

Additionally, *B. cereus* strains can produce a variety of other enterotoxins and virulence factors that seem to have minor influence on their pathogenicity, e. g. phospholipases C [140], several hemolysins including the β-channel forming 45.6 kDa hemolysin II [141], 24.4 kDa hemolysin III [142], cereolysin O (= hemolysin I) [143] or the InhA (immune inhibitor A) metalloproteases [144-146]. The *entFM* gene encoding enterotoxin FM was discovered in 1997 [147]. The enterotoxin T (*bceT*) in *B. cereus* B-4ac showed Vero cell cytotoxicity, caused fluid accumulation in ligated mouse ileal loops and was lethal to mice upon injection [148]. The zinc metalloprotease NprA, which is also designated NprB or Npr599 in *B. anthracis*, has been shown to degrade collagen, fibronectin and laminin and thus contributes to pathogenicity [149, 150]. The two-component Cereolysin AB, consisting of tandemly arranged genes of phospholipase C (*cerA*) and sphingomyelinase (*cerB*), is another hemolytic toxin of *B. cereus* [151]. The phosphatidylcholine-preferring phospholipase C is toxic to rabbit retinal tissue *in vitro*, causes retinal necrosis *in vivo* and is therefore one of the causative agents of fulminant *B. cereus* endophthalmitis [152]. The enzyme sphingomyelinase (SMase) appears in procaryotes as well as in eucaryotes and hydrolyzes sphingomyelin (SM) to ceramide and phosphocholine leading to the selective destruction of erythrocyte membranes [153]. The toxins Hbl, phospholipase C and SMase were also shown to have synergistic effects. Their combined activity enhances the lysis of sheep, bovine, swine and human erythrocytes [154]. SMase additionally supplements Nhe cytotoxicity *in vitro* as well as *in vivo*, which indicates that additional virulence factors might play a more important role than previously thought [155].

1.3.5 Regulation of enterotoxin expression

The pleiotropic transcriptional activator PlcR (Phospholipase C Regulator) is the most important virulence regulator in *B. cereus* and a key component in adaptation to (host) environment, which controls expression of extracellular virulence factors, e.g. toxins, enzymes and bacteriocins [127, 156]. Several other proteins were discovered that are under control of PlcR, such as two-component sensors, chemotaxis proteins, transporters, cytoplasmic regulators and cell wall biogenesis proteins. The promoters of genes regulated by PlcR show a highly conserved palindromic recognition site ($TATGNAN_4TNCATA$), which occurs in the promoter region of the enterotoxin genes *nhe, hbl* [127] and *cytK* [111]. It was shown that toxicity varies greatly between strains and sole presence of enterotoxin genes is not sufficient for classification of a *B. cereus* strain as toxic or apathogenic [122, 157]. This led to the conclusion that enterotoxin expression has to be strain-specific differentially regulated. Promoter regions of both *nhe* and *hbl* are exceptionally long and allow binding of several regulators that control enterotoxin expression in *B. cereus* in addition to PlcR [156, 158-161]:

The fumarate nitrate reduction regulatory protein (Fnr), a member of the Crp (cyclic AMP-binding protein) / Fnr family of transcriptional regulators, is produced as an apoform and its monomers are able to bind to specific sequences in the promoter regions of *fnr*, *resDE*, *plcR*, *nhe* and *hbl* [159]. This suggests that the regulation of virulence factor and enterotoxin expression is linked to the availability of oxygen. It was shown that *fnr* expression is downregulated in the presence of O_2 and nitrate which are both able to serve as terminal electron acceptors of respiratory pathways [162].

ResDE is a redox-sensitive signal transduction and two component system consisting of the membrane-bound histidine sensor kinase ResE and the cytoplasmatic response regulator ResD [163]. The ResD monomer interacts with target DNA to activate transcription of e.g. *fnr*, *resDE, nhe, hbl* and *plcR* depending on its phosphorylation state under oxygen limiting conditions [158, 164].

Both *nhe* and *hbl* expression are additionally regulated by catabolite repression, while *cytK* is expressed independent of CcpA-mediated catabolite control [160].

However, enterotoxin expression is not only regulated by the availability of carbon sources and oxygen, but also by nitrogen sources and the general energetic cell status. CodY is a global regulator of adaptation to unfavorable environments, sensing nutrient availability through interaction with GTP and the branched-chain amino acids (BCAAs) isoleucine, leucine and valine [165-167]. The CodY regulon contains several hundred genes including genes for motility, chemotaxis, catabolism, degradation enzymes and virulence [166, 168, 169]. CodY controls the expression of virulence genes via activation of the PlcR regulon [170].

Furthermore, it was shown that enterotoxin genes are part of the SinR-SinI regulon, which leads to a differential, tightly controlled expression of the toxins in a subpopulation of cells during biofilm formation [161].

1.4 Research objectives

B. cereus sensu lato species affiliation is dependent on mobile elements for medical purposes (anthrax toxins, the *B. cereus* emetic toxin and insecticidal toxins of *B. thuringiensis* are plasmid-bound) and not consistent with genomic data. *Bacillus cereus* is a highly heterogeneous taxon containing strains that are used as probiotics, while others are opportunistic pathogens causing severe infections or food-borne illness (emesis or diarrhea).

To analyze potential factors which determine *B. cereus* strain-dependent virulence, a selection of low and highly enterotoxic *B. cereus* strains was sequenced and assembled *de novo*. These sequence data provide the basis for comparative analysis of virulence genes and their distribution by duplication and lateral gene transfer (Chapter 3.1). As an indispensable prerequisite for this task whole-genome based phylogeny of *B. cereus* sensu lato strains as well as stable MLSA (multilocus sequence analysis) phylogeny was constructed from seven housekeeping genes of the *B. cereus* core genome.

Although all known *B. cereus* strains possess the genes for at least one enterotoxin (Nhe), their toxicity varies from innocuous to lethal. Hence, virulence regulation is far more complex than mere presence or absence of toxin genes. Further objectives of this study were therefore (i) to determine strain-specific and media-dependent differences in enterotoxin expression using bioluminescent reporter strains, (ii) to examine in detail the promoter regions of *nhe, hbl* and *cytK* to gain insight into the complex interplay of the various global regulators controlling their expression, and (iii) to characterize the influence of host epithelial cell factors on enterotoxin transcription (Chapter 3.2).

2. Material and Methods

2.1 Bacterial strains, plasmids, media and growth conditions

2.1.1 Strains and growth conditions

All *Bacillus cereus* sensu lato strains analyzed for phylogenetic studies are listed in Table S1. *B. cereus* strains were grown in either plate count (PC) liquid medium at 150 rpm or on PC plates at 30 °C for sequencing and in lysogeny broth (LB) medium for cloning purposes. For long-time storage, 850 µl of an overnight culture were mixed with 150 µl 87 % (v/v) glycerol (sterile) and frozen at -80 °C. Overnight cultures were inoculated from agar plates or angular agar and grown under shaking (150 rpm) in 3 ml medium supplemented (if necessary) with the appropriate antibiotic. Cloning steps were performed in *E. coli* (Table 1) grown in LB medium at 150 rpm or on LB agar plates at 37 °C.

Table 1: *Bacillus cereus* and *Escherichia coli* strains used for cloning and protein overexpression.

Strain	Genotype	Characteristics	Reference
Bacillus cereus			
NVH 0075-95	Wildtype	Stew with vegetables, food poisoning, Norway, 1995	[114]
INRA C3	Wildtype	Pasteurized carrot, Vaucluse, France, 1996	[139]
Escherichia coli			
TOP10	F^- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	General cloning host	Invitrogen
INV110	F^- {*traΔ36 proAB lac*I*q lacZΔM15*} *rps*L (StrR) *thr leu end*A *thi*-1 *lacY galK galT ara ton*A *tsx dam dcm supE44 Δ(lac-proAB) Δ(mcr*C-*mrr)102*::Tn*10* (TetR)	Methylase-deficient cloning host	Invitrogen
BL21 (DE3)	*B* F^- *omp*T *hsd*S_B(r_B^- m_B^-) *gal dcm* (DE3)	Expression of recombinant proteins	Novagen

2.1.2 Vectors and plasmids

All plasmids used and constructed in this study are given in Table 2.

Table 2: Vectors and recombinant plasmids used in this study.
pXen1-derivatives are described extensively in Fig. 18A and 19A and were transformed into their respective *B. cereus* host strain NVH 0075-95 or INRA C3.

Vector	Genotype and characteristics	Reference
pXen1	Gram-negative and Gram-positive shuttle vector, promoter-less luciferase (*luxABCDE*) operon for Gram-positive hosts, Amp^r, Cm^r	[171]
pXen1 [P*nhe52/lux*]	pXen1 containing the complete promoter region of the *B. cereus* NVH 0075-95 (WSBC 10552) *nhe* operon	This study
pXen1 [P*nhe52-s/lux*]	pXen1 containing a shortened promoter region of the *B. cereus* NVH 0075-95 (WSBC 10552) *nhe* operon	This study
pXen1 [P*nhe/lux*]	pXen1 containing the complete promoter region of the *B. cereus* INRA C3 (WSBC 10563) *nhe* operon	This study
pXen1 [P*nhe-s1/lux*]	pXen1 containing a shortened promoter region of the *B. cereus* INRA C3 (WSBC 10563) *nhe* operon	This study
pXen1 [P*nhe-s2/lux*]	pXen1 containing a shortened promoter region of the *B. cereus* INRA C3 (WSBC 10563) *nhe* operon	This study
pXen1 [P*nhe-s3/lux*]	pXen1 containing a shortened promoter region of the *B. cereus* INRA C3 (WSBC 10563) *nhe* operon	This study
pXen1 [P*nhe*-Δ5'UTR*/lux*]	pXen1 containing promoter region of the *B. cereus* INRA C3 (WSBC 10563) *nhe* operon with internal deletion	This study
pXen1 [P*hbl/lux*]	pXen1 containing the complete promoter region of the *B. cereus* INRA C3 (WSBC 10563) *hbl* operon	This study
pXen1 [P*hbl*-Δ5'UTR-up*/lux*]	pXen1 containing the promoter region of the *B. cereus* INRA C3 (WSBC 10563) *hbl* operon with internal deletion (5'UTR, upstream part)	This study
pXen1 [P*hbl*- Δ5'UTR-down*/lux*]	pXen1 containing the promoter region of the *B. cereus* INRA C3 (WSBC 10563) *hbl* operon with internal deletion (5'UTR, downstream part)	This study
pET28b(+)	IPTG-inducible *E. coli* expression vector for N- or C-terminal His_6-tag fusion proteins, T7*lac* promoter, Kan^r	Novagen
pET28-*codY*C3	Promoter-less *codY* of *B. cereus* INRA C3with N-terminal His_6-tag in pET28b(+), Kan^r	This study
pET28-*codY*CVUAS	Promoter-less *codY* of *B. cytotoxicus* CVUAS 2833 with N-terminal His_6-tag in pET28b(+), Kan^r	This study

2.1.3 Media

All media were prepared with ddH_2O and autoclaved for 15 min at 121 °C. When agar plates were prepared, 15 g/l agar was added to the medium before autoclaving. To study promoter activity several media with different amino acids, carbohydrates and mineral contents were used. Modified MOD medium [172, 173] was prepared with ddH_2O and autoclaved for 10 min at 110 °C. Stock solutions of 2 M glucose and trace elements were prepared with ddH_2O, sterilized by filtration (0.22 µm pore size), and added to the cooled MOD medium. To obtain MOD + 1 % casamino acids (CAA) and MOD + 1 % tryptone, additional 10 g/l CAA and tryptone, respectively, were autoclaved with the MOD medium components. CGY medium was prepared in a volume of 900 ml ddH_2O. After autoclaving 100 ml of filter sterilized 10 % glucose were added. Contents of all media used in this study are listed below.

LB:
10 g/l tryptone (peptone from casein)
5 g/l yeast extract
10 g/l NaCl

CGY:
20 g/l CAA
6 g/l yeast extract
2 g/l $(NH_4)_2SO_4$
14 g/l K_2HPO_4
6 g/l KH_2PO_4
1 g/l $Na_3C_6H_5O_7$ x 2 H_2O
2 g/l $MgSO_4$ x 7 H_2O
100 ml 10 % D-glucose

LB + 2 % CAA:
20 g/l CAA
5 g/l yeast extract
10 g/l NaCl

PC:
5 g/l tryptone
1 g/l D-glucose
2.5 g/l yeast extract

GSM (gut simulating medium):
RPMI 1640 liquid medium (Biochrom FG 1215)
1 % D-glucose
2 % CAA

cGSM (conditioned gut simulating medium):
filter sterilized GSM incubated for 22 h (37 °C, 7 % CO_2) on 14 d differentiated Caco-2 cells

MOD:

6 g/l $(NH_4)_2SO_4$
1 g/l K_2HPO_4 (1000x)
0.04 g/l $MgSO_4$ x 7 H_2O
2 g/l L-glutamic acid
0.39 g/l L-glycine
0.91 g/l L-valine
0.91 g/l L-threonine
0.4 g/l L-methionine
0.36 g/l L-histidine
0.46 g/l L-arginine
0.91 g/l L-aspartic acid
0.04 g/l L-cysteine
0.7 g/l L-isoleucine
1.37 g/l L-leucine
0.28 g/l L-phenylalanine
1.18 g/l L-lysine
0.66 g/l L-serine
0.042 g/l L-tyrosine
10 ml/l 2 M D-glucose

MOD trace elements: (1000x)

675 mg/l $FeCl_2$ x 6 H_2O
50 mg/l $MnCl_2$ x 4 H_2O
30 mg/l Na_2MoO_4 x 2 H_2O
275 mg/l $CaCl_2$
85 mg/l $ZnCl_2$
30 mg/l $CoCl_2$ x 6 H_2O
40 mg/l $CuSO_4$
24 mg/l $NaSeO_4$

The following antibiotics were added to media, if necessary:

Ampicillin	Amp^{120}	(120 µg/ml)
Chloramphenicol	Cm^{5}	(5 µg/ml)
Kanamycin	Kan^{50}	(50 µg/ml)
Spectinomycin	Spc^{200}	(200 µg/ml)

Chemicals used in this study were purchased from AppliChem GmbH, Carl Roth GmbH, Merck KGaA or Sigma-Aldrich Co. unless stated otherwise. For preparation of MOD medium chemicals with the highest purification grade available were used.

2.2 Transformation of bacteria

2.2.1 Transformation of *E. coli*

To prepare $CaCl_2$ competent *E. coli* cells [174], bacteria were grown in 100 ml LB medium to a cell density (OD_{600}) of 0.35 to 0.5. Cells were harvested (2700 g, 4 °C for 10 min) and incubated on ice for 10 min. Pellets were sequentially washed in 0.1 M $MgCl_2$ and 50 mM $CaCl_2$ and then incubated on ice for 30 min. After centrifugation they were resuspended in 1 ml 50 mM $CaCl_2$ + 15 % (v/v) glycerol and incubated on ice for 60 min. Aliquots of 100 µl were frozen in liquid nitrogen and stored at -80 °C. Competent *E. coli* were transformed with 50 – 100 ng DNA. Plasmid was added to the cells and the mixture was incubated on ice for 30 min. A heat shock treatment at 42 °C for 60 s followed. After cooling on ice, 900 µl LB medium were added and cells were regenerated for 1 h (37 °C, 150 rpm). 100 µl of cell suspension was plated on LB containing the appropriate antibiotic and incubated over night at 37 °C. In case of low transformation efficiency, cells were pelletized (1 min, 6000 g), supernatant discarded, pellet resuspended in the remaining liquid and plated on LB containing the appropriate antibiotic.

2.2.2 Transformation of *B. cereus*

To prepare electro competent *B. cereus*, cells were grown in 100 ml LB + 2 % (w/v) glycin to an OD_{600} of 0.4 to 0.7. After harvest (3500 g, 4 °C for 10 min), cells were washed sequentially in ice cold 2.5 %, 5 % and 10 % (v/v) glycerol. At last they were resuspended in 1 ml 10 % (v/v) glycerol, aliquots of 100 µl were frozen in liquid nitrogen and stored at -80 °C. Competent *B. cereus* cells were transformed with 1 µg DNA. The plasmid was added to the cells on ice and the mixture was pipetted into precooled and sterile electroporation cuvettes. Cells were electroporated in pre-cooled electroporation cuvettes with 2 mm gap (Peqlab) using the Gene Pulser™ (BioRad) at 2.0 kV, 25 µF, 200 Ω and 5000 V/cm. 900 µl LB medium was added and cells were regenerated for 2.5 – 3 h (30 °C, 150 rpm). Cells were centrifuged (1 min, 6000 g), supernatant discarded and the pellet resuspended in the remaining liquid. After plating on LB-plates containing the respective antibiotic, plates were incubated over night at 30 °C.

2.3 *De novo* sequencing and assembly of bacterial genomes

2.3.1 NGS sample preparation and Sequencing

DNA libraries were prepared using the TruSeq® DNA PCR-Free Sample Preparation Kit. The TruSeq® protocol was optimized (DNA shearing and fragment size selection) to improve assembly quality [175].

DNA libraries were sequenced on the Illumina MiSeq® system according to manufacturer´s instructions. The reagent kits used for sequencing of respective strains are indicated in Table 9.

2.3.2 Quality control of read data and assembly

Quality of raw sequencing data was assessed by FastQC (www.bioinformatics.bbsrc.ac.uk/projects/fastqc/), followed by quality filtering and trimming of reads with the NGS QC Toolkit v.2.3.2 [176]. Individual settings for each sequenced genome are summarized in Table 9. The program KmerGenie v.1.5924 [177] was used to calculate the best k-mer value for assembly from k = 23 to the maximal possible k = 223 in the default two passes. Contigs produced by the short read sequence assembler ABySS v.1.3.7 [178] (minimum scaffold and contig length 500 bp) were further quality controlled by QUAST v.2.2 [179] (using only contigs > 500 bp).

2.4 Sequencing of bacterial transcriptomes

2.4.1 Sample preparation and sequencing

Bacteria were grown stagnant in T75 cell culture flasks at 37 °C and 7 % CO_2 for 2 h in either GSM or cGSM (see 2.1) by Nadja Jeßberger (Lehrstuhl für Hygiene und Technologie der Milch, Tierärztliche Fakultät, LMU München). Cells were harvested in 5 ml aliquots and cell pellets snap-frosted in liquid nitrogen and stored at -80 °C. Total bacterial RNA was isolated as described in section 2.6.5, a minimum amount of 25 µg digested with RQ1 DNase I (Promega) and cleaned up by chloroform extraction. mRNA was purified and prepared for RT-PCR (Ferrari *et al.,* in preparation). cDNA libraries were prepared using the TruSeq™ Small RNA Kit and protocol. Libraries (15 pM) were sequenced single end (read length 50 bp) on the Illumina MiSeq® system without PhiX control DNA using the MiSeq® Reagent Kit v2 (50-cycles).

2.4.2 Analysis of transcriptome sequencing data

Quality of raw sequencing data was assessed by FastQC (www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Read data were aligned to their reference genome and its plasmids (.fasta files) using the web-server Galaxy [180] and the Bowtie2 [181] algorithm for short read alignment. The resulting SAM files were sorted, indexed and converted to BAM files as described previously [182]. Based on the GenBank file's annotations the number of reads overlapping a gene was calculated in Artemis v.16.0.0 [183]. All protein-coding and RNA-coding genes were downloaded from the NCBI FTP database (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/) as .ptt and .rnt files, respectively. Genes with less than 10 mapping reads were excluded. The Bioconductor package edgeR [184] was used for differential gene expression analysis as described previously [185]. Transcriptomes were determined in duplicates and only genes differentially expressed in both analyses were considered to be validly regulated.

2.5 *In silico* sequence analyses

2.5.1 Identification of virulence genes in *de novo* sequenced and assembled genomes

The newly generated contigs were aligned to reference genomes (*B. cereus* ATCC 14579, *B. cereus* F837/76 and *B. cytotoxicus* NVH 391-98) with progressiveMauve v.2.3.1 [186] and *nheABC, hblCDAB, cytK, plcR, adk, ccpA, glpF, glpT, panC, pta* and *pycA* were identified according to already annotated features. A second comparison of single contigs with CloneManager Suite 7 (Sci-Ed Software) was used to confirm gene locations and to control start, end and length of genes of interest.

2.5.2 Confirmation of the duplication of enterotoxin operons

Duplications of enterotoxin operons were found in some of the newly sequenced strains. These findings had to be verified to exclude the existence of duplications due to mis-assemblies. Alignment and coverage analyses were performed by Christopher Huptas (Lehrstuhl für Mikrobielle Ökologie, Zentralinstitut für Ernährungs- und Lebensmittelforschung, TU München). To confirm the presence of the second nhe_a operon trimmed and quality filtered read data was aligned separately against each of the suspicious *nhe* operons as well as the contigs on which the operons are located on. Read alignment was performed using BWA v.0.7.12 [187]. Subsequently, the resulting SAM files were converted into BAM format, whereby reads not mapping to the reference or not being part of a primary alignment to the reference were discarded. SAM file conversion and filtering was carried out using the 'view' utility of the SAMtools package v.0.1.18 [188]. Filtered BAM files served as input for the 'genomeCoverageBed' utility of the BEDTools suite v.2.17.0 [189] to obtain per-base sequencing depths of respective references in BED files. Sequencing depth histograms (BED files) were used to calculate each reference's median coverage. To confirm the presence of the second hbl_a operon the same approach of read remapping and filtering as for the *nhe* operons was applied. In addition to mapping reads to *hbl*,

hbl$_a$ and the contigs the operons are located on, reads were also mapped to an artificial sequence construct separating each strain's version of *hbl* and *hbl*$_a$ by a sequence of 5,000 'N' characters. The ratio of coverages cov_{operon} to cov_{contig} was used to show how well operons fit to their genomic backgrounds (contigs). The presence of uniquely mapping reads and the ratio of cov_{operon} to $cov_{construct}$ for each individual operon was used to prove the separate existence of the duplication.

2.5.3 Collection of data from databases

Further gene sequences and genomes were downloaded from NCBI and Patric databases until June 2014. All *B. cereus* sensu lato genomic sequences available at that time were scanned for the *nhe* operon using NCBI's BLAST. Since *B. anthracis* is known to be highly clonal, only five representative *B. anthracis* strains were included. A total of 223 strains containing *nheABC* were found and further analyzed for the presence of seven housekeeping genes. The genes *adk* (adenylate kinase), *ccpA* (catabolite control protein A), *glpF* (glycerol uptake facilitator), *glpT* (glycerol-3-phosphate transporter), *panC* (pantoate-β-alanine ligase), *pta* (phosphate acetyltransferase) and *pycA* (pyruvate carboxylase) were chosen to calculate a species tree. These housekeeping genes have already been selected for MLST [62], because they are scattered over the entire chromosome (http://mlstoslo.uio.no/cgi-bin/mlstdb/mlstdbnet4.pl?dbase=optimized&page=scheme-optimized&file =bcereusgrp_isolates.xml). Due to draft status and partially insufficient sequence quality of genome sequences from the databases, selected housekeeping gene sequences could not be identified for all 223 *B. cereus* sensu lato strains. Hence, the final set of strains was reduced to 142 *B. cereus* sensu lato strains (Table S1). Sequences of the final set are available in the NCBI database either completed or as draft genome. Accession numbers for the *de novo* assembled genomes reported in this work are given in Table 9.

2.5.4 Detection of potential recombination events

Putative recombination events within enterotoxin operons and a graphical representation of differences within multiple alignments were calculated by RDP3 [190]. In a very conservative approach only statistically proven recombination events according to the following criteria were considered: a recombination event i) was detected by at least three of the programs implemented in RDP3, ii) showed a maximum average p-value of < 0.05 and iii) both parental sequences are part of the investigated strainset.

2.5.5 Determination of phylogenetic relationships based on single gene and whole genome sequences

2.5.5.1 Average nucleotide identity

Pairwise average nucleotide identity (ANIb) of 142 *B. cereus* sensu lato genomes was calculated by Christopher Huptas for all possible distinct pairs according to the algorithm described before [191]. For calculation the script ANI.pl (by Jiapeng Chen) available at https://github.com/chjp/ANI was used taking one strain of a pair as query and the other one as reference and vice versa. The resulting two ANIb values for each pair were averaged and these 10011 values served as input for a neighbor-network computed by SplitsTree4 (version 4.13.1) [87].

2.5.5.2 Whole genome phylogenetic analyses

The newly generated contig data were compared to available genomes in a fragmented alignment approach using the program Gegenees 2.2.1 [192]. The fragmented all-all comparison was performed with BLAST 2.2.29+, at a step size of 200 bp and a comparison step size of 100 bp. Thus, contigs smaller than 200 bp were excluded. Resulting data are displayed as heatplot of similarity by Gegenees or as neighbor-network by SplitsTree4 calculated from an exported distance matrix.

Single nucleotide polymorphisms (SNPs) were detected in the entire genomes of the *B. cereus* sensu lato strain set using the program kSNP3 v.3.0 [193] at k-mer size 21 (determined by Kchooser, implemented in kSNP3). The most distant cluster, the two *B. cytotoxicus* strains (phylogenetic cluster VII) had to be excluded to obtain an FCK (fraction of k-mers present in all genomes) value > 0.1, which is necessary for adequate SNP detection efficiency. The resulting core SNP matrix (includes only SNP loci that are present in all investigated strains [194]) was basis for the calculation of a phylogenetic tree with MEGA6 [195] (Maximum likelihood method, Tamura-Nei model [196], uniform rates, using all sites, bootstrap 1000).

2.5.5.3 Phylogenetic trees

Multiple DNA sequence alignments were generated online with ClustalΩ [197] and used to compare genes. Alignments of (concatenated) genes served as input for MEGA6 [195]. The species tree of *B. cereus* sensu lato strains was built from concatenated DNA sequences of seven housekeeping genes. The genes *adk, ccpA, glpF, glpT, panC, pta,* and *pycA* were taken from the optimized MLST scheme developed by Tourasse *et al.* [62], because they are evenly distributed over the entire *B. cereus* chromosome. In contrast to the MLST-approach, entire genes were used in this comparison to display relationships as accurately as possible. To compare the operons *nheABC* and *hblCDAB* concatenated gene sequences without intergenic regions were used. Phylogenetic trees of genes or concatenated genes were calculated in MEGA6 using the maximum likelihood method based on the Tamura-Nei model with

a discrete Gamma distribution, permitting some invariant sites (TN93+G+I). This is the determined ideal substitution model ('Find Best DNA/Protein Models (ML)' function of MEGA6) for the species tree dataset containing seven concatenated housekeeping genes of 142 strains and was applied to the calculation of all phylogenetic trees. Maximum likelihood is one of the tree searching methods that utilize discrete data instead of clustering methods, such as neighbor joining, that calculate phylogenetic trees solely on the basis of distance (percent sequence difference) between sequences [198]. All sites were used and reliability was tested by 1000 bootstrap repeats. Branchpoints with bootstrap values ≥ 70 are considered reliable. The phylogenetic trees were not rooted to a certain outgroup to allow comparison of subsets and toxin genes that only occur in some strains. For multiple sequence alignment and calculation of phylogenetic trees from nucleotide sequences it was set as a prerequisite that all input sequences have the same length, because they are automatically cut during calculation (deletions within sequences were not removed). Thus, sequences were shortened to the 'lowest common denominator' to ensure that all sequences of an alignment start and end together. While the housekeeping genes were of the same length in all strains, the toxin genes did not show uniform lengths and required cutting of sequence ends. To ensure comparability, the same settings were used for the calculation of all phylogenetic trees. To further control the reliability of the maximum likelihood approach, all phylogenetic tree calculations were repeated using the neighbor joining and minimum evolution methods implemented in MEGA6.

2.5.6 Bioinformatic analysis of *nhe*, *hbl* and *cytK* 5' intergenic regions

Multiple sequence alignments were calculated using ClustalΩ (https://www.ebi.ac.uk/Tools/msa/clustalo/) and sequence conservation was graphically represented as logo (http://weblogo.threeplusone.com/create.cgi). The following strains were included in the comparative analysis of intergenic regions.

Table 3: List of representative *B. cereus* sensu lato strains for comparison of intergenic regions.
All strains possess the *nhe* operon, while strains marked by an x are included in the respective analysis.

			Analysis of intergenic regulatory region:			
Number	**Cluster**	**Strain name**	***nhe***	***hbl***	***cytK-1***	***cytK-2***
3	III	*B. cereus* NVH 0075-95	x			
5	III	*B. cereus* HWW 274-2	x			x
15	IV	*B. cereus* ATCC 14579	x	x		x
28	III	*B. anthracis* Ames Ancestor	x			
56	II	*B. cereus* BAG6X1-1		x		
57	VII	*B. cytotoxicus* NVH 391-98			x	
65	IV	*B. thuringiensis* HD-771	x	x		
67	IV	*B. thuringiensis* IBL 200				x
69	V	*B. thuringiensis* MC28		x		
71	IV	*B. thuringiensis* s. berliner ATCC 10792	x	x		x
73	III	*B. thuringiensis* s. finitimus YBT-020		x		
74	IV	*B. thuringiensis* s. huazhongensis BGSC 4BD1		x		
80	III	*B. thuringiensis* s. pulsiensis BGSC 4CC1				x
87	VI	*B. weihenstephanensis* WSBC 10204	x	x		
91	III	*B. cereus* F837/76	x	x		
112	VII	*B. cytotoxicus* NVH 883/00			x	
117	V	*B. cereus* Rock3-28	x	x		
121	V	*B. cereus* Rock4-18				x
123	IV	*B. cereus* VD014				x
131	IV	*B. cereus* VD156	x	x		
139	III	*B. cereus* MHI 86	x			x
140	II	*B. cereus* MHI 226	x			
141	III	*B. cereus* SDA KA 96	x	x		x
144	II	*B. cereus* 14294-3 (M6)	x	x		x
150	VI	*B. mycoides* DSM 2048	x	x		
155	VI	*B. cereus* BAG5X1-1	x	x		
156	II	*B. cereus* BAG5X2-1		x		
180	V	*B. cereus* VD115	x			
201	IV	*B. cereus* BAG1X2-2				x
211	II	*B. cereus* BAG2O-3	x			
236	IV	*B. cereus* #17	x	x		x
238	IV	*B. cereus* RIVM BC 964	x	x		
241	IV	*B. cereus* INRA C3	x	x		x
242	III	*B. cereus* F528/94	x	x		
243	II	*B. cereus* RIVM BC 126	x	x		

245	IV	*B. cereus* 6/27/S	x	x		x
247	IV	*B. cereus* F3175/03	x	x		
248	III	*B. cereus* F3162/04	x			
249	VII	*B. cytotoxicus* CVUAS 2833			x	
288	IV	*B. bombysepticus* str. Wang				x
289	V	*B. toyonensis* BCT-7112	x	x		

The web-based program ORF Finder (http://www.bioinformatics.org/sms2/orf_find.html) was used to detect any potential open reading frames (ORFs) embedded in the 5'UTRs of enterotoxin genes. The 5'UTR RNA sequences were further analyzed for similarity to known RNA families with Rfam v.12.0 [199] and potential RNA secondary structures were calculated at default settings with Mfold v.4.6 [200].

2.6 Nucleic acid techniques

2.6.1 Isolation of genomic DNA

Genomic DNA from *B. cereus* was isolated using a modified CTAB (cetyl trimethylammonium bromide) DNA preparation method [201]. *B. cereus* overnight cultures were pelleted and solved in 567 µl TE buffer (pH 8.0, sterile filtrated). Cells were then lysed using a Fastprep® 24 instrument (M. P. Biomedicals, 0.1 mm zirconia beads). Proteins in the supernatant were denatured by adding 30 µl 10 % (w/v) SDS and 3 µl proteinase K (20 mg/ml) and incubated at 37 °C for 3 h. Then, 100 µl 5 M NaCl were added and carefully mixed. 80 µl CTAB/NaCl (10 % CTAB in 0.7 M NaCl) were added, mixed carefully and incubated at 65 °C for 30 min to complex nucleic acids. To separate proteins from nucleic acids phenol chloroform extraction was performed. Nucleic acids were precipitated by adding 0.6 volumes of ice-cold isopropyl alcohol at 4 °C for at least 15 min. Samples were washed twice in ice-cold 70 % (v/v) ethanol, air-dried and dissolved in sterile ddH_2O at 4 °C overnight. RNA digestion with 20 µg RNAse for 30 min at 37 °C followed by a second phenol chloroform extraction ensured that only genomic DNA remained. Samples were washed twice in ice-cold 70 % (v/v) ethanol, air-dried and dissolved in sterile ddH_2O at 4 °C overnight. Quality of DNA isolation was tested by agarose gel electrophoresis and spectroscopy (Nanodrop™ spectrometer). DNA concentration for NGS was determined using the Qubit™ dsDNA High Sensitivity Assay (Qubit® 2.0 fluorometer).

2.6.2 Polymerase chain reaction (PCR)

The *Pfu* DNA polymerase (Promega) was used for high fidelity DNA amplification for cloning purposes. All other PCRs were performed with *GoTaq* G2 DNA polymerase (Promega). Colony PCR was used to detect inserts or plasmids after transformation. Cell material was dissolved in 200 µl ddH_2O, lysed using the Fastprep® 24 instrument, centrifuged (2 min, 15700 g), and 2 µl of the supernatant were used as PCR template. Annealing temperatures depended on the PCR type and primer melt temperatures. Primers used in this study are listed in tables 4 – 8. dNTPs were purchased from Thermo Scientific and mixed (20 mM, 5 mM each). PCR reactions were performed in a Primus 25 advanced® thermocycler (Peqlab).

***Pfu* PCR (50 µl volume)**		**Program**		
40.5 µl	ddH_2O	Initial denaturation	94 °C, 2 min	
5 µl	*Pfu* 10x buffer	Denaturation	94 °C, 30 s	30
1 µl (50 pmol)	Forward primer	Annealing	T_m – 5 °C, 30 s	
1 µl (50 pmol)	Reverse primer	Elongation	72 °C, 2 min/kb	
1 µl	dNTPs	Final elongation	72 °C, 8 min	
0.5 µl	*Pfu* polymerase			

***GoTaq* PCR (50 µl volume)**		**Program**		
32 µl	ddH_2O	Initial denaturation	95 °C, 2 min	
10 µl	*GoTaq* 5x buffer	Denaturation	95 °C, 30 s	30
1 µl (50 pmol)	Forward primer	Annealing	T_m – 5 °C, 30 s	
1 µl (50 pmol)	Reverse primer	Elongation	72 °C, 1 min/kb	
1 µl	dNTPs	Final elongation	72 °C, 2 min	
0.2 µl	*GoTaq* G2 polymerase			

2.6.3 Plasmid isolation

Plasmids were isolated from 6 ml overnight cultures using the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich). To isolate plasmids from *B. cereus*, cells were pelletized, resuspended according to manufacturer's instructions and disrupted by bead beating in a Fastprep® 24 instrument. DNA concentrations were determined by the Nanodrop™ spectrometer.

2.6.4 Enzymatic modification and purification of DNA

Restriction modification of DNA was used to insert fragments orientation dependent into the plasmid. Since all fragments were inserted via two different restriction sites, dephosphorylation was not necessary. Insert DNA (1 µg) and vector DNA (3 µg) was digested sequentially at 37 °C for 2 h with 3 – 9 U of restriction enzymes (Fermentas, Thermo Scientific) using the appropriate buffer system. After each digest DNA was purified using the E.Z.N.A Cycle Pure Kit (Omega Bio-tek). If necessary, fragments were purified from agarose gels with the E.Z.N.A Gel Extraction Kit (Omega Bio-tek) or concentrated using the vacuum concentrator 5301 (Eppendorf). Ligation of vector (100 ng) and insert was performed in molar ratios of 1:3 to 1:6 according to fragment lengths using T4 DNA ligase (Fermentas, Thermo Scientific) at room temperature for at least 1 h. Ligation preparations were transformed immediately as described in section 2.2.

2.6.5 Isolation of RNA

Bacteria were cultured in CGY medium [202] supplemented with 1 % (w/v) glucose as described previously [157] to an OD_{600} of 4 or for 6 h. Six ml of the culture were harvested, the cell pellets were snap-frosted in liquid nitrogen and stored at -80 °C. Total RNA was isolated from *B. cereus* using the Trizol RNA isolation method [157]. Cells were resuspended in 1 ml Trizol reagent (Sigma-Aldrich) and disrupted using ZnSilica bead beating in a Fastprep® 24 (MP Biomedicals, settings: 6.5, 45 s, three times). 200 µl chloroform were added, followed by 2 min incubation on ice and centrifugation (12 000 g, 15 min, 4 °C). Nucleic acids were precipitated by adding 500 µl isopropanol to the aqueous supernatant and centrifugation (12 000 g, 10 min, 4 °C). After two washing steps in 1 ml ice cold 70 % ethanol, nucleic acid was dried and solved in nuclease-free water. Subsequently, DNA was digested (RQ1 DNase, Promega) as described previously [157] and RNA was tested for residual DNA contamination by PCR using primers 16S_for and 16S_rev (Table 5, section 2.6.7). RNA purity and quantity was determined by measurement of absorption at 260 nm and 280 nm (Nanodrop™ spectrometer).

2.6.6 Reverse Transcriptase – PCR

First strand synthesis was performed using the qScript™ cDNA Supermix (Quanta Biosciences). Subsequent PCR (annealing temperature 56 °C) contained 50 pmol of each primer (Table 4), 5 µl green GoTaq® G2 reaction buffer, 0.05 mM of each dNTP, 62.5 mM $MgCl_2$ and 0.5 U GoTaq® G2 DNA Polymerase (Promega) in a volume of 25 µl. Amplification of the *rrn* 16S rRNA gene transcript served as a positive control, nuclease-free H_2O as a negative control. RT-PCR results were visualized on 2 % (w/v) agarose gels.

Table 4: Primer used in RT-PCR.

Primer	Target	Sequence [5‘ – 3‘]	Source
16Sf	Inner fragment of 16S rRNA gene *rrn*	GAC GTC AAA TCA TCA TGC C	This work
16Sr		GAT TCC AGC TTC ATG TAG G	
nheAaf	Inner fragment of *nheA*$_a$	CTA GTA AAG TTA GCA GAG CG	This work
nheAar		TTT CTT TTG GTA GAG CTA GAA G	
nheBaf	Inner fragment of *nheB*$_a$	TTA TAT TGC ATC GTC GGT TG	This work
nheBar		TTA TCT GCT GCT GCG ATG	
nheCaf	Inner fragment of *nheC*$_a$	CTA GAT AAC GTG GTG GC	This work
nheCar		TTC CGT TTT ATT TTT GGC ATC	
nheA_qRT_for	Inner fragment of *nheA*	AAG TAC AAA GCA TCC AAG AGA	Viktoria Krey
nheA_qRT_rev		ACA ATA TCT CCA CTT GAT CCT T	
nheB_qRT_for	Inner fragment of *nheB*	GTG AAA CAA GCT CCA GTT C	[157]
nheB_qRT_rev		AAA GCG TAC AGA TCC ATT ACT	
nheC_qRT_for	Inner fragment of *nheC*	GCA AAT GCA GAA A/CGA GAA AT	Viktoria Krey
nheC_qRT_rev		CCT ACT GTA TAC CAT TGA TTT GA	

2.6.7 Determination of transcription start sites

DNA-free RNA was used as template for 5'RACE according to manufacturer's instructions (5'RACE system for rapid amplification of cDNA ends, version 2.0, Invitrogen). All primers needed to detect transcription start sites of *nhe* and *hbl* are listed below.

Table 5: Primer used in 5'RACE.

Primer	Sequence [5' → 3']	Reference
GSP1_hbl63	TAACTGTACATTAGGAC	This study
GSP1_nhe63	CTAATTGAATTGGATAATG	This study
AAP	GGCCACGCGTCGACTAGTACGGG**II**GGG**II**GGG**II**G	Invitrogen
AUAP	GGCCACGCGTCGACTAGTAC	Invitrogen
GSP2_hbl63	CTGCTTGAACGATTGGTGTTGCGAGAG	This study
GSP2_nhe63	CTCGTAGATACTGCTGTAACCAATAACCCTG	This study
16S_for	AGAGTTTGATCCTGGCTC	Viktoria Krey
16S_rev	CGGCTACCTTGTTACGAC	Viktoria Krey

Resulting PCR fragments were analyzed on 1 % agarose gels, controlled by sequencing (GATC Biotech) and compared to genomic data to determine transcription start sites.

2.7 Agarose gel electrophoresis

Length of DNA fragments and purity of DNA/RNA was controlled by agarose gel electrophoresis. Depending on fragment lengths either 1 % or 2 % (w/v) agarose (Biozyme, Bioline) were solved in 1x TAE buffer (50x TAE: 2 M Tris, 1 M acetic acid, 0.05 M Na_2EDTA pH 8). Gels were run in horizontal PerfectBlue™ Mini Gel Systems (Peqlab) at a constant voltage of 110 V for at least 45 min depending on fragment size. Subsequently, gels were stained with GelRed™ (Biotium) and photographed under UV light (UVsolo TS Imaging System, Biometra – Analytik Jena). The GeneRuler™ DNA Ladder Mix (Fermentas, Thermo Scientific) was loaded to estimate DNA sizes.

2.8 Construction of bioluminescent *B. cereus* reporter strains

To construct bioluminescent reporter strains the promoter region of interest was inserted into the *E. coli/Bacillus* shuttle vector pXen1 in front of the *lux* cassette (Fig. 3) via the restriction sites for EcoRI and BamHI and introduced into *B. cereus* by electroporation. The plasmid pXen1 contains the *luxABCDE* genes with ribosomal binding sites (RBS) for gram positive bacteria in front of each *lux* gene, but no promoter for the *lux* operon [171]. This system takes advantage of the *Photorhabdus luminescens lux* operon, which does not require the addition of external substrate and encodes enzymes

that are stable at elevated temperatures [203]. The emission of blue-green light is based on the oxidation of reduced riboflavin phosphate ($FMNH_2$) and a long-chain fatty aldehyde (RCHO) to its fatty acid [204]. The most common substrate of the bacterial luciferase is *n*-decyl aldehyde [205]. The genes *luxA* and *luxB* encode the luciferase subunits, while *luxCDE* are coding for the fatty acid reductase complex that catalyzes the regeneration of the aldehyde.

$$FMNH_2 + RCHO + O_2 \rightarrow FMN + H_2O + RCOOH + h\nu\ (490\ nm)$$

$$RCOOH + ATP + NADPH \rightarrow NADP + AMP + PP_i + RCHO$$

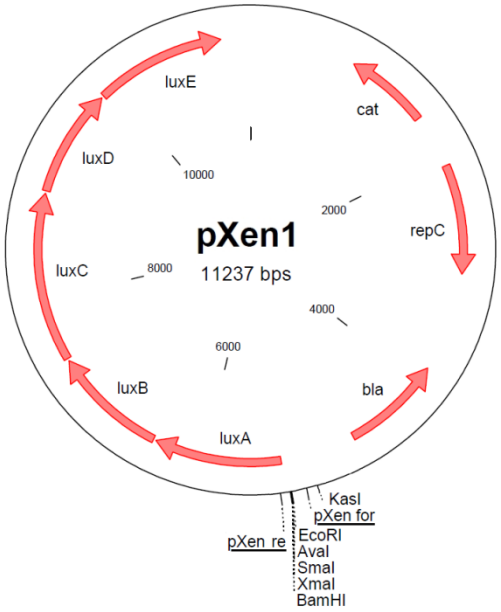


Fig. 3: Map of pXen1.

The plasmid contains the promoter-less *lux*-operon *luxABCDE* (RBS for gram-positive bacteria upstream of each *lux* gene), *cat* (chloramphenicol acetyltransferase) and *bla* (β-lactamase) genes for selection in *E. coli/B. cereus* [171]. Restriction sites of the multiple cloning site (mcs) and primer for the control of successful integration of promoter fragments are shown.

Primers were designed with Clone Manager Suite 7 and used to amplify promoter regions of interest with the *Pfu* polymerase (see section 2.6.2). The resulting DNA fragment and the reporter plasmid pXen1 were cut with restriction enzymes EcoRI and BamHI (see section 2.6.4).
The digested promoter fragment was ligated into pXen1 and the resulting plasmid was named according to the WSBC number of the *B. cereus* strain that supplied the promoter fragment, e.g. pXen1 [P*nhe52*/*lux*] containing the promoter region of NVH 0075-95 (WSBC 10552) (Table 2 and 6). The primer pair pXen for / pXen re (Table 6, Fig. 3) was used to control correct insertion and sequence within the resulting plasmids. These cloning steps were performed in *E. coli* TOP10 taking advantage of the ampicillin resistance (*bla* gene). The newly constructed promoter fusion plasmid was then transformed into the non-methylating *E. coli* INV110, isolated and introduced into *B. cereus* by electroporation. Transformation into the non-methylating *E. coli* INV110 is necessary to avoid cleavage of the plasmid by the *B. cereus* restriction system.

Table 6: Primer used in the construction of pXen1 promoter fusion plasmids.

Primer	Sequence [5´→3´]	Restric-tion site	Product and target
Nhe52 for	ATCCGAATTCTGTATACGCTATGC	EcoRI	**pXen1 [P*nhe52/lux*]**,
Nhe rev	AGTCGGATCCACTTTAACTCCTC	BamHI	554 bp promoter region
Nhe52-s for & Nhe rev	AATGAATTCACTGTATCAATGTGGG	EcoRI	**pXen1 [P*nhe52-s/lux*]**, 247 bp promoter region
Nhe for & Nhe rev	ATCCGAATTCTGTATATGCTATGC	EcoRI	**pXen1 [P*nhe/lux*]**, 554 bp promoter region
Nhe-s1 for & Nhe rev	GGAGAATTCAGCTTGAAAATAAAGGG	EcoRI	**pXen1 [P*nhe-s1/lux*]**, 406 bp promoter region
Nhe-s2 for & Nhe rev	AATGAATTCACTGTGTGAATGTGGG	EcoRI	**pXen1 [P*nhe-s2/lux*]**, 247 bp promoter region
Nhe-s3 for & Nhe rev	GCAGAATTCGTTGGGAGAGATG	EcoRI	**pXen1 [P*nhe-s3/lux*]**, 138 bp promoter region
Pnhe_A	AAAATTCATTGAACTATGCG	-	
Pnhe_B	TTTCCCGGGAGTTTTGTACTGTAAGGTC	XmaI	**pXen1 [P*nhe*-Δ5'UTR*/lux*]**,
Pnhe_C	AAACCCGGGTAATAAAATGCGATAGTAAT	XmaI	289 bp promoter region
Pnhe_D	TCATCGGCTTTAATTGATAAG	-	
Hbl for	TATCGAATTCTTATGCAATTATAC	EcoRI	**pXen1 [P*hbl/lux*]**, 898 bp
Hbl rev	CTTGGATCCGTACACTC	BamHI	promoter region
Phbl_A	ATATAAGCTTGTTATCCGCTG	-	
Phbl_B	TAGCCCGGGCTTATCCTTTCTGTCTGG	XmaI	**pXen1 [P*hbl*-Δ5'UTR*/lux*]**,
Phbl_C	AACCCCGGGAAAGGAGTGTACGGAATG	XmaI	344 bp promoter region
Phbl_D	CCTGGTATTAATGCAATTTGC	-	
Phbl-s1-B & Phbl_A	GCTCCCGGGTAATCTTGCCATAACATTG	XmaI	**pXen1 [P*hbl*-Δ5'UTR-down*/lux*]**, 612 bp promoter region
Phbl-s2-C & Phbl_D	CACCCGGGACTGTAAGCATTGGTTTATTG	XmaI	**pXen1 [P*hbl*-Δ5'UTR-up*/lux*]**, 636 bp promoter region
pXen for	GTTGGGTAACGCCAGGG	-	Different length, 99 bp
pXen re	CATAGAGAGTCCTCCTCTTG	-	without insert, multiple cloning site of pXen1

To screen clones for the presence of the plasmid, colony PCR with Go*Taq* polymerase (Promega) was performed as described in section 2.6.2. After a positive result the plasmid was isolated, a PCR with *Pfu* polymerase and pXen for / pXen re was performed and the product sequence was controlled by sequencing (GATC Biotech).

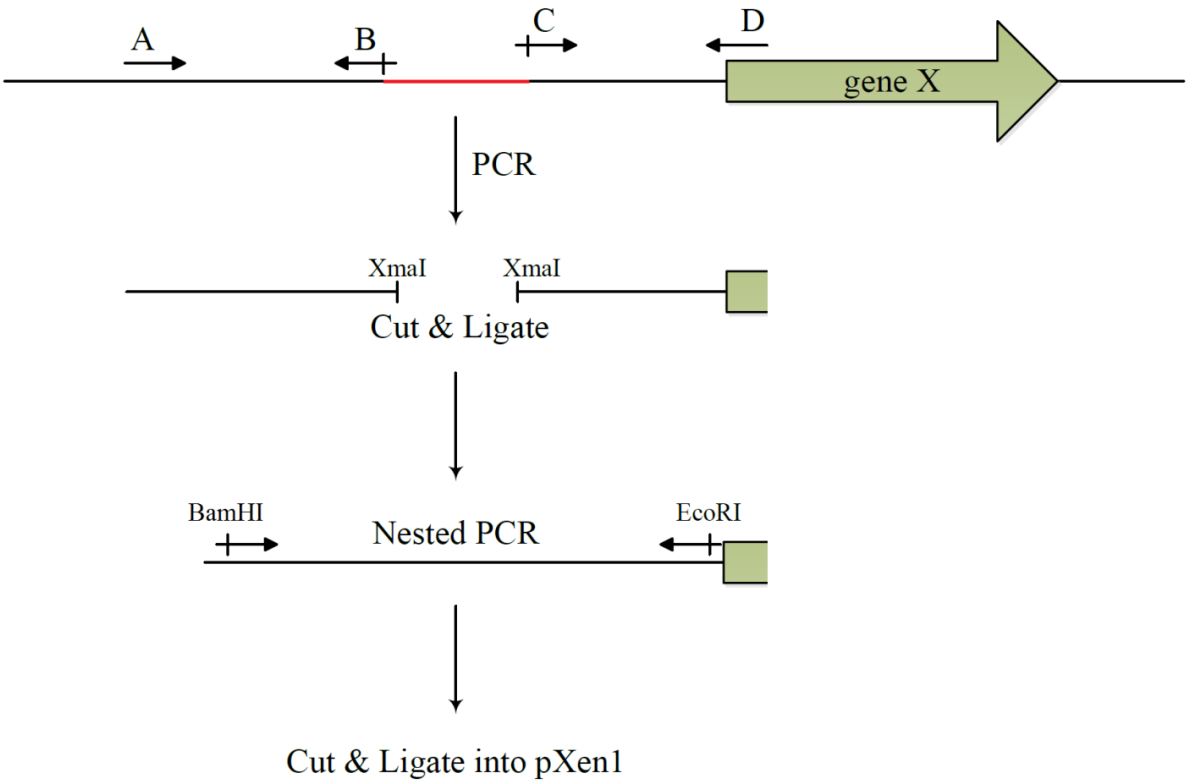


Fig. 4: Construction of pXen1 promoter fusions containing internal deletions.
Primerpairs AB and CD were used for amplification of adjacent promoter fragments. Restriction sites for XmaI were inserted via primers B and C. Cutting and ligation of the promoter fragments led to the deletion of the region marked in red. A nested PCR with primers containing restriction sites for EcoRI and BamHI allowed directional insertion of a promoter region with internal deletion into pXen1.

All pXen1 promoter fusion constructs and primers used to construct them are listed in Table 6. Promoter fusions that contain internal deletions were built by amplification of two adjacent regions and introduction of XmaI restriction sites (Fig. 4). After cutting and ligation of the fragments, a nested PCR led to the desired promoter region with internal deletion. This fragment was inserted into pXen1 as described above.

2.9 Promoter activity detection in bioluminescent reporter strains

Promoter activity was monitored under different conditions using the Victor3™ multilabel plate reader (Perkin Elmer). Bacteria were grown in pre-cultures of 3 ml of the tested medium for precisely 15 h, 150 rpm, at either 30 °C or 37 °C. For selection of pXen1 Cm^5 was added to the medium. 100 µl of the overnight culture were diluted 1:10 in medium. After vortexing, 20 µl of the dilution were added to 2 ml medium + Cm^5 in a 2 ml reaction tube and mixed thoroughly (final dilution 1:1000). A white clear-bottom microwell plate (NUNC) was filled with 200 µl per well and incubated at 800 rpm at the respective temperature. Each condition was tested four times to calculate mean values and for each strain three biological replicates were investigated. Cell density (OD_{600}) and luminescence (490 nm, 0.1 s) were measured every hour.

To monitor enterotoxin promoter activities under aerobic or gut-simulating conditions 45 ml of medium was inoculated with *B. cereus* at an OD_{600} of 0.05 from appropriate pre-cultures. When promoter activity was studied under aerobic conditions, cultures were vigorously shaken at 30 °C or 37 °C, 150 rpm in baffled 500 ml flasks. Analysis under gut-simulating conditions required unshaken cultures in T75 cell culture flasks at 37 °C and 7 % CO_2. Cell density (OD_{600}) and luminescence (490 nm, 0.1 s) were measured every hour.

2.10 Cell culture

Human Caco-2 epithelial colorectal adenocarcinoma cells were used to produce conditioned gut-simulating medium (cGSM). Caco-2 cells were grown in RPMI 1640 medium (Biochrom FG 1215) + 10 % fetal bovine serum (FBS) (Pan™ Biotech) at 37 °C and 7 % CO_2 atmosphere. Cells were splitted when they reached 90 – 100 % confluence. Cells were washed in PBS + 0.001 % EDTA (pH 7.5) and subsequently incubated in 1x Trypsin + 0.001 % EDTA (Biochrom AG) until they detached from cell culture flasks. To deactivate trypsin, medium containing FBS was added and cells were harvested at 500 g, 5 min. Cells were resuspended in fresh medium and seeded in new culture flasks in the desired amount. Cells were counted in a Neubauer improved hemocytometer (0.100 mm depth, 0.0025 mm^2, Brand).

Gut-simulating medium (GSM) comprised RPMI 1640 medium, 1 % D-glucose and 2 % CAA. To prepare conditioned GSM an initial number of $2.15*10^6$ cells were seeded in a T75 flask and allowed to differentiate for 14 days with medium changes every second day. Differentiated cells were incubated in GSM for 22 h and the resulting cGSM was sterile filtrated (0.22 µm, Millex® GP, Merck Millipore) for immediate use. Cell culture equipment was purchased from AppliChem GmbH, Biochrom AG, Carl Roth GmbH, Merck KGaA, Honeywell Riedel-deHaen Inc., Serva GmbH, Sigma-Aldrich Co. and TPP Techno Plastic Products AG.

2.11 Protein biochemistry

2.11.1 Overexpression and purification of CodY

CodY was overexpressed in *E. coli* BL21(DE3) as a soluble N-terminal His_6-tag-fusions using the plasmid pET28b(+) as described previously [170]. Primers containing restriction sites for NdeI and XhoI (Table 7) were used to construct pET28-*codY*C3 and pET28-*codY*CVUAS. Cloning steps were performed in *E. coli* TOP10 and the sequence of the resulting overexpression plasmid was controlled via sequencing with primers pET28b for and pET28b rev. Subsequently, pET28-*codY*C3 and pET28-*cod*YCVUAS was transformed into chemically competent *E. coli* BL21(DE3). Overnight cultures were diluted 1:100 and grown at 37 °C, 150 rpm in baffled flasks. When they reached exponential growth (OD_{600} 0.6 – 0.8), protein expression was induced by the addition of 1 mM IPTG. After 5 h incubation, cells were harvested (6000 g, 10 min, 4 °C), frozen in liquid nitrogen and stored at -80 °C.

Table 7: Primer used in the construction of CodY overexpression plasmids.

Primer	Sequence [5' → 3']	Restriction site	Reference
CodY-C3-for	CTTTT**CATATG**GAATTATTAGCAAAAACG	NdeI	This study
CodY-C3-rev	CTC**CTCGAG**GGAGAGTTTTTTATAAATTA	XhoI	This study
CodY-CVUAS-for	GAACTTTT**CATATG**GAATTATTAGCAAAAAC	NdeI	This study
CodY-CVUAS-rev	ATT**CTCGAG**GAAAGCTTTTTACTTACATTAG	XhoI	This study
pET28b for	GTGATGTCGGCGATATAGGC	-	This study
pET28b rev	GCTTCCTTTCGGGCTTTGTTAG	-	This study

Cells were resuspended in 5 ml (1/100 volume) native lysis buffer containing 1 mM Pefabloc SC (Merck) and 25 U/ml Benzonase endonuclease (Quiagen), disrupted with a French® Pressure Cell Press (SLM Aminco) four times at 900 psi and cellular debris was removed by double centrifugation (9000 g, 15 min, 4 °C). Soluble His-tag proteins were purified using the Äkta purifier (Amersham Biosciences) with a Frac-950 fractionator. A step-wise elution was chosen with the imidazole concentration rising from 10 mM to 83.5 mM, 304 mM, and 402 mM to a final concentration of 500 mM imidazole (100 % elution buffer). The elution fractions suspected to contain the purified His_6-CodY were controlled on a 15 % SDS-polyacrylamide gel. CodY-containing fractions were pooled and buffer composition was adjusted to buffer BS using ultrafiltration columns with a 10 kDa cut-off (Amicon Ultra-15, Merck Millipore). Protein purity was analyzed on a 15 % SDS-polyacrylamide gel and His_6-CodY was frozen in aliquots at -80 °C.

Lysis buffer (native)
50 mM NaH_2PO_4
300 mM NaCl
10 mM imidazole pH 8.0

Buffer BS
50 mM Tris-HCl (pH 7.5)
50 mM KCl
10 mM $MgCl_2$
0.5 mM Na_2EDTA (pH 8.0)
10 % glycerol

Elution buffer
50 mM NaH_2PO_4
300 mM NaCl
500 mM imidazole pH 8.0

2.11.2 Protein quantification

Total protein concentration of a His_6-CodY aliquot was determined in triplicates. The Roti®-Quant protein assay (Roth) was used with bovine serum albumin as standard (0 – 100 ng/µl) according to manufacturer's protocol in a microtiter plate. Optical density was measured with the Victor3™ multilabel plate reader (Perkin Elmer) at 600 nm.

2.11.3 Denaturing polyacrylamide gel electrophoresis

Size and purity of proteins was determined by SDS-polyacrylamide gel electrophoresis [206]. Samples (including Laemmli buffer) were heated to 100 °C for 5 min to denature proteins. Subsequent electropohoresis was performed in vertical chambers (Minigel-Twin, Biometra) at 18 mA and 350 V until the dye front ran off the gel. PageRuler™ prestained protein ladder (Thermo Scientific) was used to estimate protein sizes. Gels were stained in Coomassie staining solution overnight and destained for at least 1.5 h in destaining solution with paper towels to absorb excessive dye.

Resolving gel (15 %)
3.45 ml ddH_2O
3.75 ml 1.5 M Tris-HCl (pH 8.8)
7.5 ml Rotiphorese® Gel 30 (37.5:1)
0.15 ml 10 % (w/v) SDS
0.15 ml 10 % (w/v) APS
0.01 ml TEMED

Stacking gel (5 %)
3.7 ml ddH_2O
0.31 ml 1 M Tris-HCl (pH 6.8)
0.83 ml Rotiphorese® Gel 30 (37.5:1)
0.05 ml 10 % (w/v) SDS
0.05 ml 10 % (w/v) APS
0.01 ml TEMED

Laemmli buffer (5x)
100 mg Bromphenolblau
3.5 ml glycerol 99 %
1.5 g SDS
3.2 ml 1 M Tris-HCl pH 6.8
2.5 ml 2-Mercaptoethanol

SDS running buffer (10x)
0.25 M Tris
1.92 M glycine
1 % (w/v) SDS pH 8.3

Coomassie staining
0.25 % (w/v) Coomassie Brilliant Blue R-250
45 % (v/v) isopropanol
9 % (v/v) acetic acid
45.75 % (v/v) ddH_2O

Coomassie destaining
40 % (v/v) ethanol
10 % (v/v) acetic acid
50 % (v/v) ddH_2O

2.11.4 Gel mobility shift assay

Affinity and binding strength of His-tagged CodY to promoter regions of enterotoxin genes was tested by gel mobility shift assays in native polyacrylamide gels as described previously [170, 207]. Gels were run in vertical chambers (Minigel-Twin, Biometra) at 100 V, 30 mA for at least 2 h and stained stained with GelRed™ (Biotium) and photographed under UV light (UVsolo TS Imaging System, Biometra – Analytik Jena).

Binding buffer (10x)

50 mM Tris-HCl (pH 8.0)
750 mM KCl
2.5 mM Na_2EDTA (pH 8.0)
1 mM DTT
0.5 % (v/v) Triton X-100
62.5 % (v/v) glycerol

Binding buffer (1x)

5 mM Tris-HCl (pH 8.0)
75 mM KCl
0.25 mM Na_2EDTA (pH 8.0)
0.1 mM DTT
0.05 % (v/v) Triton X-100
6.25 % (v/v) glycerol
2 mM GTP
10 mM L-isoleucine

TBE 5x (non-denaturing)

450 mM Tris
450 mM borate
10 mM EDTA (pH 8.3)

TBE 1x (non-denaturing)

90 mM Tris
90 mM borate
2 mM EDTA (pH 8.3)
10 mM L-isoleucine
10 mM L-leucine
10 mM L-valine

Native PAA gel (10 %)

7.8 ml ddH_2O
4 ml Rotiphorese® Gel 30 (37.5:1)
3 ml 5x gel running buffer
0.2 ml 10 % (w/v) APS
0.02 ml TEMED

Table 8: Primer used in the amplification of promoter regions tested in gel mobility shift assays.

Primer names indicate target gene promoters. Product lengths refer to *B. cereus* INRA C3 or *B. cytotoxicus* CVUAS 2833. All PCR products were diluted to a concentration of 100 ng/µl.

Primer	Sequence [5' → 3']	Product length [bp]	Molarity [fmol/µl]	Reference
	***B. cereus* INRA C3**			
hbl-1-f	AATAATGATATTAGGATGTTTTGTG	307	501	This study
hbl-1-r	GATTACGATCGATAATTTACTG			
hbl-2-f	AAATTATCGATCGTAATCGAC	313	492	This study
hbl-2-r	TCAGCAAACTCCTTACTAG			
hbl-3-f	TCTAGTAAGGAGTTTGCTG	384	401	This study
hbl-3-r	ATTCCGTACACTCCTTTAC			
hbl-4-f	CAAGTTTGTAATAAACGTGTTC	334	461	This study
hbl-4-r	CCTCTCACTTCGATACTC			
hbl-5-f	GTATACACATTAATTTGTAATCATTAC	313	492	This study
hbl-5-r	GAACACGTTTATTACAAACTTG			
CytK2-f	GATAGTGATGTTGCGTTTATTGC	330	440	This study
CytK2-r	CAATCACTTCCTTTTATCTTTGTCG			
Nhe for	ATCCGAATTCTGTATATGCTATGC	568	271	This study,
Nhe rev	AGTCGGATCCACTTTAACTCCTC			see Table 6
inhA1_EMSA_F	ATGTAATTCCTCCCTAATTATCGGTC	350	427	[170]
inhA1_EMSA_R	TTGTTCATCCCTTATTTCCTCCCCTA			

Primer	Sequence [5' → 3']	Product length [bp]	Molarity [fmol/µl]	Reference
	***B. cytotoxicus* CVUAS 2833**			
CytK1-f	TCCTTTACTATTACTATCACCTCTAC	360	427	This study
CytK1-r	CATAGAATCACTTCCTATTTTGTCG			
nhe-CVUAS-F	AACCGACTTATAGGCAGCC	517	298	This study
nhe-CVUAS-R	ACTTTAATTCCTCCTAATGTATAAATTAAC			
nhe-CVUAS-rev	CTAAAAAGCATAGAAAAATGGGCC	370 (with nhe-CVUAS-F)	416	This study
nhe-CVUAS-F2	CTATGCTTTTTAGTAAATTCAAGTTTGTGTTG	295	522	This study
nhe-CVUAS-rev2	GTATTTGGAGTCATTACATTTTGTGCATAC			
inhA1-CVUAS-F	AATTTATTCCTCCCTAATTGTCGG	367	419	This study
inhA1-CVUAS-R	TTCATCCCCAGTTTTCCTCC			
	Negative control			
16SA1	GGAGGAAGGTGGGGATGACG	241	638	[208]
16SA2	ATGGTGTGACGGGCGGTGTG			

DNA fragments containing promoter regions were amplified by PCR (*GoTaq*) as described in section 2.6.2. Primer pairs and fragments ranging from 241 bp to 517 bp are listed in Table 8. A 241 bp fragment of the 16S rRNA gene *rrn* was used as negative control, since it lacks any similarity to the CodY consensus sequence. Gel electro mobility shift assays were performed as described previously [170] at 4 °C with varying amounts of protein and 100 ng target DNA. Molarity of DNA fragments is given in Table 8. The equilibrium dissociation constant K_D was estimated as described earlier [209].

3. Results

3.1 Taxonomy and virulence gene evolution of *B. cereus* sensu lato

3.1.1 *De novo* sequencing of 30 *B. cereus* sensu lato strains

In this study whole genomes of 30 *B. cereus* sensu lato strains were sequenced. 26 of them are known members of *B. cereus* sensu stricto that were either isolated from food or associated with food poisoning cases (Table S1). *B. cereus* #17 (#236) has been isolated from mouse gut (T. Clavel, personal communication) and *B. cereus* IP5832 (#237) is a commercially available probiotic strain (Bactisubtil®) [210]. Additionally, *B. mycoides* WSBC 10969 (#283), and *B. cytotoxicus* CVUAS 2833 (#249) [211] were added, because only very few genomes of these two species are publicly available.

Results of data processing, coverage, k-mer values and contig assembly are shown in Table 9. Due to a continuous optimization process sequencing obtained high quality reads (Length ≥ 80%, Q ≥ 30) that allowed *de novo* assembly to a contig number as low as 25. Assembly sizes ranged between 4.2 Mbp and 6.8 Mbp.

Table 9: *De novo* whole genome sequencing results of 30 *B. cereus* sensu lato strains.

Sequencing statistics and results of read quality filtering/trimming and contig assembly are reported.

Strain	MiSeq® reagent Kit	Quality filtering	Raw read length [bp]	# Raw reads	# Filtered & trimmed reads	Trimmed read length [bp]	Genome coverage trimmed reads (x-fold)	K-mer size [bp]	N50	# Contigs >500 bp	Assembly size [bp]	NCBI GenBank accession number
B. cereus F4430/73	Version 1, 300 cycles, paired end	Length ≥ 70%, Q ≥ 20	145	2,008,054	1,483,038	113	**32**	47	25,336	**447**	**5,577,793**	JYPK00000000
B. cereus HWW 274-2				2,837,642	1,505,012		**32**	39	32,259	**320**	**5,290,159**	JYPL00000000
B. cereus WSBC 10035	Version 2, 500 cycles, paired end	Length ≥ 70%, Q ≥ 20	180	2,996,320	1,247,284	169	**38**	75	52,377	**184**	**5,619,577**	LABS00000000
B. cereus F4429/71				3,129,982	1,491,644		**45**	67	91,497	**115**	**5,284,967**	JYPJ00000000
B. cereus 14294-3 (M6)	Version 2, 500 cycles, paired end	Length ≥ 80%, Q ≥ 30	200	5,225,132	4,625,090	189	**165**	99	875,707	**25**	**5,523,305**	JYPF00000000
B. cereus INRA A3	Version 3, 600 cycles, paired end	Length ≥ 80%, Q ≥ 30	250	13,534,264	7,187,650	239	**323**	169	189,988	**61**	**6,075,647**	LABH00000000
B. cereus INRA C3				9,841,330	4,702,936		**213**	173	348,619	**38**	**5,596,453**	LABI00000000
B. cereus RIVM BC 934				13,029,356	6,521,168		**295**	203	62,879	**207**	**6,840,916**	LABP00000000
B. cereus RIVM BC 964	Version 3, 600 cycles, paired end	Length ≥ 80%, Q ≥ 30	250	9,914,386	5,338,252	192	**194**	43	136,264	**78**	**5,815,402**	LABQ00000000
B. cereus #17				9,276,110	5,007,666	192	**182**	101	237,244	**45**	**5,852,222**	JYFW00000000
B. cereus IP5832				15,632,892	9,878,046	196	**366**	117	470,309	**28**	**5,592,318**	LABJ00000000
B. weihenstephanensis WSBC 10204				10,434,454	6,063,602	191	**219**	85	296,898	**43**	**5,655,039**	Complete genome available
B. cereus RIVM BC 126	Version 3 600 cycles, paired end	Length ≥ 80%, Q ≥ 30	250	3,742,980	2,960,928	230	**129**	183	340,103	**47**	**5,417,487**	LABO00000000
B. cereus NVH 0075-95				4,453,098	3,407,364	220	**142**	177	172,877	**92**	**6,112,682**	LABM00000000
B. cereus MHI 226				2,800,238	2,248,024	230	**98**	113	180,772	**60**	**6,233,017**	LABL00000000
B. cereus RIVM BC 90				4,690,328	3,816,892	230	**166**	183	166,503	**68**	**5,559,670**	LABN00000000
B. cereus 6/27/S				3,507,602	2,686,982	220	**112**	115	67,058	**176**	**6,771,128**	LABV00000000
B. cereus SDA KA 96				4,266,240	3,498,326	220	**146**	167	402,005	**34**	**5,335,844**	LABR00000000
B. cereus 7/27/S				3,604,194	2,965,452	225	**126**	153	259,768	**48**	**5,479,572**	LABW00000000
B. cereus MHI 86				3,970,862	3,038,394	220	**126**	63	192,413	**54**	**5,551,873**	LABK00000000
B. cereus F3175/03				2,913,134	2,235,522	220	**93**	73	204,296	**64**	**5,733,808**	JYPI00000000
B. cereus F3162/04				3,033,036	2,338,464	220	**97**	43	170,740	**77**	**5,591,156**	JZEV00000000
B. cytotoxicus CVUAS 2833				3,185,212	2,298,738	210	**91**	177	318,402	**36**	**4,127,075**	JYPG00000000
B. cereus F528/94	Version 3, 600 cycles, paired end	Length ≥ 80%, Q ≥ 30	230	7,027,494	5,982,372	219	**248**	167	355,056	**49**	**5,935,300**	JYPH00000000
B. mycoides WSBC 10969				6,040,358	4,819,360	210	**192**	167	270,494	**54**	**6,101,972**	LABT00000000
B. cereus MHI 1670	Version 3, 600 cycles, paired end	Length ≥ 80%, Q ≥ 30	250	1,707,524	1,473,124	219	**61**	89	203,445	**57**	**5,341,640**	-
B. cereus MHI 1672				1,913,274	1,583,214	213	**64**	95	189,587	**50**	**5,333,221**	-
B. cereus MHI 1761				2,915,244	2,433,744	215	**99**	113	241,231	**46**	**5,310,366**	-
B. cereus MHI 2968				2,867,850	2,331,270	206	**91**	129	204,275	**52**	**5,506,136**	-
B. cereus MHI 3233				2,862,870	2,440,608	223	**103**	113	231,424	**48**	**5,652,907**	-

Reference genome: *B. cereus* F837/76 (5,288,498 bp)

3.1.2 *B. cereus* sensu lato species affiliation

After assembly, the new sequences were screened for the presence and location of virulence determinants and housekeeping genes and used for whole-genome comparison as well as multilocus sequence analysis. A total of 223 *B. cereus* sensu lato genomes were investigated and 142 genomes were included in the phylogenetic analyses (see section 2.5.3).

3.1.2.1 Confirmation of seven phylogenetic groups within the species tree of *B. cereus* sensu lato

For taxonomical purposes, and in order to analyze horizontal gene transfer, the construction of a phylogenetic master tree which depicts the assumed "true" phylogenetic relationships of the organisms studied as correctly as possible was mandatory. An MLSA tree (Fig. 5) based on concatenated sequences of seven housekeeping genes (*adk, ccpA, glpF, glpT, panC, pta,* and *pycA*) from the *B. cereus* sensu lato core genome was constructed. The topologies of trees calculated on the basis of individual housekeeping genes were highly similar to this master tree but partially lacked resolution due to different levels of conservation (data not shown). The overall MLSA tree topology was confirmed by an analysis of whole-genome pairwise ANI comparison, visualized by a neighbor network (Fig. 6). Pairwise comparison of ANI versus pairwise distances of the seven concatenated housekeeping genes correlated nicely (Fig. 7A). These results were additionally confirmed by whole-genome SNP-based phylogeny (Fig. 8). All three methods showed highly similar tree topologies and confirmed that the MLSA tree is correctly displaying the strain phylogeny of the 142 *B. cereus* sensu lato strains.

Seven major phylogenetic clusters were recognized. Phylogenetic groups within *B. cereus* sensu lato could be distinguished with a species boundary of 94 % identity (Fig. 7). When comparing strains that do not belong to the same phylogenetic group with each other, ANI values were in the range of 80 – 94 %. These data therefore suggest that seven genomospecies exist within *B. cereus* sensu lato. However, demarcation of clusters II and III proved to be difficult. *B. cereus* F528/94 (#242, II/III, Fig. 5 – 8) could neither be affiliated reliably to cluster II, nor to cluster III with the methods used in this study (ANI ~93 %).

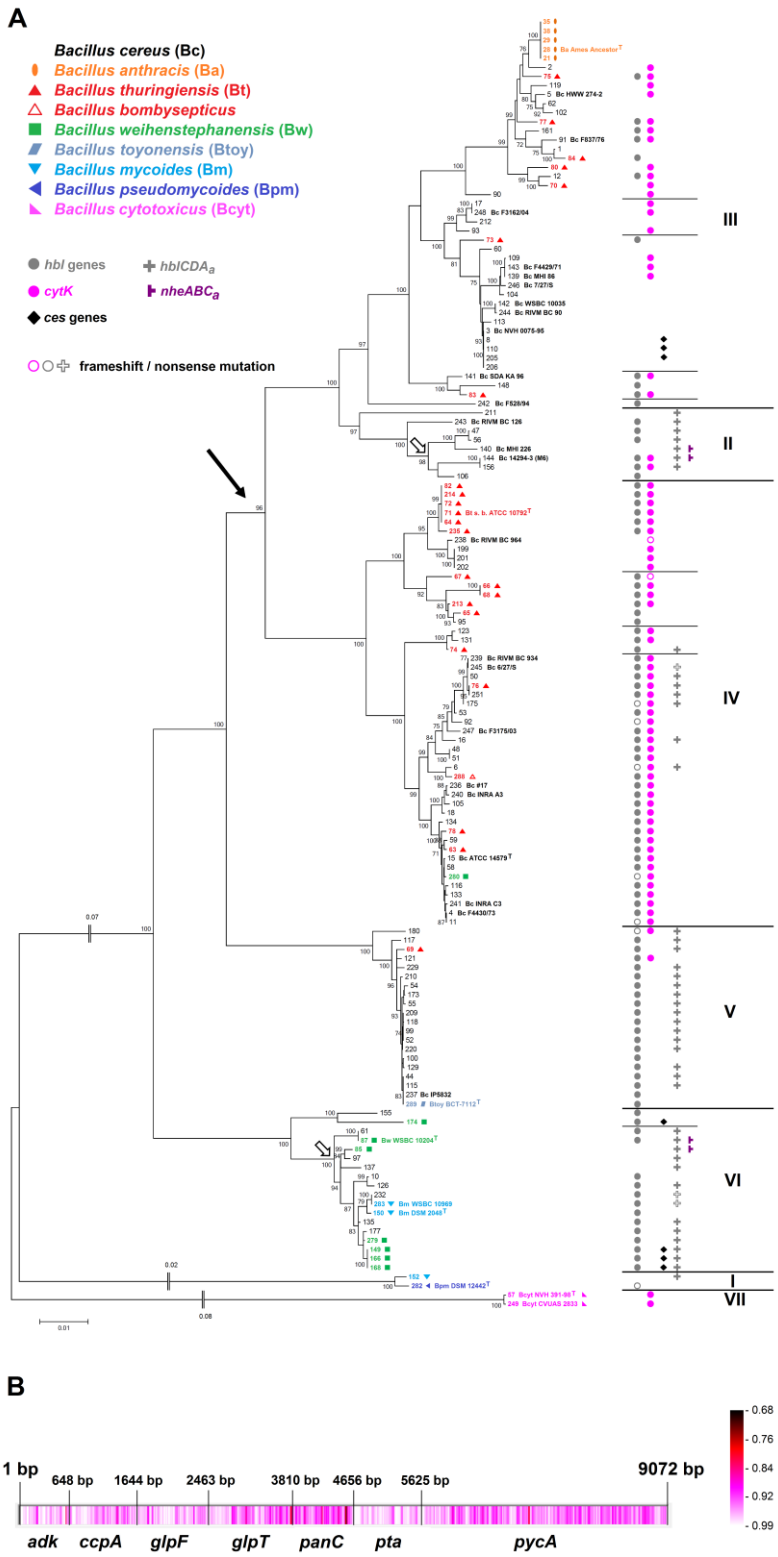


Fig. 5: MLSA-based species relationship within *B. cereus* sensu lato.

A: The phylogenetic tree (Maximum Likelihood Method) was calculated using the concatenated sequence of seven housekeeping genes from 142 *B. cereus* sensu lato strains. Seven major phylogenetic clusters and subclusters are indicated. Arrow: Suspected first appearance of *cytK-2*. Empty arrows: Suspected origin of *nhe$_a$* operons. **B:** Visualization of the sequence homology derived from a multiple sequence alignment calculated with RDP3 (see section 2.5.4). Color ranges (identity score: ≤ 1, 1 = identical in all sequences) from identical (white) to highly dissimilar (black).

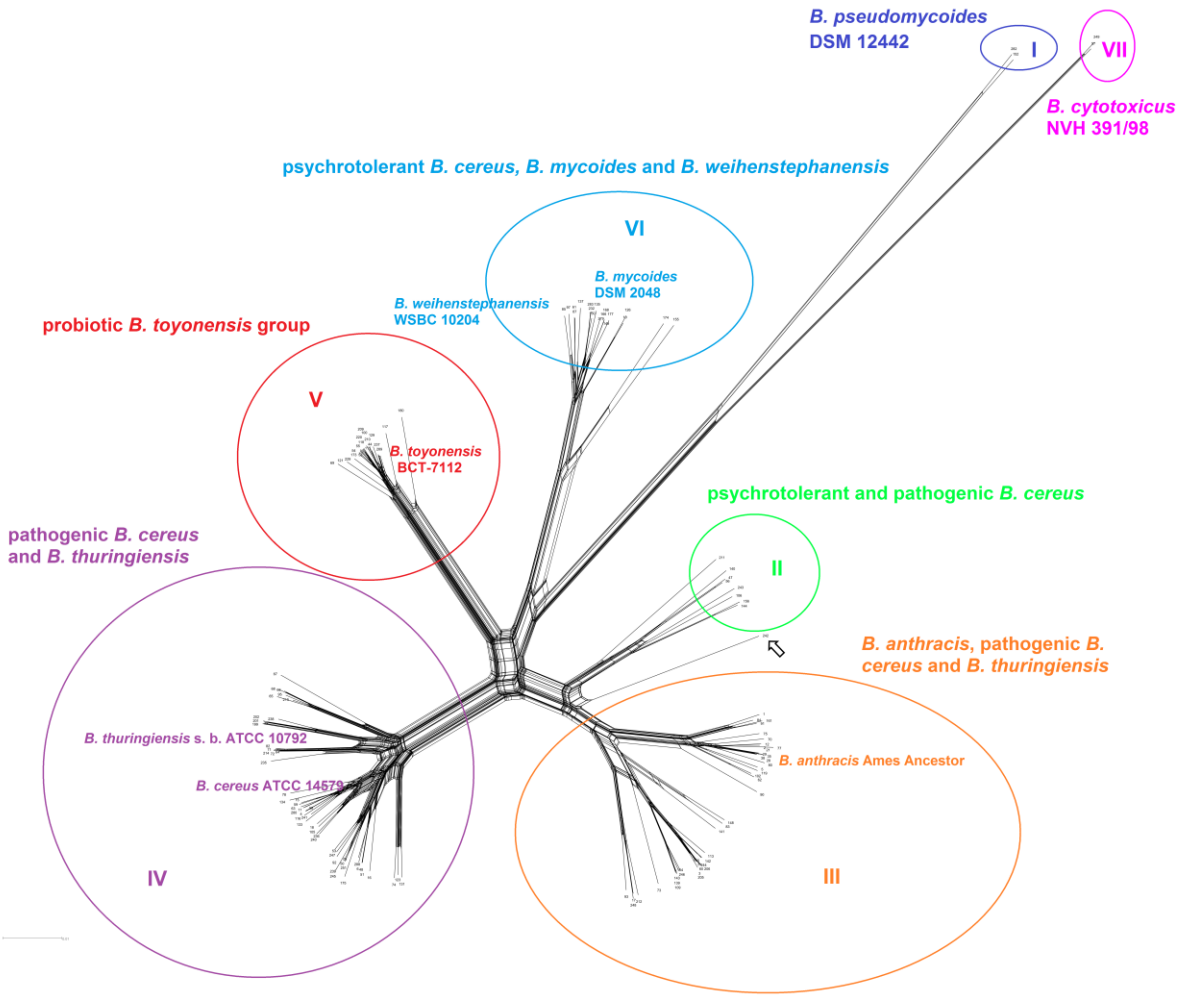


Fig. 6: ANI-based species relationship within *B. cereus* sensu lato.

Neighbor network was calculated using ANI distances of 142 *B. cereus* sensu lato genomes. Entire genomes (completed or draft) including all available plasmids were used. Phylogenetic groups are designated according to [90]. *B. cereus* F528/94 (#242), marked by an arrow, is intermediate between clusters II and III.

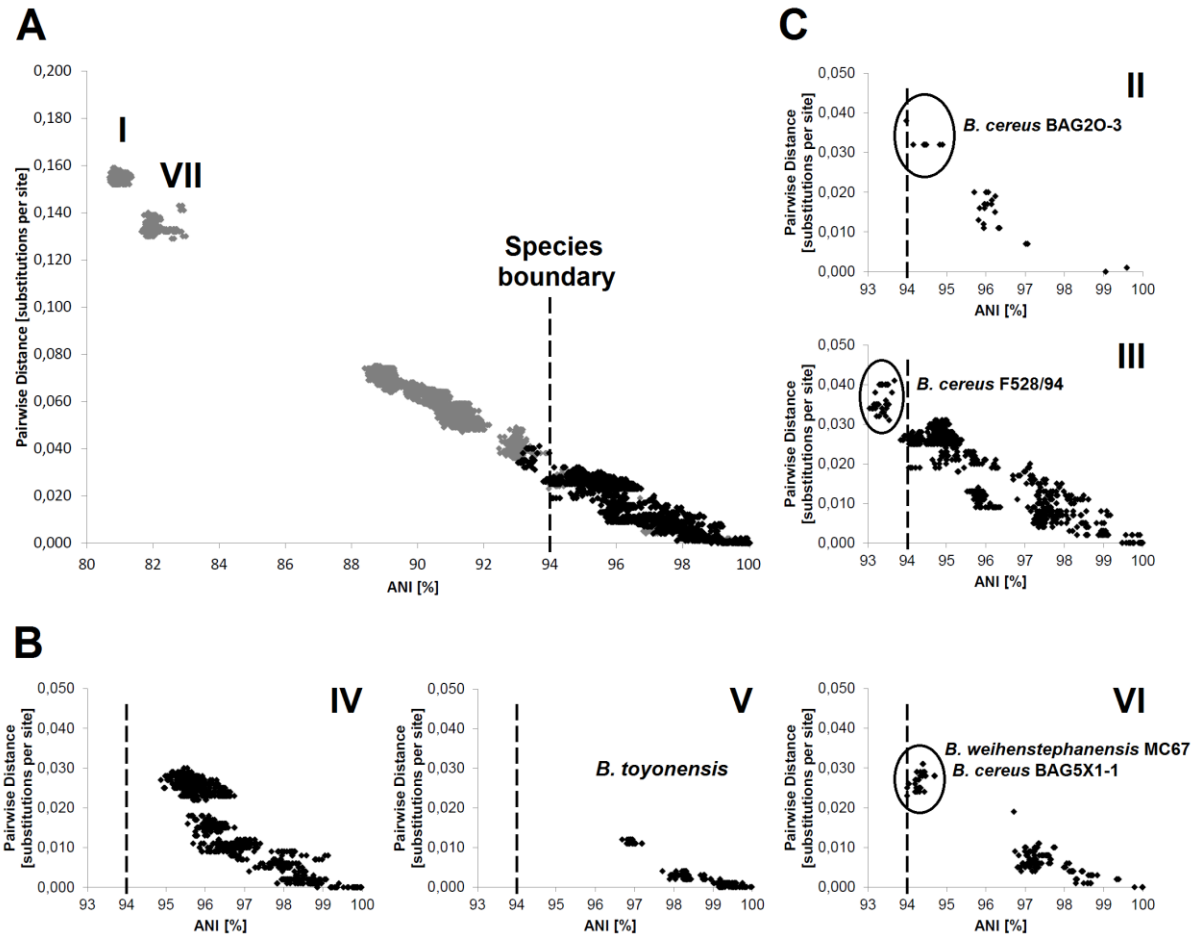


Fig. 7: Correlation of pairwise distance of concatenated housekeeping genes with whole genome ANI values.

A: Correlation of whole-genome ANI and pairwise distance of seven concatenated housekeeping genes of all 142 *B. cereus* sensu lato strains. Intra-cluster values (for clusters see Fig. 5) are depicted in black, all inter-cluster values in grey. Pearson's correlation of all values is $r^2 = 0.974$.

B and C: Intra-cluster comparison of whole-genome ANI and pairwise distance of seven concatenated housekeeping genes. Cluster affiliation (I – VII) according to species tree (Fig. 5). Strains that are difficult to sort into a distinct genomospecies are named and marked by circles. **B:** Clusters I, IV – VI can be distinguished at an ANI species boundary of ≥ 94 %. Cluster I: ANI 98.3 %, pairwise distance 0.005 substitutions per site. Cluster VII: ANI 99.5 %, pairwise distance 0.001 substitutions per site. **C:** Clusters II and III cannot be discerned on the basis of the comparisons presented: strain *B. cereus* F528/94 (#242) is affiliated to Cluster III by MLSA (Fig. 5) and whole-genome SNP analysis (Fig. 8), but to Cluster II by ANI (Fig. 6).

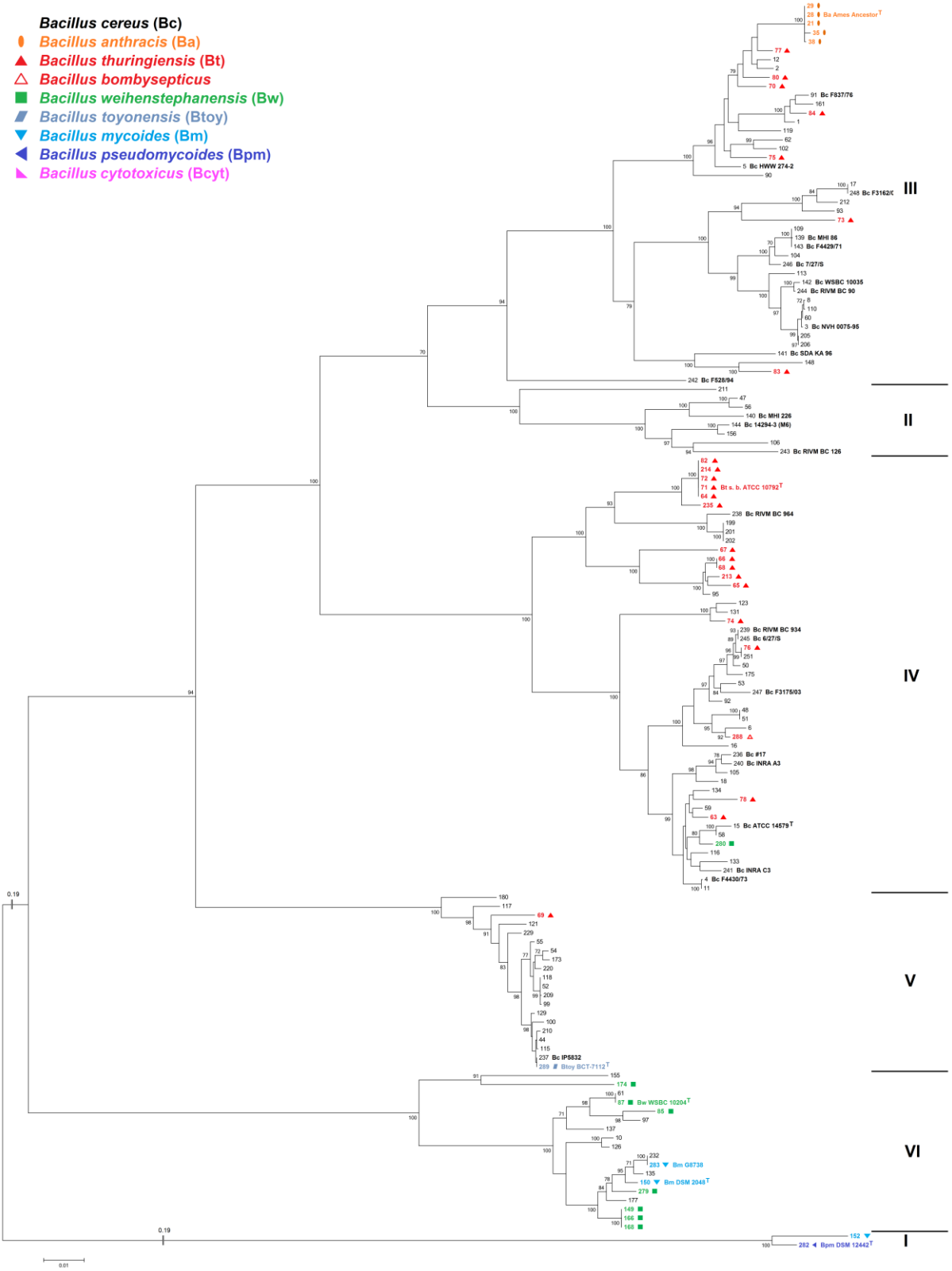


Fig. 8: SNP-based species relationship within *B. cereus* sensu lato.

The phylogenetic tree (Maximum Likelihood Method) was calculated using the genome-wide core SNP matrix (SNPs that are present in all of the analyzed genomes) of 140 *B. cereus* sensu lato genomes. Phylogenetic cluster VII is too divergent for accurate detection of core SNPs using kSNP3 and had to be excluded (see section 2.5.5.2).

Many existing species affiliations do not match the genomic relationships of the investigated strains (Fig. 5). Cluster IV contains mesophilic pathogenic *B. cereus* and *B. thuringiensis* and one *B. weihenstephanensis* strain. Cluster III comprises *B. anthracis,* emetic and non-emetic *B. cereus* and *B. thuringiensis*. While it is almost impossible to distinguish between cluster III *B. cereus* and *B. thuringiensis* isolates (Fig. 5), *B. anthracis* is an easily discernable monophyletic branch in the species tree.

However, three phylogenetic groups could be matched easily to three species. Cluster V is clearly separated from all other phylogenetic groups and includes the type strain *B. toyonensis* BCT-7112. Strains of both clusters I and VII (*B. cytotoxicus*) showed the greatest distance to all other members of *B. cereus* sensu lato. Phylogenetic analysis of housekeeping genes clustered *B. mycoides* Rock3-17 (#152, I) together with *B. pseudomycoides* DSM 12442 (type strain, #282, I) in cluster I and should therefore be considered a member of *B. pseudomycoides*.

3.1.2.2 Occurrence of virulence genes

To gain an overview over the distribution of virulence genes, 223 *B. cereus* sensu lato strains, including the *de novo* sequenced strains, were analyzed (Table S1). This analysis includes all investigated strains, while the phylogenetic analysis was reduced to 142 *B. cereus* sensu lato strains, because due to draft status and partially insufficient sequence quality of genome sequences, selected housekeeping gene sequences could not be identified for all 223 *B. cereus* sensu lato strains. All *B. cereus* strains possess the *nhe* genes. 63 % of the 223 strains were found to contain *hbl* and 34 % of these possess a second *hbl* operon (hbl_a), which are 22 % of all strains. In 2.2 % of strains a duplication of *nhe* (nhe_a) was found. *CytK-2* appeared in 41 %, *cytK-1* in 1 % and the emetic gene cluster *ces* in 5 % of all strains. *CytK-2* is far more frequent in *B. thuringiensis* strains (75 %) than in the rest of *B. cereus* sensu lato (36 %). The combined presence of *nhe, hbl* and *cytK* occurs in 30 % of the 223 *B. cereus* sensu lato strains.

3.1.3 Horizontal transfer and evolution of enterotoxin genes

3.1.3.1 Massive horizontal transfer and duplication of hbl

Hbl is not an essential operon for *B. cereus*, since only 63 % of 223 strains contain *hbl*, which appears rarely in cluster III and is absent from *B. cytotoxicus* (cluster VII) (Fig. 5). A comparison of the concatenated *hblCDAB* gene tree with a species tree consisting of all 101 *hbl*-containing strains from the set of 142 showed vastly different topologies (Fig. 9). All phylogenetic clusters except cluster V are mixed which provides evidence for massive horizontal transfer of the enterotoxin operon *hblCDAB*, both between and within phylogenetic groups. While horizontal transfer of the *hblCDAB* operon is frequent, intra-operon recombinations are extremely rare. Only five statistically significant recombination events within the set of 101 *hbl* containing strains could be detected (Table S2) and all of these occurred exclusively within their respective phylogenetic group. Intra-operon recombinations between different phylogenetic groups have probably been removed by negative selection. DNA sequence identities of *hbl* between *B. cereus* sensu lato strains are quite high (93 – 100 %). Thus, *hbl* is more conserved than housekeeping genes or *nhe* (72 – 100 % sequence identity) in *B. cereus* sensu lato (Fig. 5, 9 and 14).

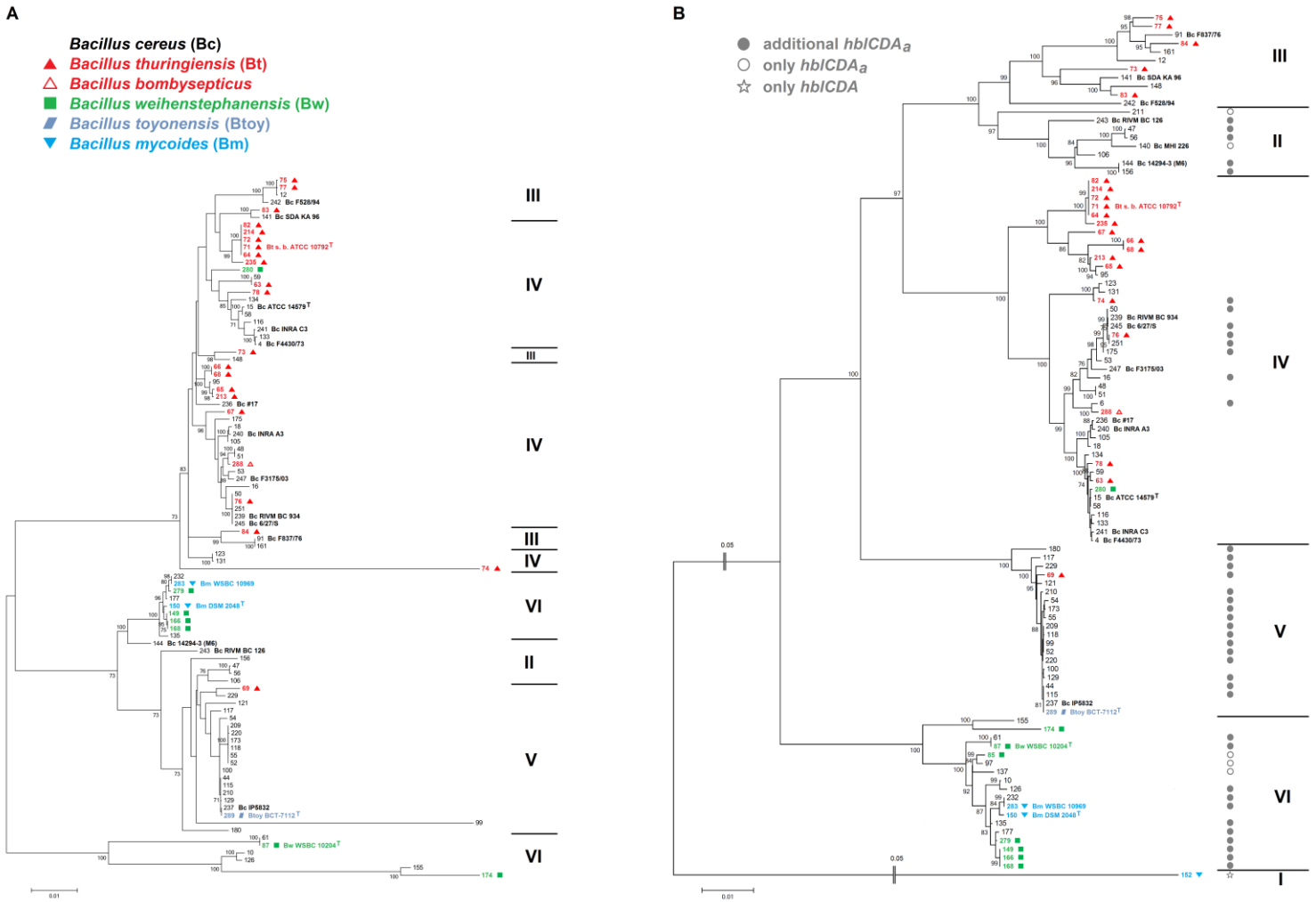


Fig. 9: *Hbl* in *B. cereus* sensu lato.

A: Phylogenetic tree (Maximum Likelihood Method) of the concatenated sequences of 94 *hblCDAB* genes. Seven strains within of 101 contain only *hblCDA* (empty circle and star) or an incomplete *hbl* operon (#6, IV) and were excluded. **B:** Phylogenetic tree based on the seven housekeeping genes of 101 *hbl*-containing *B. cereus* sensu lato strains.

It has been known that two distinct homologs of *hbl* exist in *B. cereus* [133]. Six of the *de novo* sequenced and assembled strains contain two versions of *hbl*. The existence of duplicated enterotoxin operons within the newly sequenced genomes was confirmed by comparison of coverage depths over operons and their respective background contigs as described in section 2.5.2. Results are shown and summarized in Table S3. The second *hbl* operon hbl_a is common in clusters II, V and VI and appears occasionally in cluster IV, but not at all in III and VII (Fig. 5, grey crosses; Fig. 9B, grey circles). Intra-operon recombination analysis (Table S2) revealed three significant recombination events that all include *B. cereus* MHI 226 (#140, II) as a parental sequence and took place within $hblD_a$. All $hblCDA_a$ genes cluster together, show a much more conserved tree topology than *hblCDA* and are clearly separated from *hblCDA* (Fig. 9 and 11). Furthermore, $hblCDA_a$ appears to have been deleted in various lines of the species group (Fig. 5 and Fig. 9). The topologies of *hbl* and hbl_a phylogenies are similar but not identical, which could be explained by HGT (Fig. 9 and Fig. 12). $HblCDA_a$ is as conserved as *hblCDA* (Fig. 10B) and shows an overall nucleotide sequence identity of 75 – 82 % towards the *hbl* genes, which are 89 – 100 % identical among themselves. Hbl_a are 93 – 100 % identical among each other. Six strains (#85, #97, #137, #140, #152, and #211) possess only *hblCDA*. Their version of *hbl* is homologous to $hblCDA_a$ and they may have lost *hblCDAB*. *B. mycoides* Rock3-17 (#152, I) and *B. mycoides* Rock1-4 (#151) possess an *hblCDA* that differs from both *hbl* variants described above. *B. mycoides* Rock3-17 *hblCDA* shows 80 – 82 % identity to $hblCDA_a$, but 86 – 89 % identity to *hblCDA*.

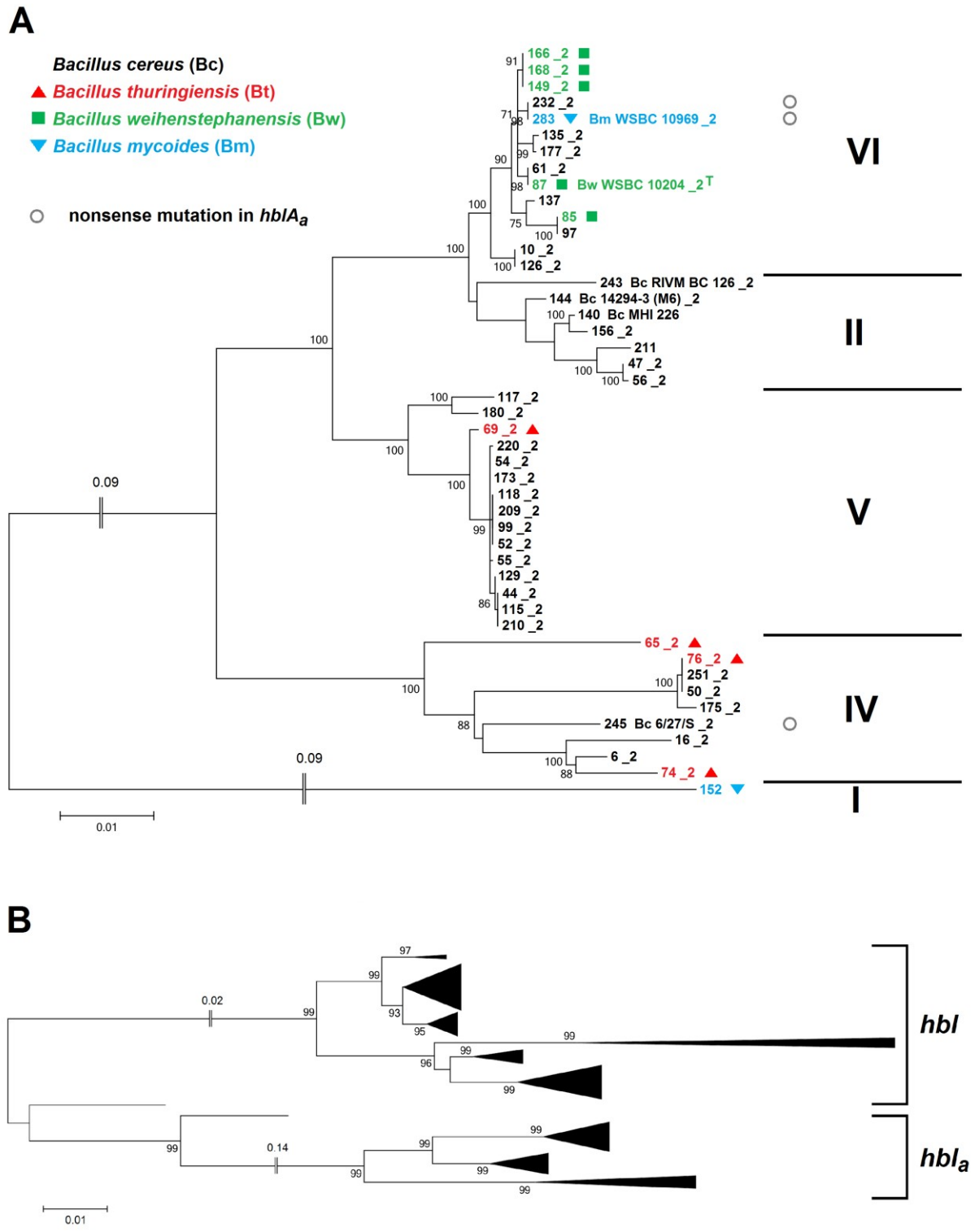
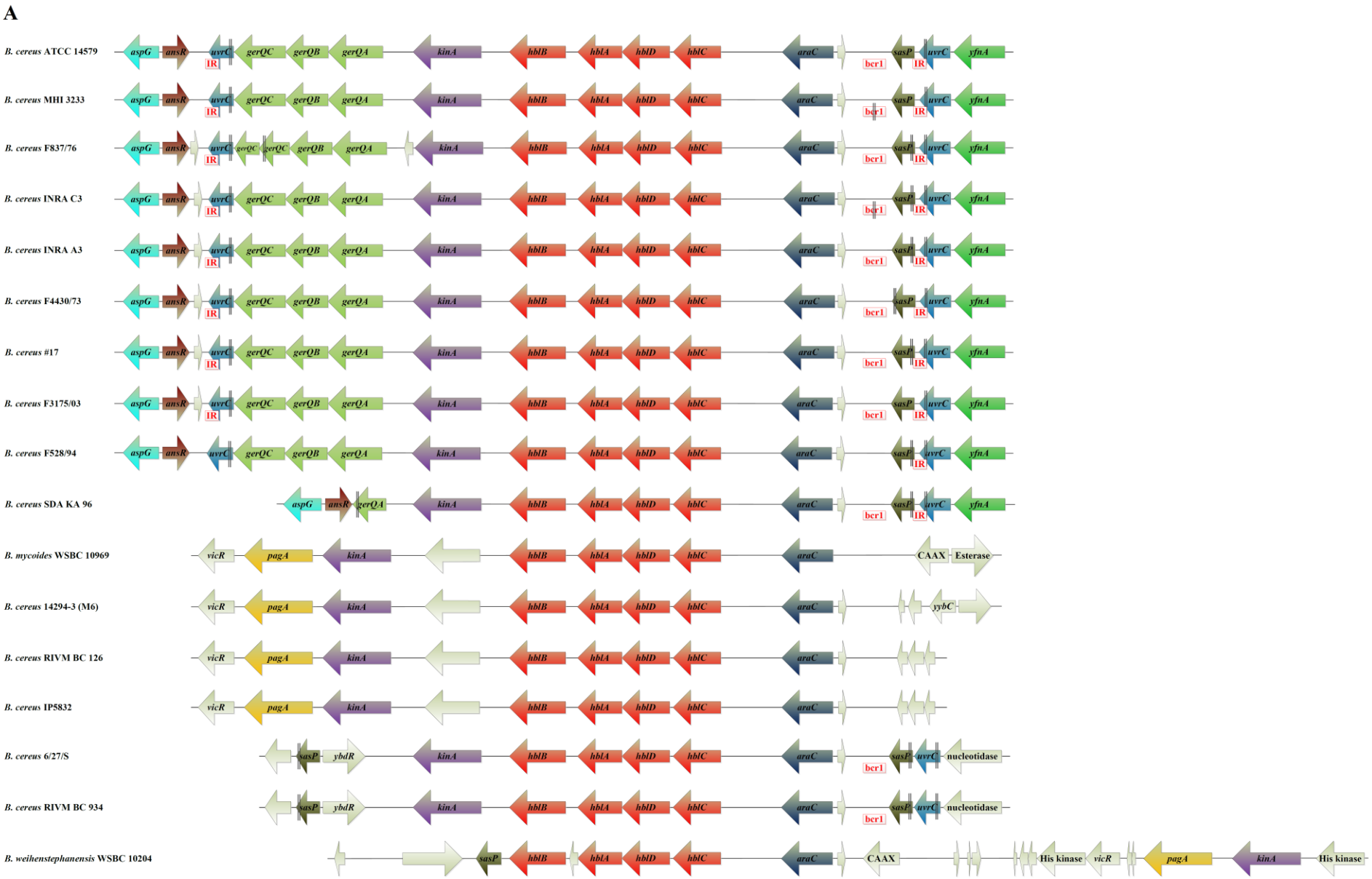


Fig. 10: *HblCDA$_a$* in *B. cereus* sensu lato.

A: Phylogenetic *hblCDA$_a$* tree (Maximum Likelihood Method) based on the concatenated sequence from 46 *B. cereus* sensu lato strains. **B:** Evidence for an ancient origin of *hbl$_a$* based on the concatenated sequence of *hblCDA* and *hblCDA$_a$* from 101 *B. cereus* sensu lato strains.

Since *hbl* occurs chromosomally as well as plasmid-bound, the immediate vicinity of *hbl* and hbl_a was analyzed with regard to potential indications of transposon activity in all 17 *hbl*-containing strains that were sequenced in this study. A comparison of putative transposon regions including *hbl* or hbl_a is shown in Figure 11. Sequence analysis and annotation with RAST [212] revealed that half of the *hblCDAB* operons are inserted within the *uvrC* gene as described earlier [99], but in the rest neither insertion sites nor length of the inserted region or adjacent genes are conserved. The lowest common denominator of inserted regions from 18 *B. cereus* sensu lato strains consists only of a transcriptional regulator gene of the *araC* family and *hblCDAB* itself. Inverted repeats (IR and bcr1) [99, 213], which mark the insertion site interrupting *uvrC* as telltale signs of transposons, could not be found in half of the investigated strains. A transposase gene could only be detected in the vicinity of hbl_a of *B. cereus* 6/27/S (#245, IV), but not adjacent to *hbl* (Fig. 11). Studied hbl_a are located close to antibiotic resistance genes and do not contain known inverted repeats. Furthermore, the gene *pagA* encoding a protective antigen similar to a gene located on the *B. anthracis* pXO1 virulence plasmid, has been inserted into the chromosome of *B. cereus, B. mycoides* and *B. weihenstephanensis* (Fig. 11) in close proximity to the *hbl* operon. This is an example for recombination between plasmid and chromosomal DNA. BLAST analysis showed that *hbl* duplications occur chromosomally as well as plasmid-bound and, hence, are mobile within *B. cereus* sensu lato. Hbl_a is located on plasmid in the eight strains *B. thuringiensis* serovar kurstaki YBT-1520 (pBMB293), *B. thuringiensis* serovar kurstaki HD-1 (pBMB299), *B. thuringiensis* serovar chinensis CT-43 (pCT281), *B. thuringiensis* serovar thuringiensis IS5056 (pIS56-285), *B. thuringiensis* serovar galleriae HD-29 (pBMB267), *B. thuringiensis* H18-1 (pHS18-1), *B. thuringiensis* YC-10 (pYC1) and *B. thuringiensis* serovar tolworthi (pKK2). These strains comprise ~50 % of all hbl_a containing *B. thuringiensis* strains with completed genomes available. The detection of plasmid-location is likely to increase with ongoing completion of genomes.


A
B. cereus ATCC 14579
B. cereus MHI 3233
B. cereus F837/76
B. cereus INRA C3
B. cereus INRA A3
B. cereus F4430/73
B. cereus #17
B. cereus F3175/03
B. cereus F528/94
B. cereus SDA KA 96
B. mycoides WSBC 10969
B. cereus 14294-3 (M6)
B. cereus RIVM BC 126
B. cereus IP5832
B. cereus 6/27/S
B. cereus RIVM BC 934
B. weihenstephanensis WSBC 10204
aspG
ansR
uvrC
IR
gerQC
gerQB
gerQA
kinA
hblB
hblA
hblD
hblC
araC
bcr1
sasP
yfnA
vicR
pagA
CAAX
Esterase
yybC
ybdR
nucleotidase
His kinase

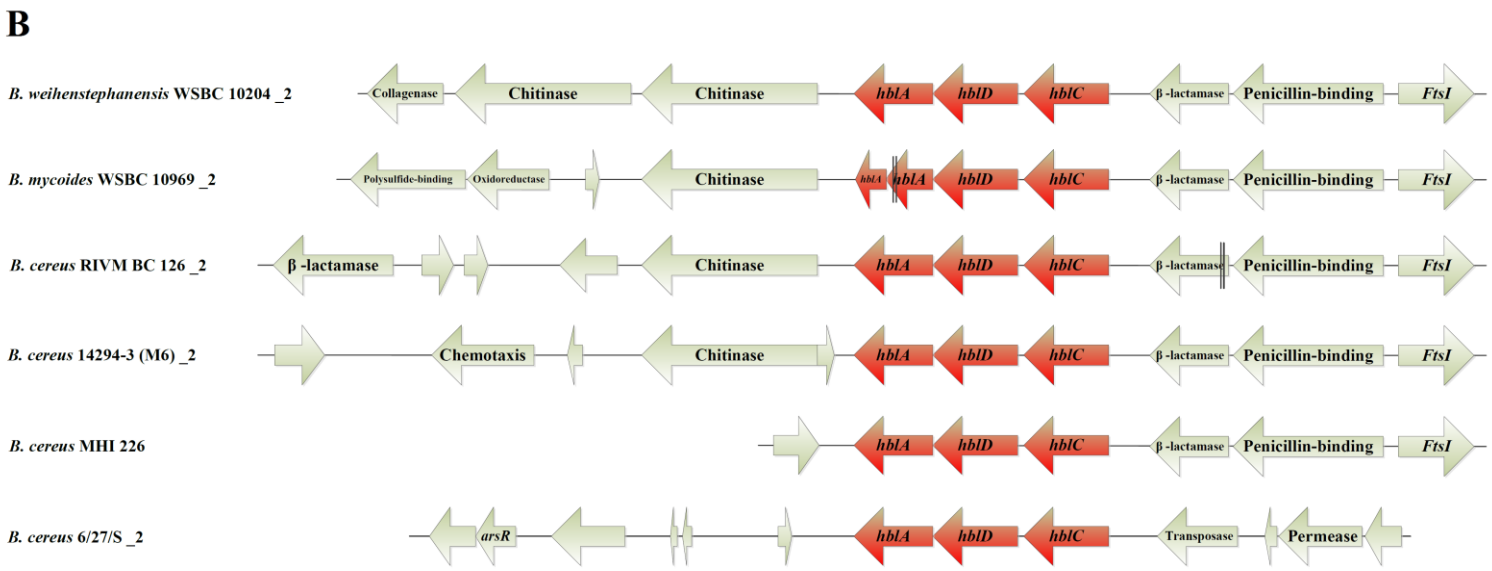


Fig. 11: Genomic organization of *hbl* operons and adjacent regions of *B. cereus* sensu lato strains.

A: 16 of the strains sequenced in this study contain *hblCDAB*, they are shown in comparison to type strain *B. cereus* ATCC 14579.

B: Six of the strains sequenced in this study contain $hblCDA_a$.

Gene denomination according to annotation with RAST [212] and previous studies [2, 99, 213]: *ansR*: *ans* (L-aspartase, L-asparaginase, malate utilization/transporter) operon repressor protein, *araC*: AraC family transcriptional regulator TrrA, *arsR*: ArsR family transcriptional regulator, *aspG*: L-asparaginase, β-lactamase: β-lactamase class A, CAAX: CAAX amino terminal protease family protein, Chemotaxis: Methyl-accepting chemotaxis protein, Esterase: Erythromycin-esterase type I, *ftsI*: cell division protein, *hblCDAB*: hemolysin BL, His kinase: Two component histidine kinase, *gerQABC*: spore germination proteins, *kinA*: sporulation kinase, nucleotidase: 5' Nucleotidase, Oxidoreductase: FAD-dependent pyridine nucleotide disulfide oxidoreductase, *pagA*: protective antigen (anthrax moiety, pXO1), Penicillin-binding: Penicillin-binding protein, Permease: Permease of the drug/metabolite transporter (DMT) family, Polysulfide-binding: Zn-dependent hydroxyacylglutathione hydrolase (polysulfide-binding protein), *sasP*: small acid soluble protein, Transposase: mobile element protein of the IS605 OrfB family transposase, *uvrC*: UvrC-like excinuclease subunit C, *vicR*: DNA-binding response regulator, *ybdR*: zinc-type alcohol dehydrogenase-like protein, *yfnA*: amino acid permease, *yybC*: uncharacterized protein, bcr1: *Bacillus cereus* repeat 1, IR: inverted repeat, ||: incomplete gene, not denominated arrow: hypothetical gene.

3.1.3.2 Horizontal transfer of cytK

The gene *cytK-2* occurs in 47 % of the strainset and only the two *B. cytotoxicus* strains (#57, #249, VII) possess *cytK-1*. In this study *cytK-2* was detected in strains of clusters II – V, but not in cluster I, VI and VII. Lateral transfer of *cytK* can be inferred from a direct comparison of the *cytK* tree with the species tree of all 68 *cytK*-containing strains (Fig. 12).

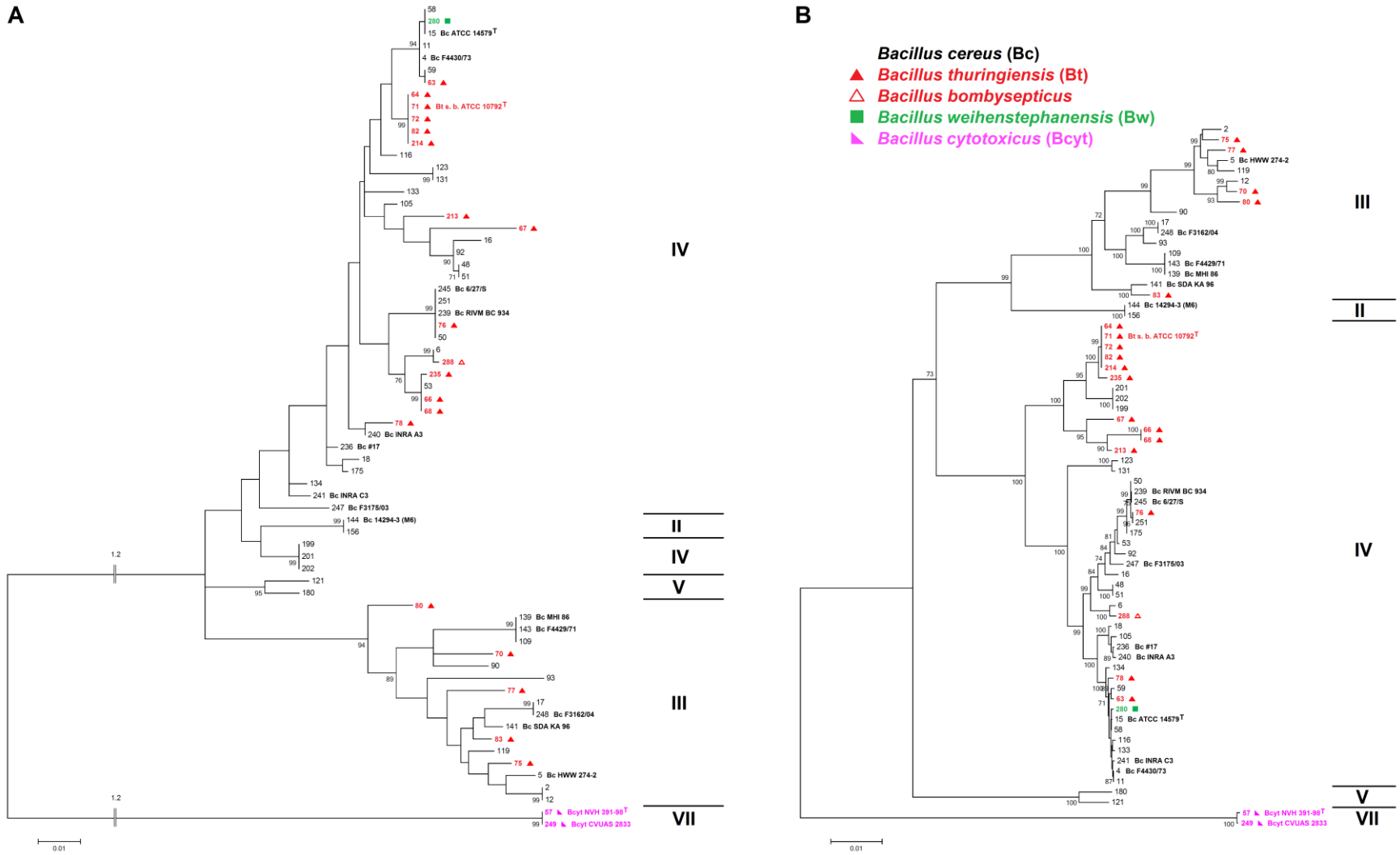


Fig. 12: *CytK* in *B. cereus* sensu lato.

A: Phylogenetic tree (Maximum Likelihood Method) based on 68 *cytK* gene sequences. **B:** Phylogenetic tree based on the concatenated sequence of seven housekeeping genes from 68 *cytK*-containing *B. cereus* sensu lato strains.

3.1.3.3 Massive horizontal transfer of plcR *and* papR

PlcR gene phylogeny of *B. cereus* sensu lato (Fig. 13) was compared with the *B. cereus* sensu lato species tree (Fig. 5). A low degree of *plcR* conservation (DNA sequence identities range from 70 % up to 100 %) is demonstrated. The surprisingly low similarity of both trees is evidence for extensive horizontal transfer of *plcR*. PapR is the quorum sensing peptide necessary for activation PlcR and part of the PlcR regulon [214]. The *papR* gene is encoded less than 100 bp downstream of *plcR* and shows a phylogeny similar to *plcR* (Fig. S1), which hints to conjoint lateral transfer. The genomes of *B. pseudomycoides* DSM 12442 (#282, I), *B. mycoides* Rock3-17 (#152, I) and *B. mycoides* Rock1-4 (#151) do not contain *papR*. Thus, in *B. pseudomycoides* the PlcR virulence regulon might be inactive leading to reduced enterotoxicity.

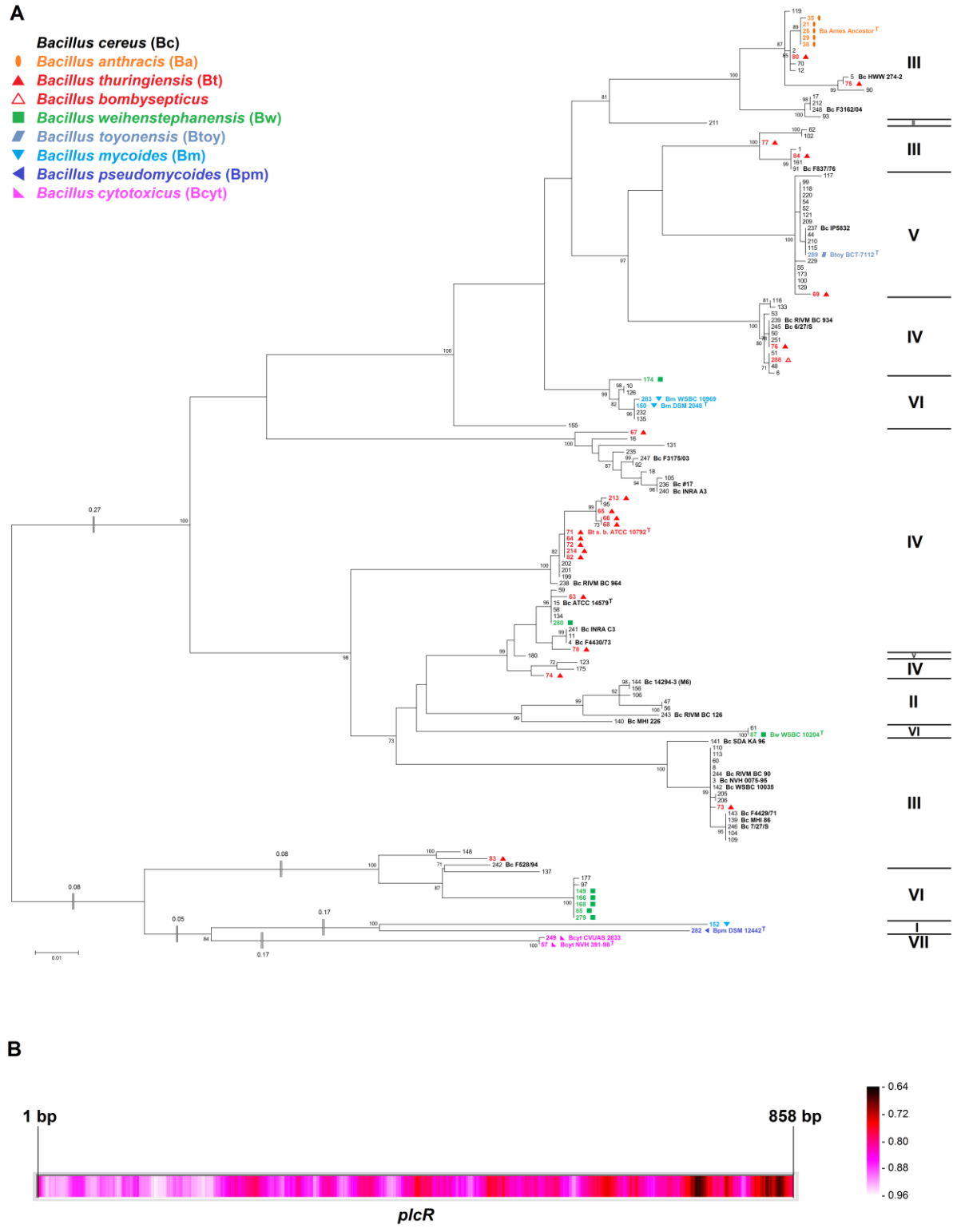


Fig. 13: *PlcR* in *B. cereus* sensu lato.

A: Phylogenetic *plcR* tree (Maximum Likelihood Method) based on the sequence from 142 *B. cereus* sensu lato strains. **B:** Visualization of the sequence homology derived from a multiple *plcR* sequence alignment calculated with RDP3 (see section 2.5.4). Color ranges (identity score: ≤ 1, 1 = identical in all sequences) from identical (white) to highly dissimilar (black).

3.1.3.4 Strictly vertical transmission and duplication of nhe

A phylogenetic tree of the concatenated *nheABC* genes was calculated (Fig. 14) and compared to the MLSA species tree (Fig. 5). Their topology is almost identical. Some strains contain more distantly related *nhe* sequences, such as *B. mycoides* and *B. pseudomycoides* (74 – 77 % sequence identity) or *B. cytotoxicus* NVH 391-98 with only 72 – 88 % DNA sequence identity to other *nhe* operons [120] (Fig. 14). No significant recombination of distantly related *nhe* operons seems to have occurred. After discovering the strictly vertical transmission pattern of *nheABC*, recombination within the *nhe* operon was investigated in more detail. In the set of 142 concatenated *nheABC* genes 21 statistically proven intra-operon recombination events could be detected, but these occurred only between closely related *nhe* regions, which keeps recombination derived variation small (Table S2). Furthermore, and most significantly, *nhe* occurs in all known *B. cereus* strains without exception, which may be evidence for an important function of the operon. It is noticable that *nheA* is the most highly conserved gene of the *nhe* operon (data not shown). While *nheB* and *nheC* could be deleted [107], no artificial knockout of *nheA* was possible. However, *B. cereus* MHI 1670, 1761 and 2968 (#291, 293, 294) contain a frameshift mutation 73 bp downstream of the translational start (Fig. 15B). Sequence analysis revealed that this natural deletion of *nheA* could be bypassed via an alternative start codon.

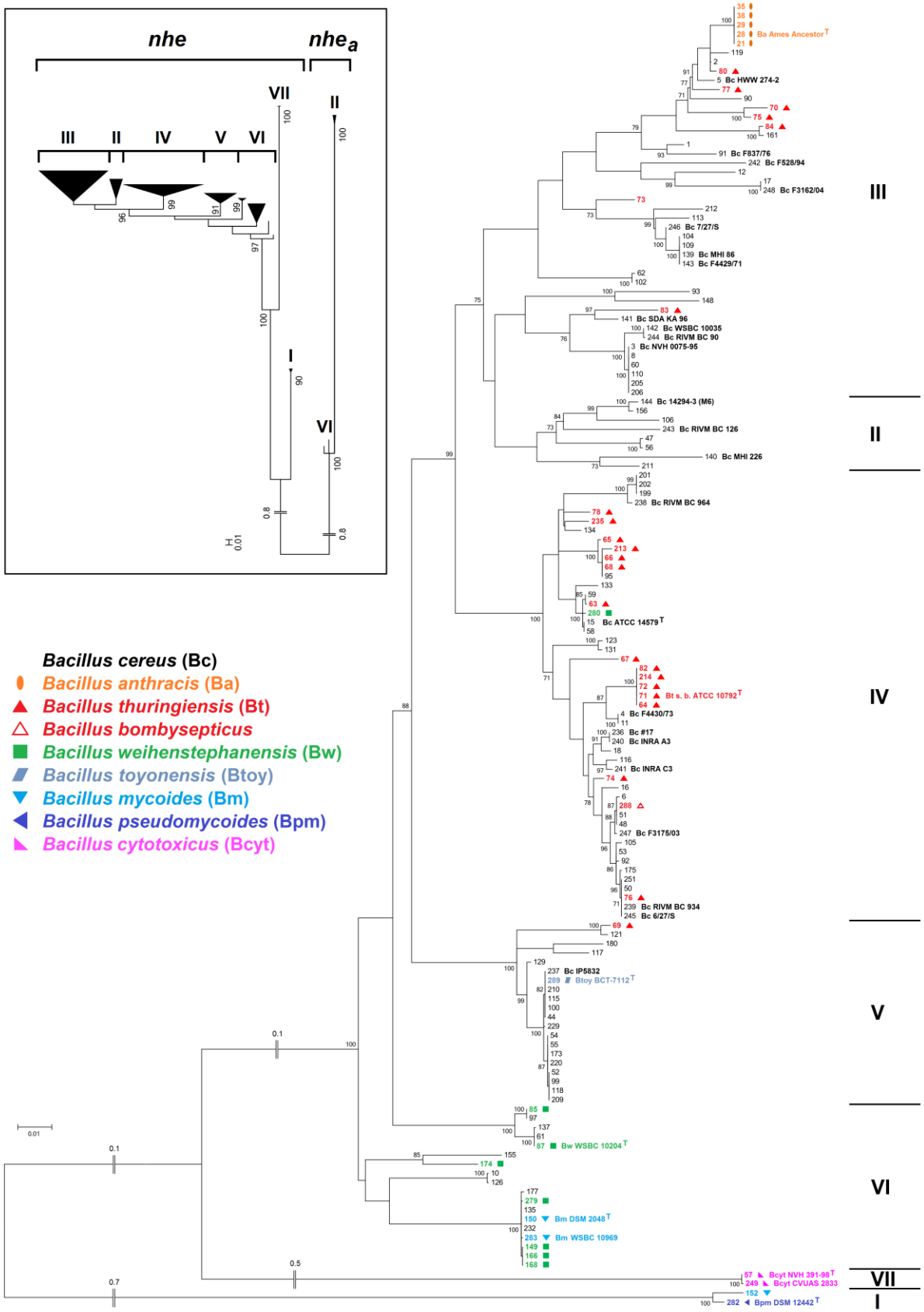


Fig. 14: *Nhe* in *B. cereus* sensu lato.

Phylogenetic *nheABC* tree (Maximum Likelihood Method) based on the concatenated sequences from 142 *B. cereus* sensu lato strains. **Inset:** Phylogenetic tree based on the concatenated sequences of *nheABC* and *nheABC*$_a$ from 142 *B. cereus* sensu lato strains, indicating an ancient origin of *nheABC*$_a$.

While no duplications of *cytK* or *plcR* could be found, a very rare second *nheABC* operon was noticed in four of the 142 *B. cereus* sensu lato strains investigated (Fig. 5 and 14). Two *B. cereus* strains (#140, #144, II) and two *B. weihenstephanensis* strains (#85, #87, VI) possess a second *nheABC* variant which was termed nhe_a. In *B. weihenstephanensis* KBAB4 (#85, VI) the nhe_a operon is part of the 417 kb plasmid pBWB401. The other three nhe_a copies could not be located on a similar plasmid, but were shown to be genuine duplications (Table S4). The nhe_a operon contains all three *nhe* genes, is actively transcribed, albeit not in all strains (Fig. 15A), and a putative PlcR recognition site upstream of $nheA_a$ is present (data not shown). Nhe_a operons differ greatly from all known *nhe* (76 - 88 % sequence identity), including *B. cytotoxicus`* *nheABC* that was until now considered the only major and most distantly related variant. Cluster II and cluster VI nhe_a are two clearly distinct variants.

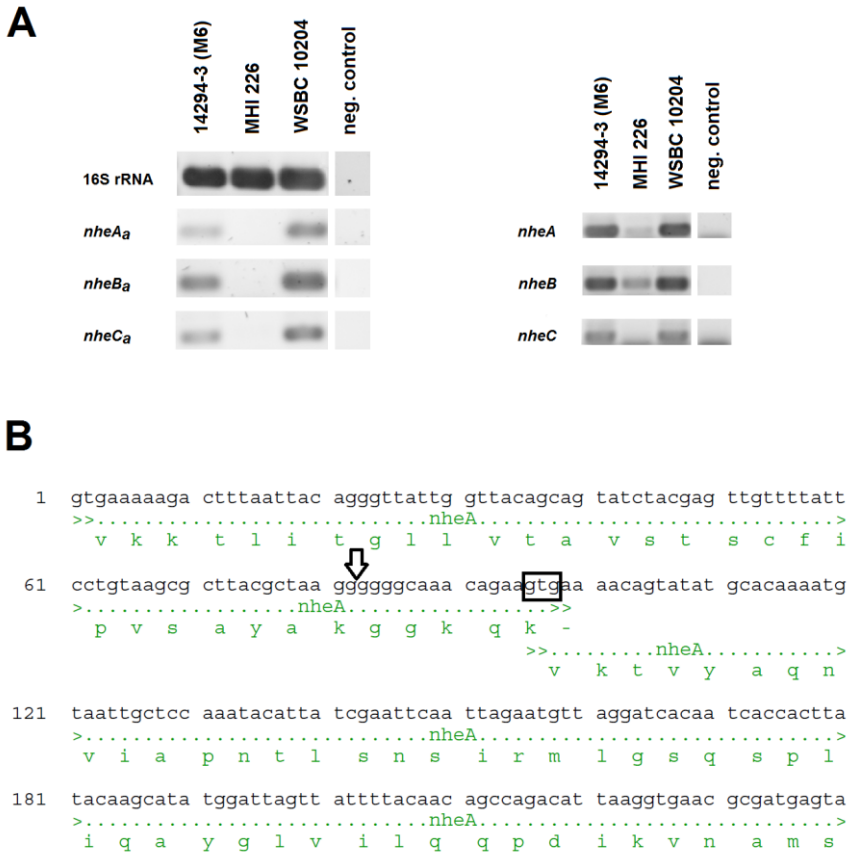


Fig. 15: Transcriptional analysis of *nhe* and nhe_a operons.

A: Transcription of *nhe* and nhe_a in *B. cereus* 14294-3 (M6), MHI 226 and *B. weihenstephanensis* WSBC 10204. Nhe_a is transcribed in *B. cereus* 14294-3 (M6) and *B. weihenstephanensis* WSBC 10204, but not in *B. cereus* MHI 226. The latter showed also a weak transcription of *nhe*. **B:** Partial *nheA* sequence of *B. cereus* MHI 1761. Frameshift (arrow, single base deletion) and a putative alternative start codon (black box) are marked. The truncated *nheA* begins with an alternative start codon (lacking 96 bp), but shows in the remaining part 98 % aa sequence identity to *nheA* of *B. cereus* NVH 0075-95 (*nhe*-reference).

3.2 Transcriptional analysis of the enterotoxin operons *nhe* and *hbl*

3.2.1 Analysis of 5' intergenic regions of *B. cereus* sensu lato enterotoxins

3.2.1.1 Length of 5' intergenic regions

To investigate the regulation of toxin expression, the 5' intergenic regions (5'IGRs) of the *cytK-1, cytK-2*, *nhe* and *hbl* operons of 142 *B. cereus* sensu lato strains (see section 3.1.2 and [215]) were compared in a multiple alignment (data not shown). Out of these, 27 *nhe,* 24 *hbl,* 3 *cytK-1* and 15 *cytK-2* toxin operons/genes of *B. cereus* strains representing the diversity of the seven phylogenetic groups of *B. cereus* sensu lato were selected and compared (Table 3, Fig. 16). The promoter containing intergenic regions of *cytK-1* and *cytK-2* (Fig. 16A/B) are relatively short (~100 bp) in comparison to the *nhe* and *hbl* 5'IGRs. However, only three *cytK-1* sequences from extremely similar *B. cytotoxicus* isolates were available, which render the comparative analysis less meaningful, but nonetheless confirm a potential clonal structure of the members of this species [30]. Some strains contain short insertions within their 5'IGRs, such as *B. pseudomycoides* DSM 12442 and other strains of phylogenetic cluster I, which possess insertions downstream of each PlcR binding site. Strains of this species have an *nhe* 5'IGR ~350 bp longer than all other *B. cereus* group strains investigated. In contrast, *B. cytotoxicus* (cluster VII) contains a *nhe* 5'IGR which is ~70 bp shorter than the other 5'IGRs, lacking the second PlcR binding site. Strains #61 and #87 (cluster VI) contain a 14 bp insertion upstream of the *hbl* ribosomal binding site (Fig. 16D). Strains of clusters III and IV, which harbor many pathogenic *B. cereus* strains, lack 11 bp close to each ResD binding site within P*hbl* (Fig. 16D). These missing regions might be used as an additional means to discern cluster III and IV from other *B. cereus* strains, but show no correlation to high toxicity, since both high and low enterotoxic strains [157] were found to harbor the deletions.

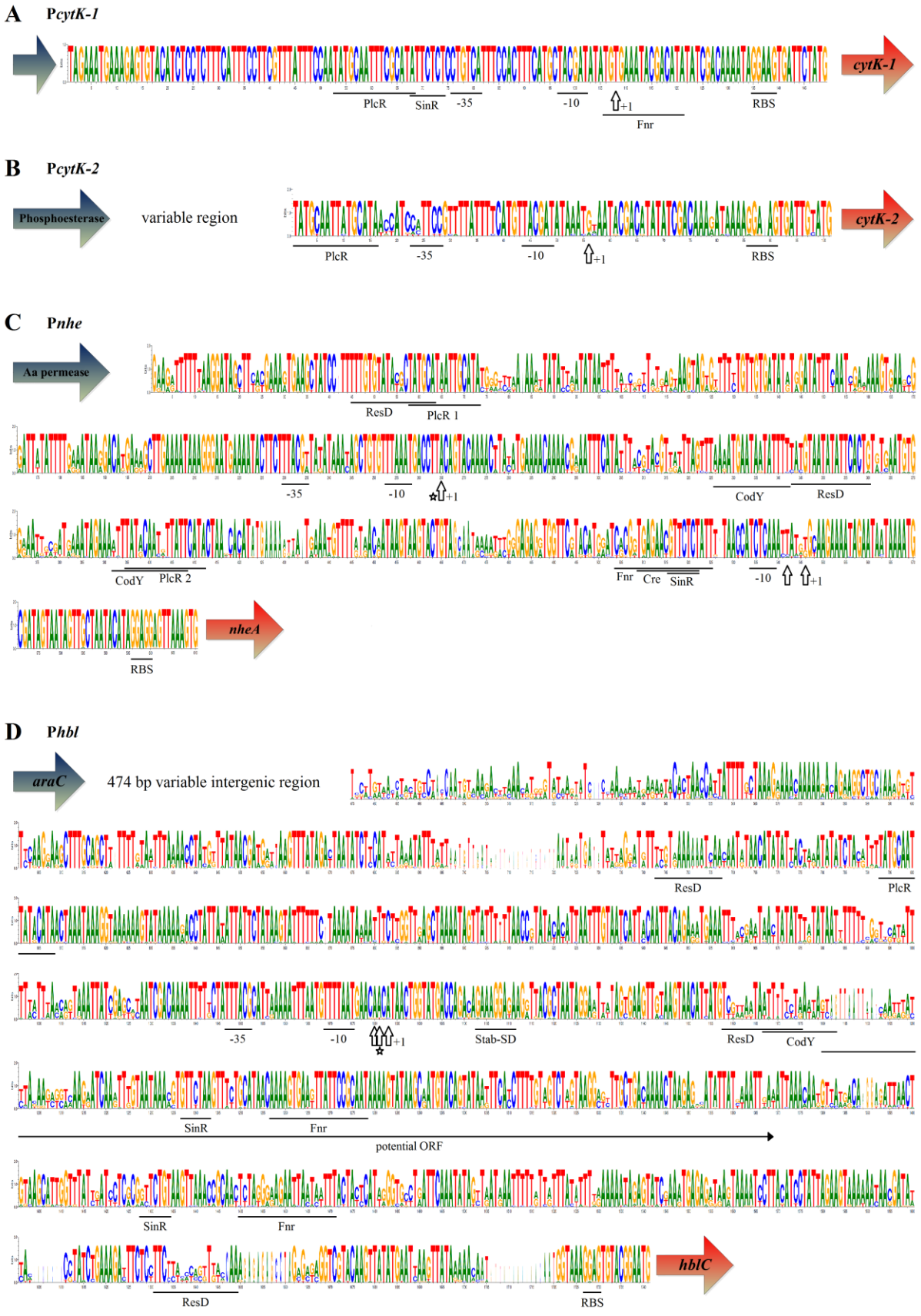

A PcytK-1
PlcR
SinR
-35
-10
+1
Fnr
RBS
cytK-1
B PcytK-2
Phosphoesterase
variable region
PlcR
-35
-10
+1
RBS
cytK-2
C Pnhe
Aa permease
ResD
PlcR 1
-35
-10
+1
CodY
ResD
CodY
PlcR 2
Fnr
Cre
SinR
-10
+1
RBS
nheA
D Phbl
araC
474 bp variable intergenic region
ResD
PlcR
-35
-10
+1
Stab-SD
ResD
CodY
SinR
Fnr
potential ORF
SinR
Fnr
ResD
RBS
hblC

Fig. 16: Structure and sequence of intergenic regions containing enterotoxin promoters in *B. cereus* sensu lato. Sequence motifs were determined by sequence comparison. Promoter regions (-35, -10) and transcription starts (+1, vertical arrow) [116, 126, 127, 216], CodY binding sites [170, 217-219], catabolite responsive element (Cre) [160], PlcR binding sites [116, 127, 156, 216], ribosomal binding site (RBS) of *hbl* [124], ResD and Fnr binding sites [158, 159, 220], SinR binding sites [221, 222], and stabilizing Shine-Dalgarno sequence (Stab-SD) [223]. Conservation of the sequences is depicted as logo and based on a multiple sequence alignment of strains representative for the seven phylogenetic groups (see Table 1 and [215]). Transcription start sites in *B. cereus* INRA C3 are marked by an asterisk. Colored arrows indicate gene function and transcriptional direction of genes. Empty arrow: hypothetical protein. *araC*: AraC family transcriptional regulator TrrA. Black arrow: transcriptional direction of a potential open reading frame (ORF).

A: *CytK-1* 5' intergenic region of 3 *B. cytotoxicus* strains. Cluster VII strains contain a non-annotated ORF upstream of *cytK-1*.

B: *CytK-2* 5' intergenic region of 15 *B. cereus* sensu lato strains. All strains are affiliated to phylogenetic clusters II – V. With the exception of the promoter, the intergenic region upstream of *cytK-2* is not conserved. 80 % of all *cytK-2* strains possess a phosphoesterase gene directly upstream of *cytK-2*. In the remaining strains an insertion of 500 – 2000 bp length separates the two genes.

C: *Nhe* 5' intergenic region of 27 *B. cereus* sensu lato strains. All clusters except I and VII contain a hypothetical amino acid permease gene upstream of *nheA*. The cluster VII (*B. cytotoxicus*) intergenic region contains the same promoter elements as the other strains with an overall sequence identity of 70 – 90 %. The intergenic region of cluster I strains (*B. pseudomycoides*) consists of the same promoter elements, but is ~350 bp longer. Thus, cluster I strains were excluded from the analysis.

D: *Hbl* 5' intergenic region of 24 *B. cereus* sensu lato strains. In 95 % of all investigated *hbl* strains *araC* appears 1600 – 1200 bp upstream of *hblC*. The intergenic region upstream of *hbl* varies in size. Presented is the entire region of which up to 500 bp are lacking in several strains. Insertions occur in clusters II, V and VI (nucleotides 1182 – 1192, 1651 – 1661) and in strains #87 (nucleotides 696 – 720, 1706 – 1720) and #155 (nucleotides 696 – 720, 1604 – 1610). A putative ORF starting with an alternative start codon (in most strains TCA or TAT) is noted. The *hbl* operon is part of a degraded transposon [215]. A transposase (pseudogene) occurs in two cluster VI strains (#155, #174) instead of *araC*.

3.2.1.2 Transcription start sites and 5' untranslated regions (5'UTRs)

A transcription start site of *nhe* in *B. cereus* strains NVH 0075-95 and NVH 1230-88 was reported to localize 66 bp and 62 bp upstream of the *nheA* startcodon [116], while in *B. thuringiensis* Bt407 the transcription start is located 331 bp upstream of *nheA* [127], potentially indicating strain-specific promoter differences. 5'RACE was used to determine transcription start sites of *nhe* and *hbl* in *B. cereus* INRA C3 and repeated once for validation. The amplified products had a length of approximately 400 bp (*nhe*) and 750 bp (*hbl*) (Fig. 17). Sequencing of the 5'RACE products revealed the transcription start sites. The transcription start of the *nhe* operon in *B. cereus* INRA C3 was shown to be identical with the one in *B. thuringiensis* Bt407 (Fig. 16C). The 5' untranslated region (5'UTR) is ~350 bp long and conserved.

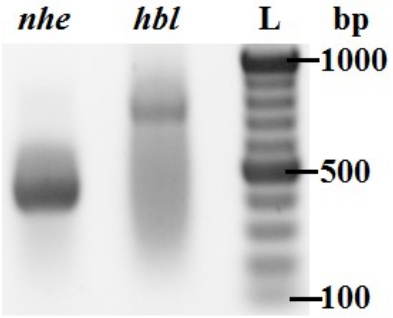


Fig. 17: 5'RACE products.
The amplified 5'RACE products were analyzed on 1 % agarose gels and sequenced to determine the transcription starts of *nhe* and *hbl* in *B. cereus* INRA C3. L = GeneRuler^{TM} DNA Ladder Mix.

The transcription start of *hbl* in *B. cereus* type strain ATCC 14579 was reported to localize 606 bp upstream of *hblC* [126], which is identical to the transcription start of *hblC* in *B. cereus* INRA C3 (this study, Fig. 16D). The transcription start of *hbl* in *B. thuringiensis* Bt 407 is located 605 bp upstream of *hblC* [127]. The 5'IGR of *hbl* shows an exceptionally long, generally conserved 5'UTR of ~660 bp. Therefore, both toxin operons harbor rather extended and conserved 5'UTRs in their 5' intergenic sequences.

3.2.1.3 Putative regulator binding sites

Putative recognition sites of all known regulators demonstrated so far to be involved in *B. cereus* enterotoxin expression have been predicted within 5'IGR alignments (Fig. 16). Putative binding sites for the same regulators were identified in the *nhe* and *hbl* intergenic regions. Most regulator sites within the *nhe*, *hbl* and *cytK* 5'IGRs have been confirmed by experimental studies during recent years: Within the *nhe* 5'IGR two PlcR binding sites are localized. Numerous Fnr sites were found in the *nhe* and *hbl* 5'IGRs [159]. Comparison with the consensus sequence and a repetition of the original *in silico* analysis confirmed one Fnr site in the *nhe* and two in the *hbl* 5'IGRs (Fig. 16). Binding sites of ResD are

conserved in *nhe* but more variable in *hbl* 5'IGRs. While both *nhe* and *hbl* expression are regulated by catabolite repression (the cre site of *hbl* is located downstream of the start codon), *cytK* expression is not controlled by CcpA [160].

The stabilizing Shine-Dalgarno sequence was located in P*hbl* and SinR (master regulator of biofilm formation) binding sites were detected within the enterotoxin promoter regions based on sequence comparison. In contrast to the tripartite enterotoxin operons *nhe* and *hbl*, both *cytK-1* and *cytK-2* contain no 5'UTR. Thus, apart from the detection of the PlcR binding site immediately upstream of the -35 element no further regulator binding site was described so far [111, 216]. Bioinformatic analysis of the 5' IGR of *cytK-1* and *cytK-2* predicted putative SinR and Fnr site in P*cytK-1* that need further experimental confirmation. Both *nhe* and *hbl* 5'IGRs contain motifs which may bind CodY (for further analysis and discussion see below).

3.2.2 Long 5' untranslated regions control expression of tripartite enterotoxins

3.2.2.1 Complete nhe *5'UTR is necessary for full promoter activity*

The functionality of the unusually long *nhe* 5'UTR was investigated using several partial deletions of 5'UTR sequences and luciferase as a reporter (Fig. 18A). All promoter fusion constructs created in this study are listed in section 2.8, Table 6. The constructs were tested for promoter activity in *B. cereus* INRA C3. Experimental analyses focused on *B. cereus* INRA C3 due to the presence of all three main enterotoxins *nhe, hbl* and *cytK-2* in this highly toxic strain [157]. The *nhe* operon was reported to contain two promoters and two PlcR binding sites [116, 127] (this study, Fig. 16C and 18A). The full-length sequence showed with $1.90*10^6$ RLU_{490}/OD_{600} the highest activity of all P*nhe* constructs while P*nhe-s1* (P*nhe* lacking PlcR binding site 1) showed a strongly decreased activity, indicating the activating role of PlcR binding site 1 by an 11x decrease of promoter activity (Fig. 18B). The highly conserved site PlcR 1 is 16 bp long and responsible for PlcR-dependent *nhe* expression. The less conserved PlcR 2 contains a 2 bp central insertion (Fig. 16C). Deletion of the upstream promoter abolished transcription (P*nhe-s2*, $1.53*10^3$ RLU_{490}/OD_{600}) indicating that the first promoter is the active and essential promoter in *B. cereus* INRA C3. Since PlcR is the major virulence regulator in *B. cereus*, a promoter fusion lacking both PlcR binding sites (P*nhe-s3*) was constructed. P*nhe-s3* activity was similar to P*nhe-s2* activity, indicating that the PlcR binding site 2 did not have an activating function on its own (Fig. 18). A promoter fusion lacking the second promoter (P*nhe*-Δ5'UTR) resulted in an activity similar to the control pXen1 (398 RLU_{490}/OD_{600}) with no active promoter. This demonstrates that this fragment of the promoter region also contains one or more essential activating regulatory elements, which might include a ResD, cre, Fnr, PlcR 2 and a putative SinR binding site.

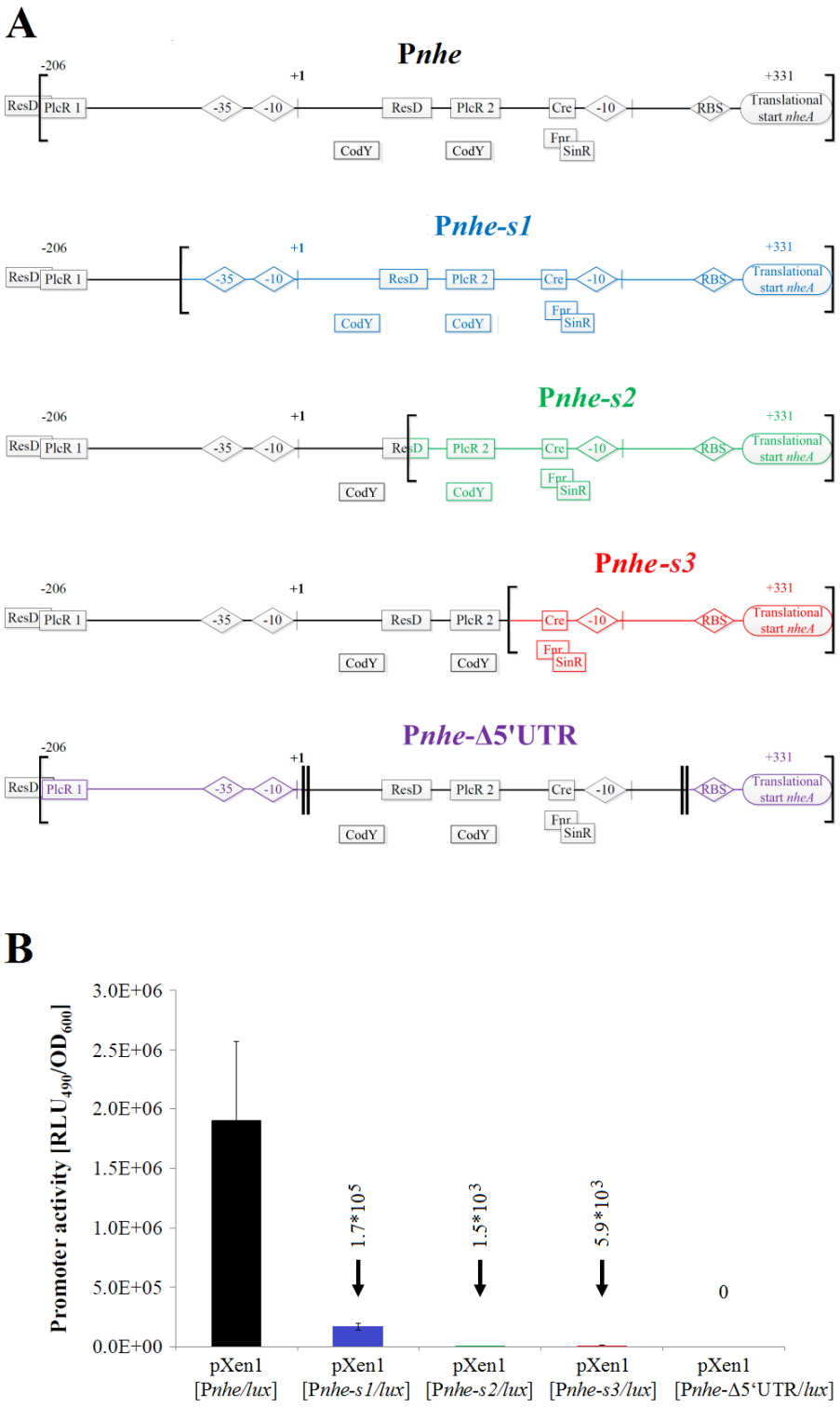


Fig. 18: Promoter activity of complete and partial *nhe* 5'IGR in *B. cereus* INRA C3.

A: P*nhe* full construct and shortened variants. Regions analyzed in promoter fusions are named and indicated by brackets, double lines enclose deletions and +1 is the transcription start site determined by 5'RACE. Promoter elements and (putative) binding sites of transcriptional regulators (compare Fig. 16) are displayed. **B:** P*nhe* promoter activities were determined in MOD minimal medium in triplicates and compared at the time of peak activity of the construct containing the entire promoter region. Luminescence signals were generated by the transcription of *lux* genes located downstream of the complete or partial 5'IGR tested for promoter activity. Negative control pXen1 without active promoter is shown in Fig. 22.

3.2.2.2 **Hbl *5'UTR represses* hbl *transcription***

To study the function of the approximately 600 bp long *hbl* 5'UTR *lux* reporter fusions including partial or complete deletion of the 5'UTR were created (Fig. 19A). The full-length construct contains the highly conserved PlcR binding site upstream of the *hbl* promoter (Fig. 16D and 19A). Transcription of the *hbl* operon has already been shown to be PlcR-dependent [127]. Deletion of the entire 5'UTR (P*hbl*-Δ5'UTR, $3.16*10^6$ RLU_{490}/OD_{600}) led to an increased promoter activity compared to the wildtype construct ($1.26*10^6$ RLU_{490}/OD_{600}, Fig. 19B). In contrast, deletion of the downstream half of the 5'UTR (P*hbl*-Δ5'UTR-down) showed the wildtype activity. Therefore, the putative ResD, SinR and Fnr binding sites in this region do not influence promoter activity under the used experimental conditions. Deletion of the upstream half of the 5'UTR (P*hbl*-Δ5'UTR-up) containing putative binding sites for CodY, ResD, SinR and Fnr led to a stimulation of transcription ($2.3*10^6$ RLU_{490}/OD_{600}), which was less pronounced compared to the deletion of the entire 5'UTR. It can be concluded that the 268 bp region designated 5'UTR-up (Fig. 19A) partially represses *hbl* transcription in *B. cereus* INRA C3.

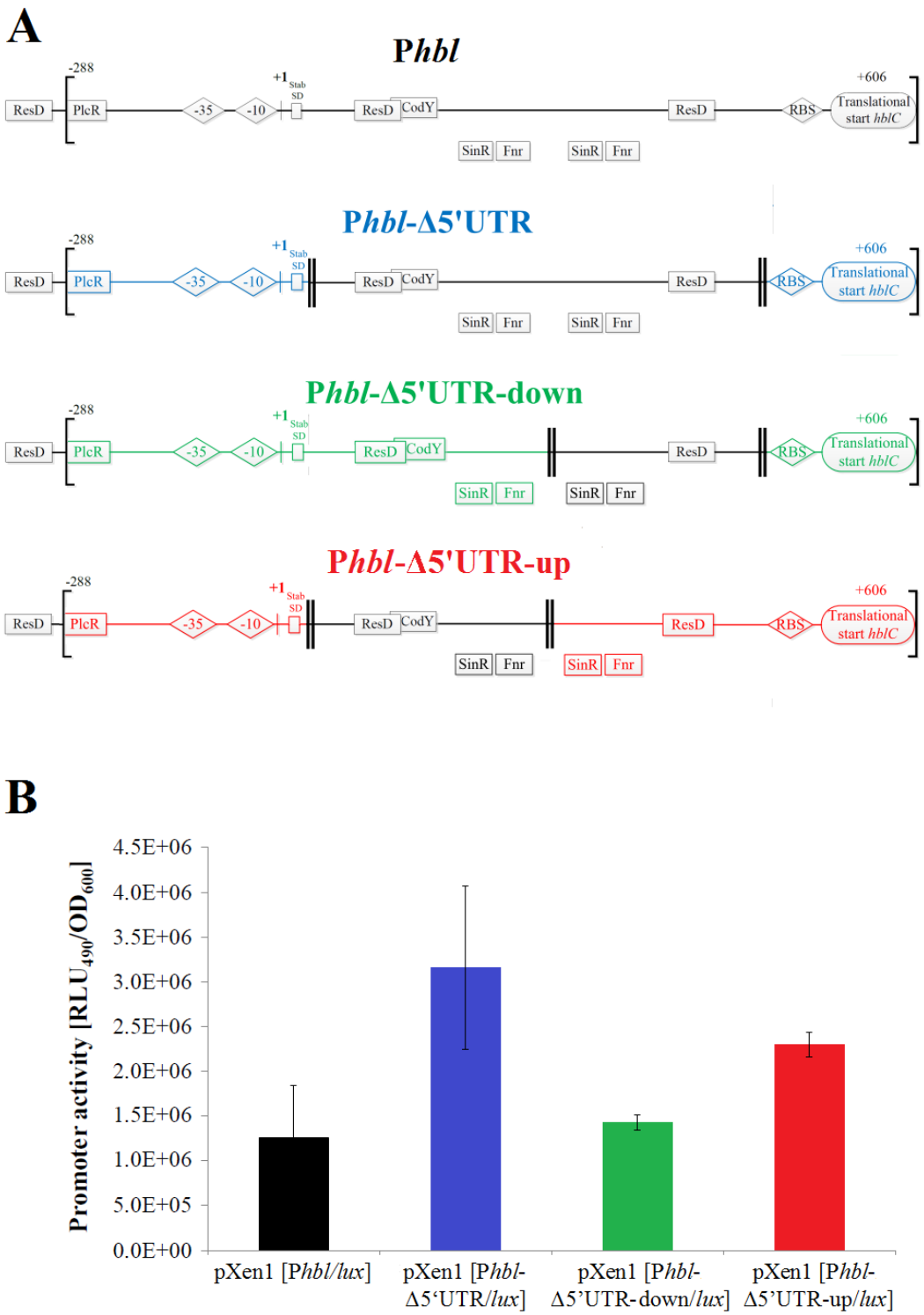


Fig. 19: Promoter activity of complete and partial *hbl* 5'UTR in *B. cereus* INRA C3.

A: P*hbl* wildtype construct and deletion variants. Regions analyzed in promoter fusions are named and indicated by brackets, double lines enclose deletions and +1 is the transcription start site determined by 5'RACE. Promoter elements and (putative) binding sites of transcriptional regulators (compare Fig. 16) are displayed. **B:** P*hbl* promoter activities were determined in MOD minimal medium in triplicates and compared at the time of peak activity of the construct containing the entire promoter region. Luminescence signals were generated by the transcription of *lux* genes indicate promoter activity of the 5'UTR variant tested. Negative control pXen1 without active promoter is shown in Fig. 22.

3.2.3 From lab to gut-simulating conditions: Increased enterotoxin promoter activity

3.2.3.1 Strain specific enterotoxin transcription is enhanced by nutrient deficiency

The tripartite enterotoxins Nhe and Hbl are suspected to be expressed in a strain-dependent manner [2]. Transcriptional activity of P*nhe* and P*hbl* regions was studied in the highly enterotoxic strains *B. cereus* INRA C3 and NVH 0075-95 [157] using promoter fusions with the luciferase genes *luxABCDE*. Promoter fusion constructs created in this study are listed in section 2.8, Table 6 and described in detail in Fig. 18A and 19A. To analyze the kinetic of enterotoxin expression in different media the three constructs containing complete promoter regions (*B. cereus* NVH 0075-95 pXen1 [P*nhe52/lux*], *B. cereus* INRA C3 pXen1 [P*nhe/lux*], *B. cereus* INRA C3 pXen1 [P*hbl/lux*]) were tested in LB and CGY medium (Fig. 20). Growth of all three strains was similar and 1 % glucose in CGY medium allowed fastest growth during exponential phase. *B. cereus* NVH 0075-95 pXen1 [P*nhe52/lux*] reached its maximal promoter activity in LB after 12 h ($2.63*10^5$ RLU_{490}/OD_{600}), while CGY medium led to a first peak of activity after 7 h but the maximal promoter activity was not reached until 26 h ($2.15*10^5$ RLU_{490}/OD_{600}, Fig. 20A). A similar *nhe* expression pattern can be observed in *B. cereus* INRA C3 pXen1 [P*nhe/lux*] (Fig. 20B). In LB medium peak activity was reached after 9 h ($4.64*10^5$ RLU_{490}/OD_{600}), in CGY after 26 h ($2.39*10^5$ RLU_{490}/OD_{600}), but *B. cereus* INRA C3 showed – especially in LB medium – higher P*nhe* activity. Thus, the kinetic studies indicate not only strain-, but also media-dependent differences in *nhe* expression. To further investigate the influence of media-components on the expression of tripartite enterotoxins, reporter strains were grown in LB + 2 % casamino acids (CAA). LB medium contains 1 % tryptone, while LB + 2 % CAA comprises acid hydrolyzed casamino acids. In the latter *B. cereus* showed retarded growth and a maximal OD_{600} lower than in LB and CGY medium, but increased promoter activity of *nhe* and *hbl* during stationary growth phase (Fig. 20A-C). The influence of free amino acids (present in MOD minimal medium and in CAA-containing media) on transcription of *nhe* and *hbl* is analyzed in detail in section 3.2.5.

The longest lag phase of ~6 h was found in MOD minimal medium. In minimal medium *nhe* promoter activity was at least 1.5x higher (*B. cereus* NVH 0075-95) and even 4.5x higher in *B. cereus* INRA C3 than in all other tested media. The kinetics of both *B. cereus* strains show in MOD medium a single peak in P*nhe* activity during late exponential phase. The same observation could be made for *hbl* promoter activity in *B. cereus* INRA C3 (Fig. 20C), which was in MOD medium 14.5x higher than in any other medium. The course of *hbl* promoter activity resembled *nhe* promoter activity, but P*nhe* activity was generally higher than P*hbl* activity in all media except MOD. Despite strain-dependent differences and the influence of amino acid or carbon sources on promoter activity and growth, nutrient deficiency was the most important trigger for enterotoxin expression with a particularly strong increase in *hbl* promoter activity.

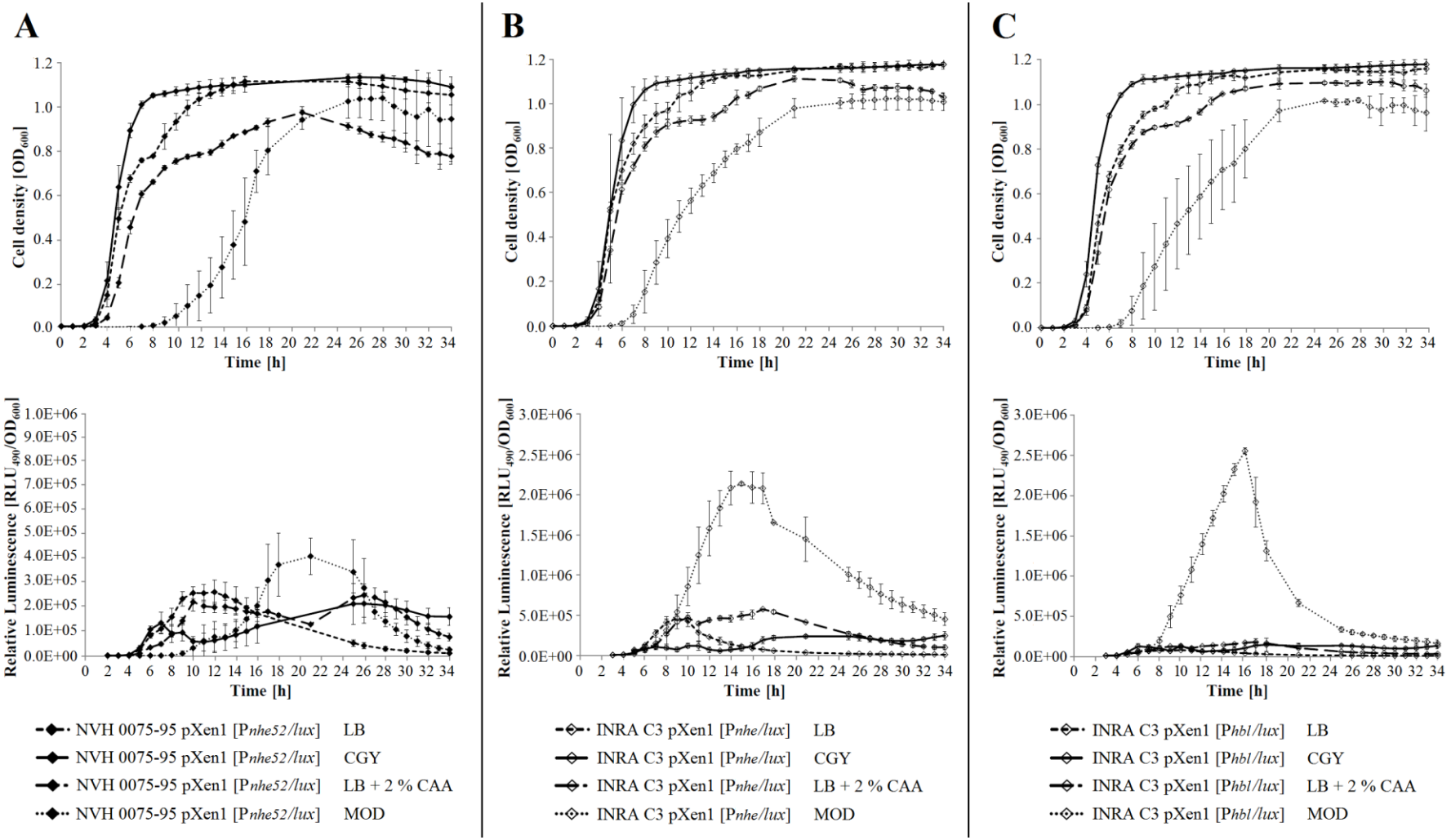


Fig. 20: Growth and promoter activity of P*nhe* and P*hbl* in *B. cereus.*

A: *B. cereus* NVH 0075-95 pXen1 [P*nhe52/lux*], **B:** *B. cereus* INRA C3 pXen1 [P*nhe/lux*], **C:** *B. cereus* INRA C3 pXen1 [P*hbl/lux*]. Experiments were performed at 37 °C, 800 rpm, mean values and standard deviations of three biological replicates are shown. Promoter activity of pXen1 without promoter was < $4.5*10^5$ RLU_{490}/OD_{600} (data not shown).

3.2.3.2 Gut-simulating conditions enhance enterotoxin promoter activity

The kinetics of enterotoxin expression shown in Fig. 20 were obtained in experiments in a volume of 200 µl and shaken at 800 rpm. In food poisoning cases *B. cereus* produces its enterotoxins during passage of the human gut. To approach a more host-like experimental set-up, volume was first raised to 45 ml under aerobic conditions. *B. cereus* NVH 0075-95 and *B. cereus* INRA C3 showed a different growth behavior at 30 °C and at 37 °C during the clinically relevant first 12 h and significantly lower enterotoxin expression in CGY medium (Fig. 21). A maximal OD_{600} of ~14 at 30 °C and ~12 at 37 °C was achieved by all strains. Despite higher cell density, promoter activities never exceeded $3.0*10^4$ RLU_{490}/OD_{600}. In Figure 18B maximal activity that occurred during 12 h of incubation is compared, which is the late peak in all aerobic cultures and the first peak in the stagnant cultures (data not shown). Under aerobic conditions at 30 °C *B. cereus* INRA C3 P*nhe* activity was up to 7x higher than P*nhe* activity of NVH 0075-95. In a volume of 45 ml *B. cereus* INRA C3 showed higher P*nhe* activity than NVH 0075-95 and the course of P*hbl* activity resembled P*nhe* in INRA C3, albeit at a lower level (data not shown) similar to the results obtained in a volume of 200 µl.

Within the human gut, aerobic conditions are rare and *B. cereus* grows in vicinity to epithelial cells which are targets to its enterotoxins. Thus, kinetics of toxin expression were determined in the gut-simulating media GSM and cGSM at 7 % CO_2 without shaking of cultures (see section 2.1). GSM-based media allow growth of *B. cereus* only to a maximal OD_{600} of ~2.7, while an OD_{600} of ~6 was reached in CGY medium under identical oxygen conditions (Fig. 21A). Nevertheless, the more restricting conditions of GSM enabled higher enterotoxin promoter activities (Fig. 21B) than in CGY medium. Secreted factors of epithelial cells (cGSM) did not lead to significant induction of enterotoxin promoters, but caused a slightly faster growth while reaching the same maximal optical density. P*nhe* activity in *B. cereus* INRA C3 was higher than in NVH 0075-95 and P*hbl* activity in INRA C3 remained a little lower than P*nhe* activity (Fig. 21B), especially at 30 °C. P*hbl* activity in *B. cereus* INRA C3 resembled P*nhe* activity at a lower level at 30 °C and at 37 °C (data not shown) in all media indicating a joint regulation of expression.

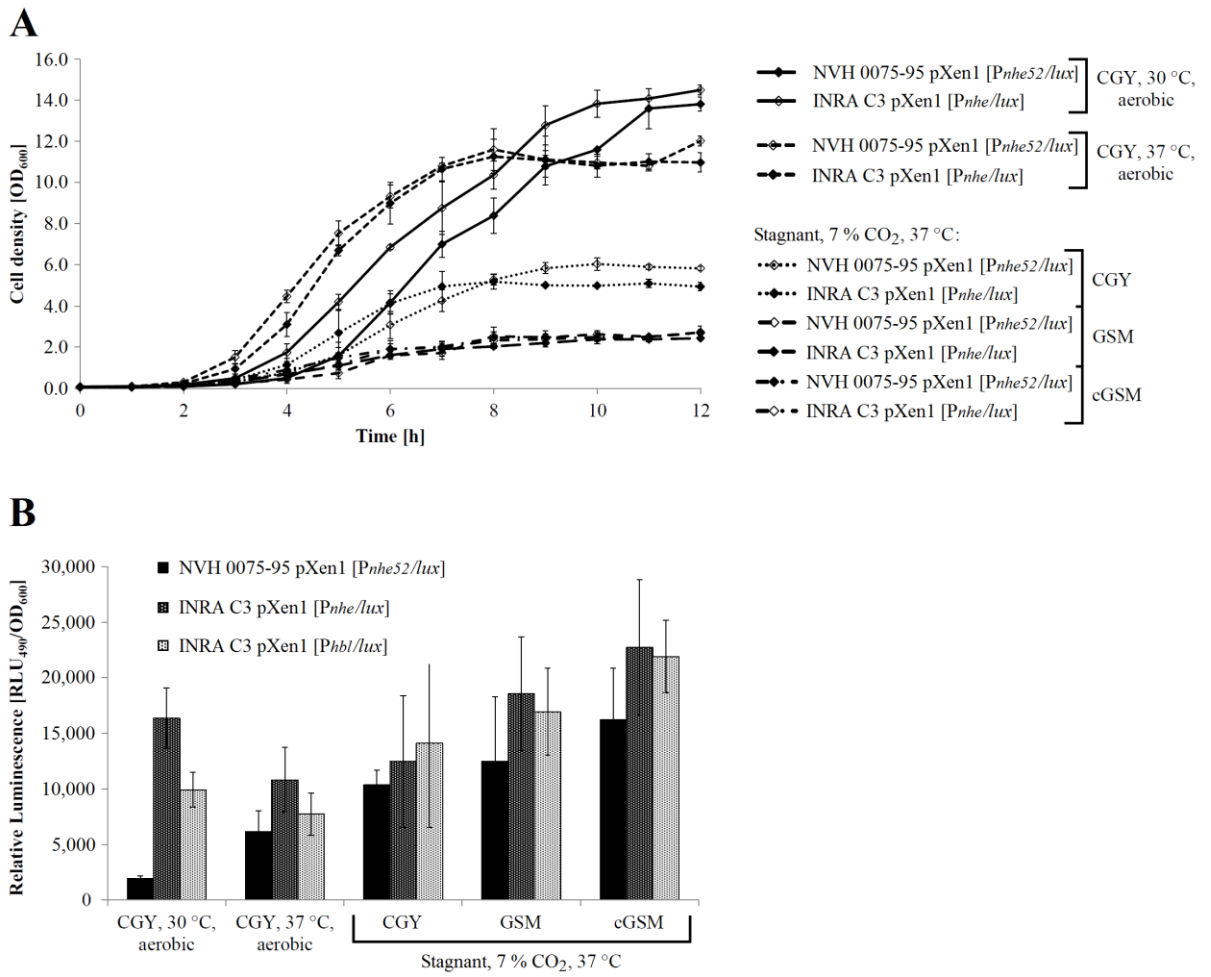


Fig. 21: *B. cereus* growth and promoter activity of P*nhe* and P*hbl* under different conditions.

A: Growth of *B. cereus* NVH 0075-95 pXen1 [P*nhe52/lux*] and *B. cereus* INRA C3 pXen1 [P*nhe/lux*] (representative for *B. cereus* INRA C3 constructs). **B:** P*nhe* and P*hbl* activity at different conditions. Maximal promoter activity occurring during 12 h incubation, mean values and standard deviations of three biological replicates are shown. Promoter activity of pXen1 without promoter was < $5.5*10^3$ RLU_{490}/OD_{600}.

3.2.4 Differentially expressed genes under gut-simulating conditions

Experiments under gut-simulating conditions showed slightly enhanced transcription of *nhe* and *hbl* in the presence of factors secreted by host epithelial cells. Transcriptomes were compared to study the global influence these host cell factors on transcription. Since a completely annotated genome is available for *B. cereus* F837/76 (Hbl reference strain), transcriptome analyses were performed with this strain, which contains the enterotoxin genes *nheABC* and *hblCDAB*, but no *cytK*. As described in section 2.4, bacteria were harvested after 2 h growth without agitation at 37 °C, 7 % CO_2 for RNA isolation, library preparation and sequencing. Transcriptomes were compared with GSM set as the reference condition. In cGSM 81 genes were upregulated and 25 genes downregulated (Table S5). 25 % of the downregulated genes encode hypothetical proteins and 20 % are insufficiently characterized. The majority of the few and only slightly downregulated genes encode ATP-dependent transporters. The

strongest downregulated gene (bcf_22175, -6x) encodes an endoribonuclease necessary for cleavage of single or double stranded RNA. 20 % of the upregulated genes encode hypothetical proteins and 11 % are not sufficiently characterized to sort them into functional categories. Most prominently upregulated (up to 64x) were genes essential for nitrate respiration, such as molybdenum cofactor biosynthesis genes and the *nar* genes (respiratory nitrate reductase). Additionally upregulated were genes that allow an enhanced acetyl-CoA formation under anaerobic conditions, such as *pflA* (pyruvate-formate lyase activating enzyme) and *pflB* (pyruvate-formate lyase). The enterotoxin genes *nheABC* were upregulated 7 – 9x, while no differential regulation of *hblCDAB* could be detected under cGSM conditions.

3.2.5 Absence of free amino acids activates enterotoxin transcription

The experiments shown in section 3.2.3.1 demonstrate that nutrient deficiency enhances transcription of *nhe* and *hbl*. Bioluminescent reporter strains containing the wildtype promoter and 5'IGRs including the upstream PlcR binding sites of *nhe* or *hbl* (P*nhe* and P*hbl*, Fig. 18A and 19A) were used to study the influence of amino acid source on enterotoxin expression in screening experiments (37 °C, 800 rpm, volume 200 µl). The amino acid source had impact on both growth and promoter activity (Fig. 20). While growth is delayed and reduced, a steep increase and a defined single maximum in promoter activity in MOD minimal medium in comparison to growth in nutrient rich media such as CGY was observed. In further experiments MOD minimal medium was supplemented with either tryptone (oligopeptides) or CAA (free amino acids) (Fig. 22). In MOD medium supplemented with 1 % CAA *nhe* and *hbl* enterotoxin promoter activity was only slightly lower than the activities achieved in unsupplemented MOD. Free amino acids are also present in MOD, albeit in lower amounts, confirming that nutrient deficiency is the most determining factor for enhanced enterotoxin promoter activity. When *B. cereus* INRA C3 was grown in MOD minimal medium supplemented with 1 % tryptone, promoter activities of P*nhe* and P*hbl* were 3x and 7x lower than in MOD minimal medium, respectively. Maximal promoter activity occurred during stationary phase when easily accessible amino acids were already depleted. These results point to an activation of enterotoxin promoter activity during unfavorable conditions, such as the absence of easily metabolizable amino acids.

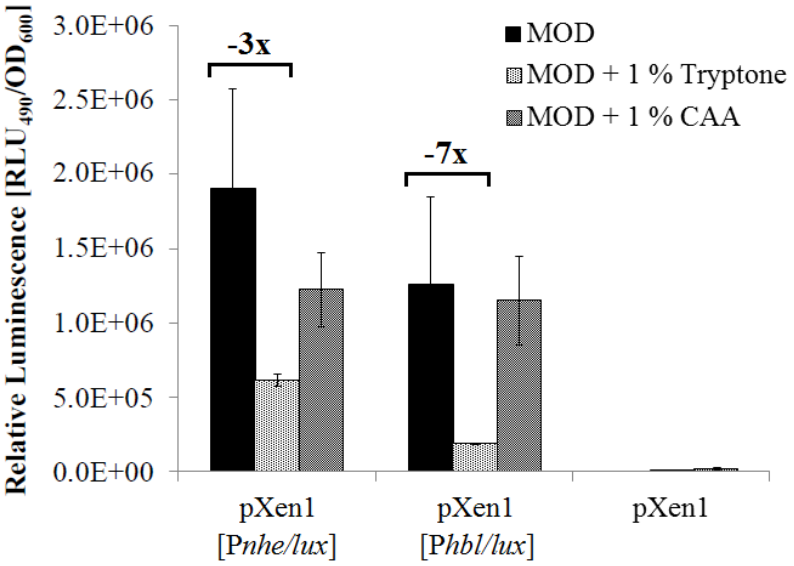


Fig. 22: Maximal enterotoxin promoter activities of *B. cereus* INRA C3 in different media.
Promoter activities were determined in triplicates and peak activities were compared. Luminescence signals generated by an active transcription of the *lux* genes are proportional to the activity of the promoter region tested. P*nhe* activity. MOD: 17 h, MOD + 1 % tryptone: 14 h, MOD + 1 % CAA: 15 h. P*hbl* activity. MOD: 16 h, MOD + 1 % tryptone: 13 h, MOD + 1 % casamino acids (CAA): 13 h. Multiplication of promoter activity in MOD + 1 % tryptone is displayed in comparison to MOD minimal medium.

3.2.6 CodY represses enterotoxin promoter activity

3.2.6.1 CodY binds as repressor to nhe *and* hbl *promoters*

Since the presented promoter activity studies hinted to a strong activation of enterotoxin transcription after depletion of free amino acids, the affinity of the BCAA-dependent regulator CodY to enterotoxin promoter regions was analyzed. Due to its size of over 900 bp the *hbl* 5'IGR was divided in five fragments and each tested in electro mobility shift assays (Fig. 23A). Comparison with the consensus sequence [219] identified putative binding sites with more than one mismatch to the consensus sequence in all tested sequences (Fig. 23A and S2). *In vitro* DNA affinity tests revealed that CodY shows a low affinity to fragments P*hbl*-1, -3 and -4 with an estimated dissociation constant K_D of around 700 nM. The most promising potential repressor binding site ATTTTT**C**GAATAGTCTA (with only one mismatch) was found in P*hbl*-2 and P*hbl*-5 downstream of the transcription start site (Fig. 16D and 23A, indicated by an ellipse). Both fragments P*hbl*-2 and P*hbl*-5 were bound by CodY (Fig. 23B) with an estimated K_D of <250 nM. To verify the positive gel shift results for P*hbl*-2 and P*hbl*-5, the *inhA1* promoter region was tested, which was already shown to be regulated by CodY [170]. Analysis of P*inhA1* resulted in a K_D of ~250 nM. Thus, CodY binds most probably to P*hbl* at a conserved CodY binding site 84 bp downstream of the transcription start site indicated by an ellipse in Fig. 23A.

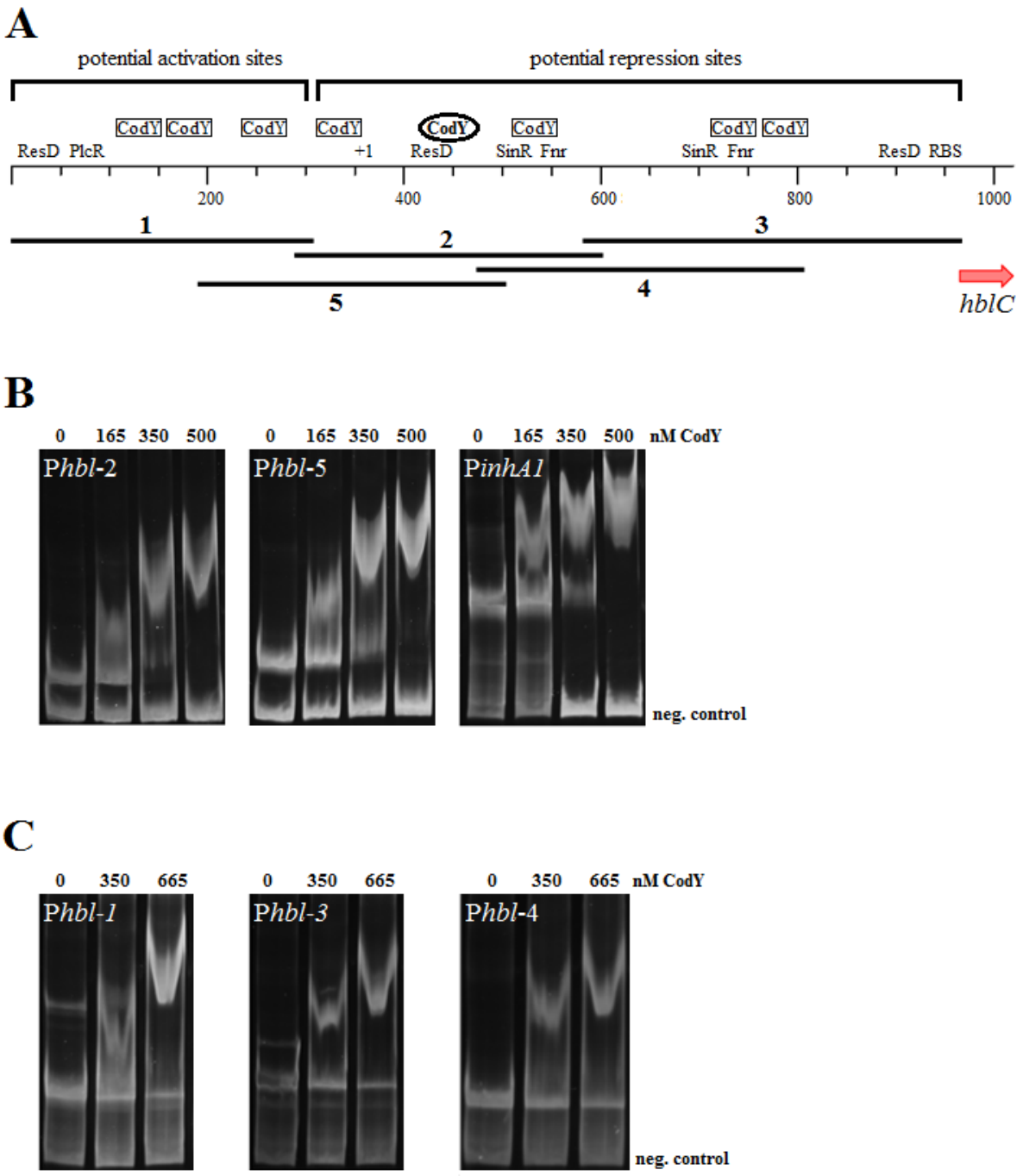


Fig. 23: Determination of CodY affinity to the *hbl* 5'IGR by gel mobility analysis.

A: 5'IGR fragments used in gel mobility analysis. The portrayed region of *B. cereus* INRA C3 P*hbl* is 966 bp long. All potential CodY binding sites found by an *in silico* analysis are indicated. Sites that contain two or more mismatches to the consensus sequence [219] are boxed. The CodY binding site indicated by an ellipse contains only one mismatch to the consensus sequence. **B and C:** Gel mobility shift assays of CodY binding to the *hbl* 5'IGRs. Reactions contained 100 ng DNA (401 – 501 fmol) and CodY concentrations are indicated with respect to the monomer. Negative control: 241 bp fragment amplified from the 16S rRNA gene *rrn*. **B:** Fragments P*hbl*-2, P*hbl*-5 and positive control P*inhA1* are bound by CodY (K_D ~250 nM). **C:** Fragments P*hbl*-1, P*hbl*-3 and P*hbl*-4 are bound with low affinity (K_D ~700 nM).

CodY also binds the *nhe* 5'IGR of *B. cereus* INRA C3 and *B. cytotoxicus* CVUAS 2833 at K_D ~125 nM and ~330 nM, respectively (Fig. 24). *CodY* and *nhe* operon sequences of *B. cytotoxicus* CVUAS 2833 are identical to the type strain *B. cytotoxicus* NVH 391-98, suggesting the same binding affinity. Additionally, CodY affinity to the *cytK* promoter regions was analyzed, but neither specific interactions with P*cytK-1* nor with P*cytK-2* were detected (K_D values >1000 nM, data not shown).

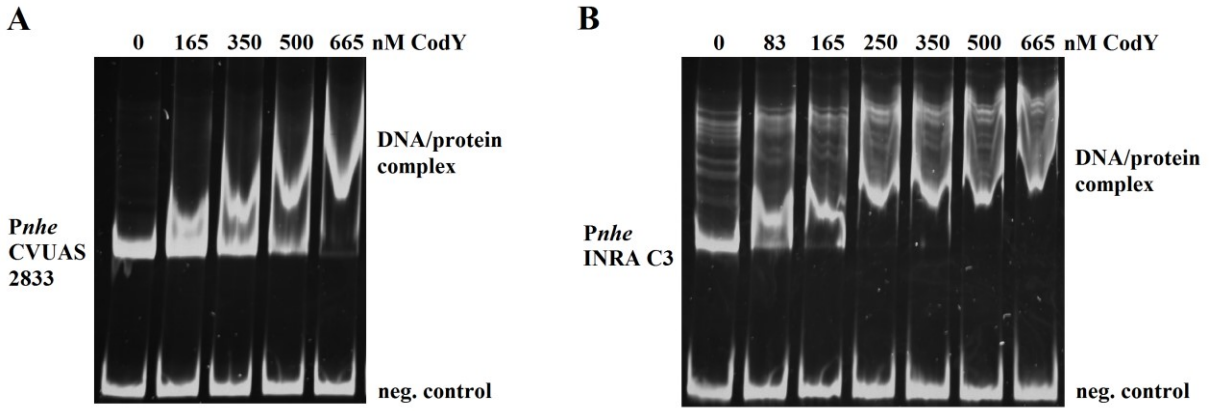


Fig. 24: Determination of CodY affinity to the *nhe* 5'IGR by gel mobility analysis.
A: Gel electro mobility shift assay of the *nhe* 5'IGR from *B. cytotoxicus* CVUAS 2833 with CodY: K_D ~330 nM. **B:** Gel electro mobility shift assay of the *nhe* 5'IGR from *B. cereus* INRA C3 with CodY: K_D ~125 nM. Reactions contained 100 ng DNA (298 and 271 fmol) and CodY concentrations are indicated with respect to the monomer. Negative control: 241 bp fragment amplified from the 16S rRNA gene *rrn.*

3.2.6.2 Fine-tuning of enterotoxin expression via conservation of regulator binding sites

The previous experiments are evidence for direct interaction between CodY and the *nhe* and *hbl* 5'UTRs. *B. cereus* INRA C3 and *B. cytotoxicus* CVUAS 2833 *nhe* 5'UTRs were positive in *in vitro* mobility shift assays (Fig. 24), but *B. cereus* F4810/72 was negative [170]. Whithin the *nhe* 5'UTR three putative binding sites with potentially repressive function were localized in the three *B. cereus* sensu lato strains tested so far (Fig. S3). One of the binding sites is identical in all three *nhe* 5'IGRs. Differences in the second binding site might cause the observed strain-specific deviation of binding affinity. The third potential binding site [170] occurs in *B. cereus*, but not in *B. cytotoxicus* leading to the conclusion that it plays a marginal role in CodY-mediated repression of P*nhe* activity. Both relevant sites are marked in Fig. 16C. DNA fragments containing only one of the two sites were negative in gel shift experiments (data not shown). Thus, more than one target site might be necessary to allow effective binding of CodY. The two putative CodY binding sites in the *nhe* 5'IGR and the confirmed site in the *hbl* 5'IGR were found in almost all of the 142 investigated *B. cereus* sensu lato strains (Fig. 25A). They show maximally one nucleotide mismatch to the *B. subtilis* consensus sequence [219], but a comparison reveals considerable variability with only one completely conserved T at position 14 and an almost conserved

A at position 10. In contrast, PlcR binding sites within the enterotoxin promoters are considerably higher conserved (Fig. 25B).

This suggests that CodY-mediated regulation depends on the individual target sequence and the consequential binding affinity, while regulation of enterotoxin expression via PlcR is less strains-specific and primarily controlled by quorum sensing and the activity of the PlcR-PapR complex. All putative ResD, Fnr, CodY, cre and the PlcR 2 sites are also variable (Fig. 16), perhaps indicating that enterotoxin transcription is controlled differently in each *B. cereus* strain.

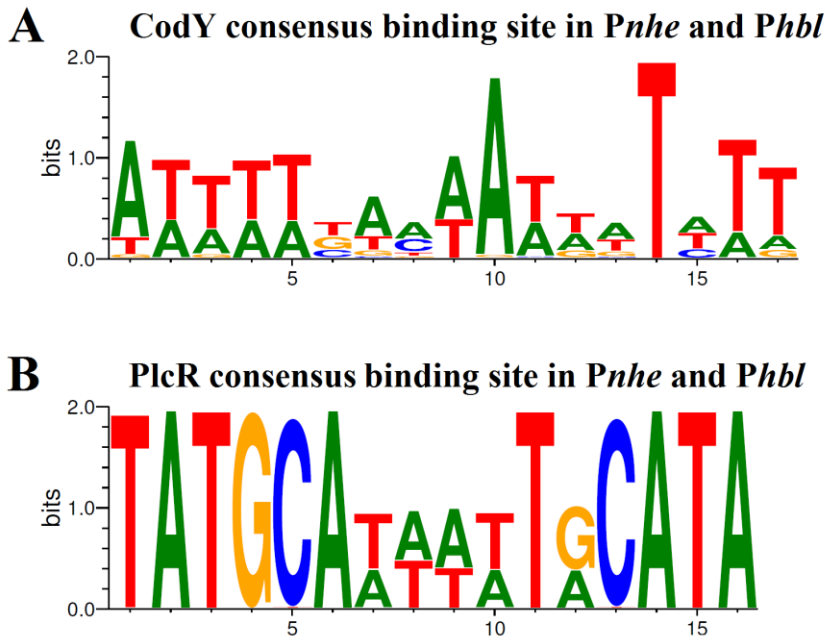


Fig. 25: CodY and PlcR consensus binding sequences in *B. cereus* sensu lato enterotoxins.

Conservation of the consensus sites is depicted as logo based on the sequence comparison of *nhe* and *hbl* 5'IGRs of 142 *B. cereus* sensu lato strains. Strainlist and detailed cluster affiliation are described in Table S1. **A:** CodY consensus sequence of three CodY binding sites in 142 strains (based on 379 sequences: one potential site in 142 *nhe* 5'IGRs, one potential site in 140 *nhe* 5'IGRs (not present in the two cluster I *B. pseudomycoides* strains) and one site in 97 *hbl* 5'IGRs). This consensus sequence is highly similar to the CodY consensus binding sequence in *B. subilis* [219]. **B:** PlcR consensus binding sequence found in all 142 *nhe* and *hbl* 5'IGRs (based on 239 sequences: 142 *nhe* 5'IGRs (PlcR 1), 97 *hbl* 5'IGRs).

4. Discussion

4.1 Inconsistent taxonomy of *B. cereus* sensu lato and virulence gene distribution

In this study 30 *B. cereus* sensu lato strains with different enterotoxic potential (for details see [157]) were sequenced and assembled *de novo* (Tables 9 and S1). It has been discussed controversially whether well-adapted pathogenic bacteria generally contain smaller genomes (due to less variable selection pressure) than environmental isolates [224]. Assembly sizes of the newly sequenced strains ranged from 4.2 (*B. cytotoxicus* CVUAS 2833) to 6.8 Mbp (*B. cereus* RIVM BC 934), which fits to genome sizes of already sequenced members of *B. cereus* sensu lato: *B. anthracis* 5.0 – 5.5 Mbp, *B. cytotoxicus* 4.1 Mbp, and *B. mycoides* 5.6 – 6.1 Mbp. However, several exceptions like the enterotoxic Nhe reference strain *B. cereus* NVH 0075-95 (6.1 Mbp) support the notion that genome size does not correlate with pathogenicity.

B. cereus sensu lato currently consists of the eight species *B. cereus, B. thuringiensis, B. anthracis, B. toyonensis, B. weihenstephanensis, B. mycoides, B. pseudomycoides* and *B. cytotoxicus.* Three additional species of the *B. cereus* group, *B. gaemokensis* [225], *B. manliponensis* [226] and *B. bingmayongensis* [227] have been discovered in Korea and China, but were excluded from this study since genome sequences were not yet available. This study suggests that *B. bombysepticus* [228] is a *B. thuringiensis* strain due to its insecticidal capacity and its shared localization in the species tree (Fig. 5). The existing *B. cereus* species affiliation is inconsistent with genomic strain relationships (Fig. 5). The most striking example were *B. thuringiensis* isolates scattered among clusters III to V. Techniques typically used to demarcate species, such as DDH, 16S rDNA sequencing and a combination of phenotypic characteristics, reach their limits when applied to *B. cereus* sensu lato, since the strains are closely related and many of the distinguishing virulence factors are encoded on mobile genetic elements. *B. cereus, B. thuringiensis* and *B. anthracis* are suspected to be one species on the basis of very close genomic relatedness (≥ 99.7 % 16S rRNA sequence identity) [27]. Analog, some strains of *B. mycoides, B. cereus* and *B. thuringiensis* show high genomic similarity to *B. weihenstephanensis* [28, 229]. Since the number of available strains and genomic data is constantly increasing, new methods have been developed for taxonomic distinction. MLST, AFLP and MLEE schemes that determine *B. cereus* phylogeny based on similarity of several genomic loci [60, 64, 97, 98] as well as BLAST distance phylogeny or ANI approaches based on entire genomes [43, 50, 230] are now used.

The phylogenetic analysis of this study is based on a MLSA species tree of 142 *B. cereus* sensu lato strains calculated from seven housekeeping genes (Fig. 5), which correlates nicely with whole-genome ANI (Fig. 6 and 7) and SNP-based phylogeny (Fig. 8). All three methods resulted in seven phylogenetic clusters which are consistent with the seven major groups found previously by comparison of molecular data (AFLP, 16S rRNA gene and *panC* analysis) and temperature tolerance [90]. A combination of 16S/23S rRNA gene and *gyrB* analysis [231] as well as MLST, MLEE and AFLP data [64] additionally confirmed the division of *B. cereus* sensu lato in seven clusters. Based on a more limited MLST approach, using partial gene sequences only as well as a different set of genes than applied in this MLSA

analysis, Cardazzo *et al.* [98] concluded that reticulate evolution of housekeeping genes should be an important factor of *B. cereus* evolution. Based on whole genome and MLSA analysis limited reticulate evolution within the seven housekeeping genes cannot be excluded, but this does not seem to mask the strain phylogeny of the *B. cereus* strain set. The whole genome analysis revealed that these seven phylogenetic groups are separated by at least 94 % ANI – with the exception of phylogenetic clusters II and III (Fig. 7). It has been proposed to use average nucleotide identity between genomes for bacterial species delineation [48]. A boundary of 94 – 96 % ANI corresponding to ~70 % DDH similarity was proposed [191, 232].

The distribution of the strains investigated in this study confirms the observation [27, 43, 90, 91] that many existing species affiliations do not match the genomic relationships (Fig. 5), e.g. an isolate designated as *B. weihenstephanensis* (#280) is found in cluster IV among *B. cereus* and *B. thuringiensis* strains. The presence of a unique *cspA* signature is described as specific for psychrotolerant *B. cereus* sensu lato (*B. mycoides, B. pseudomycoides* and *B. weihenstephanensis*) [90, 232], but the psychrotolerant *cspA*-signature could not be detected in *B. weihenstephanensis* FSL R5-860 (#280, IV). Thus, this strain is likely to be incorrectly classified as *B. weihenstephanensis*.

The seven major clusters can be divided into several subgroups, such as a branch of *B. cereus* and *B. thuringiensis* strains with genomic similarity to *B. anthracis* in cluster III or the two branches within cluster VI (Fig. 5). Due to its impact on agriculture and food industry, *B. cereus* sensu lato taxonomy is an interesting issue that was investigated by a combination of MLSA, 16S rRNA gene analysis and GBDP in 224 strains [230] yielding highly similar results. A large study that included 13,151 prokaryotic genomes in an analysis using genome-wide ANI (gANI) and the alignment fraction (AF) [50] between two genomes also confirmed the phylogenetic clusters found within *B. cereus* sensu lato. Both studies confirm the clusters and subclusters as depicted in the MLSA species tree. The corresponding clusters are listed in detail in Table S1. While clusters II and III show little evolutionary distance and cannot be distinguished reliably with any of the mentioned genomic methods, *B. toyonensis* BCT-7112 (#289, V, type strain [43]) is part of the distinct phylogenetic cluster V (ANI boundary > 96 %). The recently described species *B. toyonensis* BCT-7112 differs from other *B. cereus* sensu lato type strains at an ANI < 92 % and shows a distinct peptidoglycan diamino acid pattern [43]. Both *B. toyonensis* BCT-7112 and the newly sequenced *B. cereus* IP5832 (#237, V) are commercially available probiotics, Toyocerin® and Bactisubtil® [210], respectively. Thus, the almost clonally related *Bacillus* strains of cluster V might also be feasible as probiotics for animal feeding and all isolates belonging to this group should be renamed as members of *B. toyonensis*. However, the distribution of enterotoxin genes (Fig. 5) hints that cluster V strains might be potential human pathogens until their toxicity or safety is reliably established.

B. mycoides Rock1-4 (#151, I) and *B. mycoides* Rock3-17 (#152, I) were found together with *B. pseudomycoides* DSM 12442 (type strain, #282, I) in cluster I. Based on these data both strains should be renamed as *B. pseudomycoides* after confirmation by fatty acid profiling. The same conclusion was reached independently by Varghese et al. [50].

In cluster VI several strains (#149, 166, 168 and 174) containing the emetic *ces* cluster were shown to be *B. weihenstephanensis* [37, 38] and should be named accordingly. Interestingly, none of the ten emetic strains within the set of 223 *B. cereus* sensu lato strains – which appear only in clusters III (mesophilic *B. cereus*) and VI (psychrotolerant *B. weihenstephanensis*) – contains the *cytK* gene. Since *cytK* (especially *cytK-1* [30]) occurs mostly in thermotolerant strains (clusters III, IV and VII), this phenomenon might be linked to temperature preferences and also hints to a connection between psychrotolerance and the emetic toxin cereulide. *CytK-2* is far more frequent in *B. thuringiensis* strains (75 %) than in the rest of *B. cereus* sensu lato (36 %). These results match a study from 2006 that investigated 74 uncharacterized *B. thuringiensis* strains. All of them harbored the *nhe* genes, 74 % *hbl* and 73 % *cytK-2*, displaying about the same potential to cause diarrhea as *B. cereus* [137]. In 30 % of the investigated 223 *B. cereus* sensu lato strains all three enterotoxin operons *nhe, hbl* and *cytK* were found. The combined presence of *nhe, hbl* and *cytK* occurs more often among diarrheal (63 %) than among food-borne *B. cereus* sensu stricto strains (33 %) [118]. Despite an uneven distribution of enterotoxin genes among the phylogenetic clusters no group can be considered innocuous according to their toxin profile. Strains of clusters I and V might be low toxic due to the absence of *papR* and subsequent inactivation of the PlcR virulence regulon (cluster I) or due to known probiotic characteristics of at least some strains within the almost clonal cluster V. Still, no reliable correlation between high- and low-enterotoxic strains and their current species affiliation and toxin propagation could be found. These results fuel the need for development of additional typing methods. Distribution, variation and abundancy of additional virulence factors, such as sphingomyelinase, neutralproteases, immune inhibitors A or hemolysins, could be promising indicators for *B. cereus* sensu lato enterotoxicity.

4.2 Evolution of enterotoxin genes is only constrained by preservation of gene functionality

Horizontal gene transfer is mediated by plasmids or by recombination processes, which can occur ‘randomly’ or via transposable elements. Detection of horizontal gene transfer and recombination has always been difficult, due to incomplete and insufficient datasets as well as limitations of algorithms and substitution models. Nevertheless, certain criteria have been applied to deduce evidence of HGT from nucleotide sequence comparisons. The distribution of genes or phenotypes and the suspicious occurrence of a divergent variant within the considered cluster are the most important information [88, 233]. Additionally, atypical base composition, codon usage, oligonucleotide frequencies or number of open reading frames hint to lateral transfer between distantly related species [88]. When comparing bacteria as closely related as *B. cereus* sensu lato, most of these approaches are not feasible [89]. Thus, the comparison of phylogenetic tree topology is the most intuitive approach to detection of HGT between closely related strains. Phylogenetic trees represent the species or gene evolution only if

orthologous genes are compared which have not been transferred laterally. In contrast, paralogs result from gene duplications [83]. After several generations it becomes increasingly difficult to recognize paralogs, especially when gene loss is involved. However, *hbl* and *nhe* operons proved to be similar enough for distinction from their more dissimilar duplications.

4.2.1 Opposing evolution of *hbl*, *nhe* and their duplicates

4.2.1.1 Highly specific interaction of Hbl components restricts recombination

The *hbl* operon was found in 63 % of all 223 investigated strains, the duplication in 22 %. This strain set contains 26 published *B. thuringiensis* strains, 92 % of them contain *hbl* and 23 % hbl_a. A large study including 616 *B. cereus* and *B. thuringiensis* strains revealed the increased abundancy of *hbl* (87 %) among *B. thuringiensis* isolates as well [234]. Furthermore, *hbl* is chromosomally encoded while hbl_a is occasionally yet solely in *B. thuringiensis* located on plasmid. These results confirm the distinctly separate evolution of *hbl* and hbl_a (section 3.1.3.1, Fig. 10). The duplication of *hblCDAB* as well as the subsequent loss of *hblB* must be an unique event which occurred early in the evolution of *B. cereus* sensu lato since all $hblCDA_a$ genes cluster together, show a much more conserved tree topology and are clearly separated from *hblCDA* (Fig. 9 and 11) despite the fact that they are scattered over five phylogenetic clusters. The hypothesis that the duplication of *hblCDA* must be an ancient and unique event could be confirmed by its occurrence in all phylogenetic groups but III (loss of hbl_a at furcation of II and III) and VII (Fig. 5). Some strains of cluster IV probably acquired hbl_a via HGT. Deletion of $hblCDA_a$ in clusters II, V and VI indicates that a second copy of *hbl* might have low impact on the organisms` fitness (Fig. 5 and Fig. 9). Several strains were found that possess only hbl_a (#85, #97, #137, #140, #152, and #211), which seems to have assumed the function of *hbl*. *B. mycoides* Rock3-17 (#152, I) and *B. mycoides* Rock1-4 (#151) contain a variant of *hblCDA* different from both known versions. An obvious conclusion would be that these two strains have no hbl_a, but their *hblCDAB* developed independently and lost *hblB*.

The hbl_a operon consists of the genes *hblCDA*, while *hbl* comprises *hblCDA* and the *hblB* gene, which is not transcribed at a detectable level since the *hblCDA* mRNA appears to terminate within the *hblB* gene [126]. However, *hblB* was shown to possess its own transcription start site and both transcript and gene product HblB' have been detected [235]. These authors noted that HblB' might have the same function as HblB (encoded by *hblA*), but is expressed independently of PlcR. *HblB* could have been generated by duplication of a part of *hblA* and fusion with an open reading frame (ORF) in the 3` end [236]. Økstad *et al.* also speculate that the entire *hbl* operon might have arisen from one ancestor gene since the genes *hblCDAB* and their products have retained a high sequence similarity. The entire *hbl* operon sequence is extremely high conserved [2] (Fig. 9) yet frequently and freely exchanged among *B. cereus* sensu lato, while intra-operon recombinations are very rare. Possibly due to directional selection $hblCDA_a$ is as conserved as *hblCDA* (Fig. 10B). Both, high sequence conservation and rare

intra-operon recombination, suggest that the interaction between Hbl components might be quite specific, thus constraining sequence variation. This assumption is supported by experimental studies of the interaction of Hbl and the duplicated Hbl_a proteins. Both operons encode a functional toxin. However, despite their high similarity, not all Hbl / Hbl_a components are interchangeable to form functional toxins [133].

It has been speculated that *hblCDAB* is part of a large 18 kb transposon [2, 99, 236]. The analysis of the putative transposon region in the newly sequenced strains (Fig. 11) revealed that neither insertion sites nor length of the inserting region or adjacent genes are conserved. Thus, one may speculate that the *hbl* operon is part of a highly degraded transposon which is in most cases not functional anymore. The *B. anthracis* virulence gene *pagA* was detected in proximity to the *hbl* operon of *B. cereus, B. mycoides* and *B. weihenstephanensis* strains, proving that recombination between virulence plasmids and the bacterial chromosome occurs frequently.

4.2.1.2 Putative vital function determines evolution of nhe

Not only Hbl components/genes show similarity among each other, but also NheA, B and C were found to share high amino acid identities. In fact, a comparison of all six Nhe and Hbl proteins revealed 18 % to 44 % amino acid identity, which led to the suggestion that they may share a common ancestor gene [2]. *Nhe* occurs in all known *B. cereus* sensu lato strains, is transmitted strictly vertical and shows the same degree of conservation as housekeeping genes with few intra-operon recombination events that occurred only between closely related *nhe* regions. This may be evidence for an important function of the operon. Deletion of the entire *nhe* operon was shown to be impossible, so far only *nheC* could be deleted successfully and a frameshift (single base deletion) was introduced into *nheB* after 966 bp [107]. It is noticeable that *nheA* is the most highly conserved gene of the *nhe* operon (data not shown). Thus, *nheA* might contain a second vital function besides toxin formation. While Nhe components and their interactions are currently analyzed extensively, all studies have been focused on the pore-forming activity [115, 117, 237, 238]. However, in NheA an enlarged β-tongue structure was found [238] which, despite its similarity to ClyA and HblB, might be involved in a protein-protein interaction associated with an unknown function. If such a function exists, balancing selection may have resulted in co-evolution of *nhe* with a (yet unknown) interacting factor. Such an interaction would impose certain constraints on the NheA structure, allowing recombination between closely related *nhe* copies but leading to the observed absence of horizontal gene transfer between distantly related strains in order to prevent a loss of fitness. While the existence of such a hypothetical function can be predicted based on this phylogenetic analysis, further experimental investigation is necessary, because three growing *B. cereus* strains (#291 MHI 1670, # 293 MHI 1761 and #294 MHI 2968) are known to contain a natural deletion of the *nhe* operon [115]. Genomes of these strains were sequenced within this study and revealed that all three contain the same frameshift mutation 73 bp downstream of the *nheA* startcodon

(section 3.1.3.4, Fig. 15B). Sequence analysis also demonstrated that a second start codon 96 bp downstream of the original is present in this *nheA* variant. Strains containing this N-terminally truncated NheA showed no toxicity and were negative in NheA enzyme immuno assays [115]. However, the monoclonal antibody used to detect NheA [115, 239] binds to a C-terminal epitope (A. Didier, personal communication), which is located in close proximity to the N-terminus [238]. N-terminal truncation might cause a structural change that prevents effective binding of the NheA antibody. The essential function could nevertheless still be active within the remaining part of the protein, since none of the three strains contains a copy of *nhe* (*nhe*$_a$) to assume the speculated essential function.

A second copy of *nhe*, named *nhe*$_a$, was discovered in four of the 142 *B. cereus* sensu lato strains included in the phylogenetic analyses. *Nhe*$_a$ exists in clusters II and VI (Fig. 5 and 14) which both appear to be actively transcribed (Fig. 15A). *Nhe*$_a$ might have resulted from two relatively recent, but separate HGT events into two strains of clusters II and VI, since cluster II and cluster VI *nhe*$_a$ are two clearly distinct variants. The donor strains harboring the two *nhe*$_a$ versions have not yet been identified, but sequence comparison shows that their *nhe*$_a$ must have separated very early in the evolution of the *B. cereus* group (Fig. 14, inset). Apparently, the *nhe*$_a$ operon is not stably integrated in the genome since several strains in both phylogenetic groups II and VI seem to have lost it shortly after acquisition (Fig. 5). Since the collection of data for this analysis had ended, new genomes were published or updated and *nhe*$_a$ could be identified in another four strains. *B. thuringiensis* Al Hakam and *B. mycoides* 219298 *nhe*$_a$ show 95 % identity to cluster VI *nhe*$_a$. The *nhe*$_a$ copies on plasmid of *B. cereus* FM1 and *B. thuringiensis* serovar tolworthi plasmid pKK1 show 96 – 97 % identity to cluster II *nhe*$_a$. Thus, *nhe*$_a$ appears in 2.0 % of the currently available 409 *B. cereus* sensu lato strains (11/2015) and three of the known eight *nhe*$_a$ are plasmid bound. The detection of plasmid-location is likely to increase with an ongoing completion of genomes. *B. thuringiensis* Al Hakam is part of cluster III but contains a chromosomal *nhe*$_a$ similar to cluster VI strains, which strongly hints to HGT. While *nhe*$_a$ is rare and phylogenetically unstable within *B. cereus* sensu lato, *hbl*$_a$ seems to be an established and functional toxin. Both *nhe*$_a$ and *hbl*$_a$ might be starting points for the evolution of new pore-forming enterotoxins in *B. cereus* sensu lato, analogous to the suspected development of *hbl* and *nhe* themselves from an ancient ancestor by gene duplication [2].

4.2.2 Evolutionary origin of *cytK-2*

The third diarrhea causing agent is the single-component chromosomally encoded toxin CytK. CytK is a hemolytic, dermonecrotic and β-barrel pore-forming enterotoxin [111]. Two variants of CytK are known. 89 % of the amino acid sequence of CytK-2 is identical to CytK-1 and CytK-2 is also able to form pores in planar lipid bilayers, but it shows only 20 % of CytK-1 toxicity possibly due to a smaller pore size [134]. The *CytK-2* gene appears in clusters II – V, which might have acquired the toxin gene via lateral transfer prior to splitting into clusters II, III and IV (Fig. 5, marked by an arrow) from the *B. cytotoxicus* phylogenetic line where the CytK ancestor may have originated. Subsequently, *cytK-2* may have been lost in some strains of clusters II and III. In contrast, a recent horizontal transfer of *cytK-2* to a few cluster V strains seems to have occurred (Fig. 5).

4.2.3 Selective conservation of *plcR* allows efficient adaptation to environmental changes

While both tripartite enterotoxin operons *nhe* and *hbl* exist in duplicates, no duplication of the virulence regulator gene *plcR* could be found. All tested virulence genes *hbl, cytK, plcR* and *papR* – except *nhe* – showed evidence for horizontal transfer within *B. cereus* sensu lato, but the most extensive lateral transfer and least sequence conservation was noticed in *plcR* (section 3.1.3.3, Fig. 13). The pleiotropic transcriptional activator PlcR (Phospholipase C Regulator) regulates the expression of many virulence factors in *B. cereus* [127]. Expression of *nhe, hbl* [127] and *cytK* [111] is controlled by PlcR. The gene phylogeny of *plcR* hints to lateral transfer within 52 *B. cereus, B. anthracis* and *B. thuringiensis* strains [97]. In this study, *plcR* gene phylogeny of 142 strains of all species of the *B. cereus* group (Fig. 13) was compared with the *B. cereus* sensu lato species tree (Fig. 5) and showed very low similarity. These grossly conflicting topologies may be a result of both frequent horizontal transfer and rapid divergent evolution of *plcR* driven by a variety of environmental selection pressures. The small peptide PapR binds and activates PlcR as part of a quorum sensing system. *PapR* phylogeny was found to be similar to *plcR,* but strains of cluster I do not contain *papR*, which might cause reduced pathogenicity. In the activated form PlcR-PapR binds to its recognition site and initiates transcription of virulence genes. It seems likely that the protein-complex binds to its target DNA as a dimer or tetramer since the recognition site consists of an inverted repeat [214, 240]. The DNA binding domain helix-turn-helix motif is located in the N-terminal part of the active PlcR protein, the regulatory domain in the C-terminal region [240]. Greatest sequence variation is found in the 3` end of *plcR* (Fig. 13B). While binding to the recognition site requires a highly conserved protein structure, the regulatory function obviously tolerates higher sequence variation. The regulatory domains of PlcR are derived from TPR (**t**etratrico**p**eptide **r**epeats)-domains, which are important protein-protein interaction modules [240]. PlcR was experimentally inactivated in *B. cereus* and *B. thuringiensis* by a number of mutations [241] including deletions, additions, nonsense mutations, high diversity, and mutations in *papR* or *opp*, genes necessary for PlcR activation. Inactivation of the PlcR regulon and a subsequent non-hemolytic phenotype also occur

naturally. The PlcR regulon comprises at least 45 positively controlled genes that encode mainly secreted proteins including enzymes, toxins, autolysins, sensors, and cell wall components [156]. Inactivation of PlcR impairs not just hemolysis, but also results in atypical growth on chromogenic selective media (based on PlcR regulated enzymes) [242]. This shows that adaptation by controlling PlcR regulon activity leads to different competitiveness in response to complex environmental signals. In *B. anthracis* a nonsense mutation after 642 bp leads to a truncated and inactive protein [127]. It has been hypothesized that in *B. anthracis* active PlcR interferes with the virulence regulator AtxA (encoded on pXO1) and sporulation [243]. Another study showed that the inactive PlcR causes downregulation of many virulence determinants (secreted proteases, phospholipases and hemolysins) and subsequent lack of hemolytic activity, but does not impair sporulation in *B. anthracis* [244]. Thus, reasons for PlcR inactivation in *B. anthracis* are not clear yet. In *B. cereus* biovar *anthracis* a C-terminal frameshift results in a four amino acids longer protein that completely inactivates the *plcR* regulon [245]. The *B. cereus* biovar *anthracis plcR* mutation also occurs in *B. cereus* ISP3191 (#102, III). C-terminal *plcR* gene extensions appear in *B. cytotoxicus* (+ 9 bp), *B. pseudomycoides* (+ 33 bp) and *B. weihenstephanensis* (+ 6 bp), but their function is yet unknown. Highly variable regulatory domains of PlcR as starting point for evolution could lead to a change in PapR-binding stability and subsequently influence structural changes as well as activation of transcription. Hence, this might be a way of allowing fast adaptation to changing environmental or host conditions by modulating transcriptional activation of specific PlcR-controlled genes without unnecessary inactivation of the entire regulon.

This study shows that recombination within *B. cereus* sensu lato is limited only by preservation of gene/protein functionality. Consequently, the pathogenic potential of (psychrotolerant) environmental strains or probiotic strains can change rapidly with a single and simple exchange of genetic material. This observation may render the current risk assessment strategies questionable.

4.3 Strain- and media-dependent enterotoxin transcription

4.3.1 Enterotoxin gene transcription is highly variable

4.3.1.1 Strain-specific enterotoxin expression

In this study *nhe* promoter activity of two highly toxic *B. cereus* strains [157], *B. cereus* NVH 0075-95 (cluster III) and *B. cereus* INRA C3 (cluster IV), was compared. Transcription of the latter was higher in all media and conditions tested, especially in minimal medium and at 30 °C under aerobic conditions (Fig. 20 and 21). These results indicate strain-dependent expression of *nhe* both under screening conditions and in a more host-like and clinically relevant experimental set-up. Not only the tripartite enterotoxin Nhe is expressed in a strain-specific manner, CytK-1 was also found to be expressed variably. While *B. cytotoxicus* NVH 391-98 and INRA AF2 were highly toxic, NVH 883/00 produced little CytK-1 and was non-cytotoxic [120]. Enterotoxin production was enhanced under anaerobic conditions in the strain *B. cereus* F4430/73 [246]. *B. cereus* NVH 1230-88, however, did not produce enterotoxins under anaerobic conditions [247], which confirms the strain-dependent expression of enterotoxins. In the strain *B. cereus* F4430/73 Hbl production was in minimal medium supplemented with glucose lower than Nhe production [248]. The *hbl* transcriptional kinetic in *B. cereus* INRA C3 determined in this study resembles that of *nhe* at a lower level, but increased *hbl* transcription during nutrient deficiency is more pronounced than *nhe* transcription. This indicates a joint general regulation of enterotoxin transcription, which has already been proven since expression of both *nhe* and *hbl* is controlled by the virulence regulator PlcR and the redox regulators Fnr and ResD [156, 162, 163]. It has to be noted that *hbl* expression seems to be more affected by Fnr than *nhe* expression [249, 250], hinting to an individual fine-tuning of enterotoxin transcription.

In contrast to *B. cereus* NVH 0075-95, enterotoxin transcription of *B. cereus* INRA C3 was more active in CGY medium at 30 °C than at 37 °C (Fig. 21). This trend was even more pronounced at the translational level at 32 °C. The NheB titre of *B. cereus* NVH 0075-95 was 8x higher than the titre of *B. cereus* INRA C3 [139], hinting to an additional translational regulation. While aerobic conditions at 32 °C have been considered the optimal for *B. cereus* growth and toxin production for a long time [172, 251], this additionally indicates strain-specific temperature optima for enterotoxin expression. Several studies have already discovered enterotoxic *B. cereus* strains to possess higher but varying toxic potentials at 32 °C or 37 °C in comparison to lower temperatures in a clearly strain-dependent manner [252-254].

Nutrient and oxygen restricted conditions (GSM and cGSM, stagnant cultures) led to increased enterotoxin promoter activity (Fig. 21). It can be concluded that each *B. cereus* strain regulates transcription of enterotoxins individually and has its own optimum temperature, medium-composition and oxygen level for maximal promoter activity.

4.3.1.2 Media-dependent enterotoxin expression: influence of the nutrient-sensitive regulator CodY

The course and level of enterotoxin transcription was shown to be different in each medium tested (Fig. 20). This clearly media-dependent expression has been noticed before [255]. Glucose catabolite repression controls transcription of *nhe* and *hbl*, while the presence of sucrose or fructose enhanced their expression [160, 248, 256]. In contrast to the tripartite enterotoxins, *cytK* is expressed independently of CcpA-mediated catabolite control [160]. Enterotoxicity is additionally influenced by the amino acid source [247], most likely via the branched-chain amino acid (BCAA) sensing pleiotropic transcriptional repressor CodY [166, 257]. During rapid growth the nutrient sensor CodY represses more than a hundred genes, while during nutrient-deficiency CodY-mediated repression is released [167]. Experiments of this study show that the absence of free amino acids caused enhanced transcription of enterotoxins in *B. cereus* INRA C3 (Fig. 22). MOD minimal medium was supplemented with either tryptone or casamino acids. Tryptone represents enzymatically digested casein and is a mixture of differently sized oligopeptides [258] which are less accessible to *B. cereus*, but thus longer available. Casamino acids, in contrast, consist of acid hydrolyzed casein resulting in vitamin-free amino acids [259] which are fast and easily metabolized. BCAAs [166] and GTP [165] activate the nutrient-sensitive repressor CodY. Thus, CodY-dependent control should be longer active in MOD + 1 % tryptone, which still contains free amino acids from the degradation of oligopeptides. The quorum sensing virulence regulator system PlcR-PapR is known to be controlled by CodY via import of the signaling peptide PapR [260] and the enterotoxin genes are part of the PlcR regulon [110]. In contrast to previous studies, which found the PlcR regulon activated by CodY in an emetic *B. cereus* as well as an enterotoxic *B. cereus* and an *B. thuringiensis* strain [170, 260, 261], results of this study indicate repression of the enterotoxin transcription by CodY. *In silico* comparison of the CodY binding consensus sequence with the *nhe* and *hbl* promoter regions revealed several potential binding sites downstream of the promoter, hinting at a direct repression by CodY. While no interaction between CodY and the *nhe* promoter region of the emetic strain *B. cereus* F4810/72 could be shown [170], a possible direct regulation of *hbl* expression by CodY has not yet been analyzed.

This study provides evidence for interaction of CodY with both *nhe* and *hbl* intergenic regions at sites of potential repressive functions (see section 3.2.6). A high-affinity binding site within the 5'UTR of *hbl* was confirmed in gel mobility shift assays (Fig. 23), wheras three potential CodY binding sites were located in the *nhe* 5'UTR (Fig. S3). Strong affinity to CodY seems to depend on the sequence of the second binding site (see section 3.2.6.2). The presented results additionally show that while affinity of CodY to these sites might be different, binding to more than one target seems to be necessary for effective repression of *nhe* transcription. A comparison of the binding sites in *nhe* and *hbl* promoter regions and the consensus sequence of CodY binding sites in *B. subtilis* [219] revealed considerable variations (Fig. 25A). This low degree of conservation indicates that CodY-mediated transcriptional regulation is influenced by the regulators' affinity to its target sites. Variations at almost all positions within the binding motif have previously been shown to cause different affinity to CodY [262]. The

DNA-binding C-terminal domain of CodY is highly conserved, which is also true for the N-terminal cofactor binding domain responsible for dimer formation [263] that shows only a slightly greater variability (Fig. S4). The strain-specific binding and a suspicious lack of CodY binding site conservation support the hypothesis that the binding site sequence may play a role in the fine-tuning of enterotoxin transcription. In contrast, PlcR binding sites within the enterotoxin promoters reveal a much higher conservation (Fig. 25B). The sequence of the DNA-binding N-terminal domain of PlcR is conserved, while the regulatory C-terminal regions are variable to allow adaptation to changing conditions [215, 240]. Therefore, PlcR-mediated activation of enterotoxin transcription is controlled by protein activity and environmental factors, while CodY-mediated repression additionally may depend on intrinsic strain-specific 5'UTR sequences.

As the major virulence regulator in *B. cereus*, PlcR is responsible for activation of enterotoxin transcription under unfavorable conditions [156]. While PlcR and putative SinR binding sites are highly conserved, all putative ResD, Fnr, CodY, cre and the PlcR 2 sites are more variable, suggesting that both oxygen and nutrient levels affect enterotoxin transcription differently in each *B. cereus* strain. Fine-tuning of enterotoxin transcription and response to specific condition changes might be controlled by other regulators in a strain-specific manner via affinity to their varied binding motifs.

4.3.2 Host epithelial cell factors have a limited influence on transcription of virulence factors

The previously discussed transcriptional experiments focused solely on the transcription of the tripartite enterotoxin operons *nhe* and *hbl*. Under gut-simulating conditions both *nhe* and *hbl* transcription are enhanced in comparison to lab conditions (see section 3.2.3.2). Within the human gut not only enterotoxin gene expression but also transcription of other virulence genes should increase. In transcriptome comparisons the influence of factors secreted by host epithelial cells was studied (see section 3.2.4).

Gut-simulating medium (GSM) was conditioned by 22 h incubation on human Caco-2 cells, then sterile filtrated and used to grow *B. cereus* F837/76. Medium composition of the resulting cGSM might have been slightly altered by the epithelial cells in comparison to unconditioned GSM. *B. cereus* F837/76 grew to an OD_{600} of 0.14 in GSM and 0.38 in cGSM until harvest. This transcriptome analysis showed 106 differentially expressed genes in cGSM in comparison to GSM. The comparison of transcriptomes with/without factors secreted by human epithelial cells revealed an upregulation of some virulence genes but particularly of genes essential to nitrate respiration and acetyl-CoA formation in the presence of host cell factors (Table S5). Thus, after 2 h of growth the transcriptomes display that under cGSM conditions nutrients and especially oxygen were starting to become limiting, leading to enhanced expression of genes necessary for alternative energy production.

ResDE and Fnr activate virulence and nitrate respiratory gene expression under oxygen-limiting conditions [158, 164, 264, 265], which was also demonstrated in other *B. cereus* strains by

transcriptional profiling [266]. In accordance with measured optical density, bacteria seem to have been harvested from cGSM at the transition between exponential growth and the first impacts of limitation due to faster growth than in the reference condition. The effect of oxygen limitation was greatest, as the more than 60x upregulation of *nar* nitrate reductase genes show. NarGHI is a membrane-bound nitrate reductase necessary for respiratory and assimilatory nitrogen metabolism that requires a bis-molybdopterin guanine dinucleotide cofactor [267]. Molybdenum cofactor biosynthesis genes were also upregulated up to 60-fold under cGSM conditions. *B. cereus* utilizes nitrate respiration under anaerobic conditions [162] as long as a sufficient amount of nitrate is present. It is conceivable that cGSM provided more nitrate than GSM due to incubation on eukaryotic epithelial cells. Alternatively, the faster growth induced by secreted factors of epithelial cells forced the bacteria to activate genes for nitrate respiration earlier.

It was shown that enterotoxin gene expression is induced by high cell density via PlcR and substrate limitations, but not by oxygen deprivation [268]. In this study, Nhe components were upregulated 7 – 9x in cGSM, but *hbl* and other virulence factors were not stronger expressed, indicating and confirming a differential regulation of the enterotoxin expression by interplay of a variety of environmental factors. Additionally, gene expression of energy consuming proteins such as ATPases was found to be downregulated. In summary, the presence of host cell factors did not significantly enhance transcription of virulence factors but led to faster growth and a shorter lag phase. Comparison of transcriptomes therefore seems to display the starting activation of gene expression needed during oxygen deficiency in cGSM.

Enterotoxin production was previously shown to be slightly enhanced in cGSM in a strain-specific manner. All nine tested *B. cereus* strains start toxin production in cGSM faster and earlier than in GSM (Nadja Jeßberger, unpublished). However, the presented results are evidence that secreted factors of host epithelial cells are not sufficient for continued enhanced enterotoxin transcription. It can be assumed that direct contact between bacteria and host cells is the missing trigger. *B. cereus* is capable of adhesion to Caco-2 cells followed by cytotoxicity toward the epithelial cells and even internalization [269]. Since *B. cereus* virulence is mainly controlled via the PlcR-PapR quorum-sensing system, adhesion could cause higher bacterial density, which induces cytotoxicity or direct cell-cell contact might trigger the release of virulence factors [269]. Furthermore, enterotoxin expression might additionally be regulated on translational level, because high toxin transcript levels not necessarily occur in high toxin producing strains [157]. Further research and experiments should therefore investigate enterotoxin transcription as well as production in direct contact with host cells.

4.3.3 Potential functions of *nhe* and *hbl* 5' intergenic regions

With a size of several hundred base pairs it is possible that the intergenic sequences upstream of *nhe* and *hbl* might encode small proteins. There is increasing evidence that intergenic regions in prokaryotes code for unknown small proteins [270]. No obvious ORFs are present upstream of *nhe*, but the 5'UTR of *hbl* contains a putative ORF of varying size (180 – 192 nucleotides, Fig. 16D). A BLASTP analysis did not reveal any homology to proteins of known function, thus the expression and function of this ORF remains to be studied.

5'UTRs can contain temperature sensitive RNA thermometers [271] or metabolite sensitive riboswitches [272]. In *Listeria monocytogenes* several virulence genes were found to be post-transcriptionally regulated by long 5'UTRs [273]. The *prfA* 5'UTR is a thermosensor allowing transcription of the transcriptional activator at 37 °C and blocking it at lower temperatures [274]. The activating and temperature-independent function of listerial *actA* and *hly* 5'UTRs was shown [275, 276], but the mechanism of expression enhancement is not yet clear. Recently, several repeat regions that might encode novel riboswitches have been identified in *B. cereus* [277], but none of them is located in the 5'UTR of *nhe* or *hbl*. Analysis of the 5'UTR sequences revealed no similarities with known RNA families and no obviously conserved RNA secondary structures. Nonetheless, the long 5'IGRs could encode yet unknown functions or they interact with different regulators to allow for differential expression of enterotoxin genes.

In promoter fusions with the *lux* operon the (partial) 5'IGRs were investigated to determine their function in *B. cereus* INRA C3 (section 3.2.2). The PlcR binding site 1 upstream of the two putative promoters is essential for *nhe* expression (Fig. 18). But a deletion of the 5'UTR also abolished transcriptional activity. Thus, transcription of *nhe* strongly depends on the presence of the entire IGR, indicating the necessity of a concerted interaction of all regulatory elements therein to trigger *nhe* expression. In contrast to *nhe*, deletion of the *hbl* 5'UTR caused increased transcription (Fig. 18). The 268 bp region designated 5'UTR-up is at least partially responsible for the repression of *hbl* transcription. This region contains (putative) binding sites for ResD, CodY, Fnr and SinR (Fig. 16 and 19A). One or several of these regulators acts as a repressor for *hbl* transcription under the conditions tested.

4.4 Various global regulators control enterotoxin expression

The mere presence of a virulence or toxin gene is not sufficient to prove actual toxicity [122, 139]. It was recently shown that complex regulatory processes control the expression of enterotoxins in *B. cereus* and cause highly variable toxin quantities [157]. A variety of global and pleiotropic regulators are involved in the complex regulation of enterotoxin expression in *B. cereus* in response to changing environmental and nutritional conditions. Therefore, classification of *B. cereus* strains as high or low toxic turned out to be highly complicated and depends on their environment as much as on their genetic

background [157]. This paragraph aims to give an overview of the interplay of global regulators known to be involved in enterotoxin expression.

B. cereus is a facultative anaerobe bacterium, which responds to anaerobiosis (as found e.g. in the human intestine) with increased enterotoxin production controlled by the redox regulator Fnr and the ResDE signal transduction system [162, 163, 249]. The fumarate nitrate reduction regulatory protein (Fnr), a member of the Crp (cyclic AMP-binding protein) / Fnr family of transcriptional regulators and homologue to the catabolite activator protein of Gram-negatives, is produced as an apoform and its monomers are able to bind to specific sequences in the promoter regions of *fnr*, *resDE*, *plcR*, *nhe* and *hbl* [159]. It was shown that *fnr* expression is downregulated in the presence of O_2 and nitrate which are both able to serve as terminal electron acceptors of respiratory pathways [162]. Fnr is additionally involved in carbohydrate-dependent catabolite repression [250, 278, 279]. Enterotoxin expression is activated by Fnr, especially in media containing fructose or sucrose [250]. Several Fnr sites were found in the *nhe* and *hbl* 5'IGRs and hypothesized to facilitate the interplay of transcriptional activation and repression by Fnr [159]. *Fnr* expression is controlled by ResDE [164]. ResDE is a redox-sensitive signal transduction and two component system consisting of the membrane-bound histidine sensor kinase ResE and the cytoplasmatic response regulator ResD [163]. The ResD monomer interacts with target DNA to activate transcription of e.g. *fnr*, *resDE, nhe, hbl* and *plcR* depending on its phosphorylation state under oxygen limiting conditions [158, 164]. It was proposed that ResE, a membrane-associated kinase/phosphatase (reduced phosphatase activity under anaerobic conditions), phosphorylates the response regulator ResD according to oxygen levels, which in turn activates the ResDE regulon [280, 281]. Oxygen-dependent expression is mediated by ResD and Fnr, while the two redox systems interact not only directly with a DNA recognition site, but are also capable of interaction with each other [158] and formation of a ternary complex with the virulence regulator PlcR [282].

PlcR (Phospholipase C Regulator) is the pleiotropic transcriptional regulator of the PlcR-PapR quorum-sensing system that activates virulence genes such as phospholipases, proteases, hemolysins and enterotoxins [127, 214]. Several other proteins were discovered that are under control of PlcR, such as two-component sensors, chemotaxis proteins, transporters, cytoplasmic regulators and cell wall biogenesis proteins. Expression of PlcR is autoregulated and activated at the onset of stationary phase [283]. The promoters of PlcR regulated genes share a highly conserved palindromic recognition site ($TATGNAN_4TNCATA$), which can be found in the promoter region of *nhe, hbl* [127] and the promoter of *cytK* [111]. Within the *nhe* 5'IGR two PlcR recognition sites were found previously [116, 127] (Fig. 16C). This study presents evidence that PlcR 1 is responsible for PlcR-dependent *nhe* expression, while the less conserved PlcR 2 [156] is one of the necessary elements for full P*nhe* activity, but not sufficient for full activation of *nhe* transcription.

The enterotoxin genes are part of the SinR-SinI regulon, which leads to a differential, tightly controlled expression of *hbl* in a subpopulation of cells during biofilm formation [161]. Similarities between *nhe* and *hbl* 5'IGRs allow the suspicion that both enterotoxins might be under control of the SinR-SinI

agonist-antagonist pair [284]. Biofilm formation is under control of several transition phase regulators, including SinR (master regulator of biofilm formation: SinR-SinI regulon controls transition from planktonic to sessile state), Spo0A (early sporulation transcription factor: activates sporulation and suppresses AbrB-regulon), AbrB (transition state regulator: suppressing stationary phase gene expression) and the virulence regulator PlcR [161, 221, 285-287]. The latter acts as a repressor of biofilm formation [287]. Spo0A represses expression of both *plcR* [288] and *abrB* [289]. While *AbrB* and *spo0A* are no targets of CodY, the nutrient and energy state sensing regulator is positively controlling *plcR* expression as well as biofilm formation [170, 260, 290, 291].

The pleiotropic repressor CodY is activated by the interaction with BCAAs [166] and GTP [165]. Thus, CodY is deactivated in response to low energy state and nitrogen availability and absence of the active CodY leads to a decreased transcription of the PlcR regulon [170, 260, 261]. In addition to the indirect control of enterotoxin expression via PlcR, this study provides evidence that CodY directly binds to *hbl* and *nhe* 5'UTRs at a repressor site in the investigated strains *B. cereus* INRA C3 and *B. cytotoxicus* CVUAS 2833. However, *nhe* expression was previously shown to be not directly regulated by CodY in an emetic *B. cereus* strain [170]. This confirms strain-specific regulation of enterotoxicity in *B. cereus*. With regard to the number of regulators and conditions that influence enterotoxin expression, the effect of CodY-mediated repression might additionally be modulated *in vivo* and be highly media-dependent.

The catabolite control protein A (CcpA) controls the efficiency of glucose metabolism in Gram-positives [160, 292]. Both *nhe* and *hbl* expression are additionally regulated by catabolite repression (the cre site of *hbl* is located downstream of the start codon), while *cytK* is expressed independent of CcpA-mediated catabolite control [160]. Both CodY and CcpA repress enterotoxin expression under nutrient rich conditions and allow virulence factor production at the beginning of stationary phase. Complex formation between CcpA and CodY was found especially in presence of glucose (when CcpA binds DNA) [293], which suggests a cooperation between both regulators. CcpA-mediated glucose repression was also found to control the *resABCDE* operon [294], closing the circle between all mentioned factors involved in enterotoxin expression. While regulators (CodY, Fnr, ResD) bind directly to enterotoxin promoters, they all additionally control promoter activity via PlcR, which is the central switch in enterotoxin expression according to temperature, cell density, life cycle and nutrient availability.

Furthermore, enterotoxic *B. cereus* strains were found to possess higher but varying toxic potentials at 32 °C or 37 °C in comparison to lower temperatures in a clearly strain-dependent manner [252-254]. In contrast, the psychrotolerant *B. weihenstephanensis* showed reduced toxicity at elevated temperatures while expression of the entire PlcR regulon was highest at 15 °C [295-297] leading to the conclusion that temperature-dependent expression of enterotoxin genes might not be regulated directly, but more likely mediated by PlcR. This further indicates that PlcR is a global regulator and a key component in the adaptation to (host) environment [156].

These observations demonstrate that the unusually long promoter regions of *nhe* and *hbl* allow concomitant interaction of several global regulators and facilitate differential regulation of enterotoxin

transcription. Thus, it is possible that deletion of 5'UTR parts led to a strong decrease of *nhe* promoter activity and enhanced *hbl* promoter activity not just due to the loss of binding sites but also due to impeded interactions between regulators.

5. Conclusion

This study showed that seven concatenated housekeeping gene sequences depict species relationships in *B. cereus* sensu lato as accurately as whole-genome comparisons. Many potentially probiotic *B. cereus* strains form a clearly distinguishable phylogenetic line within *B. cereus* sensu lato (Cluster V, ANI boundary > 96 %), which is proposed to designate as *B. toyonensis* named according to the type strain. However, species affiliation of strains is contradictory in many cases. The phylogenetic analysis of this study, therefore, calls for a reassessment of this group's taxonomy. Lateral transfer of virulence genes *hbl, cytK, nhe* and *plcR* within *B. cereus* sensu lato appears to be constrained only by preservation of gene function. The evolution of the three *B. cereus* enterotoxin operons is shaped unexpectedly different, which leads to the hypothesis that the strictly vertical transmission of *nhe* operons is caused by a second, unknown but fitness relevant function of *nhe*. Determination of this function will be an important step to discover the relevance of the Nhe protein complex, which occurs solely within *B. cereus* sensu lato. Furthermore, ancient diversification of *nhe* and *hbl* operons and propagation of hbl_a suggest a potential to develop new enterotoxin variants. The distribution of pathogenicity factors and frequent recombination among *B. cereus* sensu lato phylogenetic groups should be taken into account during risk assessment of the currently valid species of this group, especially concerning probiotic *B. toyonensis* and *B. mycoides* strains.

Evidence for a high, CodY-mediated *nhe* and *hbl* promoter activity under nutrient, especially amino acid limiting conditions is presented in this study. While PlcR is the main virulence activator in *B. cereus* sensu lato, CodY may to be used for a strain specific fine-tuning of enterotoxin transcription via repression in response to specific environmental conditions. The unusually long promoter regions of *nhe* and *hbl* might be important for a concomitant interaction of several global regulators. However, the actual enterotoxin synthesis in *B. cereus* is rarely consistent with transcriptional activity and is, moreover, highly strain-specific [26, 157]. It might be speculated that the 5'UTRs, in addition, interfere with post-transcriptional and/or translational processes [159] to modify the efficiency of enterotoxin production according to the environmental conditions prevalent in the human intestine. Further experimental investigations should also include the impact of host epithelial cells and the gut microbiota to provide a reliable assessment of *B. cereus* sensu lato enterotoxicity.

6. Literature

1. Frankland GC, Frankland PF: **Studies on Some New Micro-Organisms Obtained from Air**. *Philosophical Transactions of the Royal Society B: Biological Sciences* 1887, **178**(0):257-287.
2. Stenfors Arnesen LP, Fagerlund A, Granum PE: **From soil to gut: *Bacillus cereus* and its food poisoning toxins**. *FEMS microbiology reviews* 2008, **32**(4):579-606.
3. Duc LH, Hong HA, Barbosa TM, Henriques AO, Cutting SM: **Characterization of *Bacillus* Probiotics Available for Human Use**. *Applied and Environmental Microbiology* 2004, **70**(4):2161-2171.
4. Hong HA, Duc le H, Cutting SM: **The use of bacterial spore formers as probiotics**. *FEMS microbiology reviews* 2005, **29**(4):813-835.
5. Zou Y, Hu M, Lv Y, Wang Y, Song H, Yuan YJ: **Enhancement of 2-keto-gulonic acid yield by serial subcultivation of co-cultures of *Bacillus cereus* and *Ketogulonicigenium vulgare***. *Bioresource technology* 2013, **132**:370-373.
6. Blackburn MB, Martin PA, Kuhar D, Farrar RR, Jr., Gundersen-Rindal DE: **Crystalliferous *Bacillus cereus* group bacteria from a Maryland hardwood forest are dominated by psychrotolerant strains**. *MicrobiologyOpen* 2014, **3**(4):578-584.
7. Hoornstra D, Andersson MA, Teplova VV, Mikkola R, Uotila LM, Andersson LC, Roivainen M, Gahmberg CG, Salkinoja-Salonen MS: **Potato crop as a source of emetic *Bacillus cereus* and cereulide-induced mammalian cell toxicity**. *Appl Environ Microbiol* 2013, **79**(12):3534-3543.
8. Ruusunen M, Salonen M, Pulkkinen H, Huuskonen M, Hellstrom S, Revez J, Hanninen ML, Fredriksson-Ahomaa M, Lindstrom M: **Pathogenic bacteria in Finnish bulk tank milk**. *Foodborne pathogens and disease* 2013, **10**(2):99-106.
9. Dierick K, Van Coillie E, Swiecicka I, Meyfroidt G, Devlieger H, Meulemans A, Hoedemaekers G, Fourie L, Heyndrickx M, Mahillon J: **Fatal family outbreak of *Bacillus cereus*-associated food poisoning**. *Journal of clinical microbiology* 2005, **43**(8):4277-4279.
10. Martinelli D, Fortunato F, Tafuri S, Cozza V, Chironna M, Germinario C, Pedalino B, Prato R: **Lessons learnt from a birthday party: a *Bacillus cereus* outbreak, Bari, Italy, January 2012**. *Ann Ist Super Sanita* 2013, **49**(4):391-394.
11. Slaten DD, Oropeza RI, Werner SB: **An outbreak of *Bacillus cereus* food poisoning--are caterers supervised sufficiently**. *Public health reports* 1992, **107**(4):477-480.
12. Helgason E, Caugant DA, Olsen I, Kolsto AB: **Genetic structure of population of *Bacillus cereus* and *B. thuringiensis* isolates associated with periodontitis and other human infections**. *Journal of clinical microbiology* 2000, **38**(4):1615-1622.
13. Stevens MP, Elam K, Bearman G: **Meningitis due to *Bacillus cereus*: A case report and review of the literature**. *Can J Infect Dis Med Microbiol* 2012, **23**(1):e16-19.
14. Barrie D, Wilson JA, Hoffman PN, Kramer JM: ***Bacillus cereus* meningitis in two neurosurgical patients: an investigation into the source of the organism**. *The Journal of infection* 1992, **25**(3):291-297.
15. David DB, Kirkby GR, Noble BA: ***Bacillus cereus* endophthalmitis**. *The British journal of ophthalmology* 1994, **78**(7):577-580.
16. Miyata J, Tasaka S, Miyazaki M, Yoshida S, Naoki K, Sayama K, Asano K, Fujiwara H, Ohkusu K, Hasegawa N *et al*: ***Bacillus cereus* Necrotizing Pneumonia in a Patient with Nephrotic Syndrome**. *Internal Medicine* 2013, **52**(1):101-104.
17. Ehling-Schulz M, Fricker M, Grallert H, Rieck P, Wagner M, Scherer S: **Cereulide synthetase gene cluster from emetic *Bacillus cereus*: structure and location on a mega virulence plasmid related to *Bacillus anthracis* toxin plasmid pXO1**. *BMC microbiology* 2006, **6**:20.
18. Posfay-Barbe KM, Schrenzel J, Frey J, Studer R, Korff C, Belli DC, Parvex P, Rimensberger PC, Schappi MG: **Food poisoning as a cause of acute liver failure**. *Pediatr Infect Dis J* 2008, **27**(9):846-847.
19. Shiota M, Saitou K, Mizumoto H, Matsusaka M, Agata N, Nakayama M, Kage M, Tatsumi S, Okamoto A, Yamaguchi S *et al*: **Rapid detoxification of cereulide in *Bacillus cereus* food poisoning**. *Pediatrics* 2010, **125**(4):e951-955.

20. Shinagawa K: **Analytical methods for *Bacillus cereus* and other *Bacillus* species**. *International journal of food microbiology* 1990, **10**(2):125-141.
21. Ehling-Schulz M, Fricker M, Scherer S: ***Bacillus cereus*, the causative agent of an emetic type of food-borne illness**. *Molecular nutrition & food research* 2004, **48**(7):479-487.
22. Messelhausser U, Frenzel E, Blochinger C, Zucker R, Kampf P, Ehling-Schulz M: **Emetic *Bacillus cereus* are more volatile than thought: recent foodborne outbreaks and prevalence studies in Bavaria (2007-2013)**. *BioMed research international* 2014, **2014**:465603.
23. EFSA, ECDC: **The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2011**. *EFSA Journal* 2013, **11**(4):250.
24. EFSA, ECDC: **The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2012**. *EFSA Journal* 2014, **12**(2):320.
25. EFSA, ECDC: **The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2013**. *EFSA Journal* 2015, **13**(1):165.
26. Ceuppens S, Boon N, Uyttendaele M: **Diversity of *Bacillus cereus* group strains is reflected in their broad range of pathogenicity and diverse ecological lifestyles**. *FEMS microbiology ecology* 2013, **84**(3):433-450.
27. Ash C, Farrow JA, Dorsch M, Stackebrandt E, Collins MD: **Comparative analysis of *Bacillus anthracis, Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA**. *International journal of systematic bacteriology* 1991, **41**(3):343-346.
28. Soufiane B, Cote JC: ***Bacillus weihenstephanensis* characteristics are present in *Bacillus cereus* and *Bacillus mycoides* strains**. *FEMS microbiology letters* 2013, **341**(2):127-137.
29. Nakamura LK: ***Bacillus pseudomycoides* sp. nov**. *International journal of systematic bacteriology* 1998, **48 Pt 3**:1031-1035.
30. Guinebretiere MH, Auger S, Galleron N, Contzen M, De Sarrau B, De Buyser ML, Lamberet G, Fagerlund A, Granum PE, Lereclus D *et al*: ***Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* Group occasionally associated with food poisoning**. *International journal of systematic and evolutionary microbiology* 2013, **63**(Pt 1):31-40.
31. Damgaard PH, Granum PE, Bresciani J, Torregrossa MV, Eilenberg J, Valentino L: **Characterization of *Bacillus thuringiensis* isolated from infections in burn wounds**. *FEMS immunology and medical microbiology* 1997, **18**(1):47-53.
32. Kuroki R, Kawakami K, Qin L, Kaji C, Watanabe K, Kimura Y, Ishiguro C, Tanimura S, Tsuchiya Y, Hamaguchi I *et al*: **Nosocomial Bacteremia Caused by Biofilm-Forming *Bacillus cereus* and *Bacillus thuringiensis***. *Internal Medicine* 2009, **48**(10):791-796.
33. Koch R: **Die Ätiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des *Bacillus Anthracis***. 1876.
34. Hoffmaster AR, Ravel J, Rasko DA, Chapman GD, Chute MD, Marston CK, De BK, Sacchi CT, Fitzgerald C, Mayer LW *et al*: **Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(22):8449-8454.
35. Brezillon C, Haustant M, Dupke S, Corre JP, Lander A, Franz T, Monot M, Couture-Tosi E, Jouvion G, Leendertz FH *et al*: **Capsules, Toxins and AtxA as Virulence Factors of Emerging *Bacillus cereus* Biovar *anthracis***. *PLoS neglected tropical diseases* 2015, **9**(4):e0003455.
36. Lechner S, Mayr R, Francis KP, Pruss BM, Kaplan T, Wiessner-Gunkel E, Stewart GS, Scherer S: ***Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group**. *International journal of systematic bacteriology* 1998, **48 Pt 4**:1373-1382.
37. Thorsen L, Hansen BM, Nielsen KF, Hendriksen NB, Phipps RK, Budde BB: **Characterization of emetic *Bacillus weihenstephanensis*, a new cereulide-producing bacterium**. *Applied and Environmental Microbiology* 2006, **72**(7):5118-5121.
38. Hoton FM, Fornelos N, N'Guessan E, Hu X, Swiecicka I, Dierick K, Jaaskelainen E, Salkinoja-Salonen M, Mahillon J: **Family portrait of *Bacillus cereus* and *Bacillus weihenstephanensis* cereulide-producing strains**. *Environmental microbiology reports* 2009, **1**(3):177-183.

39. Nakamura K, Jackson MA: **Clarification of the Taxonomy of *Bacillus mycoides***. *International journal of systematic bacteriology* 1995, **45**(1):4.
40. Beattie SH, Williams AG: **Detection of toxigenic strains of *Bacillus cereus* and other *Bacillus* spp. with an improved cytotoxicity assay**. *Letters in applied microbiology* 1999, **28**(3):221-225.
41. Rau J, Perz R, Klittich G, Contzen M: **[Cereulide forming presumptive *Bacillus cereus* strains from food--differentiating analyses using cultural methods, LC-MS/MS, PCR, and infrared spectroscopy in consideration of thermotolerant isolates]**. *Berliner Und Münchener Tierärztliche Wochenschrift* 2009, **122**(1-2):11.
42. Contzen M, Hailer M, Rau J: **Isolation of *Bacillus cytotoxicus* from various commercial potato products**. *International journal of food microbiology* 2014, **174**:19-22.
43. Jimenez G, Urdiain M, Cifuentes A, Lopez-Lopez A, Blanch AR, Tamames J, Kampfer P, Kolsto AB, Ramon D, Martinez JF *et al*: **Description of *Bacillus toyonensis* sp. nov., a novel species of the *Bacillus cereus* group, and pairwise genome comparisons of the species of the group by means of ANI calculations**. *Systematic and applied microbiology* 2013, **36**(6):383-391.
44. Mayr E: **Speciation phenomena in birds**. *The American Naturalist* 1940, **74**(752):29.
45. Rossello-Mora R, Amann R: **The species concept for prokaryotes**. *FEMS microbiology reviews* 2001, **25**(1):39-67.
46. Thompson CC, Chimetto L, Edwards RA, Swings J, Stackebrandt E, Thompson FL: **Microbial genomic taxonomy**. *BMC genomics* 2013, **14**:913.
47. Kim M, Oh HS, Park SC, Chun J: **Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes**. *International journal of systematic and evolutionary microbiology* 2014, **64**(Pt 2):346-351.
48. Richter M, Rossello-Mora R: **Shifting the genomic gold standard for the prokaryotic species definition**. *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(45):19126-19131.
49. Rossello-Mora R: **Updating prokaryotic taxonomy**. *Journal of bacteriology* 2005, **187**(18):6255-6257.
50. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A: **Microbial species delineation using whole genome sequences**. *Nucleic Acids Res* 2015, **43**(14):6761-6771.
51. Henz SR, Huson DH, Auch AF, Nieselt-Struwe K, Schuster SC: **Whole-genome prokaryotic phylogeny**. *Bioinformatics* 2005, **21**(10):2329-2335.
52. Konstantinidis KT, Tiedje JM: **Genomic insights that advance the species definition for prokaryotes**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(7):2567-2572.
53. Konstantinidis KT, Tiedje JM: **Towards a genome-based taxonomy for prokaryotes**. *Journal of bacteriology* 2005, **187**(18):6258-6264.
54. Sentausa E, Fournier PE: **Advantages and limitations of genomics in prokaryotic taxonomy**. *Clin Microbiol Infect* 2013, **19**(9):790-795.
55. Mayden RL: **Species: The units of diversity**. In: *A Hierarchy of Species Concepts: The Denouement in the Saga of the Species Problem.* Edited by Claridge MF, Dawah HA, Wilson MR: Chapman and Hall; 1997: 381-423.
56. Hull DL: **Are species really individuals?** *Systematic biology* 1976, **25**(2):17.
57. Wiley EO: **The evolutionary species concept reconsidered**. *Syst Zool* 178, **27**(1):10.
58. Ward DM: **A natural species concept for prokaryotes**. *Curr Opin Microbiol* 1998, **1**(3):271-277.
59. Whitman WB, Coleman DC, Wiebe WJ: **Prokaryotes: the unseen majority**. *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(12):6578-6583.
60. Hill KK, Ticknor LO, Okinaka RT, Asay M, Blair H, Bliss KA, Laker M, Pardington PE, Richardson AP, Tonks M *et al*: **Fluorescent Amplified Fragment Length Polymorphism Analysis of *Bacillus anthracis, Bacillus cereus*, and *Bacillus thuringiensis* Isolates**. *Applied and Environmental Microbiology* 2004, **70**(2):1068-1080.

61. Carlson CR, Caugant DA, Kolsto AB: **Genotypic Diversity among *Bacillus cereus* and *Bacillus thuringiensis* Strains**. *Appl Environ Microbiol* 1994, **60**(6):1719-1725.
62. Tourasse NJ, Helgason E, Okstad OA, Hegna IK, Kolsto AB: **The *Bacillus cereus* group: novel aspects of population structure and genome dynamics**. *Journal of applied microbiology* 2006, **101**(3):579-593.
63. Helgason E, Tourasse NJ, Meisal R, Caugant DA, Kolsto AB: **Multilocus Sequence Typing Scheme for Bacteria of the *Bacillus cereus* Group**. *Applied and Environmental Microbiology* 2004, **70**(1):191-201.
64. Tourasse NJ, Okstad OA, Kolsto AB: **HyperCAT: an extension of the SuperCAT database for global multi-scheme and multi-datatype phylogenetic analysis of the *Bacillus cereus* group population**. *Database : the journal of biological databases and curation* 2010, **2010**:baq017.
65. Maiden MC: **Multilocus sequence typing of bacteria**. *Annual review of microbiology* 2006, **60**:561-588.
66. Ehling-Schulz M, Messelhausser U: ***Bacillus* "next generation" diagnostics: moving from detection toward subtyping and risk-related strain profiling**. *Frontiers in microbiology* 2013, **4**:32.
67. Versalovic J, Schneider M, De Bruijn FJ, Lupski JR: **Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction**. *Methods in molecular and cellular biology* 1994, **5**(1):15.
68. Ehling-Schulz M, Svensson B, Guinebretiere MH, Lindbäck T, Andersson M, Schulz A, Fricker M, Christiansson A, Granum PE, Märtlbauer E *et al*: **Emetic toxin formation of *Bacillus cereus* is restricted to a single evolutionary lineage of closely related strains**. *Microbiology* 2005, **151**(Pt 1):183-197.
69. Park SH, Kim HJ, Kim JH, Kim TW, Kim HY: **Simultaneous detection and identification of *Bacillus cereus* group bacteria using multiplex PCR**. *J Microbiol Biotechnol* 2007, **17**(7):1177-1182.
70. Ogawa H, Fujikura D, Ohnuma M, Ohnishi N, Hang'ombe BM, Mimuro H, Ezaki T, Mweene AS, Higashi H: **A Novel Multiplex PCR Discriminates *Bacillus anthracis* and Its Genetically Related Strains from Other *Bacillus cereus* Group Species**. *PloS one* 2015, **10**(3):e0122004.
71. Yang IC, Shih DY, Huang TP, Huang YP, Wang JY, Pan TM: **Establishment of a novel multiplex PCR assay and detection of toxigenic strains of the species in the *Bacillus cereus* group**. *J Food Prot* 2005, **68**(10):2123-2130.
72. Forghani F, Kim JB, Oh DH: **Enterotoxigenic Profiling of Emetic Toxin- and Enterotoxin-Producing *Bacillus cereus*, Isolated from Food, Environmental, and Clinical Samples by Multiplex PCR**. *Journal of food science* 2014, **79**(11):M2288-2293.
73. Ehling-Schulz M, Guinebretiere MH, Monthan A, Berge O, Fricker M, Svensson B: **Toxin gene profiling of enterotoxic and emetic *Bacillus cereus***. *FEMS microbiology letters* 2006, **260**(2):232-240.
74. Zhang Z, Feng L, Xu H, Liu C, Shah NP, Wei H: **Detection of viable enterotoxin-producing *Bacillus cereus* and analysis of toxigenicity from ready-to-eat foods and infant formula milk powder by multiplex PCR**. *Journal of dairy science* 2015.
75. Jain R, Rivera MC, Moore JE, Lake JA: **Horizontal Gene Transfer in Microbial Genome Evolution**. *Theoretical Population Biology* 2002, **61**(4):489-495.
76. Jain R, Rivera MC, Moore JE, Lake JA: **Horizontal gene transfer accelerates genome innovation and evolution**. *Molecular biology and evolution* 2003, **20**(10):1598-1602.
77. Ochman H: **Lateral and oblique gene transfer**. *Current opinion in genetics & development* 2001, **11**(6):616-619.
78. Bengtsson-Palme J, Boulund F, Fick J, Kristiansson E, Larsson DG: **Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India**. *Frontiers in microbiology* 2014, **5**:648.
79. Ochman H, Lawrence JG, Groisman EA: **Lateral gene transfer and the nature of bacterial innovation**. *Nature* 2000, **405**(6784):299-304.
80. de la Cruz F, Davies J: **Horizontal gene transfer and the origin of species: lessons from bacteria**. *Trends in microbiology* 2000, **8**(3):128-133.

81. Garcia-Vallve S, Romeu A, Palau J: **Horizontal gene transfer in bacterial and archaeal complete genomes**. *Genome research* 2000, **10**(11):1719-1725.
82. Koonin EV, Makarova KS, Aravind L: **Horizontal gene transfer in prokaryotes: quantification and classification**. *Annual review of microbiology* 2001, **55**:709-742.
83. Koonin EV: **Orthologs, paralogs, and evolutionary genomics**. *Annual review of genetics* 2005, **39**:309-338.
84. Spratt BG, Hanage WP, Feil EJ: **The relative contributions of recombination and point mutation to the diversification of bacterial clones**. *Curr Opin Microbiol* 2001, **4**(5):602-606.
85. Nordborg M, Tavare S: **Linkage disequilibrium: what history has to tell us**. *Trends in genetics : TIG* 2002, **18**(2):83-90.
86. Posada D, Crandall KA: **Evaluation of methods for detecting recombination from DNA sequences: computer simulations**. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(24):13757-13762.
87. Huson DH, Bryant D: **Application of phylogenetic networks in evolutionary studies**. *Molecular biology and evolution* 2006, **23**(2):254-267.
88. Ragan MA: **Detection of lateral gene transfer among microbial genomes**. *Current opinion in genetics & development* 2001, **11**(6):620-626.
89. Lawrence JG, Ochman H: **Reconciling the many faces of lateral gene transfer**. *Trends in microbiology* 2002, **10**(1):1-4.
90. Guinebretiere MH, Thompson FL, Sorokin A, Normand P, Dawyndt P, Ehling-Schulz M, Svensson B, Sanchis V, Nguyen-The C, Heyndrickx M *et al*: **Ecological diversification in the *Bacillus cereus* Group**. *Environmental microbiology* 2008, **10**(4):851-865.
91. Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolsto AB: ***Bacillus anthracis, Bacillus cereus,* and *Bacillus thuringiensis* - one species on the basis of genetic evidence**. *Appl Environ Microbiol* 2000, **66**(6):2627-2630.
92. Hu X, Swiecicka I, Timmery S, Mahillon J: **Sympatric soil communities of *Bacillus cereus* sensu lato: population structure and potential plasmid dynamics of pXO1- and pXO2-like elements**. *FEMS microbiology ecology* 2009, **70**(3):344-355.
93. Hu X, van der Auwera G, Timmery S, Zhu L, Mahillon J: **Distribution, diversity, and potential mobility of extrachromosomal elements related to the *Bacillus anthracis* pXO1 and pXO2 virulence plasmids**. *Applied and Environmental Microbiology* 2009, **75**(10):3016-3028.
94. Thorsen L, Budde BB, Henrichsen L, Martinussen T, Jakobsen M: **Cereulide formation by *Bacillus weihenstephanensis* and mesophilic emetic *Bacillus cereus* at temperature abuse depends on pre-incubation conditions**. *International journal of food microbiology* 2009, **134**(1-2):133-139.
95. Gonzalez JM, Jr., Brown BJ, Carlton BC: **Transfer of *Bacillus thuringiensis* plasmids coding for delta-endotoxin among strains of *B. thuringiensis* and *B. cereus***. *Proceedings of the National Academy of Sciences of the United States of America* 1982, **79**(22):6951-6955.
96. Helgason E, Caugant DA, Lecadet MM, Chen Y, Mahillon J, Lovgren A, Hegna I, Kvaloy K, Kolsto AB: **Genetic diversity of *Bacillus cereus/B. thuringiensis* isolates from natural sources**. *Curr Microbiol* 1998, **37**(2):80-87.
97. Ko KS, Kim JW, Kim JM, Kim W, Chung SI, Kim IJ, Kook YH: **Population structure of the *Bacillus cereus* group as determined by sequence analysis of six housekeeping genes and the *plcR* Gene**. *Infect Immun* 2004, **72**(9):5253-5261.
98. Cardazzo B, Negrisolo E, Carraro L, Alberghini L, Patarnello T, Giaccone V: **Multiple-locus sequence typing and analysis of toxin genes in *Bacillus cereus* food-borne isolates**. *Applied and Environmental Microbiology* 2008, **74**(3):850-860.
99. Han CS, Xie G, Challacombe JF, Altherr MR, Bhotika SS, Brown N, Bruce D, Campbell CS, Campbell ML, Chen J *et al*: **Pathogenomic sequence analysis of *Bacillus cereus* and *Bacillus thuringiensis* isolates closely related to *Bacillus anthracis***. *Journal of bacteriology* 2006, **188**(9):3382-3390.
100. Ticknor LO, Kolsto AB, Hill KK, Keim P, Laker MT, Tonks M, Jackson PJ: **Fluorescent Amplified Fragment Length Polymorphism Analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* Soil Isolates**. *Applied and Environmental Microbiology* 2001, **67**(10):4863-4873.

101. Sorokin A, Candelon B, Guilloux K, Galleron N, Wackerow-Kouzova N, Ehrlich SD, Bourguet D, Sanchis V: **Multiple-locus sequence typing analysis of *Bacillus cereus* and *Bacillus thuringiensis* reveals separate clustering and a distinct population structure of psychrotrophic strains**. *Applied and Environmental Microbiology* 2006, **72**(2):1569-1578.
102. Granum PE, Lund T: ***Bacillus cereus* and its food poisoning toxins**. *FEMS microbiology letters* 1997, **157**(2):223-228.
103. Clavel T, Carlin F, Lairon D, Nguyen-The C, Schmitt P: **Survival of *Bacillus cereus* spores and vegetative cells in acid media simulating human stomach**. *Journal of applied microbiology* 2004, **97**(1):214-219.
104. Fagerlund A, Lindbäck T, Granum PE: ***Bacillus cereus* cytotoxins Hbl, Nhe and CytK are secreted via the Sec translocation pathway**. *BMC microbiology* 2010, **10**:304.
105. Mueller M, Grauschopf U, Maier T, Glockshuber R, Ban N: **The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism**. *Nature* 2009, **459**(7247):726-730.
106. Peraro MD, van der Goot FG: **Pore-forming toxins: ancient, but never really out of fashion**. *Nature reviews Microbiology* 2016, **14**(2):77-92.
107. 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**. *Microbiology* 2008, **154**(Pt 3):693-704.
108. Zhu K, Didier A, Dietrich R, Heilkenbrinker U, Waltenberger E, Jessberger N, Martlbauer E, Benz R: **Formation of small transmembrane pores: An intermediate stage on the way to *Bacillus cereus* non-hemolytic enterotoxin (Nhe) full pores in the absence of NheA**. *Biochemical and biophysical research communications* 2015.
109. Wallace AJ, Stillman TJ, Atkins A, Jamieson SJ, Bullough PA, Green J, Artymiuk PJ: ***E. coli* hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy**. *Cell* 2000, **100**(2):265-276.
110. Ramarao N, Sanchis V: **The pore-forming haemolysins of *Bacillus cereus*: a review**. *Toxins* 2013, **5**(6):1119-1139.
111. Lund T, De Buyser ML, Granum PE: **A new cytotoxin from *Bacillus cereus* that may cause necrotic enteritis**. *Molecular microbiology* 2000, **38**(2):254-261.
112. Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE: **Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore**. *Science* 1996, **274**(5294):1859-1866.
113. Hardy SP, Lund T, Granum PE: **CytK toxin of *Bacillus cereus* forms pores in planar lipid bilayers and is cytotoxic to intestinal epithelia**. *FEMS microbiology letters* 2001, **197**(1):47-51.
114. Lund T, Granum PE: **Characterisation of a non-haemolytic enterotoxin complex from *Bacillus cereus* isolated after a foodborne outbreak**. *FEMS microbiology letters* 1996, **141**(2-3):151-156.
115. Lindbäck T, Hardy SP, Dietrich R, Sodring M, Didier A, Moravek M, Fagerlund A, Bock S, Nielsen C, Casteel M *et al*: **Cytotoxicity of the *Bacillus cereus* Nhe enterotoxin requires specific binding order of its three exoprotein components**. *Infect Immun* 2010, **78**(9):3813-3821.
116. Lindbäck T, Fagerlund A, Rodland MS, Granum PE: **Characterization of the *Bacillus cereus* Nhe enterotoxin**. *Microbiology* 2004, **150**(Pt 12):3959-3967.
117. Heilkenbrinker U, Dietrich R, Didier A, Zhu K, Lindbäck T, Granum PE, Märtlbauer E: **Complex formation between NheB and NheC is necessary to induce cytotoxic activity by the three-component *Bacillus cereus* Nhe enterotoxin**. *PloS one* 2013, **8**(4):e63104.
118. Guinebretiere MH, Broussolle V, Nguyen-The C: **Enterotoxigenic Profiles of Food-Poisoning and Food-Borne *Bacillus cereus* Strains**. *Journal of clinical microbiology* 2002, **40**(8):3053-3056.
119. Guinebretiere MH, Velge P, Couvert O, Carlin F, Debuyser ML, Nguyen-The C: **Ability of *Bacillus cereus* group strains to cause food poisoning varies according to phylogenetic affiliation (groups I to VII) rather than species affiliation**. *Journal of clinical microbiology* 2010, **48**(9):3388-3391.

120. Fagerlund A, Brillard J, Furst R, Guinebretiere MH, Granum PE: **Toxin production in a rare and genetically remote cluster of strains of the *Bacillus cereus* group**. *BMC microbiology* 2007, **7**:43.
121. Moravek M, Dietrich R, Buerk C, Broussolle V, Guinebretiere MH, Granum PE, Nguyen-The C, Märtlbauer E: **Determination of the toxic potential of *Bacillus cereus* isolates by quantitative enterotoxin analyses**. *FEMS microbiology letters* 2006, **257**(2):293-298.
122. Dietrich R, Moravek M, Burk 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.
123. Granum PE, O'Sullivan K, Lund T: **The sequence of the non-haemolytic enterotoxin operon from *Bacillus cereus***. *FEMS microbiology letters* 1999, **177**(2):225-229.
124. Ryan P, Macmillan JD, Zilinskas BA: **Molecular cloning and characterization of the genes encoding the L1 and L2 components of hemolysin BL from *Bacillus cereus***. *J Bacteriol* 1997, **179**(8):5.
125. Beecher DJ, Macmillan JD: **Characterization of the components of hemolysin BL from *Bacillus cereus***. *Infect Immun* 1991, **59**(5):1778-1784.
126. Lindbäck T, Okstad OA, Rishovd AL, Kolsto AB: **Insertional inactivation of *hblC* encoding the L2 component of *Bacillus cereus* ATCC 14579 haemolysin BL strongly reduces enterotoxigenic activity, but not the haemolytic activity against human erythrocytes**. *Microbiology* 1999, **145 (Pt 11)**:3139-3146.
127. Agaisse H, Gominet M, Okstad OA, Kolsto AB, Lereclus D: **PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis***. *Molecular microbiology* 1999, **32**(5):1043-1053.
128. Turnbull PC, Kramer JM, Jorgensen K, Gilbert RJ, Melling J: **Properties and production characteristics of vomiting, diarrheal, and necrotizing toxins of *Bacillus cereus***. *Am J Clin Nutr* 1979, **32**(1):219-228.
129. Thompson NE, Ketterhagen MJ, Bergdoll MS, Schantz EJ: **Isolation and some properties of an enterotoxin produced by *Bacillus cereus***. *Infect Immun* 1984, **43**(3):887-894.
130. Beecher DJ, MacMillan JD: **A novel bicomponent hemolysin from *Bacillus cereus***. *Infect Immun* 1990, **58**(7):2220-2227.
131. Dietrich R, Fella C, Strich S, Märtlbauer E: **Production and characterization of monoclonal antibodies against the hemolysin BL enterotoxin complex produced by *Bacillus cereus***. *Appl Environ Microbiol* 1999, **65**(10):4470-4474.
132. Beecher DJ, Wong AC: **Tripartite hemolysin BL from *Bacillus cereus.* Hemolytic analysis of component interactions and a model for its characteristic paradoxical zone phenomenon**. *The Journal of biological chemistry* 1997, **272**(1):233-239.
133. Beecher DJ, Wong AC: **Tripartite haemolysin BL: isolation and characterization of two distinct homologous sets of components from a single *Bacillus cereus* isolate**. *Microbiology* 2000, **146 (Pt 6)**:1371-1380.
134. Fagerlund A, Ween O, Lund T, Hardy SP, Granum PE: **Genetic and functional analysis of the *cytK* family of genes in *Bacillus cereus***. *Microbiology* 2004, **150**(Pt 8):2689-2697.
135. Castiaux V, Liu X, Delbrassinne L, Mahillon J: **Is Cytotoxin K from *Bacillus cereus* a bona fide enterotoxin?** *International journal of food microbiology* 2015, **211**:79-85.
136. Beecher DJ, Wong AC: **Identification and analysis of the antigens detected by two commercial *Bacillus cereus* diarrheal enterotoxin immunoassay kits**. *Appl Environ Microbiol* 1994, **60**(12):4614-4616.
137. Swiecicka I, van der Auwera GA, Mahillon J: **Hemolytic and nonhemolytic enterotoxin genes are broadly distributed among *Bacillus thuringiensis* isolated from wild mammals**. *Microbial ecology* 2006, **52**(3):544-551.
138. Krause N, Moravek M, Dietrich R, Wehrle E, Slaghuis J, Martlbauer E: **Performance characteristics of the Duopath(R) Cereus Enterotoxins assay for rapid detection of enterotoxinogenic *Bacillus cereus* strains**. *International journal of food microbiology* 2010, **144**(2):322-326.
139. 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 : official journal of the International Society on Toxinology* 2014, **77**:49-57.

140. Drobniewski FA: ***Bacillus cereus*** **and related species**. *Clinical microbiology reviews* 1993, **6**(4):324-338.
141. Baida G, Budarina ZI, Kuzmin NP, Solonin AS: **Complete nucleotide sequence and molecular characterization of hemolysin II gene from *Bacillus cereus***. *FEMS microbiology letters* 1999, **180**(1):7-14.
142. Baida GE, Kuzmin NP: **Cloning and primary structure of a new hemolysin gene from *Bacillus cereus***. *Biochimica et biophysica acta* 1995, **1264**(2):151-154.
143. Kreft J, Berger H, Hartlein M, Muller B, Weidinger G, Goebel W: **Cloning and expression in *Escherichia coli* and *Bacillus subtilis* of the hemolysin (cereolysin) determinant from *Bacillus cereus***. *Journal of bacteriology* 1983, **155**(2):681-689.
144. Fedhila S, Gohar M, Slamti L, Nel P, Lereclus D: **The *Bacillus thuringiensis* PlcR-Regulated Gene *inhA2* Is Necessary, but Not Sufficient, for Virulence**. *Journal of bacteriology* 2003, **185**(9):2820-2825.
145. Jung SC, Paik HR, Kim MS, Baik KS, Lee WY, Seong CN, Choi SK: **InhA-like protease secreted by *Bacillus* sp. S17110 inhabited in turban shell**. *J Microbiol* 2007, **45**(5):402-408.
146. Guillemet E, Cadot C, Tran SL, Guinebretiere MH, Lereclus D, Ramarao N: **The InhA metalloproteases of *Bacillus cereus* contribute concomitantly to virulence**. *Journal of bacteriology* 2010, **192**(1):286-294.
147. Asano SI, Nukumizu Y, Bando H, Iizuka T, Yamamoto T: **Cloning of novel enterotoxin genes from *Bacillus cereus* and *Bacillus thuringiensis***. *Appl Environ Microbiol* 1997, **63**(3):1054-1057.
148. Agata N, Ohta M, Arakawa Y, Mori M: **The *bceT* gene of *Bacillus cereus* encodes an enterotoxic protein**. *Microbiology* 1995, **141 (Pt 4)**:983-988.
149. Chung MC, Popova TG, Millis BA, Mukherjee DV, Zhou W, Liotta LA, Petricoin EF, Chandhoke V, Bailey C, Popov SG: **Secreted neutral metalloproteases of *Bacillus anthracis* as candidate pathogenic factors**. *The Journal of biological chemistry* 2006, **281**(42):31408-31418.
150. Cadot C, Tran SL, Vignaud ML, De Buyser ML, Kolsto AB, Brisabois A, Nguyen-The C, Lereclus D, Guinebretiere MH, Ramarao N: **InhA1, NprA, and HlyII as candidates for markers to differentiate pathogenic from nonpathogenic *Bacillus cereus* strains**. *Journal of clinical microbiology* 2010, **48**(4):1358-1365.
151. Gilmore MS, Cruz-Rodz AL, Leimeister-Wachter M, Kreft J, Goebel W: **A *Bacillus cereus* cytolytic determinant, cereolysin AB, which comprises the phospholipase C and sphingomyelinase genes: nucleotide sequence and genetic linkage**. *Journal of bacteriology* 1989, **171**(2):744-753.
152. Beecher DJ, Olsen TW, Somers EB, Wong AC: **Evidence for contribution of tripartite hemolysin BL, phosphatidylcholine-preferring phospholipase C, and collagenase to virulence of *Bacillus cereus* endophthalmitis**. *Infect Immun* 2000, **68**(9):5269-5276.
153. Oda M, Takahashi M, Matsuno T, Uoo K, Nagahama M, Sakurai J: **Hemolysis induced by *Bacillus cereus* sphingomyelinase**. *Biochimica et biophysica acta* 2010, **1798**(6):1073-1080.
154. Beecher DJ, Wong AC: **Cooperative, synergistic and antagonistic haemolytic interactions between haemolysin BL, phosphatidylcholine phospholipase C and sphingomyelinase from *Bacillus cereus***. *Microbiology* 2000, **146 Pt 12**:3033-3039.
155. Doll VM, Ehling-Schulz M, Vogelmann R: **Concerted action of sphingomyelinase and non-hemolytic enterotoxin in pathogenic *Bacillus cereus***. *PloS one* 2013, **8**(4):e61404.
156. Gohar M, Faegri K, Perchat S, Ravnum S, Okstad OA, Gominet M, Kolsto AB, Lereclus D: **The PlcR virulence regulon of *Bacillus cereus***. *PloS one* 2008, **3**(7):e2793.
157. Jessberger N, Krey VM, Rademacher C, Böhm ME, Mohr AK, Ehling-Schulz M, Scherer S, Märtlbauer E: **From genome to toxicity: a combinatory approach highlights the complexity of enterotoxin production in *Bacillus cereus***. *Frontiers in microbiology* 2015, **6**:560.
158. Esbelin J, Armengaud J, Zigha A, Duport C: **ResDE-dependent regulation of enterotoxin gene expression in *Bacillus cereus*: evidence for multiple modes of binding for ResD and interaction with Fnr**. *Journal of bacteriology* 2009, **191**(13):4419-4426.
159. Esbelin J, Jouanneau Y, Armengaud J, Duport C: **ApoFnr binds as a monomer to promoters regulating the expression of enterotoxin genes of *Bacillus cereus***. *Journal of bacteriology* 2008, **190**(12):4242-4251.

160. van der Voort M, Kuipers OP, Buist G, de Vos WM, Abee T: **Assessment of CcpA-mediated catabolite control of gene expression in *Bacillus cereus* ATCC 14579**. *BMC microbiology* 2008, **8**:62.
161. Fagerlund A, Dubois T, Okstad OA, Verplaetse E, Gilois N, Bennaceur I, Perchat S, Gominet M, Aymerich S, Kolsto AB *et al*: **SinR controls enterotoxin expression in *Bacillus thuringiensis* biofilms**. *PloS one* 2014, **9**(1):e87532.
162. Zigha A, Rosenfeld E, Schmitt P, Duport C: **The redox regulator Fnr is required for fermentative growth and enterotoxin synthesis in *Bacillus cereus* F4430/73**. *Journal of bacteriology* 2007, **189**(7):2813-2824.
163. Duport C, Zigha A, Rosenfeld E, Schmitt P: **Control of enterotoxin gene expression in *Bacillus cereus* F4430/73 involves the redox-sensitive ResDE signal transduction system**. *Journal of bacteriology* 2006, **188**(18):6640-6651.
164. Nakano MM, Zuber P, Glaser P, Danchin A, Hulett FM: **Two-component regulatory proteins ResD-ResE are required for transcriptional activation of *fnr* upon oxygen limitation in *Bacillus subtilis***. *Journal of bacteriology* 1996, **178**(13):3796-3802.
165. Ratnayake-Lecamwasam M, Serror P, Wong KW, Sonenshein AL: ***Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels**. *Genes & development* 2001, **15**(9):1093-1103.
166. Shivers RP, Sonenshein AL: **Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids**. *Molecular microbiology* 2004, **53**(2):599-611.
167. Sonenshein AL: **CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria**. *Curr Opin Microbiol* 2005, **8**(2):203-207.
168. Handke LD, Shivers RP, Sonenshein AL: **Interaction of *Bacillus subtilis* CodY with GTP**. *Journal of bacteriology* 2008, **190**(3):798-806.
169. Majerczyk CD, Dunman PM, Luong TT, Lee CY, Sadykov MR, Somerville GA, Bodi K, Sonenshein AL: **Direct targets of CodY in *Staphylococcus aureus***. *Journal of bacteriology* 2010, **192**(11):2861-2877.
170. Frenzel E, Doll V, Pauthner M, Lucking G, Scherer S, Ehling-Schulz M: **CodY orchestrates the expression of virulence determinants in emetic *Bacillus cereus* by impacting key regulatory circuits**. *Molecular microbiology* 2012, **85**(1):67-88.
171. Francis KP, Yu J, Bellinger-Kawahara C, Joh D, Hawkinson MJ, Xiao G, Purchio TF, Caparon MG, Lipsitch M, Contag PR: **Visualizing pneumococcal infections in the lungs of live mice using bioluminescent *Streptococcus pneumoniae* transformed with a novel gram-positive *lux* transposon**. *Infect Immun* 2001, **69**(5):3350-3358.
172. Glatz BA, Goepfert JM: **Defined conditions for synthesis of *Bacillus cereus* enterotoxin by fermenter-grown cultures**. *Appl Environ Microbiol* 1976, **32**(3):400-404.
173. Rosenfeld E, Duport C, Zigha A, Schmitt P: **Characterization of aerobic and anaerobic vegetative growth of the food-borne pathogen *Bacillus cereus* F4430/73 strain**. *Canadian journal of microbiology* 2005, **51**(2):149-158.
174. Sambrook J, Russell DW: **Molecular Cloning: A Laboratory Manual**, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001.
175. Huptas C, Scherer S, Wenning M: **Optimized Illumina PCR-free library preparation for bacterial whole genome sequencing and analysis of factors influencing de novo assembly**. *BMC research notes* 2016, **9**(1):269.
176. Patel RK, Jain M: **NGS QC Toolkit: a toolkit for quality control of next generation sequencing data**. *PloS one* 2012, **7**(2):e30619.
177. Chikhi R, Medvedev P: **Informed and automated k-mer size selection for genome assembly**. *Bioinformatics* 2014, **30**(1):31-37.
178. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I: **ABySS: a parallel assembler for short read sequence data**. *Genome research* 2009, **19**(6):1117-1123.
179. Gurevich A, Saveliev V, Vyahhi N, Tesler G: **QUAST: quality assessment tool for genome assemblies**. *Bioinformatics* 2013, **29**(8):1072-1075.
180. Goecks J, Nekrutenko A, Taylor J, Galaxy T: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences**. *Genome biology* 2010, **11**(8):R86.

181. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nature methods* 2012, **9**(4):357-359.
182. Landstorfer R, Simon S, Schober S, Keim D, Scherer S, Neuhaus K: **Comparison of strand-specific transcriptomes of enterohemorrhagic *Escherichia coli* O157:H7 EDL933 (EHEC) under eleven different environmental conditions including radish sprouts and cattle feces**. *BMC genomics* 2014, **15**:353.
183. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation**. *Bioinformatics* 2000, **16**(10):944-945.
184. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics* 2010, **26**(1):139-140.
185. Muhlig A, Behr J, Scherer S, Muller-Herbst S: **Stress response of *Salmonella enterica* serovar *typhimurium* to acidified nitrite**. *Appl Environ Microbiol* 2014, **80**(20):6373-6382.
186. Darling AE, Mau B, Perna NT: **progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement**. *PloS one* 2010, **5**(6):e11147.
187. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics* 2009, **25**(14):1754-1760.
188. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.
189. Quinlan AR: **BEDTools: The Swiss-Army Tool for Genome Feature Analysis**. *Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]* 2014, **47**:11 12 11-11 12 34.
190. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P: **RDP3: a flexible and fast computer program for analyzing recombination**. *Bioinformatics* 2010, **26**(19):2462-2463.
191. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM: **DNA-DNA hybridization values and their relationship to whole-genome sequence similarities**. *International journal of systematic and evolutionary microbiology* 2007, **57**(Pt 1):81-91.
192. Agren J, Sundstrom A, Hafstrom T, Segerman B: **Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups**. *PloS one* 2012, **7**(6):e39107.
193. Gardner SN, Slezak T, Hall BG: **kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome**. *Bioinformatics* 2015, **31**(17):2877-2878.
194. Gardner SN, Hall BG: **When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes**. *PloS one* 2013, **8**(12):e81760.
195. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6: Molecular Evolutionary Genetics Analysis version 6.0**. *Molecular biology and evolution* 2013, **30**(12):2725-2729.
196. Tamura K, Nei M: **Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees**. *Molecular biology and evolution* 1993, **10**(3):512-526.
197. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J *et al*: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega**. *Molecular systems biology* 2011, **7**:539.
198. Baldauf SL: **Phylogeny for the faint of heart: a tutorial**. *Trends in genetics : TIG* 2003, **19**(6):345-351.
199. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J *et al*: **Rfam 12.0: updates to the RNA families database**. *Nucleic Acids Res* 2015, **43**(Database issue):D130-137.
200. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction**. *Nucleic Acids Research* 2003, **31**(13):3406-3415.
201. Murray MG, Thompson WF: **Rapid isolation of high molecular weight plant DNA**. *Nucleic Acids Res* 1980, **8**(19):4321-4325.
202. Beecher DJ, Wong AC: **Improved purification and characterization of hemolysin BL, a hemolytic dermonecrotic vascular permeability factor from *Bacillus cereus***. *Infect Immun* 1994, **62**(3):980-986.

203. Francis KP, Joh D, Bellinger-Kawahara C, Hawkinson MJ, Purchio TF, Contag PR: **Monitoring bioluminescent *Staphylococcus aureus* infections in living mice using a novel *luxABCDE* construct**. *Infect Immun* 2000, **68**(6):3594-3600.
204. Meighen EA: **Molecular biology of bacterial bioluminescence**. *Microbiol Rev* 1991, **55**(1):123-142.
205. Hollis RP, Lagido C, Pettitt J, Porter AJ, Killham K, Paton GI, Glover LA: **Toxicity of the bacterial luciferase substrate, *n*-decyl aldehyde, to *Saccharomyces cerevisiae* and *Caenorhabditis elegans***. *FEBS Lett* 2001, **506**(2):140-142.
206. Laemmli UK: **Cleavage of structural proteins during the assembly of the head of bacteriophage T4**. *Nature* 1970, **227**(5259):680-685.
207. Frenzel E: **Regulation of the biosynthesis of the food-borne *Bacillus cereus* toxin cereulide**. TUM; 2011.
208. Martineau F, Picard FJ, Roy PH, Ouellette M, Bergeron MG: **Species-specific and ubiquitous DNA-based assays for rapid identification of *Staphylococcus epidermidis***. *Journal of clinical microbiology* 1996, **34**(12):2888-2893.
209. Belitsky BR, Sonenshein AL: **Roadblock repression of transcription by *Bacillus subtilis* CodY**. *Journal of molecular biology* 2011, **411**(4):729-743.
210. Hoa NT, Baccigalupi L, Huxham A, Smertenko A, Van PH, Ammendola S, Ricca E, Cutting SM: **Characterization of *Bacillus* Species Used for Oral Bacteriotherapy and Bacterioprophylaxis of Gastrointestinal Disorders**. *Applied and Environmental Microbiology* 2000, **66**(12):5241-5247.
211. Böhm ME, Huptas C, Krey VM, Scherer S: **Draft Genome Sequence of *Bacillus cytotoxicus* CVUAS 2833, a Very Close Relative to Type Strain NVH 391-98 Isolated from a Different Location**. *Genome announcements* 2015, **3**(4).
212. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M *et al*: **The RAST Server: rapid annotations using subsystems technology**. *BMC genomics* 2008, **9**:75.
213. Okstad OA, Hegna I, Lindbäck T, Rishovd AL, Kolsto AB: **Genome organization is not conserved between *Bacillus cereus* and *Bacillus subtilis***. *Microbiology* 1999, **145 (Pt 3)**:621-631.
214. Slamti L, Lereclus D: **A cell-cell signaling peptide activates the PlcR virulence regulon in bacteria of the *Bacillus cereus* group**. *The EMBO journal* 2002, **21**(17):4550-4559.
215. 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 evolutionary biology* 2015, **15**:246.
216. Brillard J, Lereclus D: **Comparison of cytotoxin *cytK* promoters from *Bacillus cereus* strain ATCC 14579 and from a *B. cereus* food-poisoning strain**. *Microbiology* 2004, **150**(Pt 8):2699-2705.
217. den Hengst CD, Curley P, Larsen R, Buist G, Nauta A, van Sinderen D, Kuipers OP, Kok J: **Probing direct interactions between CodY and the *oppD* promoter of *Lactococcus lactis***. *Journal of bacteriology* 2005, **187**(2):512-521.
218. Wray LV, Jr., Fisher SH: ***Bacillus subtilis* CodY operators contain overlapping CodY binding sites**. *Journal of bacteriology* 2011, **193**(18):4841-4848.
219. Belitsky BR, Sonenshein AL: **Genome-wide identification of *Bacillus subtilis* CodY-binding sites at single-nucleotide resolution**. *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**(17):7026-7031.
220. Geng H, Zhu Y, Mullen K, Zuber CS, Nakano MM: **Characterization of ResDE-dependent *fnr* transcription in *Bacillus subtilis***. *Journal of bacteriology* 2007, **189**(5):1745-1755.
221. Kearns DB, Chu F, Branda SS, Kolter R, Losick R: **A master regulator for biofilm formation by *Bacillus subtilis***. *Molecular microbiology* 2005, **55**(3):739-749.
222. Chu F, Kearns DB, Branda SS, Kolter R, Losick R: **Targets of the master regulator of biofilm formation in *Bacillus subtilis***. *Molecular microbiology* 2006, **59**(4):1216-1228.
223. Agaisse H, Lereclus D: **STAB-SD: a Shine-Dalgarno sequence in the 5' untranslated region is a determinant of mRNA stability**. *Molecular microbiology* 1996, **20**(3):633-643.

224. Novichkov PS, Wolf YI, Dubchak I, Koonin EV: **Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes**. *Journal of bacteriology* 2009, **191**(1):65-73.
225. Jung MY, Paek WK, Park IS, Han JR, Sin Y, Paek J, Rhee MS, Kim H, Song HS, Chang YH: ***Bacillus gaemokensis* sp. nov., isolated from foreshore tidal flat sediment from the Yellow Sea**. *J Microbiol* 2010, **48**(6):867-871.
226. Jung MY, Kim JS, Paek WK, Lim J, Lee H, Kim PI, Ma JY, Kim W, Chang YH: ***Bacillus manliponensis* sp. nov., a new member of the *Bacillus cereus* group isolated from foreshore tidal flat sediment**. *J Microbiol* 2011, **49**(6):1027-1032.
227. Liu B, Liu GH, Hu GP, Sengonca C, Lin NQ, Tang JY, Tang WQ, Lin YZ: ***Bacillus bingmayongensis* sp. nov., isolated from the pit soil of Emperor Qin's Terra-cotta warriors in China**. *Antonie van Leeuwenhoek* 2014, **105**(3):501-510.
228. Cheng T, Lin P, Jin S, Wu Y, Fu B, Long R, Liu D, Guo Y, Peng L, Xia Q: **Complete Genome Sequence of *Bacillus bombysepticus*, a Pathogen Leading to *Bombyx mori* Black Chest Septicemia**. *Genome announcements* 2014, **2**(3).
229. Soufiane B, Baizet M, Cote JC: **Multilocus sequence analysis of *Bacillus thuringiensis* serovars *navarrensis, bolivia* and *vazensis* and *Bacillus weihenstephanensis* reveals a common phylogeny**. *Antonie van Leeuwenhoek* 2013, **103**(1):195-205.
230. Liu Y, Lai Q, Goker M, Meier-Kolthoff JP, Wang M, Sun Y, Wang L, Shao Z: **Genomic insights into the taxonomic status of the *Bacillus cereus* group**. *Scientific reports* 2015, **5**:14082.
231. Bavykin SG, Lysov YP, Zakhariev V, Kelly JJ, Jackman J, Stahl DA, Cherni A: **Use of 16S rRNA, 23S rRNA, and *gyrB* gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms**. *Journal of clinical microbiology* 2004, **42**(8):3711-3730.
232. Francis KP, Mayr R, von Stetten F, Stewart GS, Scherer S: **Discrimination of psychrotrophic and mesophilic strains of the *Bacillus cereus* group by PCR targeting of major cold shock protein genes**. *Appl Environ Microbiol* 1998, **64**(9):3525-3529.
233. Silva JC, Loreto EL, Clark JB: **Factors that affect the horizontal transfer of transposable elements**. *Current issues in molecular biology* 2004, **6**(1):57-71.
234. Ngamwongsatit P, Buasri W, Pianariyanon P, Pulsrikarn C, Ohba M, Assavanig A, Panbangred W: **Broad distribution of enterotoxin genes (*hblCDA*, *nheABC, cytK,* and *entFM*) among *Bacillus thuringiensis* and *Bacillus cereus* as shown by novel primers**. *International journal of food microbiology* 2008, **121**(3):352-356.
235. Clair G, Roussi S, Armengaud J, Duport C: **Expanding the known repertoire of virulence factors produced by *Bacillus cereus* through early secretome profiling in three redox conditions**. *Molecular & cellular proteomics : MCP* 2010, **9**(7):1486-1498.
236. Okstad OA, Gominet M, Purnelle B, Rose M, Lereclus D, Kolsto AB: **Sequence analysis of three *Bacillus cereus* loci carrying PIcR-regulated genes encoding degradative enzymes and enterotoxin**. *Microbiology* 1999, **145 (Pt 11)**:3129-3138.
237. Phung D, Granum PE, Dietrich R, Martlbauer E, Hardy SP: **Inhibition of cytotoxicity by the Nhe cytotoxin of *Bacillus cereus* through the interaction of dodecyl maltoside with the NheB component**. *FEMS microbiology letters* 2012, **330**(2):98-104.
238. 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.
239. Didier A, Dietrich R, Gruber S, Bock S, Moravek M, Nakamura T, Lindbäck T, Granum PE, Martlbauer E: **Monoclonal antibodies neutralize *Bacillus cereus* Nhe enterotoxin by inhibiting ordered binding of its three exoprotein components**. *Infect Immun* 2012, **80**(2):832-838.
240. Declerck N, Bouillaut L, Chaix D, Rugani N, Slamti L, Hoh F, Lereclus D, Arold ST: **Structure of PlcR: Insights into virulence regulation and evolution of quorum sensing in Gram-positive bacteria**. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(47):18490-18495.

241. Slamti L, Perchat S, Gominet M, Vilas-Boas G, Fouet A, Mock M, Sanchis V, Chaufaux J, Gohar M, Lereclus D: **Distinct mutations in PlcR explain why some strains of the *Bacillus cereus* group are nonhemolytic**. *Journal of bacteriology* 2004, **186**(11):3531-3538.
242. Fricker M, Reissbrodt R, Ehling-Schulz M: **Evaluation of standard and new chromogenic selective plating media for isolation and identification of *Bacillus cereus***. *International journal of food microbiology* 2008, **121**(1):27-34.
243. Mignot T, Mock M, Robichon D, Landier A, Lereclus D, Fouet A: **The incompatibility between the PlcR- and AtxA-controlled regulons may have selected a nonsense mutation in *Bacillus anthracis***. *Molecular microbiology* 2001, **42**(5):1189-1198.
244. Sastalla I, Maltese LM, Pomerantseva OM, Pomerantsev AP, Keane-Myers A, Leppla SH: **Activation of the latent PlcR regulon in *Bacillus anthracis***. *Microbiology* 2010, **156**(Pt 10):2982-2993.
245. Klee SR, Brzuszkiewicz EB, Nattermann H, Bruggemann H, Dupke S, Wollherr A, Franz T, Pauli G, Appel B, Liebl W *et al*: **The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids**. *PloS one* 2010, **5**(7):e10986.
246. Duport C, Thomassin S, Bourel G, Schmitt P: **Anaerobiosis and low specific growth rates enhance hemolysin BL production by *Bacillus cereus* F4430/73**. *Archives of microbiology* 2004, **182**(1):90-95.
247. Beattie SH, Williams AG: **Growth and diarrhoeagenic enterotoxin formation by strains of *Bacillus cereus in vitro* in controlled fermentations and in situ in food products and a model food system**. *Food microbiology* 2002, **19**:11.
248. Ouhib O, Clavel T, Schmitt P: **The production of *Bacillus cereus* enterotoxins is influenced by carbohydrate and growth rate**. *Curr Microbiol* 2006, **53**(3):222-226.
249. Zigha A, Rosenfeld E, Schmitt P, Duport C: **Anaerobic cells of *Bacillus cereus* F4430/73 respond to low oxidoreduction potential by metabolic readjustments and activation of enterotoxin expression**. *Archives of microbiology* 2006, **185**(3):222-233.
250. Messaoudi K, Clavel T, Schmitt P, Duport C: **Fnr mediates carbohydrate-dependent regulation of catabolic and enterotoxin genes in *Bacillus cereus* F4430/73**. *Research in microbiology* 2010, **161**(1):30-39.
251. Turnbull PC: ***Bacillus cereus* toxins**. *Pharmacol Ther* 1981, **13**(3):453-505.
252. Fermanian C, Lapeyre C, Fremy JM, Claisse M: **Production of diarrheal toxin by selected strains of *Bacillus cereus***. *International journal of food microbiology* 1996, **30**(3):345-358.
253. Fermanian C, Lapeyre C, Fremy JM, Claisse M: **Diarrhoeal toxin production at low temperature by selected strains of *Bacillus cereus***. *The Journal of dairy research* 1997, **64**(4):551-559.
254. Stenfors Arnesen L, Granum PE, Buisson C, Bohlin J, Nielsen-LeRoux C: **Using an insect model to assess correlation between temperature and virulence in *Bacillus weihenstephanensis* and *Bacillus cereus***. *FEMS microbiology letters* 2011, **317**(2):196-202.
255. Garcia-Arribas ML, Kramer JM: **The effect of glucose, starch, and pH on growth, enterotoxin and haemolysin production by strains of *Bacillus cereus* associated with food poisoning and non-gastrointestinal infection**. *International journal of food microbiology* 1990, **11**(1):21-33.
256. Ouhib-Jacobs O, Lindley ND, Schmitt P, Clavel T: **Fructose and glucose mediates enterotoxin production and anaerobic metabolism of *Bacillus cereus* ATCC14579^{T}**. *Journal of applied microbiology* 2009, **107**(3):821-829.
257. Guedon E, Serror P, Ehrlich SD, Renault P, Delorme C: **Pleiotropic transcriptional repressor CodY senses the intracellular pool of branched-chain amino acids in *Lactococcus lactis***. *Molecular microbiology* 2001, **40**(5):1227-1239.
258. Wang J, Su Y, Jia F, Jin H: **Characterization of casein hydrolysates derived from enzymatic hydrolysis**. *Chem Cent J* 2013, **7**(1):62.
259. Nolan RA: **Amino acids and growth factors in vitamin-free casamino acids**. *Mycologia* 1971, **63**(6):1231-1234.
260. Slamti L, Lemy C, Henry C, Guillot A, Huillet E, Lereclus D: **CodY Regulates the Activity of the Virulence Quorum Sensor PlcR by Controlling the Import of the Signaling Peptide PapR in *Bacillus thuringiensis***. *Frontiers in microbiology* 2015, **6**:1501.

261. Lindbäck T, Mols M, Basset C, Granum PE, Kuipers OP, Kovacs AT: **CodY, a pleiotropic regulator, influences multicellular behaviour and efficient production of virulence factors in *Bacillus cereus***. *Environmental microbiology* 2012, **14**(8):2233-2246.
262. Belitsky BR, Sonenshein AL: **Contributions of multiple binding sites and effector-independent binding to CodY-mediated regulation in *Bacillus subtilis***. *Journal of bacteriology* 2011, **193**(2):473-484.
263. Levdikov VM, Blagova E, Joseph P, Sonenshein AL, Wilkinson AJ: **The structure of CodY, a GTP- and isoleucine-responsive regulator of stationary phase and virulence in gram-positive bacteria**. *The Journal of biological chemistry* 2006, **281**(16):11366-11373.
264. Geng H, Zuber P, Nakano MM: **Regulation of Respiratory Genes by ResD–ResE Signal Transduction System in *Bacillus subtilis***. 2007, **422**:448-464.
265. Reents H, Gruner I, Harmening U, Bottger LH, Layer G, Heathcote P, Trautwein AX, Jahn D, Hartig E: ***Bacillus subtilis* Fnr senses oxygen via a [4Fe-4S] cluster coordinated by three cysteine residues without change in the oligomeric state**. *Molecular microbiology* 2006, **60**(6):1432-1445.
266. Passalacqua KD, Varadarajan A, Byrd B, Bergman NH: **Comparative transcriptional profiling of *Bacillus cereus* sensu lato strains during growth in CO_2-bicarbonate and aerobic atmospheres**. *PloS one* 2009, **4**(3):e4904.
267. Richardson DJ, Berks BC, Russell DA, Spiro S, Taylor CJ: **Functional, biochemical and genetic diversity of prokaryotic nitrate reductases**. *Cellular and molecular life sciences : CMLS* 2001, **58**(2):165-178.
268. van der Voort M, Abee T: **Transcriptional regulation of metabolic pathways, alternative respiration and enterotoxin genes in anaerobic growth of *Bacillus cereus* ATCC 14579**. *Journal of applied microbiology* 2009, **107**(3):795-804.
269. Minnaard J, Lievin-Le Moal V, Coconnier MH, Servin AL, Perez PF: **Disassembly of F-actin cytoskeleton after interaction of *Bacillus cereus* with fully differentiated human intestinal Caco-2 cells**. *Infect Immun* 2004, **72**(6):3106-3112.
270. Neuhaus K, Landstorfer R, Fellner L, Simon S, Schafferhans A, Goldberg T, Marx H, Ozoline O, Rost B, Kuster B *et al*: **Translatomics combined with transcriptomics and proteomics reveals novel functional, recently evolved orphan genes in *Escherichia coli* O157:H7 (EHEC).** *BMC genomics* 2016, **in press**.
271. Narberhaus F, Waldminghaus T, Chowdhury S: **RNA thermometers**. *FEMS microbiology reviews* 2006, **30**(1):3-16.
272. Winkler WC, Breaker RR: **Regulation of bacterial gene expression by riboswitches**. *Annual review of microbiology* 2005, **59**:487-517.
273. Loh E, Gripenland J, Johansson J: **Control of *Listeria monocytogenes* virulence by 5'-untranslated RNA**. *Trends in microbiology* 2006, **14**(7):294-298.
274. Johansson J, Mandin P, Renzoni A, Chiaruttini C, Springer M, Cossart P: **An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes***. *Cell* 2002, **110**(5):551-561.
275. Wong KK, Bouwer HG, Freitag NE: **Evidence implicating the 5' untranslated region of *Listeria monocytogenes actA* in the regulation of bacterial actin-based motility**. *Cellular microbiology* 2004, **6**(2):155-166.
276. Shen A, Higgins DE: **The 5' untranslated region-mediated enhancement of intracellular listeriolysin O production is required for *Listeria monocytogenes* pathogenicity**. *Molecular microbiology* 2005, **57**(5):1460-1473.
277. Kristoffersen SM, Tourasse NJ, Kolsto AB, Okstad OA: **Interspersed DNA repeats bcr1-bcr18 of *Bacillus cereus* group bacteria form three distinct groups with different evolutionary and functional patterns**. *Molecular biology and evolution* 2011, **28**(2):963-983.
278. Busby S, Ebright RH: **Transcription activation by catabolite activator protein (CAP)**. *Journal of molecular biology* 1999, **293**(2):199-213.
279. Körner H, Sofia HJ, Zumft WG: **Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs**. *FEMS microbiology reviews* 2003, **27**(5):559-592.

280. Nakano MM, Zhu Y: **Involvement of ResE phosphatase activity in down-regulation of ResD-controlled genes in *Bacillus subtilis* during aerobic growth**. *Journal of bacteriology* 2001, **183**(6):1938-1944.
281. Nakano MM, Zhu Y, Haga K, Yoshikawa H, Sonenshein AL, Zuber P: **A mutation in the 3-phosphoglycerate kinase gene allows anaerobic growth of *Bacillus subtilis* in the absence of ResE kinase**. *Journal of bacteriology* 1999, **181**(22):7087-7097.
282. Esbelin J, Jouanneau Y, Duport C: ***Bacillus cereus* Fnr binds a [4Fe-4S] cluster and forms a ternary complex with ResD and PlcR**. *BMC microbiology* 2012, **12**:125.
283. Lereclus D, Agaisse H, Gominet M, Salamitou S, Sanchis V: **Identification of a *Bacillus thuringiensis* gene that positively regulates transcription of the phosphatidylinositol-specific phospholipase C gene at the onset of the stationary phase**. *Journal of bacteriology* 1996, **178**(10):2749-2756.
284. Bai U, Mandic-Mulec I, Smith I: **SinI modulates the activity of SinR, a developmental switch protein of *Bacillus subtilis*, by protein-protein interaction**. *Genes & development* 1993, **7**(1):139-148.
285. Hamon MA, Lazazzera BA: **The sporulation transcription factor Spo0A is required for biofilm development in *Bacillus subtilis***. *Molecular microbiology* 2001, **42**(5):1199-1209.
286. Hamon MA, Stanley NR, Britton RA, Grossman AD, Lazazzera BA: **Identification of AbrB-regulated genes involved in biofilm formation by *Bacillus subtilis***. *Molecular microbiology* 2004, **52**(3):847-860.
287. Hsueh YH, Somers EB, Lereclus D, Wong AC: **Biofilm formation by *Bacillus cereus* is influenced by PlcR, a pleiotropic regulator**. *Appl Environ Microbiol* 2006, **72**(7):5089-5092.
288. Lereclus D, Agaisse H, Grandvalet C, Salamitou S, Gominet M: **Regulation of toxin and virulence gene transcription in *Bacillus thuringiensis***. *International Journal of Medical Microbiology* 2000, **290**(4-5):295-299.
289. Perego M, Spiegelman GB, Hoch JA: **Structure of the gene for the transition state regulator, *abrB*: regulator synthesis is controlled by the *spo0A* sporulation gene in *Bacillus subtilis***. *Molecular microbiology* 1988, **2**(6):689-699.
290. Molle V, Nakaura Y, Shivers RP, Yamaguchi H, Losick R, Fujita Y, Sonenshein AL: **Additional Targets of the *Bacillus subtilis* Global Regulator CodY Identified by Chromatin Immunoprecipitation and Genome-Wide Transcript Analysis**. *Journal of bacteriology* 2003, **185**(6):1911-1922.
291. Hsueh YH, Somers EB, Wong AC: **Characterization of the *codY* gene and its influence on biofilm formation in *Bacillus cereus***. *Archives of microbiology* 2008, **189**(6):557-568.
292. Stenz L, Francois P, Whiteson K, Wolz C, Linder P, Schrenzel J: **The CodY pleiotropic repressor controls virulence in gram-positive pathogens**. *FEMS immunology and medical microbiology* 2011, **62**(2):123-139.
293. Wünsche A, Hammer E, Bartholomae M, Völker U, Burkovski A, Seidel G, Hillen W: **CcpA forms complexes with CodY and RpoA in *Bacillus subtilis***. *The FEBS journal* 2012, **279**(12):2201-2214.
294. Choi SK, Saier MH, Jr.: **Mechanism of CcpA-mediated glucose repression of the *resABCDE* operon of *Bacillus subtilis***. *Journal of molecular microbiology and biotechnology* 2006, **11**(1-2):104-110.
295. Rejasse A, Gilois N, Barbosa I, Huillet E, Bevilacqua C, Tran S, Ramarao N, Stenfors Arnesen LP, Sanchis V: **Temperature-dependent production of various PlcR-controlled virulence factors in *Bacillus weihenstephanensis* strain KBAB4**. *Appl Environ Microbiol* 2012, **78**(8):2553-2561.
296. Stenfors LP, Mayr R, Scherer S, Granum PE: **Pathogenic potential of fifty *Bacillus weihenstephanensis* strains**. *FEMS microbiology letters* 2002, **215**(1):47-51.
297. Stenfors Arnesen LP, O'Sullivan K, Granum PE: **Food poisoning potential of *Bacillus cereus* strains from Norwegian dairies**. *International journal of food microbiology* 2007, **116**(2):292-296.
298. Zwick ME, Joseph SJ, Didelot X, Chen PE, Bishop-Lilly KA, Stewart AC, Willner K, Nolan N, Lentz S, Thomason MK *et al*: **Genomic characterization of the *Bacillus cereus* sensu lato species: backdrop to the evolution of *Bacillus anthracis***. *Genome research* 2012, **22**(8):1512-1524.

299. van der Auwera GA, Feldgarden M, Kolter R, Mahillon J: **Whole-Genome Sequences of 94 Environmental Isolates of *Bacillus cereus* Sensu Lato**. *Genome announcements* 2013, **1**(5).
300. Rasko DA, Altherr MR, Han CS, Ravel J: **Genomics of the *Bacillus cereus* group of organisms**. *FEMS microbiology reviews* 2005, **29**(2):303-329.
301. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME: **Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis***. *Journal of bacteriology* 2000, **182**(10):2928-2936.
302. Chun JH, Hong KJ, Cha SH, Cho MH, Lee KJ, Jeong DH, Yoo CK, Rhie GE: **Complete genome sequence of *Bacillus anthracis* H9401, an isolate from a Korean patient with anthrax**. *Journal of bacteriology* 2012, **194**(15):4116-4117.
303. Haggblom MM, Apetroaie C, Andersson MA, Salkinoja-Salonen MS: **Quantitative Analysis of Cereulide, the Emetic Toxin of *Bacillus cereus*, Produced under Various Conditions**. *Applied and Environmental Microbiology* 2002, **68**(5):2479-2483.
304. Rasko DA, Ravel J, Okstad OA, Helgason E, Cer RZ, Jiang L, Shores KA, Fouts DE, Tourasse NJ, Angiuoli SV *et al*: **The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1**. *Nucleic Acids Res* 2004, **32**(3):977-988.
305. Ji F, Zhu Y, Ju S, Zhang R, Yu Z, Sun M: **Promoters of crystal protein genes do not control crystal formation inside exosporium of *Bacillus thuringiensis* ssp. *finitimus* strain YBT-020**. *FEMS microbiology letters* 2009, **300**(1):11-17.
306. Hernandez E, Ramisse F, Ducoureau JP, Cruel T, Cavallo JD: ***Bacillus thuringiensis* subsp. *konkukian* (serotype H34) superinfection: case report and experimental evidence of pathogenicity in immunosuppressed mice**. *Journal of clinical microbiology* 1998, **36**(7):2138-2139.
307. Cachat E, Barker M, Read TD, Priest FG: **A *Bacillus thuringiensis* strain producing a polyglutamate capsule resembling that of *Bacillus anthracis***. *FEMS microbiology letters* 2008, **285**(2):220-226.
308. Challacombe JF, Altherr MR, Xie G, Bhotika SS, Brown N, Bruce D, Campbell CS, Campbell ML, Chen J, Chertkov O *et al*: **The complete genome sequence of *Bacillus thuringiensis* Al Hakam**. *Journal of bacteriology* 2007, **189**(9):3680-3681.
309. Turnbull PC, Nottingham JF, Ghosh AC: **A severe necrotic enterotoxin produced by certain food, food poisoning and other clinical isolates of *Bacillus cereus***. *British journal of experimental pathology* 1977, **58**(3):273-280.
310. Agata N, Ohta M, Yokoyama K: **Production of *Bacillus cereus* emetic toxin (cereulide) in various foods**. *International journal of food microbiology* 2002, **73**(1):23-27.
311. Xiong Z, Jiang Y, Qi D, Lu H, Yang F, Yang J, Chen L, Sun L, Xu X, Xue Y *et al*: **Complete genome sequence of the extremophilic *Bacillus cereus* strain Q1 with industrial applications**. *Journal of bacteriology* 2009, **191**(3):1120-1121.
312. Swiecicka I, Fiedoruk K, Bednarz G: **The occurrence and properties of *Bacillus thuringiensis* isolated from free-living animals**. *Letters in applied microbiology* 2002, **34**(3):194-198.
313. He J, Shao X, Zheng H, Li M, Wang J, Zhang Q, Li L, Liu Z, Sun M, Wang S *et al*: **Complete genome sequence of *Bacillus thuringiensis* mutant strain BMB171**. *Journal of bacteriology* 2010, **192**(15):4074-4075.
314. Lereclus D, Arantes O, Chaufaux J, Lecadet M: **Transformation and expression of a cloned delta-endotoxin gene in *Bacillus thuringiensis***. *FEMS microbiology letters* 1989, **51**(1):211-217.
315. He J, Wang J, Yin W, Shao X, Zheng H, Li M, Zhao Y, Sun M, Wang S, Yu Z: **Complete genome sequence of *Bacillus thuringiensis* subsp. *chinensis* strain CT-43**. *Journal of bacteriology* 2011, **193**(13):3407-3408.
316. Hoffmaster AR, Novak RT, Marston CK, Gee JE, Helsel L, Pruckler JM, Wilkins PP: **Genetic diversity of clinical isolates of *Bacillus cereus* using multilocus sequence typing**. *BMC microbiology* 2008, **8**:191.
317. Wang A, Pattemore J, Ash G, Williams A, Hane J: **Draft genome sequence of *Bacillus thuringiensis* strain DAR 81934, which exhibits molluscicidal activity**. *Genome announcements* 2013, **1**(2):e0017512.

318. Murawska E, Fiedoruk K, Bideshi DK, Swiecicka I: **Complete genome sequence of *Bacillus thuringiensis* subsp. *thuringiensis* strain IS5056, an isolate highly toxic to *Trichoplusia ni***. *Genome announcements* 2013, **1**(2):e0010813.
319. Guan P, Ai P, Dai X, Zhang J, Xu L, Zhu J, Li Q, Deng Q, Li S, Wang S *et al*: **Complete genome sequence of *Bacillus thuringiensis* serovar *Sichuansis* strain MC28**. *Journal of bacteriology* 2012, **194**(24):6975.
320. Antwerpen M, Proenca DN, Ruckert C, Licht K, Kalinowski J, Hanczaruk M, Tiemann C, Grass G: **Draft genome sequence of *Bacillus anthracis* BF-1, isolated from Bavarian cattle**. *Journal of bacteriology* 2012, **194**(22):6360-6361.
321. Harrington R, Ondov BD, Radune D, Friss MB, Klubnik J, Diviak L, Hnath J, Cendrowski SR, Blank TE, Karaolis D *et al*: **Genome Sequence of the Attenuated Carbosap Vaccine Strain of *Bacillus anthracis***. *Genome announcements* 2013, **1**(1).
322. Price EP, Seymour ML, Sarovich DS, Latham J, Wolken SR, Mason J, Vincent G, Drees KP, Beckstrom-Sternberg SM, Phillippy AM *et al*: **Molecular epidemiologic investigation of an anthrax outbreak among heroin users, Europe**. *Emerging infectious diseases* 2012, **18**(8):1307-1313.
323. Ruckert C, Licht K, Kalinowski J, Espirito Santo C, Antwerpen M, Hanczaruk M, Reischl U, Holzmann T, Gessner A, Tiemann C *et al*: **Draft genome sequence of *Bacillus anthracis* UR-1, isolated from a German heroin user**. *Journal of bacteriology* 2012, **194**(21):5997-5998.
324. Miller JM, Hair JG, Hebert M, Hebert L, Roberts FJ, Jr., Weyant RS: **Fulminating bacteremia and pneumonia due to *Bacillus cereus***. *Journal of clinical microbiology* 1997, **35**(2):504-507.
325. Lereclus D, Menou G, Lecadet MM: **Isolation of a DNA sequence related to several plasmids from *Bacillus thuringiensis* after a mating involving the *Streptococcus faecalis* plasmid pAM beta 1**. *Molecular & general genetics : MGG* 1983, **191**(2):307-313.
326. Su L, Zhou T, Zhou L, Fang X, Li T, Wang J, Guo Y, Chang D, Wang Y, Li D *et al*: **Draft genome sequence of *Bacillus cereus* strain LCT-BC244**. *Journal of bacteriology* 2012, **194**(13):3549.
327. Naranjo M, Denayer S, Botteldoorn N, Delbrassinne L, Veys J, Waegenaere J, Sirtaine N, Driesen RB, Sipido KR, Mahillon J *et al*: **Sudden death of a young adult associated with *Bacillus cereus* food poisoning**. *Journal of clinical microbiology* 2011, **49**(12):4379-4381.

7. Annex

Table S1: List of 223 *B. cereus* sensu lato strains investigated in this study.

142 strains listed in the first part of the table were included in the final set and are listed according to their cluster affiliation (C1 – C7) derived from concatenated housekeeping gene species tree (Fig. 5). Presence (x) or absence (-) of toxin and regulator genes is indicated. For 81 strains listed in the second part of the table, one or more housekeeping gene(s) or *plcR* could not be identified from the genome sequence. These strains were excluded from further analyses, but are affiliated into clusters according to a whole-genome comparison with Gegenees 2.2.1 [192] (data not shown). Type strains are highlighted in bold, strains sequenced in this study are marked by an asterisk. Five *B. cereus* strains (#291 – #295) were sequenced only recently and, therefore, not part of the phylogenetic analyses.

Ba: *Bacillus anthracis*, Bb: *Bacillus bombysepticus*, Bcyt: *Bacillus cytotoxicus*, Bm: *Bacillus mycoides*, Bpm: *Bacillus pseudomycoides*, Bt: *Bacillus thuringiensis*, Btoy: *Bacillus toyonensis*, Bw: *Bacillus weihenstephanensis*, not specified name: *Bacillus cereus*.

Recently, two publications gave insight into *B. cereus* sensu lato taxonomy based on entire genomes using gANI and AF [50] and GBDP and MLSA [230]. Their results – as listed below – confirm this study and phylogenetic clusters correlate nicely with the MLSA species tree (Fig. 5).

Part 1: Final strainset

Cluster	Strain	Original name	*nhe*	*hbl*	*cytK*	*plcR*	2. *hbl*	2. *nhe*	*ces*	Source	Reference	Cluster according to [230]	Cliques according to [50]
Cluster 1 (C1)													
C1_S1	282	**Bpm DSM 12442**	x	-	-	x	-	-	-	Soil	[39]	BCG05	305
C1_S2	152	Bm Rock3-17	x	-	-	x	x	-	-	Soil, Maryland	[298]	BCG05	305
Cluster 2 (C2)													
C2_S1	47	BAG2X1-2	x	x	-	x	x	-	-	Soil, Massachusetts	[299]	BCG13	984
C2_S2	56	BAG6X1-1	x	x	-	x	x	-	-	Soil, Massachusetts	[299]	BCG13	984
C2_S3	106	MM3	x	x	-	x	-	-	-	Food	[298]	Single strain cluster	367
C2_S4	140	MHI 226 *	x	-	-	x	x	x	-	Milk and milk products, Germany	1)	Missing	Missing
C2_S5	144	14294-3 (M6) *	x	x	x	x	x	x	-	Ice cream, Germany, 2004	3) WSBC 10904	Missing	Missing
C2_S6	156	BAG5X2-1	x	x	x	x	x	-	-	Soil, Massachusetts	[299]	Single strain cluster	367
C2_S7	211	BAG2O-3	x	-	-	x	x	-	-	Soil, Massachusetts	[299]	BCG18	440
C2_S8	243	RIVM BC 126 *	x	x	-	x	x	-	-	Patients faeces, Netherlands, 1999	2)	Missing	
Cluster 3 (C3)													
C3_S1	28	**Ba str. Ames Ancestor**	x	-	-	x	-	-	-	Dead heifer, Texas	[300]	BCG01	507
C3_S2	29	Ba str. Ames	x	-	-	x	-	-	-	Laboratory strain	[301]	BCG01	507
C3_S3	21	Ba str. A0248	x	-	-	x	-	-	-	Human, USAMRIID, Ohio	sequenced by J. Craig Venter Institute	BCG01	507
C3_S4	35	Ba str. H9401	x	-	-	x	-	-	-	Clinical, cutaneous anthrax, Korea	[302]	BCG01	507
C3_S5	38	Ba str. Sterne	x	-	-	x	-	-	-	Laboratory strain	[300]	BCG01	507
C3_S6	1	03BB102	x	-	-	x	-	-	-	Dust, fatal pneumonia, Texas	Sequenced by TIGR	BCG01	507
C3_S7	2	95/8201	x	-	x	x	-	-	-	Endocarditis, UK 1995	[298]	BCG01	507
C3_S8	3	NVH 0075-95 *	x	-	-	x	-	-	-	Stew with vegetables, food poisoning, Norway, 1995	[114]	missing	missing

C3_S9	5	HWW 274-2 *	x	-	x	x	-	-	-	Milk powder, Germany, 2004	Lang-Halter, unpublished	missing	missing
C3_S10	8	AH187	x	-	-	x	-	-	x	Vomit, cooked rice, London 1972	[303]	BCG12	52-1
C3_S11	12	AH820	x	x	x	x	-	-	-	Periodontosis, Norway 1995	[91]	BCG01	507
C3_S12	17	ATCC 10987	x	-	x	x	-	-	-	Cheese spoilage, Canada 1930	[304]	BCG10	807
C3_S13	60	BDRD-ST26	x	-	-	x	-	-	-	BDRD stock strain	[298]	BCG12	52-1/52-2
C3_S14	62	biovar anthracis str. Cl	x	-	-	x	-	-	-	Chimpanzee, fatal anthrax, Cote d'Ivoire (CI) 2001	[245]	BCG01	507
C3_S15	70	Bt serovar andalousiensis BGSC 4AW1	x	-	x	x	-	-	-	Spain	[298]	BCG01	507
C3_S16	73	Bt serovar finitimus YBT-020	x	x	-	x	-	-	-	Soil, China (Huazhong Agricultural University)	[305]	single strain cluster	52-2
C3_S17	75	Bt serovar konkukian str. 97-27	x	x	x	x	-	-	-	Wound infection, french soldier in Yugoslawia 1995	[306]	BCG01	507
C3_S18	77	Bt serovar monterrey BGSC 4AJ1	x	x	x	x	-	-	-	Mexico	[307]	BCG01	507
C3_S19	80	Bt serovar pulsiensis BGSC 4CC1	x	-	x	x	-	-	-	Grain field, Pakistan	[298]	BCG01	507
C3_S20	83	Bt serovar tochigiensis BGSC 4Y1	x	x	x	x	-	-	-	Soil, Japan	[298]	BCG14	250
C3_S21	84	Bt str. Al Hakam	x	x	-	x	-	-	-	Suspected bioweapon facility, Iraq	[308]	BCG01	507
C3_S22	90	E33L	x	-	x	x	-	-	-	Zebra carcass, Namibia, 1996	[300]	BCG01	507
C3_S23	91	F837/76	x	x	-	x	-	-	-	Human, wound, postoperative infection 1976	[309]	BCG01	507
C3_S24	93	FRI-35	x	-	x	x	-	-	-	?	Sequenced by Los Alamos National	BCG10	807
C3_S25	102	ISP3191	x	-	-	x	-	-	-	Spice, Belgium	[93]	BCG01	507
C3_S26	104	m1293	x	-	-	x	-	-	-	Cream cheese	[298]	BCG12	52-1/52-2
C3_S27	109	MSX-D12	x	-	x	x	-	-	-	Antarctic concordia station and ISS	[299]	BCG12	52-1/52-2
C3_S28	110	NC7401	x	-	-	x	-	-	x	Food poisoning, chow mein, Japan	[310]	BCG12	52-1/52-2
C3_S29	113	Q1	x	-	-	x	-	-	-	Deep surface oil reseroir, China	[311]	BCG12	52-1/52-2

C3_S30	119	Rock3-42	x	-	x	x	-	-	-	Soil, Rockville in Maryland	[298]	BCG01	507
C3_S31	139	MHI 86 *	x	-	x	x	-	-	-	Infant food, Germany	1)	missing	missing
C3_S32	141	SDA KA 96 *	x	x	x	x	-	-	-	Raw milk, Sweden, 1997	[139]	missing	missing
C3_S33	142	WSBC 10035 *	x	-	-	x	-	-	-	Pasteurized milk, Germany, 1993	[139]	missing	missing
C3_S34	143	F4429/71 *	x	-	x	x	-	-	-	Vanilla pudding, Netherlands, 1971	1) MHI 1543	missing	missing
C3_S35	148	ATCC 4342	x	x	-	x	-	-	-	?	[298]	BCG14	250
C3_S36	161	BGSC 6E1	x	x	-	x	-	-	-	?	[298]	BCG01	507
C3_S37	205	IS195	x	-	-	x	-	-	x	Intestine of bank vole, Poland	[312]	BCG12	52-1
C3_S38	206	IS845/00	x	-	-	x	-	-	-	Intestine of bank vole, Poland	[312]	BCG12	52-1
C3_S39	212	F	x	-	-	x	-	-	-	Permafrost sample, 3 mio. years old	Institute of chemical biology and fundamental	BCG10	807
C3_S40	242	F528/94 *	x	x	-	x	-	-	-	Beef chow mein & rice, food poisoning, UK (PHLS), 1994	[118]	missing	missing
C3_S41	244	RIVM BC 90 *	x	-	-	x	-	-	-	Human faeces, Netherlands, 1999	2)	missing	missing
C3_S42	246	7/27/S *	x	-	-	x	-	-	-	Human faeces	1) MHI 3185	missing	missing
C3_S43	248	F3162/04 *	x	-	x	x	-	-	-	Human faeces, 2004	1) MHI 3173	missing	missing
Cluster 4 (C4)													
C4_S1	15	**ATCC 14579**	x	x	x	x	-	-	-	Air, cow-shed	[1]	BCG03	80
C4_S2	71	**Bt serovar berliner ATCC 10792**	x	x	x	x	-	-	-	Mediterranean flour moth (Ephestia kuehniella)	[298]	BCG04	80
C4_S3	4	F4430/73 *	x	x	x	x	-	-	-	Pea soup, Belgium, 1973	[246]	missing	missing
C4_S4	6	172560W	x	x	x	x	x	-	-	Burn wound	[298]	BCG03	80
C4_S5	11	AH676	x	x	x	x	-	-	-	Soil, Norway	[298]	BCG03	80
C4_S6	16	ATCC 10876	x	x	x	x	x	-	-	?	[298]	BCG03	80
C4_S7	18	B4264	x	x	x	x	-	-	-	Fatal pneumonia, blood and pleural fluid, 1969	[298]	BCG03	80

C4_S8	48	BAG3O-2	x	x	x	x	-	-	-	Soil, Massachusetts	[299]	BCG03	80
C4_S9	50	BAG3X2-2	x	x	x	x	x	-	-	Soil, Massachusetts	[299]	BCG03	80
C4_S10	51	BAG4O-1	x	x	x	x	-	-	-	Soil, Massachusetts	[299]	BCG03	80
C4_S11	53	BAG4X12-1	x	x	x	x	-	-	-	Soil, Massachusetts	[299]	BCG03	80
C4_S12	58	BDRD-Cer4	x	x	x	x	-	-	-	BDRD stock strain	[298]	BCG03	80
C4_S13	59	BDRD-ST24	x	x	x	x	-	-	-	BDRD stock strain	[298]	BCG03	80
C4_S14	63	Bt BMB171	x	x	x	x	-	-	-	Lab strain, China	[313]	BCG03	80
C4_S15	64	Bt Bt407	x	x	x	x	-	-	-	? Strain isolated by O. Arantes	[314]	BCG04	80
C4_S16	65	Bt HD-771	x	x	-	x	x	-	-	?	Sequenced by Los Alamos National	BCG17	80
C4_S17	66	Bt HD-789	x	x	x	x	-	-	-	?	Sequenced by Los Alamos National	BCG17	80
C4_S18	67	Bt IBL 200	x	x	x	x	-	-	-	Human	[298]	BCG17	80
C4_S19	68	Bt IBL 4222	x	x	x	x	-	-	-	Cat	[298]	BCG17	80
C4_S20	72	Bt serovar chinensis CT-43	x	x	x	x	-	-	-	China	[315]	BCG04	80
C4_S21	74	Bt serovar huazhongensis BGSC 4BD1	x	x	-	x	x	-	-	China	[298]	BCG03	80
C4_S22	76	Bt serovar kurstaki str. T03a001	x	x	x	x	x	-	-	Mediterranean flour moth (Ephestia kuehniella)	[298]	BCG03	80
C4_S23	78	Bt serovar pakistani str. T13001	x	x	x	x	-	-	-	Lepidoptera	[298]	BCG03	80
C4_S24	82	Bt serovar thuringiensis str. T01001	x	x	x	x	-	-	-	Mediterranean flour moth (Ephestia kuehniella)	[298]	BCG04	80
C4_S25	92	F65185	x	x	x	x	-	-	-	Open fracture, New York	[298]	BCG03	80
C4_S26	95	G9842	x	x	-	x	-	-	-	Stool, food poisoning, Nebraska 1996	[316]	BCG17	80
C4_S27	105	m1550	x	x	x	x	-	-	-	Uncooked chicken, Brazil	[298]	BCG03	80
C4_S28	116	Rock1-15	x	x	x	x	-	-	-	Soil, Rockville in Maryland	[298]	BCG03	80
C4_S29	123	VD014	x	x	x	x	-	-	-	Soil, Spain	[93]	BCG03	80

C4_S30	131	VD156	x	x	x	x	-	-	-	Soil, Abu Dhabi, UAE	[93]	BCG03	80
C4_S31	133	VD169	x	x	x	x	-	-	-	Dubai, UAE	[93]	BCG03	80
C4_S32	134	VD200	x	x	x	x	-	-	-	Water, Scotland	[93]	BCG03	80
C4_S33	175	VD133	x	x	x	x	x	-	-	Soil, Martinique	[93]	BCG03	80
C4_S34	199	BAG2O-1	x	-	x	x	-	-	-	Soil, Massachusetts	[299]	BCG04	80
C4_S35	201	BAG1X2-2	x	-	x	x	-	-	-	Soil, Massachusetts	[299]	BCG04	80
C4_S36	202	BAG1X2-1	x	-	x	x	-	-	-	Soil, Massachusetts	[299]	BCG04	80
C4_S37	213	Bt DAR 81934	x	x	x	x	-	-	-	Australia	[317]	BCG17	80
C4_S38	214	Bt serovar thuringiensis str. IS5056	x	x	x	x	-	-	-	Soil, Biebrza Nation Park, Poland	[318]	BCG04	80
C4_S39	235	Bt YBT-1518	x	x	x	x	-	-	-	Soil, China	Huazhong Agricultural University	missing	80
C4_S40	236	#17 *	x	x	x	x	-	-	-	Mouse microbiota	Obtained from Thomas Clavel	missing	missing
C4_S41	238	RIVM BC 964 *	x	-	x	x	-	-	-	Kebab, Netherlands, 2002	2)	missing	missing
C4_S42	239	RIVM BC 934 *	x	x	x	x	-	-	-	Lettuce, Netherlands, 2002	[139]	missing	missing
C4_S43	240	INRA A3 *	x	x	x	x	-	x	-	Starch, Normandie, France, 1998	[139]	missing	missing
C4_S44	241	INRA C3 *	x	x	x	x	-	-	-	Pasteurized carrot, Vaucluse, France, 1996	[139]	missing	missing
C4_S45	245	6/27/S *	x	x	x	x	x	-	-	Human faeces	1) MHI 3172	missing	missing
C4_S46	247	F3175/03 *	x	x	x	x	-	-	-	Human faeces, 2004	1) MHI 3169	missing	missing
C4_S47	251	HD73	x	x	x	x	x	-	-	?	Sequenced by Broad Institute	BCG03	80
C4_S48	280	Bw FSL R5-860	x	x	x	x	-	-	-	Pasteurized Milk	Sequenced by Cornell University	missing	missing
C4_S49	288	Bb str. Wang	x	x	x	x	-	-	-	Cadaver of silkworm larvae (Bombyx mori)	[228]	missing	missing
Cluster 5 (C5)													
C5_S1	289	**Btoy BCT-7112**	x	x	-	x	-	-	-	Purified for use as probiotic, Japan 1966	[43]	BCG09	missing

C5_S2	44	BAG1O-2	x	x	-	x	x	-	-	Soil, Massachusetts	[299]	BCG09	770
C5_S3	52	BAG4X2-1	x	x	-	x	x	-	-	Soil, Massachusetts	[299]	BCG09	770
C5_S4	54	BAG5O-1	x	x	-	x	x	-	-	Soil, Massachusetts	[299]	BCG09	770
C5_S5	55	BAG6O-1	x	x	-	x	x	-	-	Soil, Massachusetts	[299]	BCG09	770
C5_S6	69	Bt MC28	x	x	-	x	x	-	-	Forest, Sichuan China	[319]	BCG09	770
C5_S7	99	HuB2-9	x	x	-	x	x	-	-	Environmental isolate	[93]	BCG09	770
C5_S8	100	HuB5-5	x	x	-	x	-	-	-	Environmental isolate	[93]	BCG09	770
C5_S9	115	Rock1-3	x	x	-	x	x	-	-	Soil, Rockville in Maryland	[298]	BCG09	770
C5_S10	117	Rock3-28	x	x	-	x	x	-	-	Soil, Rockville in Maryland	[298]	BCG09	770
C5_S11	118	Rock3-29	x	x	-	x	x	-	-	Soil, Rockville in Maryland	[298]	BCG09	770
C5_S12	121	Rock4-18	x	x	x	x	-	-	-	Soil, Rockville in Maryland	[298]	BCG09	missing
C5_S13	129	VD148	x	x	-	x	x	-	-	Soil, Switzerland	[93]	BCG09	770
C5_S14	173	HuB4-10	x	x	-	x	x	-	-	Environmental isolate	[93]	BCG09	770
C5_S15	180	VD115	x	x	x	x	x	-	-	Soil, France	[93]	BCG09	770
C5_S16	209	VD214	x	x	-	x	x	-	-	Water, Scotland	[93]	BCG09	770
C5_S17	210	BAG2O-2	x	x	-	x	x	-	-	Soil, Massachusetts	[299]	BCG09	770
C5_S18	220	HuA2-3	x	x	-	x	x	-	-	Environmental isolate	[93]	BCG09	770
C5_S19	229	VD131	x	x	-	x	x	-	-	Soil, Martinique	[93]	BCG09	770
C5_S20	237	IP5832 *	x	x	-	x	-	-	-	Commercial probiotic	[210]	missing	missing
Cluster 6 (C6)													
C6_S1	87	**Bw WSBC 10204 ***	x	x	-	x	x	x	-	Milk	[36]	BCG02	missing
C6_S2	150	**Bm DSM 2048**	x	x	-	x	-	-	-	Soil	[298]	BCG02	777

C6_S3	10	AH621	x	x	-	x	x	-	-	Soil, Norway	[298]	BCG02	777
C6_S4	61	BDRD-ST196	x	x	-	x	x	-	-	BDRD stock strain	[298]	BCG02	777
C6_S5	85	Bw KBAB4	x	-	-	x	x	x	-	Soil	[300]	BCG02	777
C6_S6	97	HuA2-4	x	-	-	x	x	-	-	Environmental isolate	[93]	BCG02	777
C6_S7	126	VD048	x	x	-	x	x	-	-	Soil, Denmark	[93]	BCG02	777
C6_S8	135	VDM022	x	x	-	x	x	-	-	Soil, Greenland	[93]	BCG02	777
C6_S9	137	VDM062	x	-	-	x	x	-	-	Soil, Scotland	[93]	BCG02	777
C6_S10	149	Bw BtB2-4	x	x	-	x	x	-	-	Forest soil, Belgium	[38]	BCG02	777
C6_S11	155	BAG5X1-1	x	x	-	x	-	-	-	Soil, Massachusetts	[299]	BCG16	273
C6_S12	166	Bw CER057	x	x	-	x	x	-	x	Parsley, Belgium	[38]	BCG02	777
C6_S13	168	Bw CER074	x	x	-	x	x	-	x	Raw milk, Belgium	[38]	BCG02	777
C6_S14	174	Bw MC67	x	x	-	x	-	-	x	Soil, Denmark	[37]	BCG16	273
C6_S15	177	VD078	x	x	-	x	x	-	-	Soil, Greenland	[93]	BCG02	777
C6_S16	232	VDM019	x	x	-	x	x	-	-	Soil, Greenland	[93]	BCG02	777
C6_S17	279	Bw FSL H7-687	x	x	-	x	-	-	-	Pasteurized Milk	Sequenced by Cornell University	missing	missing
C6_S18	283	Bm WSBC 10969 *	x	x	-	x	x	-	-	Raw milk, Germany, 2014	This study	missing	missing
Cluster 7 (C7)													
C7_S1	57	**Bcyt NVH 391-98**	x	-	x	x	-	-	-	Vegetable puree, food poisoning, France, 1998	[111]	single strain cluster	1126
C7_S2	249	Bcyt CVUAS2833 *	x	-	x	x	-	-	-	Potato puree, food poisoning, Germany, 2007	[30]	missing	missing

1) Strain collection of the Department for Hygiene and Technology of Milk (MHI), Germany
2) Strain collection of the Rijksinstituut voor Volksgezondheid en Milieu (RIVM), Netherlands
3) Weihenstephan *Bacillus cereus* group Strain Collection (WSBC), Germany

Part 2: Additional strains

Cluster	Strain	Original name	*nhe*	*hbl*	*cytK*	*plcR*	2. *hbl*	2. *nhe*	*ces*	Source	Reference		
C6	9	AH603	x	x	-	-	-	-	-	Dairy	[298]	BCG02	777
C4	13	AH1134	x	x	x	-	-	-	-	Pediatric endophtalmitis, Oklahoma City, Dean McGee Eye Institute	Sequenced by TIGR	BCG03	80
C3	14	AND1407	x	-	-	-	-	-	x	Blackcurrant	[38]	BCG12	52-1/52-2
C3	19	Ba str. A0174	x	-	-	-	-	-	-	Canada	Sequenced by Los Alamos National Laboratory	BCG01	507
C3	20	Ba str. A0193	x	-	-	-	-	-	-	Bovine isolate, South Dakota	Sequenced by Los Alamos National	BCG01	507
C3	22	Ba str. A0389	x	-	-	-	-	-	-	Bekasi, Indonesia	Sequenced by Los Alamos National	BCG01	507
C3	23	Ba str. A0442	x	-	-	-	-	-	-	Kudu (Antelope), Kruger National Park, South Africa	Sequenced by Los Alamos National Laboratory	BCG01	507
C3	24	Ba str. A0465	x	-	-	-	-	-	-	Bovine isolate, France	Sequenced by Los Alamos National	BCG01	507
C3	25	Ba str. A0488	x	-	-	-	-	-	-	Infected cattle, UK 1935	Sequenced by Los Alamos National	BCG01	507
C3	26	Ba str. A1055	x	-	-	-	-	-	-	From Paul Keim´s laboratory	[300]	BCG01	507
C3	27	Ba str. A2012	x	-	-	-	-	-	-	Clinical, inhalational anthrax, West Palm Beach, Florida 2001	[301]	BCG01	507
C3	30	Ba str. Australia 94	x	-	-	-	-	-	-	Australia	[300]	BCG01	507
C3	31	Ba str. BF1	x	-	-	-	-	-	-	Cow carcass, Bavaria, Germany	[320]	BCG01	missing
C3	32	Ba str. CDC 684	x	-	-	x	-	-	-	?	Sequenced by J. Craig Venter Institute	BCG01	587
C3	33	Ba str. CNEVA-9066	x	-	-	-	-	-	-	France	[300]	BCG01	507
C3	34	Ba str. Carbosap	x	-	-	-	-	-	-	Italy	[321]	BCG01	507
C3	36	Ba str. Heroin Ba4599	x	-	-	-	-	-	-	Clinical, first case of anthrax outbreak, Glasgow, Scotland, Herion	[322]	BCG01	507
C3	37	Ba str. Kruger B	x	-	-	-	-	-	-	Kruger National Park, Sout Africa	[301]	BCG01	507
C3	39	Ba str. Tsiankovskii-I	x	-	-	-	-	-	-	Soviet Union	Sequenced by J. Craig Venter Institute	BCG01	507

C3	40	Ba str. UR-1	x	-	-	-	-	-	-	Clinical, injectional anthrax in a German heroin user	[323]	BCG01	507
C3	41	Ba str. Vollum	x	-	-	-	-	-	-	Occurs in the UK, Spain, Zimbabwe	[300]	BCG01	507
C3	42	Ba str. Western North America USA6153	x	-	-	-	-	-	-	?	[300]	BCG01	507
C4	45	BAG1X1-2	x	x	x	-	-	-	-	Soil, Massachusetts	[299]	BCG04	80
C6	46	BAG1X1-3	x	x	-	x	-	-	-	Soil, Massachusetts	[299]	BCG11	2
C6	49	BAG3X2-1	x	x	-	x	-	-	-	Soil, Massachusetts	[299]	BCG11	2
C3	79	Bt serovar pondicheriensis BGSC 4BA1	x	x	x	x	-	-	-	Soil, India	[298]	BCG01	507
C4	81	Bt serovar sotto str. T04001	x	x	-	x	x	-	-	Canada	[298]	BCG17	80
C3	94	G9241	x	x	x	-	-	-	-	Pneumonia, 1987	[324]	BCG14	250
C4	96	Bt s. kurstaki str. HD73	x	x	x	x	x	-	-	?	[325]	BCG03	80
C4	98	HuB1-1	x	x	x	x	-	-	-	Environmental isolate	[93]	BCG04	80
C3	101	IS075	x	-	-	-	-	-	x	Intestine of bank vole, Poland	[312]	BCG12	52-1
C3	103	LCT-BC244	x	-	-	-	-	-	-	China General Microbiological Culture Collection Center (CGMCC)	[326]	BCG12	52-1/52-2
C4	107	MSX-A1	x	x	x	-	-	-	-	Antarctic concordia station and ISS	[299]	BCG17	80
C3	108	MSX-A12	x	-	-	-	-	-	-	Antarctic concordia station and ISS	[299]	BCG12	52-1/52-2
C3	111	NVH0597-99	x	-	x	-	-	-	-	Spice mix, food poisoning outbreak, Norway	Sequenced by TIGR	BCG01	507
C2	114	R3098/03	x		-	x	-	-	-	Septicemia, UK	[298]	single strain cluster	4162
C4	120	Rock4-2	x	x	x	x	-	-	-	Soil, Rockville in Maryland	[298]	BCG03	80
C3	122	SJ1	x	-	x	-	-	-	-	?	University of Arizona	BCG01	missing
C4	124	VD022	x	x	-	-	-	-	-	Water, Belgium	[93]	BCG17	80

C4	125	VD045	x	x	x	-	-	-	-	Soil, Denmark	[93]	BCG03	80
C3	127	VD102	x	-	x	-	-	-	-	Soil, Guadeloupe	[93]	BCG12	52-1/52-2
C6	128	VD142	x	x	-	-	-	-	-	Soil, Scotland	[93]	BCG02	777
C4	130	VD154	x	x	x	-	-	-	-	Soil, Abu Dhabi, UAE	[93]	BCG03	80
C4	132	VD166	x	x	x	-	-	-	-	Dubai, UAE	[93]	BCG03	80
C6	136	VDM034	x	x	-	x	-	-	-	Soil, Spain	[93]	BCG02	777
C3	138	W	x	x	x	-	-	-	-	?	Sequenced by TIGR	BCG01	507
C6	146	AH1272	x	x	-	x	-	-	-	Amniotic fluid, Iceland	[298]	BCG11	2
C6	147	AH1273	x	x	-	x	-	-	-	Human blood, Iceland	[298]	BCG11	2
C1	151	Bm Rock1-4	x	x	-	x	-	-	-	Soil, Maryland	[298]	BCG05	305
C5	153	BAG2X1-1	x	x	-	x	-	-	-	Soil, Massachusetts	[299]	BCG19	926
C6	158	BAG6O-2	x	x	-	x	-	-	-	Soil, Massachusetts	[299]	BCG16	273
C6	160	BAG6X1-2	x	x	-	-	-	-	-	Soil, Massachusetts	[299]	single strain cluster	3455
C6	171	HuA2-1	x	x	-	-	-	-	-	Environmental isolate	[93]	BCG02	777
C6	172	HuA4-10	x	x	-	x	-	-	-	Environmental isolate	[93]	single strain cluster	3536
C6	179	VD107	x	x	-	-	-	-	-	Soil, Guadeloupe	[93]	single strain cluster	1119
C6	188	VD118	x	x	-	x	-	-	-	Soil, Guadeloupe	[93]	BCG16	273
C5	196	BAG2X1-3	x	x	-	x	-	-	-	Soil, Massachusetts	[299]	BCG19	926
C4	200	BAG1X2-3	x	-	x	x	-	-	-	Soil, Massachusetts	[299]	BCG04	80
C4	203	BAG1X1-1	x	-	x	x	-	-	-	Soil, Massachusetts	[299]	BCG04	80
C4	204	K-5975c	x	x	-	x	-	-	-	Pasta salad, fatal food poisoning, 2003, Belgium	[9]	BCG17	80
C4	207	VD140	x	-	x	-	-	-	-	Soil, Scotland	[93]	BCG03	80

C2	216	B5-2	x	x	-	-	-	-	-	Soil, China	[93]	BCG18	440
C6	217	BAG1O-1	x	x	-	-	-	-	-	Soil, Massachusetts	[299]	BCG11	2
C2	218	BAG3O-1	x	x	-	-	-	-	-	Soil, Massachusetts	[299]	BCG18	440
C4	219	BAG5X12-1	x	x	x	-	-	-	-	Soil, Massachusetts	[299]	BCG03	80
C6	221	HuA2-9	x	x	-	-	-	-	-	Environmental isolate	[93]	BCG02	777
C6	222	HuA3-9	x	-	-	-	-	-	-	Environmental isolate	[93]	BCG02	777
C4	223	HuB4-4	x	x	-	-	-	-	-	Environmental isolate	[93]	BCG17	80
C4	224	ISP2954	x	-	x	-	-	-	-	Food (durum wheat), Belgium	[93]	BCG03	80
C4	226	Schrouff	x	x	-	-	-	-	-	Milk, Belgium	[93]	BCG17	80
C4	227	TIAC219	x	x	-	-	-	-	-	Spaghetti and tomato sauce, lethal intoxication, Brussels, Belgium, 2008	[327]	BCG17	80
C6	231	VD146	x	x	-	-	-	-	-	Soil, Scotland	[93]	BCG02	777
C6	233	VDM053	x	x	-	-	-	-	-	Water, Belgium	[93]	BCG11	2
C4	234	BAG1O-3	x	-	x	-	-	-	-	Soil, Massachusetts	[299]	BCG04	80
C3	290	H3081.97	x	-	-	x	-	-	x	Environmental isolate, USA, CDC	Sequenced by J. Craig Venter Institute	BCG12	52-1/52-2
-	291	MHI 1670 *	x	-	x	x	-	-	-	Milk proteins	1)	Missing	Missing
-	292	MHI 1672 *	x	-	-	x	-	-	x	Food isolate	1) [115]	Missing	Missing
-	293	MHI 1761 *	x	-	x	x	-	-	-	Food isolate	1) [115]	Missing	Missing
-	294	MHI 2968 *	x	-	x	x	-	-	-	Infant food	1)	Missing	Missing
-	295	MHI 3233 *	x	x	x	x	-	-	-	?	1)	Missing	Missing

1) Strain collection of the Department for Hygiene and Technology of Milk (MHI), Germany

Table S2: Intra-operon recombination analysis of enterotoxin operons.

142 concatenated *nheABC* genes, 94 concatenated *hblCDAB* genes and 46 concatenated *hblCDA*$_a$ genes were investigated. Statistically proven recombination events were detected by RDP3 (see section 2.5.4). Strains containing two recombinations are highlighted in bold, strains containing three recombinations are additionally underlined. * breakpoint unclear. All breakpoints of recombination events were independent of gene boundaries.

Thirteen *nhe* intra-operon recombinations involved cluster III strains and all strains containing two recombinations are also found in cluster III. In cluster VI a group of closely related strains (#61, #85, #87, #97 and #137) contains three recombinations.

Recombinant sequence(s) *nheABC*	Breakpoints [bp]	Cluster	Minor parent	Major parent	#programs	Max. average p-value
212, **113**, 104, 109, 139, 143, 246	64 – 2256	III & II	140	73	7	1,00E-07
117, 129, 44, 100, 115, 210, 237, 289, 52, 99, 118, 209, 54, 55, 173, 220	1325 – 2202	IV & V	213	180	7	2,30E-02
62, 102	18 – 2371	III & IV	119	16	7	4,56E-03
70, 75, 90, 119, **1**	1113 – 2424	III	84	91	7	9,30E-03
113	65* – 515	III	83	246	5	3,51E-04
211, 144, 256	985 – 2446	II	140	56	7	2,89E-04
280, 15, 58, 59, 63, 4, 11, 64, 71, 72, 82, 214, 116	1995 – 3031	IV	199	240	5	7,62E-03
69, 121, 52, 99, 118, 209, 54, 55, 173, 220	87 – 1317*	V & III	117	91	5	1,52E-03
93	38 – 1954	III	148	83	7	1,07E-03
85, 97, 61, 87, 137	188 – 1294	V & III	177	93	7	1,78E-02
235, 78, 134, 133	1581 – 2212	IV	238	236	4	8,16E-04
56, 47	1059 – 2791	II & III	140	242	6	2,29E-03
83	995 – 1582	III	148	206	6	2,59E-02
174	2355 – 3030	VI	155	126	4	2,18E-03

1, 91	386 – 819	III	242	5	5	4,56E-02
73	2631 – 2907	III & IV	90	247	4	8,18E-03
141	1996* – 2790	III & IV	12	4	3	1,47E-02
85, 97, 61, 87, 137	1295* – 2008	II & IV	47	135	4	2,38E-02
91	1686 – 2034	IV & III	241	28	3	2,23E-03
10, 126	1025 – 1327	VI	149	174	7	3,67E-02
85, 97, 61, 87, 137	2009* – 2662	VI	177	10	6	4,21E-02

Recombinant sequence(s) *hblCDAB*	**Breakpoints [bp]**	**Cluster**	**Minor parent**	**Major parent**	**# programs**	**Max. average p-value**
126, 61, 87, 10	2041 – 4892	VI	166	155	7	1,04E-18
174	3681 – 4863*	VI	61	10	7	6,82E-04
229, 69, 54	674 – 1685	V	55	180	4	2,44E-03
280	2867 – 4962	IV	59	241	3	2,58E-03
236, 65, 95, 213	2138 – 3421	IV	116	16	3	3,14E-02

Recombinant sequence(s) *hblCDA$_a$*	**Breakpoints [bp]**	**Cluster**	**Minor parent**	**Major parent**	**# programs**	**Max. average p-value**
144	1325 – 2475	II & VI	232	140	6	8,27E-04
10, 126, 85, 97	2173 – 2362	II & V	180	140	3	1,47E-02
135, 177, 137, 232, 283, 61, 87, 149, 166, 168	2180 – 2362	II & V	180	140	3	1,47E-02

Table S3: Confirmation of presence of second *hbl*$_a$ operons.
Five of the *de novo* assembled strains were found to contain two versions of *hbl*, which could be discerned by ratios of cov_{operon} to $cov_{construct}$ being greater than 1.

Strain	Operon	Median construct coverage ($cov_{construct}$)	Median contig coverage (cov_{contig})	Ratio ($cov_{construct}/cov_{contig}$)	Median operon coverage (cov_{operon})	Ratio ($cov_{operon}/cov_{construct}$)
#245	*hbl*	51	50	1.02	68	1.33
	hbl$_a$	180	157	1.15	210	1.17
	hbl	58	50	1.16	68	1.17
	hblCD	202	1,587**	0.13	222	1.10
	hbl$_a$	196	157	1.25*	210	1.07
	hblCD	60	1,587**	0.04***	222	3.70***
#243	*hbl*	79	83	0.95	93	1.18
	hbl$_a$	76	74	1.03	126	1.66
#283	*hbl*	103	104	0.99	116	1.13
	hbl$_a$	108	109	0.99	156	1.44
#87	*hbl*	102	108	0.94	132	1.29
	hbl$_a$	95	108	0.88	169	1.78
#144	*hbl*	101	101	1.00	123	1.22
	hbl$_a$	102	101	1.01	157	1.54

* Reads that map uniquely to *hbl*$_a$ contain reads of *hbl*$_a$ and *hblCD*. Therefore, coverage is increased in comparison to cov_{contig}.

** Extremely high coverage over the contig (length 4105 bp) suggests that it might be (part of) a plasmid.

*** In a construct containing *hbl*$_a$ and *hblCD* few reads map uniquely against *hblCD*.

Table S3 summarizes the median coverage information obtained for each operon (cov_{operon}) and the contig (cov_{contig}) it belongs to as well as the median coverages of *hbl* and *hbl*$_a$ within artificial sequence constructs ($cov_{construct}$). Taking the ratio of $cov_{construct}$ to cov_{contig} shows that all operons fit very well to their genomic backgrounds (contigs), since respective values are close to 1. Taking the ratio of cov_{operon} to $cov_{construct}$ for each individual operon shows that median coverages obtained after remapping against operon sequences alone are higher than compared to the ones after remapping against corresponding artificial constructs, since respective ratio values are greater than 1. This can be explained by the fact that within each artificial construct reads are preferentially forming primary alignments (best hits) to the operon (*hbl* or *hbl*$_a$) where they naturally are originating from. In contrast, when mapping against individual operon sequences alone (no construct!), a substantial fraction of reads originating from *hbl*$_a$

are aligning to *hbl* as well, but only due to the missing possibility of forming a better alignment with hbl_a (since it is not present). This observation accounts also in vice versa direction.
The third copy of *hbl* in strain #245 is due to an assembly error. On the one hand, an extremely low ratio ($cov_{construct}/cov_{contig}$) of 0.04 was found. On the other hand, there are almost no reads mapping uniquely to the third *hbl* copy, revealing it as a mis-assembled second copy of hbl_a.

Table S4: Confirmation of presence of second nhe_a operons.
Three of the *de novo* assembled strains were found to contain nhe_a, which is discerned from *nhe* by its uniquely mapping reads.

Strain	Operon	Median operon coverage (cov_{operon})	Median contig coverage (cov_{contig})	Ratio ($cov_{operon}/cov_{contig}$)	# Unique reads	# Combined reads
#87	*nhe*	155	108	1.44*	2,996	40
	nhe_a	153	144	1.06	3,183	
#144	*nhe*	123	127	0.97	9,893**	4
	nhe_a	235	231	1.02	4,800	
#140	*nhe*	59	62	0.95	950	8
	nhe_a	66	66	1.00	1,106	

* The *nhe* operon maybe be located within the wrong contig due to an unexpected high ratio ($cov_{operon}/cov_{contig}$) of 1.44. However, examination of read sets mapping uniquely either to *nhe* or nhe_a unambiguously show that both versions are present within the genome of strain #87.
** Unusually high read number is caused by a ~40x higher coverage over the intergenic region (2474 – 2634 bp) between *nheB* and *nheC*, which may be due to a duplication of this region into a plasmid with high copy numbers.

Table S4 summarizes the median coverage information obtained for each operon (cov_{operon}) and the contig (cov_{contig}) it belongs to. Taking the ratio of cov_{operon} to cov_{contig} shows that all operons fit very well to their genomic backgrounds (contigs), since respective values are close to 1. Significant read pile-up beyond genomic backgrounds of individual operons is not observable, since nhe_a diverged from *nhe* to a point that its reads do not align to *nhe* anymore (and vice versa). Examination of read sets mapping to nhe_a and *nhe* indeed shows that most reads either map to nhe_a or *nhe* (# unique reads) with only a small number of reads mapping to both copies (# combined reads).

Table S5: Comparison of *B. cereus* F837/76 transcriptomes after 2 h growth in cGSM vs. GSM.
81 genes (16 hypothetical proteins, 20 %) were found to be upregulated at least 2x and 20 genes (5 hypothetical proteins, 25 %) to be downregulated at least -2x. Hypothetical proteins were excluded from the following list. Coding genes are named according to RefSeq.ptt files downloaded from the NCBI FTP database and functional allocation was performed according to COGs from the NCBI COG database and general protein information from the Uniprot database.

C	Energy production and conversion
E	Amino acid transport and metabolism
F	Nucleotide transport and metabolism
H	Coenzyme transport and metabolism
J	Translation, ribosomal structure and biogenesis
K	Transcription
M	Cell wall/membrane/envelope biogenesis
N	Cell motility
O	Post-translational modification, protein turnover, and chaperones
P	Inorganic ion transport and metabolism
Q	Secondary metabolites biosynthesis, transport, and catabolism
T	Signal transduction mechanisms
V	Defense mechanisms
R	General function prediction only
S	Function unknown

Gene	Transcriptional change	Functional category	Protein
upregulated			
bcf_10510	63.78	CP	respiratory nitrate reductase subunit beta
bcf_10515	61.96	CP	respiratory nitrate reductase subunit delta
bcf_10545	61.81	H	Molybdenum cofactor biosynthesis protein MoaE
bcf_10550	54.32	H	molybdenum cofactor biosynthesis protein MoaD
bcf_10555	38.78	P	Nitrate/nitrite transporter
bcf_10540	34.81	H	Molybdopterin biosynthesis protein MoeA
bcf_09585	23.99	CO	transport ATP-binding protein CydD
bcf_02495	22.76	O	pyruvate formate-lyase activating enzyme
bcf_10505	20.46	CP	respiratory nitrate reductase subunit alpha
bcf_20640	17.71	E	glutamine transport ATP-binding protein GlnQ
bcf_02860	17.16	R (O?)	collagenase
bcf_10575	17.02	H	sirohydrochlorin ferrochelatase
bcf_10535	16.76	H	Molybdopterin biosynthesis protein MoeB
bcf_20645	16.04	E	glutamine ABC transporter permease
bcf_10580	15.49	H	Uroporphyrinogen-III methyltransferase
bcf_17865	15.05	F	Ribonucleotide reductase of class III (anaerobic), large subunit

bcf_23215	14.24	E	argininosuccinate lyase
bcf_03465	13.00	M	Broad-substrate range phospholipase C
bcf_26115	9.82	S (M?)	integral membrane protein
bcf_24070	9.46	C	Cytochrome d ubiquinol oxidase subunit II
bcf_17875	9.43	F	Ribonucleotide reductase of class III (anaerobic), large subunit
bcf_10585	9.41	PQ	Nitrite reductase (NAD(P)H) small subunit
bcf_09570	9.37	C	Cytochrome d ubiquinol oxidase subunit I
bcf_02490	8.94	C	Pyruvate formate-lyase
bcf_23220	8.68	E	argininosuccinate synthase
bcf_09270	8.61	V	Enterotoxin C
bcf_09195	8.17	R	radical SAM protein
bcf_20650	7.75	ET	amino acid ABC transporter substrate-binding protein
bcf_24065	7.67	C	Cytochrome d ubiquinol oxidase subunit I
bcf_09265	7.51	V	putative non-hemolytic enterotoxin lytic component L1
bcf_09820	7.13	S (V?)	S-layer protein
bcf_09260	6.53	V	Non-hemolytic enterotoxin A
bcf_12035	6.49	H	protoporphyrinogen oxidase
bcf_15970	6.42	V?	Zinc metalloproteinase
bcf_01305	6.35	R (K?)	oligopeptide ABC transporter substrate-binding protein OppA
bcf_13360	6.15	V	bacillolysin
bcf_02940	5.90	NT	methyl-accepting chemotaxis protein
bcf_10530	5.77	H	Molybdenum cofactor biosynthesis protein MoaA
bcf_16995	5.33	PC	hydroxylamine reductase
bcf_08565	5.17	R	Oxidoreductase ucpA
bcf_27065	4.85	M	D-alanyl-D-alanine carboxypeptidase
bcf_10590	4.78	C	Nitrite reductase (NAD(P)H) large subunit
bcf_09215	4.21	M	Peptidase, M23/M37 family
bcf_06425	4.16	H	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase
bcf_04940	4.13	S (V?)	S-layer protein
bcf_21115	4.04	E	L-cystine uptake protein TcyP
bcf_06435	3.69	R	putative oxidoreductase
bcf_18660	3.68	R (O?)	collagenase
bcf_16330	3.64	V	Thiol-activated cytolysin
bcf_02785	3.59	O	Periplasmic thiol-disulfide interchange protein DsbA
bcf_05420	3.46	H	Protoporphyrinogen IX oxidase HemY
bcf_05415	3.41	H	Ferrochelatase, protoheme ferro-lyase
bcf_24050	3.40	T	S-ribosylhomocysteine lyase
bcf_17525	3.29	M	N-acetylmuramoyl-L-alanine amidase

bcf_22330	3.28	H	porphobilinogen deaminase
bcf_04435	3.25	E?	L-cystine ABC transporter ATP-binding protein TcyC
bcf_07165	3.21	C?	glyoxylate reductase
bcf_04485	3.19	Q	Alpha-acetolactate decarboxylase
bcf_17515	3.02	R	lipoprotein, NLP/P60 family
bcf_17495	2.99	Phage	Holin
bcf_08965	2.97	R (K?)	oligopeptide ABC transporter substrate-binding protein OppA
bcf_18435	2.87	K	DeoR family transcriptional regulator
bcf_04425	2.86	E?	L-cystine ABC transporter substrate binding protein TcyA
bcf_06345	2.82	K	PadR family transcriptional regulator
bcf_11365	2.73	R	Alcohol dehydrogenase

downregulated

bcf_22175	-6.03	VJ	endoribonuclease L-PSP
bcf_00955	-5.49	E	Arginase
bcf_00825	-3.61	R (P?)	ATPase component of general energizing module of ECF transporters
bcf_20010	-3.56	C	3-ketoacyl-CoA thiolase
bcf_00835	-3.35	R (P?)	Transmembrane component of ECF transporters
bcf_23790	-3.34	CP	Na+/H+ antiporter
bcf_20705	-3.23	CJ	Phosphate butyryltransferase
bcf_00830	-3.20	R (P?)	ATPase component of general energizing module of ECF transporters
bcf_00840	-3.12	J	tRNA pseudouridine synthase A
bcf_01515	-2.96	O	putative molecular chaperone
bcf_20695	-2.79	C	Butyrate kinase
bcf_03390	-2.78	E (P?)	Oligopeptide transport ATP-binding protein OppD
bcf_03385	-2.76	E (P?)	Oligopeptide transport ATP-binding protein OppF
bcf_01520	-2.73	J	ribosomal-protein-S18p-alanine acetyltransferase
bcf_01510	-2.66	R	ATPase

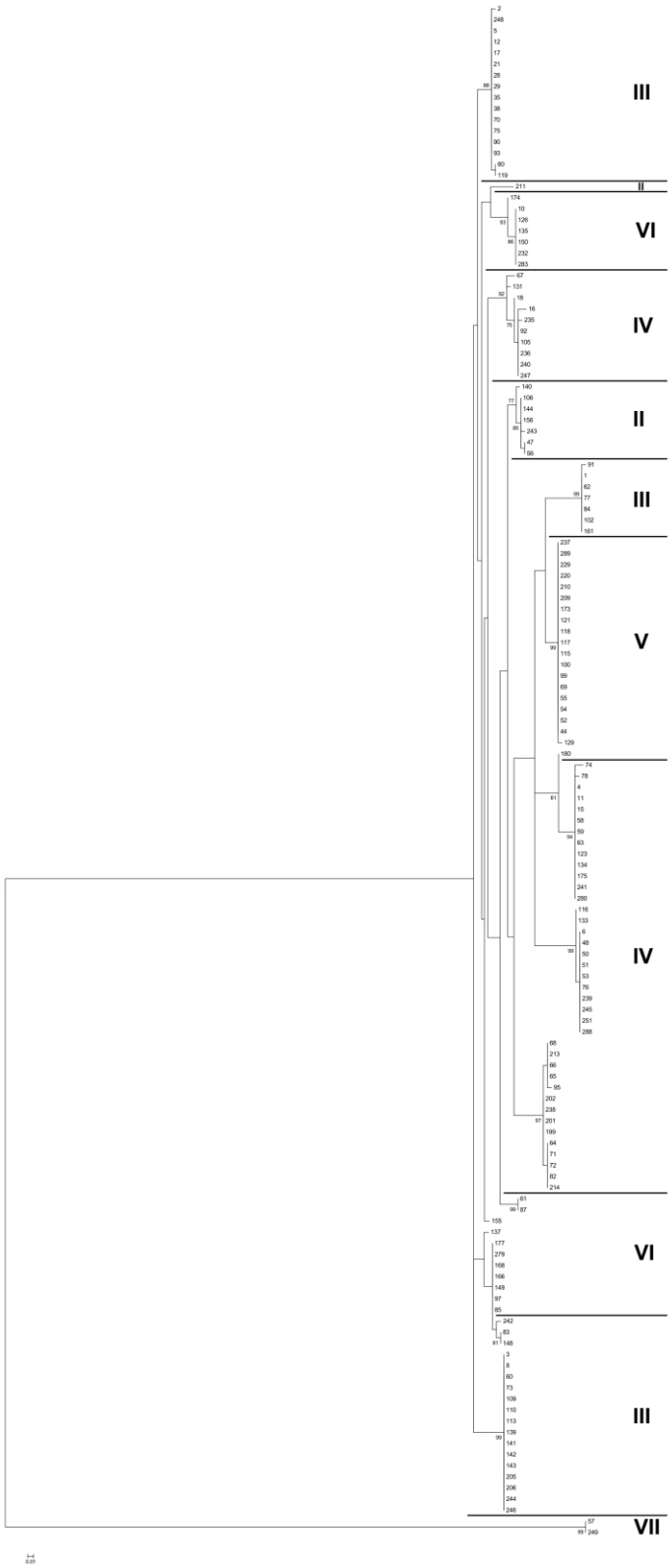


Fig. S1: *PapR* in *B. cereus* sensu lato.

Phylogenetic *papR* tree (Maximum Likelihood Method) based on the sequence (length 144 bp) of 138 *B. cereus* sensu lato strains. Strains #152 and #282 (*B. pseudomycoides*, cluster I) do not possess *papR*. Strains #104 and #212 have been excluded from the phylogenetic analysis due to incomplete sequence data (*papR* sequences contain Ns). Both resolution and bootstrap values are low due to short sequence length. Comparison with the *plcR* tree (Fig. 13) reveals generally similar topology.

Fig. S2: *Hbl* 5'IGR of *B. cereus* INRA C3.

Translation starts and potential CodY binding sites are indicated. Yellow: Sites with ≥2 mismatches, red: sites with 1 mismatch in comparison to the consensus sequence [219]. Binding site with positive results in EMSA is underlined.

> *Hbl* 5'IGR INRA C3 (966 bp)

ATAATGATATTAGGATGTTTTGTGAAAAAATCAACAATATAACATATATTACTAAATATA
TCTACATTTTATGCAATTATACATAACTAAATAAAGGTAAAAAAGTATAAAAAGACCTAT
TATATTATTCTATAAGTATTTTTTCTAAAATAAAATTTCTCGGTTGAGCTAAAATAGTTA
TTTTTAACCGTATACACATTAATTTGTAATCATTACAATTACAGAAATGAAATTTACGGA
TAACTATATTTGTATAATTTTTTTCGGTTCATATTTTTATTTAACAGTAAATTATCGATC
GTAATCGACAAAATTTTTCTATTTACGCATTAAAAATTTAATGTTTTAATGAACAACATA
ACTGGTATGACCAGACAGAAAGGATAAGGTTACGCTAATAGGAATTATAGTGAAGTTGTA
AGTAACATTATGTTGAAAATATTTTTCGAATAGTCTATTTATTTACAAGAGGTCAAGAAT
CAAGTTTGTAATAAACGTGTTCTAAGTTTCTGCATAACAAAAGTGAAGTTATTCCGCAAT
AAAAGTATAAGCGATGTACAGTATAATTTTACCTTTTTTAGTCTAGTAAGGAGTTTGCTG
ATAAAACTAAGAGTAATATTATTAAATTTAAATTAAACAATGTTATGGCAAGATTAACTG
TAAGCATTGGTTTATTGATTCTCGCGGTTCTGTAAGTTAAACCGCAATTCTAGGGAAGAA
TTACACATTTACTATTCATAGGGTGCTTGATTCAAATATAGTTAATAAATTTTTGTTATT
TATATTTGAAAAAATAGAGTATCGAAGTGAGAGGTAAGTAAAATCCTTACATTCTATTAG
AAGTAAAAAATAACGATATTATCCTATCTGAAAGATTTCTCCTTCTTAATCAGTTTACAA
AGAGAGAGGTCATACAAGTTATATGAATTAAGTTATAAAAACAATGGTAAAGGAGTGTAC
GGAATG *hblC*

P*hbl* parts tested in gel mobility shift experiments:

>1

AATAATGATATTAGGATGTTTTGTGAAAAAATCAACAATATAACATATATTACTAAATAT
ATCTACATTTTATGCAATTATACATAACTAAATAAAGGTAAAAAAGTATAAAAAGACCTA
TTATATTATTCTATAAGTATTTTTTCTAAAATAAAATTTCTCGGTTGAGCTAAAATAGTT
ATTTTTAACCGTATACACATTAATTTGTAATCATTACAATTACAGAAATGAAATTTACGG
ATAACTATATTTGTATAATTTTTTTCGGTTCATATTTTTATTTAACAGTAAATTATCGAT
CGTAATC

>2

AAATTATCGATCGTAATCGACAAAATTTTTCTATTTACGCATTAAAAATTTAATGTTTTA
ATGAACAACATAACTGGTATGACCAGACAGAAAGGATAAGGTTACGCTAATAGGAATTAT
AGTGAAGTTGTAAGTAACATTATGTTGAAAATATTTTTCGAATAGTCTATTTATTTACAA
GAGGTCAAGAATCAAGTTTGTAATAAACGTGTTCTAAGTTTCTGCATAACAAAAGTGAAG
TTATTCCGCAATAAAAGTATAAGCGATGTACAGTATAATTTTACCTTTTTTAGTCTAGTA
AGGAGTTTGCTGA

>3

TCTAGTAAGGAGTTTGCTGATAAAACTAAGAGTAATATTATTAAATTTAAATTAAACAAT
GTTATGGCAAGATTAACTGTAAGCATTGGTTTATTGATTCTCGCGGTTCTGTAAGTTAAA
CCGCAATTCTAGGGAAGAATTACACATTTACTATTCATAGGGTGCTTGATTCAAATATAG
TTAATAAATTTTTGTTATTTATATTTGAAAAAATAGAGTATCGAAGTGAGAGGTAAGTAA
AATCCTTACATTCTATTAGAAGTAAAAAATAACGATATTATCCTATCTGAAAGATTTCTC
CTTCTTAATCAGTTTACAAAGAGAGAGGTCATACAAGTTATATGAATTAAGTTATAAAAA
CAATGGTAAAGGAGTGTACGGAAT

>4

CAAGTTTGTAATAAACGTGTTCTAAGTTTCTGCATAACAAAAGTGAAGTTATTCCGCAAT
AAAAGTATAAGCGATGTACAGTATAATTTTACCTTTTTTAGTCTAGTAAGGAGTTTGCTG
ATAAAACTAAGAGTAATATTATTAAATTTAAATTAAACAATGTTATGGCAAGATTAACTG
TAAGCATTGGTTTATTGATTCTCGCGGTTCTGTAAGTTAAACCGCAATTCTAGGGAAGAA
TTACACATTTACTATTCATAGGGTGCTTGATTCAAATATAGTTAATAAATTTTTGTTATT
TATATTTGAAAAAATAGAGTATCGAAGTGAGAG

>5

GTATACACATTAATTTGTAATCATTACAATTACAGAAATGAAATTTACGGATAACTATAT
TTGTATAATTTTTTTCGGTTCATATTTTTATTTAACAGTAAATTATCGATCGTAATCGAC
AAAATTTTTCTATTTACGCATTAAAAATTTAATGTTTTAATGAACAACATAACTGGTATG
ACCAGACAGAAAGGATAAGGTTACGCTAATAGGAATTATAGTGAAGTTGTAAGTAACATT
ATGTTGAAAATATTTTTCGAATAGTCTATTTATTTACAAGAGGTCAAGAATCAAGTTTGT
AATAAACGTGTTC

Fig. S3: *Nhe* 5'IGRs of *B. cereus* INRA C3, *B. cereus* F4810/72 and *B. cytotoxicus* CVUAS 2833.
Potential CodY binding sites and translation starts are indicated in yellow. The first CodY binding site is identical in the three strains. A potential CodY binding site was located in P*nhe* of *B. cereus* F4810/72 (grey underlined), but no binding of CodY to P*nhe* F4810/72 could be detected [170]. This site is not present in P*nhe* CVUAS 2833. P*nhe* INRA C3 and P*nhe* CVUAS 2833 were positive in EMSA experiments. CodY binding affinity might be controlled by the sequence of the second binding site, which is different in each strain.
P*nhe* of *B. cereus* INRA C3 (directly controlled by CodY) and F4810/72 (not directly controlled by CodY) share 89 % nucleotide sequence identity.

>P*nhe* INRA C3
ATCCTTTTTGTGTATATGCTATGCATAATTGCATACGGGGTTAAAAATTATATCAATATA
ATTTTACCGGTTGTTTGCAAGTAGTGTTTTCTGTTGTGATATGTAGATATTTCAATCGAT
AAAGTGAATCGGATTATATTTTGAAATAAGGACATGAAAGCTTGAAAATAAAGGGAATGA
AAATACTTCTTTACGTTATATAAACAGCTGTGTTTAAATGACCTTACAGTACAAAACTTA
TATGAAAACAAAACGAAATTTCATATTTACGTATGTTGTTTGGTTTAAAATGAATAATAT
TTTTATGTAATATATTCACTGTGTGAATGTGGGAATTTCGATGAAATAGAAAATTTATAC
AATATTATTCATACTAAACACATATGAAAAGTTTTATAACAATAAGTAAGTACTGTAGCA
ATAACAGTTGGGAGAGATGGTTCATACATAGATCACGGTGAGAACGTTCTCTATTTTAAC
CATCTCAAATTATGGGCAAGAAAATAGAATAATAAAATGCGATAGTAATAGTTGCTAATA
CATAGGAGGAGTTAAAGTG *nheA*

>P*nhe* F4810/72
CTTTACGAAAGTGAAGCTATCCTTTTTGTGTATACGCTATGCATAATTGCATACGAGTCT
GAAAGATATATTGATATAATTTTTGTGTCTGTGAGTAAGTAGAGGTTTTTGTTGTGATAT
ATGGATATTTCAATTGAAAAAGTAAAGTGGATTATATTTTAAAATAAAGACATGAAAGCT
TGAAAATAAAGGGAATGAAAATACTTCTTTACGTTATATAAATAGCTGTGTTTAAATGAC
CTTACAGTACAAAACTTATATGAAAACAAAACGGAATTTCATATTTATGTATGTGAATTA
GTTTAAAATGAATAATATTTTTATGTAATATATTCACTGTATCAATGTGGGAAATAAGAT
GAAATAGAAAATTTATGCAATGTTATTCATACTAACCACATATGAAAAGTTTTATAACGA
TAAGTAAGTACTGTAGCAATAAAATTTGGGAGAGATGGTTCGTGTGTGAATCGCGGTGAG
AACGTTCTCTATTTTAACCATCTCAAACTATGTGCAAGAAAATAGAATAATAAAATGCGA
TAGTAATAGTTACTAATACATAGGAGGAGTTAAAGTG *nheA*

>P*nhe* CVUAS 2833
GTAGCTGCCTATAAGTCGGTGAAACCGACTTATAGGCAGCCTTTTTTAATTTTATAA
ACAGAAGAGACTATGCACAATTACATATTCGCGAAATAATTATTTCAATATAACACTATC
TCTTGTGGAGAAACAAGCTTTTTTACTATTATGCATAGATATTTTAAATTGAAAAATCTT
GATTACTTTATTTTAAAATAAGGAGATAAATCATTGGAGATAAAGGGAAATAAAATAGTT
CTTTACGTTAATAAGGCGCTGTGTTTAAATAATTTTACAGTACAGAACTTGAATGAAAAA
CAAAAGATTGTTTTCTACGATTTTGCTAAAAATGAATAATATTTTCATGTAATTTATGCA
ATGCGTAAATGGGCCCATTTTTCTATGCTTTTTAGTAAATTCAAGTTTGTGTTGTATTTT
CAATTTCGTAAATAGGAAGAGAGTGCAAGGGCTGAATGTTAATTGTCCCTATTTATGAAG
GGTAACAACTGAATCTGTATAAAAAATAAATTGTTAATTTATACATTAGGAGGAATTAAA
GTG *nheA*

Fig. S4: Alignment of CodY amino acid sequences.

Amino acid sequences from *B. subtilis* subsp. *subtilis* 6051-HGW, *B. cytotoxicus* CVUAS 2833, *B. cereus* INRA C3 and *B. cereus* F4810/72 were aligned and functional domains [263, 292] are marked. **Δ:** important for isoleucine ligand binding, **°:** important for dimer formation, **•**: important for DNA-binding.

B. subtilis CodY shares an overall amino acid sequence identity of ~80 % with *B. cereus* CodY: 81 % F4810/72, 82 % INRA C3, 83 % CVUAS 2833. Segments **β2 to β3 and β3 to β4 are responsible for BCAA ligand binding**. Segments **β8 to α9 (aa 203 – 226) are the HTH-domain.** N-terminal cofactor binding domain (variable) and C-terminal DNA-binding domain (highly conserved) are present within a single polypeptide chain. N-terminal GAF (cGMP-stimulated phosphodiesterases, adenylate cyclases and a bacterial transcription regulator FhlA, α1 – α5) domain is responsible for BCAA-binding and could also provide GTP-binding. Differences between strains are highlighted. All differences except 251 are located in the N-terminal region. Exchanges at positions 94, 133, 147 and 251 lead to a change in amino acid polarity and charge.

Acknowledgments

I want to express my sincere gratitude to Prof. Dr. Siegfried Scherer for the chance to carry out my doctoral studies at the Chair for Microbial Ecology, ZIEL Research Center for Food and Nutrition. I especially appreciate his support, encouragement and help in drafting the publications. Despite his busy schedule he always had time for my questions ☺

I particularly thank Viktoria Krey for her support, vast knowledge and optimistic personality. She was a great supervisor and colleague, constantly encouraged and supported me and is a good friend.

I thank all my colleagues for the great time and work climate, especially my office-roommates Andrea Rütschle and Katharina Springer. I would also like to thank Christopher Huptas for his contribution to parts of this work, his help with all bioinformatics and NGS problems – and of course his two little guards, who always ensured that no unauthorized person entered his office!

Special thanks go to Scharifa Bornschier, Christine Braig and Romy Wecko for their excellent technical assistance as well as the funny and off-topic talks in between. I additionally thank my students Felix Behr, Stephan Rambichler, Johanna Reiter and Astrid Zinz for their contributions to this work.

I want to thank all cooperation partners, especially Nadja Jeßberger, with whom we worked together on the project of identification and risk assessment of enteropathogenic *B. cereus*. Thanks for the helpful advice and the great teamwork.

Of course I thank Nicoletta Nolle and Lea Fellner for every hour we spent together in- and outside of work, our vacations, sports and movie evenings were a great pleasure!

Last but not least, I deeply appreciate the support, interest and patience of my parents and my brother. Thank you for listening to my rambling!

CURRICULUM VITAE

Maria-Elisabeth Böhm

Born in Augsburg, Germany, 30 June 1987

Nationality: German

EDUCATION

Ph.D. (Dr. rer. nat.) in Microbiology 05/2012 – 03/2016

Technical University Munich, Chair for Microbial Ecology

Thesis "Comparative phylogenetic and transcriptional analysis of *Bacillus cereus* sensu lato enterotoxin genes *nhe, hbl* and *cytK" (Advisor: Prof. Dr. S. Scherer)*

Diploma in Biology 10/2006 – 03/2012

University of Ulm, Institute of Microbiology and Biotechnology

Thesis "Modification of the pentosephosphate pathway in *Corynebacterium glutamicum* mutant strains for the enhanced production of succinate" *(Advisor: Prof. Dr. B. Eikmanns)*

Abitur 09/1997 – 06/2006

Dominikus Ringeisen Gymnasium Ursberg

WORKING EXPERIENCE

Internship 08/2010 – 10/2010

Microbiological, biochemical and chemical routine diagnostics of food

Eurofins Laborservices GmbH, Labor Dr. Hallermayer, Augsburg, Germany