

# Discrimination among *Bacillus thuringiensis* H serotypes, serovars and strains based on 16S rRNA, *gyrB* and *aroE* gene sequence analyses

Brahim Soufiane · Jean-Charles Côté

Received: 7 July 2008 / Accepted: 18 September 2008 / Published online: 7 October 2008  
© Her Majesty the Queen in Right of Canada 2008

**Abstract** Our aim was to investigate the capability of each of three genes, 16S rRNA, *gyrB* and *aroE*, to discriminate, first, among *Bacillus thuringiensis* H serotypes; second, among *B. thuringiensis* serovars from the same H serotype; and third, among *B. thuringiensis* strains from the same serovar. The 16S rRNA, *gyrB* and *aroE* genes were amplified from 21 *B. thuringiensis* H serotypes and their nucleotide sequences determined. Additional strains from four *B. cereus* sensu lato species were included for comparison purposes. These sequences were pair-wise compared and phylogenetic relationships were revealed. Each of the three genes under study could discriminate among *B. thuringiensis* H serotypes. The *gyrB* and *aroE* genes showed a discriminatory power among *B. thuringiensis* H serotypes up to nine fold greater than that of the 16S rRNA gene. The *gyrB* gene was retained for subsequent analyses to discriminate *B. thuringiensis* serovars from the same H serotype and to discriminate strains from same serovar. A total of 42 *B. thuringiensis* strains, which encompassed 25 serovars from 12 H serotypes, were analyzed. The *gyrB*

gene nucleotide sequences were different enough as to be sufficient to discriminate among *B. thuringiensis* serovars from the same H serotype and among *B. thuringiensis* strains from the same serovar.

**Keywords** 16S rRNA gene · *aroE* gene · *Bacillus thuringiensis* · Discriminatory power · *gyrB* gene · H serotypes · Serovars

## Introduction

The genus *Bacillus* encompasses the Gram-positive, rod-shaped, aerobic, spore-forming bacteria (Claus and Berkeley 1986). The so-called *Bacillus cereus* sensu lato group comprises six highly related species: *B. cereus*, *Bacillus anthracis*, *Bacillus thuringiensis*, *Bacillus mycoides*, *Bacillus pseudomycoides* and *Bacillus weihenstephanensis*. The species *B. thuringiensis* is distinguished from the other *Bacillus* species by the synthesis, upon sporulation, of a proteinaceous parasporal inclusion body, the crystal. In some strains, this crystal has been shown to be specifically toxic to some insect pests. Over the last century, some *B. thuringiensis* strains have been developed as biological insecticides for the control of lepidopteran, dipteran or coleopteran pests, respectively (Höfte and Whiteley 1989; Glare and O’Callaghan 2000). Several screening programs have been established in the hope of isolating novel *B. thuringiensis* strains expressing novel insecticidal,

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s10482-008-9285-4) contains supplementary material, which is available to authorized users.

---

B. Soufiane · J.-C. Côté (✉)  
Agriculture and Agri-Food Canada, Research Centre,  
430 Gouin Blvd, Saint-Jean-sur-Richelieu, QC,  
Canada J3B 3E6  
e-mail: cotejc@agr.gc.ca

pesticidal and biological activities. As a result, by 1996, 50,000 *B. thuringiensis* strains were kept in various collections worldwide (Sanchis et al. 1996). H serotyping, the immunological reaction to the bacterial flagellar antigen, has been established as the method of choice for the classification of the wide diversity of *B. thuringiensis* strains (de Barjac and Bonnefoi 1962). At least 69 *B. thuringiensis* H serotypes and 80 serological varieties (serovars), have now been characterized (Lecadet et al. 1999). H serotyping, however, is limited in its capability to distinguish strains from the same H serotype by the rarity of antigenic subfactors and will not distinguish strains from the same serovars.

The 16S rRNA gene (which codes for the RNA component of the ribosome small subunit) is the golden standard for single-gene phylogenetic analyses (Woese 1987; Woese et al. 1990). The *gyrB* gene (a house-keeping gene encoding the subunit B protein of DNA gyrase, topoisomerase type I) (Yamamoto and Harayama 1995) and *aroE* gene (a house-keeping gene encoding the shikimate dehydrogenase) (Maiden et al. 1998) have also been used to complement the 16S rRNA gene in bacterial phylogenetic analysis. Both house-keeping genes are present in most bacteria, are not transmitted horizontally, and their evolutionary rate is higher than that of the 16S rRNA (Yamamoto and Harayama 1995; Maiden et al. 1998). Recently, the *gyrB* gene has proven useful at distinguishing between members of the *B. cereus* sensu lato group (Bavykin et al. 2004; Ko et al. 2004; La Duc et al. 2004; Park et al. 2007).

We set to investigate the capability of each gene, 16S rRNA, *gyrB* and *aroE*, to discriminate among, first, *B. thuringiensis* H serotypes; second, serovars from the same H serotype; and third, strains from the same serovar.

We report here the 16S rRNA, *gyrB* and *aroE* genes nucleotide sequences from *B. thuringiensis* strains which encompass 21 H serotypes. Sequences from 16 additional strains from four species in the *Bacillus cereus* group sensu lato (*B. cereus*, *B. anthracis*, *B. mycoides* and *B. weihenstephanensis*) were also included for comparison purposes. These sequences were pair-wise compared and phylogenetic relationships were revealed. The level of resolution of each gene nucleotide sequence in discriminating among *B. thuringiensis* H serotypes is discussed. The *gyrB* gene was retained for the subsequent

analyses on *B. thuringiensis* serovars from the same H serotype and strains from the same serovar.

## Materials and methods

### Bacterial strains, culture conditions and DNA isolation

The 74 bacterial strains used in this study and their sources are listed in Tables 1 and 3. The 40 strains included in Table 1 encompass 24 *B. thuringiensis* strains from 21 H serotypes, four *B. cereus*, three *B. anthracis*, four *B. mycoides* and five *B. weihenstephanensis* strains. The 42 *B. thuringiensis* strains included in Table 3 encompass 12 H serotypes and 25 serovars. A *B. thuringiensis* serovar is a serological variety that share common flagellar antigens, including, when present, common distinct antigenic subfactors. Hence, whereas most *B. thuringiensis* H serotypes contain a single serovar, some have been further divided into two or more serovars (Table 3). A *B. thuringiensis* H serotype is designated for the letter H followed by a number. A serovar is given a latine name (Lecadet et al. 1999). Eight *B. thuringiensis* strains are common to Tables 1 and 3. The genomes had been fully sequenced for eight of the 74 bacterial strains: the three *B. anthracis* strains, three of the four *B. cereus* strains, *B. thuringiensis* serovar *konkukian* and *B. weihenstephanensis* strain KBAB4. Their sequences were available from GenBank (Table 1; <http://www.ncbi.nlm.nih.gov/>). The total DNAs from two *B. weihenstephanensis* strains, WSBC 10001 and WSBC 10365, were kindly provided by Dr. Monika Ehling-Schulz, Technical University of Munich, Germany. These ten strains were not cultured in the present study. The other 64 bacterial strains were cultured and their total DNAs isolated as described previously (Xu and Côté 2003, 2006). *E. coli* strain TOP10 (Invitrogen Canada, Burlington, ON, Canada) was used for cloning PCR fragments. It was cultured as described elsewhere (Xu and Côté 2003, 2006).

### PCR amplification

#### 16S rRNA gene

The 16S rRNA gene was amplified with primer pair 16S-F1/-R1: 16S-F1 (5'-GAACGCTGGCGGCGTGC

**Table 1** *Bacillus* species and strains used in Figs. 3–5

Species	H-antigen	Source or strain <sup>a</sup>	GenBank accession no. <sup>b</sup>			
			Entire genome	<i>aroE</i>	<i>gyrB</i>	16S rRNA
<i>Bacillus thuringiensis</i> serovar <i>finitimus</i>	2	BGSC 4B1		EF210214	EF210248	EF210290
<i>Bacillus thuringiensis</i> serovar <i>finitimus</i>	2	BGSC 4B2 <sup>c</sup>		EF210215	EF210249	EF210291
<i>Bacillus thuringiensis</i> serovar <i>alesti</i>	3a,3c	IEBC-T03 001 <sup>c</sup>		EF210230	EF210264	EF210300
<i>Bacillus thuringiensis</i> serovar <i>fukuokaensis</i>	3a,3d,3e	IEBC-T03C001		EF210245	EF210279	EF210301
<i>Bacillus thuringiensis</i> serovar <i>sotto</i>	4a,4b	IEBC-T04 001 <sup>c</sup>		EF210225	EF210259	EF210309
<i>Bacillus thuringiensis</i> serovar <i>kenyae</i>	4a,4c	IEBC-T04B001		EF210223	EF210257	EF210315
<i>Bacillus thuringiensis</i> serovar <i>entomocidus</i>	6	IEBC-T06 001 <sup>c</sup>		EF210221	EF210255	EF210312
<i>Bacillus thuringiensis</i> serovar <i>ostriniaie</i>	8a,8c	IEBC-T08A001 <sup>c</sup>		EF210241	EF210275	EF210314
<i>Bacillus thuringiensis</i> serovar <i>tolworthi</i>	9	IEBC-T09 001 <sup>c</sup>		EF210243	EF210277	EF210288
<i>Bacillus thuringiensis</i> serovar <i>thompsoni</i>	12	IEBC-T12 001 <sup>c</sup>		EF210226	EF210260	EF210310
<i>Bacillus thuringiensis</i> serovar <i>tohokuensis</i>	17	IEBC-T17 001 <sup>c</sup>		EF210242	EF210276	EF210294
<i>Bacillus thuringiensis</i> serovar <i>shandongensis</i>	22	IEBC-T22 001 <sup>c</sup>		EF210234	EF210268	EF210287
<i>Bacillus thuringiensis</i> serovar <i>konkukian</i>	34	97-27 <sup>c</sup>				
<i>Bacillus thuringiensis</i> serovar <i>seoulensis</i>	35	IEBC-T35 001 <sup>c</sup>	AE017355	EF210224	EF210258	EF210286
<i>Bacillus thuringiensis</i> serovar <i>oswaldocruzi</i>	38	IEBC-T38 001 <sup>c</sup>		EF210231	EF210265	EF210284
<i>Bacillus thuringiensis</i> serovar <i>higo</i>	44	IEBC-T44 001 <sup>c</sup>		EF210222	EF210256	EF210308
<i>Bacillus thuringiensis</i> serovar <i>roskildiensis</i>	45	IEBC-T45 001 <sup>c</sup>		EF210233	EF210267	EF210304
<i>Bacillus thuringiensis</i> serovar <i>wratislaviensis</i>	47	IEBC-T47 001 <sup>c</sup>		EF210227	EF210261	EF210311
<i>Bacillus thuringiensis</i> serovar <i>poloniensis</i>	54	IEBC-T54 001 <sup>c</sup>		EF210246	EF210281	EF210305
<i>Bacillus thuringiensis</i> serovar <i>zhaodongensis</i>	62	IEBC-T62 001 <sup>c</sup>		EF210244	EF210278	EF210289
<i>Bacillus thuringiensis</i> serovar <i>bolivia</i>	63	IEBC-T63 001 <sup>c</sup>		EF210236	EF210270	EF210299
<i>Bacillus thuringiensis</i> serovar <i>pulsiensis</i>	65	IEBC-T65 001 <sup>c</sup>		EF210232	EF210266	EF210285
<i>Bacillus thuringiensis</i> serovar <i>graciosensis</i>	66	IEBC-T66 001 <sup>c</sup>		EF210240	EF210226	EF210302
<i>Bacillus thuringiensis</i> serovar <i>vazensis</i>	67	IEBC-T67 001 <sup>c</sup>		EF210235	EF210269	EF210293
<i>Bacillus anthracis</i>		Ames Ancestor	AE017334			
<i>Bacillus anthracis</i>		Ames	AE016879			
<i>Bacillus anthracis</i>		Sterne	AE017225			
<i>Bacillus cereus</i>		ATCC 14579	AE016877			
<i>Bacillus cereus</i>		ATCC 10987	AE017194			
<i>Bacillus cereus</i>		E33L	CP000001			
<i>Bacillus cereus</i>		BGSC 6A17		EF210217	EF210251	EF210297
<i>Bacillus mycoides</i>		BGSC 6A19		EF210218	EF210252	EF210306
<i>Bacillus mycoides</i>		BGSC 6A13		EF210216	EF210250	EF210296
<i>Bacillus mycoides</i>		BGSC 6A20		EF210219	EF210253	EF210313
<i>Bacillus mycoides</i>		ATCC 6462		EF210229	EF210263	EF210295
<i>Bacillus weihenstephanensis</i>		WSBC 10001		EF210220	EF210254	EF210307
<i>Bacillus weihenstephanensis</i>		WSBC 10365		EF210247	EF210280	EF210283
<i>Bacillus weihenstephanensis</i>		CCM 4965		EF210228	EF210262	EF210282
<i>Bacillus weihenstephanensis</i>		CCM 4966		EF210239	EF210273	EF210298
<i>Bacillus weihenstephanensis</i>		KBAB4	CP000903			

<sup>a</sup> IEBC, International Entomopathogenic *Bacillus* Centre, Institut Pasteur, Paris, France; BGSC, *Bacillus* Genetic Stock Center, The Ohio State University, Columbus, OH, USA; ATCC, American Type Culture Collection, Rockville, MD, USA; WSBC, *B. weihenstephanensis* genomic DNA was kindly provided by the Microbial Ecology Group, Technical University of Munich, Freising, Germany; CCM, Czech Collection of Microorganisms, Masaryk University Brno, Brno, Czech Republic

<sup>b</sup> GenBank accession number for 16S rRNA, *gyrB* and *aroE* genes

<sup>c</sup> Indicates strains used in the calculation of the discriminatory power among *Bacillus thuringiensis* H serotypes

CTAA-3') and 16S-R1 (5'-GCACCTTCGATACG GCTACC-3'). Primers were designed based on the consensus sequence of the 16S rRNA gene in *B. thuringiensis* serovar *konkukian*, *B. halodurans*, *B. licheniformis*, *B. cereus* E33L and *B. anthracis* Ames Ancestor, all publicly available from GenBank. DNA was amplified for 30 cycles at a denaturing temperature of 95°C for 60 s, annealing at 63°C for 45 s, and extension at 72°C for 45 s.

#### *gyrB* gene

The *gyrB* gene was amplified with primer pair *gyrB*-F1/-R1: *gyrB*-F1 (5'-ATGGAACAAAAGCAAATG CA-3') and *gyrB*-R1 (5'-TTAAATATCAAGGTT TTTCA-3'). Primers were designed based on the consensus sequence of the *gyrB* gene in *B. thuringiensis* serovar *konkukian*, and *B. cereus* strains E33L, ATCC 10987 and ATCC 14579, and *B. anthracis* strains Ames Ancestor, Ames and Sterne. A second primer pair *gyrB*-F2/-R2: *gyrB*-F2 (5'-ACWCGTATG CGTGARYTRGC-3') and *gyrB*-R2 (5'-CCTTGYTT TGCWGAWCCDCC-3') was developed to amplify the central region of *gyrB* gene. These primers were designed based on the partial sequence of the *gyrB* gene determined with the *gyrB*-F1/-R1 primer pair. With the *gyrB*-F1/-R1 primer pair, DNA was amplified for 30 cycles at a denaturing temperature of 95°C for 30 s, annealing at 46°C for 30 s, and extension at 72°C for 30 s. With the *gyrB*-F2/-R2 primer pair, DNA was amplified for 30 cycles at a denaturing temperature of 95°C for 45 s, annealing at 50°C for 30 s, and extension at 72°C for 60 s.

#### *aroE* gene

The *aroE* gene was amplified with primer pair *aroE*-F1/-R1: *aroE*-F1 (5'-ATCGGAAATCCAATTGG ACA-3') and *aroE*-R1 (5'-CCTGTCCACATTTCAA AYG-3'). Primers were designed based on the consensus sequence of the *aroE* gene in *B. thuringiensis* serovar *konkukian*, *B. cereus* strains E33L, ATCC 10987 and ATCC 14579, and *B. anthracis* strains Ames Ancestor, Ames and Sterne. DNA was amplified for 30 cycles at denaturing temperature of 95°C for 30 s, annealing at 47°C for 30 s, and extension at 72°C for 60 s.

Positions of primers respective to the *B. thuringiensis* serovar *konkukian* 16S rRNA, *gyrB* and *aroE*

genes are shown in Fig. 1a, b and c, respectively. All amplified products were visualized on a 0.7% agarose gel and stored at -20°C.

#### DNA cloning and sequencing

Cloning of the amplified products, selection of transformants, isolation and analyses of recombinant plasmids, and determination of the nucleotide sequences were done as described previously (Xu and Côté 2006). The 16S rRNA, *gyrB* and *aroE* genes sequences of *B. thuringiensis* serovar *konkukian*; *B. cereus* strains ATCC 14579, ATCC 10987, E33L; *B. weihenstephanensis* strain KBAB4 and *B. anthracis* strains Ames Ancestor, Ames and Sterne were retrieved directly from GenBank.

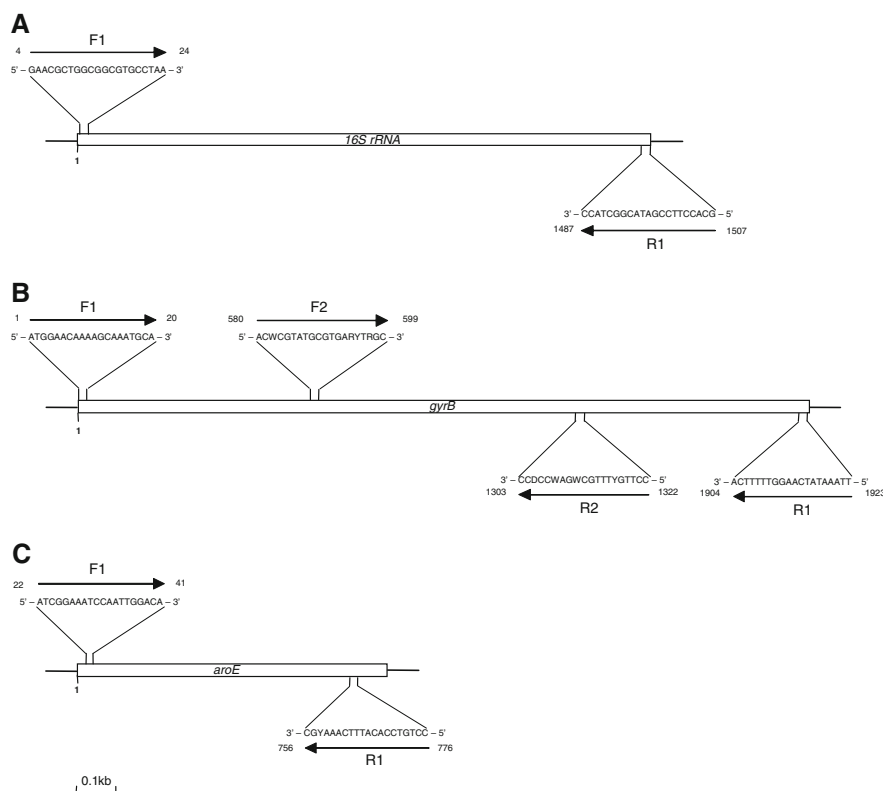
#### Percentage sequence similarity among *B. thuringiensis* H serotypes

The multiple alignments of the nucleotide sequences of the orthologous 16S rRNA, *gyrB* and *aroE* genes for 21 *B. thuringiensis* H serotypes listed in Table 1 were done using Clustal W (Thompson et al. 1994). Pairwise comparisons revealed respective percentage sequence similarities. The discriminatory power (DP) was defined here based on the median value of percentage sequence similarities among *B. thuringiensis* at the H serotype level. Lower median values reflected lower percentage sequence similarities and, consequently, higher discriminating power among *B. thuringiensis* serovars.

#### Phylogenetic analysis

The DNA sequences were analyzed using MEGA (Molecular Evolutionary Genetics Analysis; version 3.1) (Kumar et al. 2004). The 16S rRNA, *gyrB*, and *aroE* nucleotide sequences from 24 *B. thuringiensis* strains, four *B. cereus*, four *B. mycoides*, five *B. weihenstephanensis* and three *B. anthracis* strains were aligned using Clustal W (Thompson et al. 1994), a neighbor-joining tree was constructed (Saitou and Nei 1987), bootstrapped using 1,000 random samples of sites from the alignment, displayed and printed with Tree Explorer, all using MEGA.

**Fig. 1** Map of the *Bacillus thuringiensis* serovar *konkukian* 16S rRNA (a), *gyrB* (b) and *aroE* (c) genes. The number 1 refers to the first nucleotide of the three gene coding regions. Arrows indicate the orientations and positions of the primer pairs used for amplification. For the 16S rRNA gene, the primer pair 16S-F1/R1 is at nucleotide positions 4-24/1487-1507. For the *gyrB* gene, the primer pairs *gyrB*-F1/R1 and *gyrB*-F2/R2 are at nucleotide positions 1-20/1904-1923 and 580-599/1303-1322, respectively. For the *aroE* gene, the primer pair *aroE*-F1/R1 is at nucleotide positions 22-41/756-776



Nucleotide sequence accession number

Sequence data from this study have been deposited with GenBank under accession no. EF210214 to EF210315 and EU761164 to EU761197.

**Results**

16S rRNA, *gyrB* and *aroE* genes polymorphism

The nucleotide sequences of the 16S rRNA, *gyrB* and *aroE* genes were analyzed in the 40 bacterial strains listed in Table 1. These included 24 *B. thuringiensis* strains from 21 H serotypes, and 16 strains from four other species in the *Bacillus cereus* group sensu lato. Although most *B. thuringiensis* strains were selected from different H serotypes, some serovars were, however, selected purposely from same H serotype (i.e. H3 serovars *alesti* and *fukuokaensis*, and H4 serovars *sotto* and *kenyae*), and some strains from same serovar (i.e. H2 serovar *finitimus*, strains BGSC 4B1 and BGSC 4B2) in the hope of obtaining a first

indication of the level of resolution each gene nucleotide sequence could provide. The nucleotide sequences of the entire genomes were publicly available for 8 of the 40 bacterial strains: *B. thuringiensis* serovar *konkukian*, three *B. cereus* strains (ATCC 14579, ATCC 10987 and E33L), *B. weihenstephanensis* strain KBAB4 and the three *B. anthracis* under study. Their 16S rRNA, *gyrB* and *aroE* gene nucleotide sequences were retrieved directly from GenBank. All three genes were amplified, cloned and their nucleotide sequences determined for the 32 other bacterial strains.

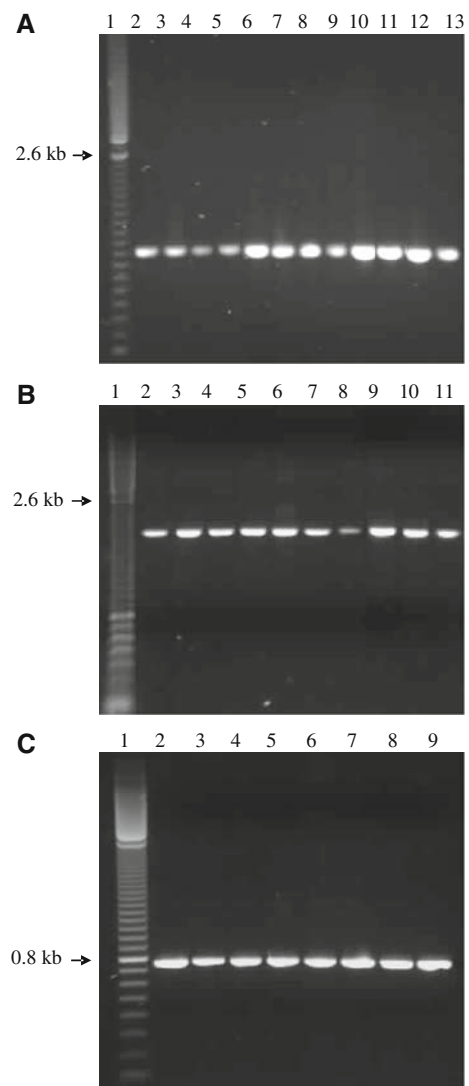
A single pair of primers, 16S-F1/R1 (Fig 1a) proved sufficient to amplify the 16S rRNA gene from all 32 bacterial strains. All strains yielded a single amplicon of 1.5 kbp (later determined at 1,461 bp) in length, primers not included. Likewise, a first pair of primers, *gyrB*-F1/R1 (Fig 1b), proved sufficient to amplify the *gyrB* gene from all 32 bacterial strains. Here also, all strains yielded a single amplicon of 1.9 kbp (1,881 bp) in length, primers not included. To facilitate the determination of the nucleotide sequence of the amplicon central region, a second

**Fig. 2** Agarose gel electrophoresis of 16S rRNA (a), *gyrB* (b) and *aroE* (c) gene amplification products in selected strains using specific primers. **a** Amplification products using the 16S-F1/-R1 primer pair. Lanes: 1, molecular weight marker, 100-bp DNA ladder; 2, *Bacillus thuringiensis* serovar *tolworthi*; 3, *B. thuringiensis* serovar *tohokuensis*; 4, *B. thuringiensis* serovar *vazensis*; 5, *B. thuringiensis* serovar *graciosensis*; 6, *B. thuringiensis* serovar *roskildiensis*; 7, *B. thuringiensis* serovar *polonensis*; 8, *B. mycooides* ATCC 6462; 9, *B. mycooides* BGSC 6A13; 10, *B. mycooides* BGSC 6A19; 11, *B. cereus* BGSC 6A17; 12, *B. weihenstephanensis* CCM 4965; 13, *B. weihenstephanensis* CCM 4966. **b** Amplification products using the *gyrB*-F1/-R1 primer pair. Lanes: 1, Molecular weight marker, 100-bp DNA ladder; 2, *Bacillus thuringiensis* serovar *tolworthi*; 3, *B. thuringiensis* serovar *tohokuensis*; 4, *B. thuringiensis* serovar *vazensis*; 5, *B. thuringiensis* serovar *roskildiensis*; 6, *B. thuringiensis* serovar *polonensis*; 7, *B. mycooides* ATCC 6462; 8, *B. mycooides* BGSC 6A13; 9, *B. cereus* BGSC 6A17; 10, *B. weihenstephanensis* CCM 4965; 11, *B. weihenstephanensis* CCM 4966. **c** Amplification products using the *aroE*-F1/-R1 primer pair. Lanes: 1, Molecular weight marker, 100-bp DNA ladder; 2, *Bacillus thuringiensis* serovar *tolworthi*; 3, *B. thuringiensis* serovar *tohokuensis*; 4, *B. thuringiensis* serovar *vazensis*; 5, *B. thuringiensis* serovar *roskildiensis*; 6, *B. thuringiensis* serovar *polonensis*; 7, *B. mycooides* ATCC 6462; 8, *B. weihenstephanensis* CCM 4965; 9, *B. weihenstephanensis* CCM 4966

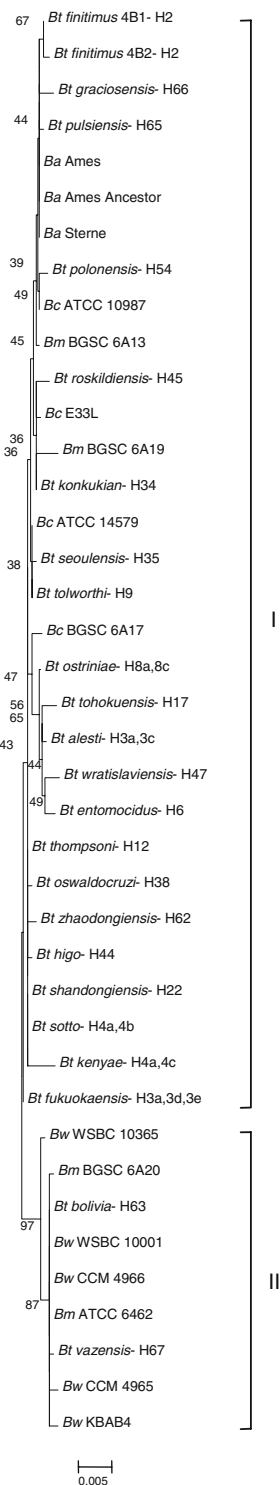
primer pair, *gyrB*-F2/-R2 (Fig 1b), was designed for the amplification of the central region. Finally, a single pair of primers, *aroE*-F1/-R1 (Fig 1c), proved sufficient to amplify the *aroE* gene from all 32 bacterial strains. Here also, all strains yielded a single amplicon of 0.7 kbp (714 bp) in length, primers not included. Subsets of the amplification products for the 16S rRNA, *gyrB* and *aroE* genes for selected bacterial strains are shown in Fig. 2a, b and c, respectively.

16S rRNA, *gyrB* and *aroE* genes nucleotide sequences-inferred phylogenetic analyses

Three bootstrapped neighbor-joining trees were generated from the alignment of the nucleotide sequences of the 16S rRNA, *gyrB* and *aroE* gene amplicons. They are presented in Figs. 3–5, respectively. Figure 3 shows the phylogenetic relationships between the 16S rRNA gene nucleotide sequences for the 40 bacterial species and strains shown in Table 1. Two clusters, I and II, were revealed at the 99% nucleotide sequence identities. Species-wise, the three *B. anthracis* strains are grouped together in cluster I and appear undistinguishable from each other. The four *B. cereus* strains are in cluster I. Two



of the four *B. mycooides* strains, BGSC 6A13 and BGSC 6A19, are located in cluster I whereas the other two, BGSC 6A20 and ATCC 6462, are located in cluster II. The five *B. weihenstephanensis* strains are found in cluster II. All *B. thuringiensis* H serotypes are found in cluster I, with the exceptions of H63 (serovar *bolivia*) and H67 (serovar *vazensis*) which are found in cluster II alongside the *B. weihenstephanensis* strains. Two H serotypes, H3 and H4, contained two serovars each. The two H3 serovars, *fukuokaensis* and *alesti*, and the two H4 serovars, *kenyae* and *sotto*, are all found in cluster I, on different locations. Not only can *B. thuringiensis* H serotypes be distinguished based on 16S rRNA gene nucleotide sequences, but the same is true for



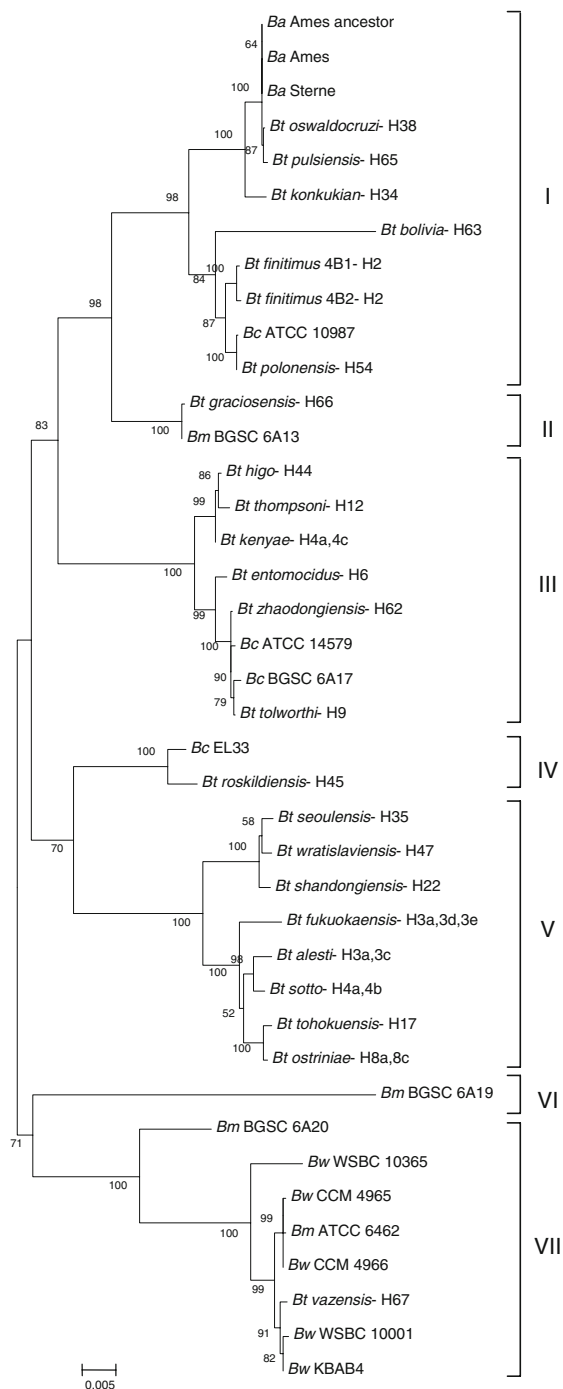
**Fig. 3** Bootstrapped neighbor-joining tree of *Bacillus thuringiensis* (*Bt*), *B. cereus*, *B. mycooides* and *B. weihenstephanensis* generated from the alignment of 16S rRNA nucleotide sequences. Bootstrap values higher than 30% are indicated. Roman numerals, I and II, refer to the two clusters revealed at the 99% nucleotide sequence identities. The horizontal bar represents 0.005% differences in nucleotide identities. Abbreviations: *Bt*: *Bacillus thuringiensis*; *Bc*: *B. cereus*; *Bm*: *B. mycooides*; *Bw*: *B. weihenstephanensis*

strains BGSC 4B1 and BGSC 4B2) are, as expected, found next to each other in cluster I, yet different enough as to be distinguishable.

Figure 4 shows the phylogenetic relationships between the *gyrB* gene nucleotide sequences for the same 40 bacterial species and strains (Table 1). Seven clusters, I–VII, were revealed at the 95% nucleotide sequence identities. Species-wise, here also the three *B. anthracis* strains are grouped together in cluster I and appear undistinguishable from each other. The four *B. cereus* strains are found in three clusters. Strains ATCC 14579 and BGSC 6A17 are grouped together in cluster III while the other two, ATCC 10987 and E33L, are located in clusters I and IV, respectively. The four *B. mycooides* strains are found in three clusters. Strains BGSC 6A20 and ATCC 6462 are grouped together in cluster VII while the other two, BGSC 6A13 and BGSC 6A19, are located in cluster II and VI, respectively. The five *B. weihenstephanensis* strains are found in cluster VII. All *B. thuringiensis* H serotypes are distributed in clusters I–V with the exception of *B. thuringiensis* H67 (serovar *vazensis*) which is found in cluster VII alongside the *B. weihenstephanensis* strains. Cluster VI contains no *B. thuringiensis* H serotype. The two H3 serovars, *fukuokaensis* and *alesti*, are both found in cluster V. The two H4 serovars, *kenyae* and *sotto*, are found in different clusters, III and V, respectively. Here also, *B. thuringiensis* H serotypes can be distinguished based on *gyrB* gene nucleotide sequences. This is true also for serovars from the same H serotype and for strains from the same serovar as exemplified here again by both *B. thuringiensis* serovar *finitimus* strains found next to each other, in cluster I, yet different enough as to be distinguishable.

both pairs of serovars from the same H serotype studied here. Interestingly enough, the two strains from the same serovar studied here (serovar *finitimus*,

Figure 5 shows the phylogenetic relationships between the *aroE* gene nucleotide sequences for the same 40 bacterial species and strains (Table 1). Five clusters, I–V, were revealed at the 95% nucleotide



**Fig. 4** Bootstrapped neighbor-joining tree of *Bacillus thuringiensis* (*Bt*), *B. cereus*, *B. mycooides* and *B. weihenstephanensis* generated from the alignment of *gyrB* nucleotide sequences. Bootstrap values higher than 30% are indicated. Roman numerals, I–VII, refer to the seven clusters revealed at the 95% nucleotide sequence identities. The horizontal bar represents 0.005% differences in nucleotide identities. Abbreviations are as indicated in Fig. 3

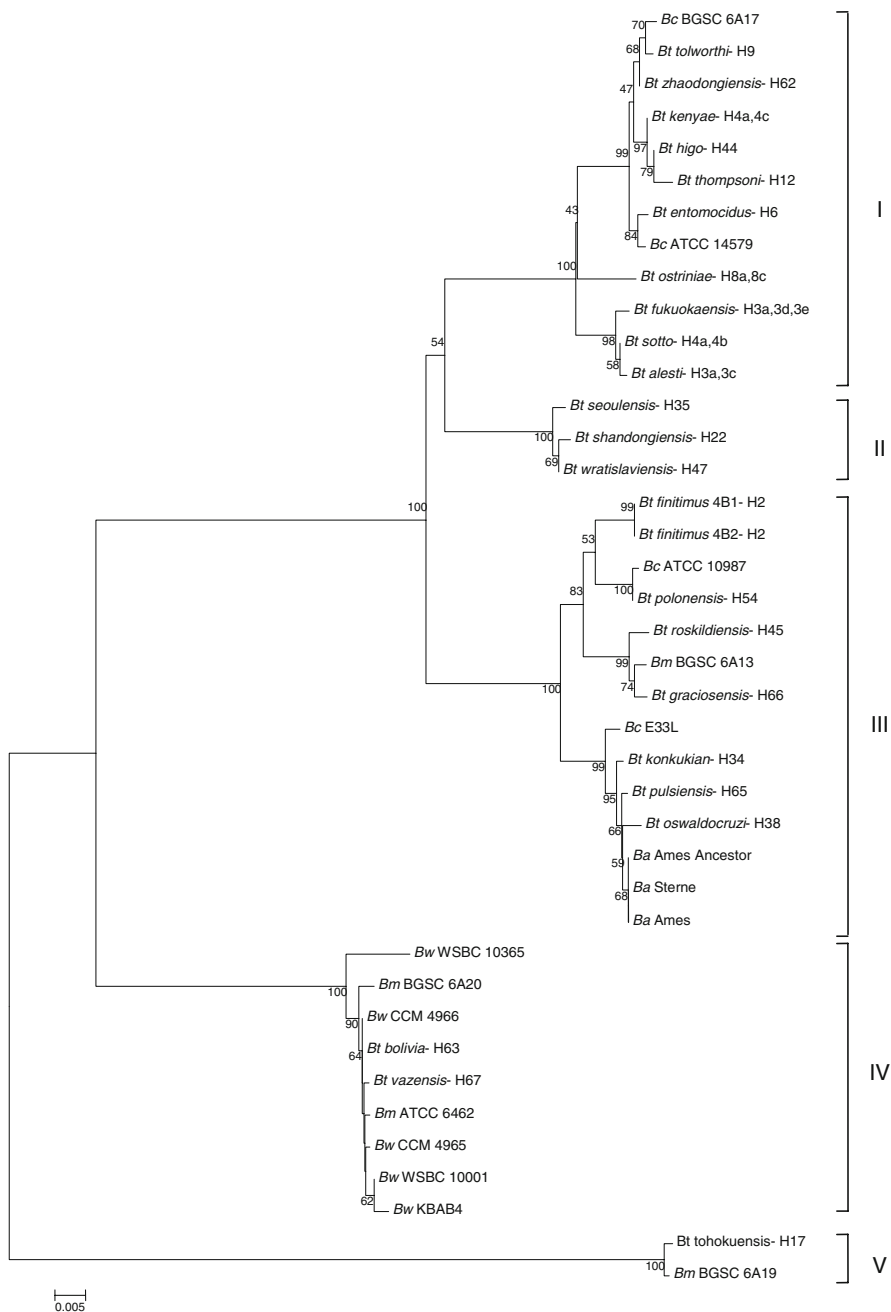
sequence identities. Species-wise, here also, the three *B. anthracis* strains are grouped together in cluster III and appear undistinguishable from each other. The four *B. cereus* strains are found in two clusters. Strains BGSC 6A17 and ATCC 14579 are located in cluster I whereas the other two, ATCC 10987 and E33L are located in cluster III. The four *B. mycooides* strains are found in three different clusters. Strains BGSC 6A13, BGSC 6A20, ATCC 6462 and BGSC 6A19 are found in clusters III, IV, IV and V, respectively. The five *B. weihenstephanensis* strains are found in cluster IV. The *B. thuringiensis* H serotypes are distributed in each of the five clusters. Interestingly, here also, *B. thuringiensis* H67 (serovar *vazensis*) is found alongside the *B. weihenstephanensis* strains. *B. thuringiensis* H63 (serovar *bolivia*) is also found in this same cluster. The two H3 serovars, *fukuokaensis* and *alesti*, are found in cluster I close to each other. The two H4 serovars, *kenya* and *sotto*, are found on separate locations in cluster I. Here also, *B. thuringiensis* H serotypes can be distinguished based on *aroE* gene nucleotide sequences. This is true also for serovars from the same H serotype. Interestingly enough, however, strains from the same serovar, here the two *B. thuringiensis* serovar *finitimus* strains, are found next to each other in cluster III and appear undistinguishable.

#### Percentage sequence similarity among *B. thuringiensis* H serotypes

Pairwise 16S rRNA, *gyrB* and *aroE* gene nucleotide sequence comparisons among the 21 *B. thuringiensis* H serotypes listed in Table 1 revealed respective percentage sequence similarities (supplementary data). The discriminatory power (DP) of the 16S rRNA gene was compared with the DP of the *gyrB* and *aroE* genes. The 16S rRNA gene showed the highest median value of percentage sequence similarity of 99%. The *gyrB* and *aroE* genes both showed lower median value of percentage sequence similarities of 91% (Table 2). The *gyrB* and *aroE* genes showed a median percentage of nucleotide sequence divergence among *B. thuringiensis* H serotypes up to nine fold greater than the one of the 16S rRNA gene. Clearly, both genes can be used to achieve deeper levels of discrimination among *B. thuringiensis* H serotypes than the 16S rRNA gene. The *gyrB* gene, however, present some advantages over the *aroE*



**Fig. 5** Bootstrapped neighbor-joining tree of *Bacillus thuringiensis* (*Bt*), *B. cereus*, *B. mycoides* and *B. weihenstephanensis* generated from the alignment of *aroE* nucleotide sequences. Bootstrap values higher than 30% are indicated. Roman numerals, I–V, refer to the five clusters revealed at the 95% nucleotide sequence identities. The horizontal bar represents 0.005% differences in nucleotide identities. Abbreviations are as indicated in Fig. 3



gene as a gene of choice for discriminating among *B. thuringiensis* lower classification ranks, serovars from same H serotype and strains from same serovar. Here, the *gyrB* gene nucleotide sequence could distinguish strains from same serovar as exemplified with the two *B. thuringiensis* serovar *finitimus* strains under study. The *aroE* gene nucleotide sequence

could not. But, perhaps more importantly, the *gyrB* gene is increasingly used as a macromolecule of choice for bacterial phylogenetic analysis and a database of *gyrB* sequences was implemented a few years ago (Watanabe et al. 2001). The latter will prove invaluable as more *B. thuringiensis* sequences are added.

**Table 2** Percentage nucleotide sequence similarity between 16S rRNA, *gyrB* and *aroE* genes among *Bacillus thuringiensis* H serotypes

Genes	Percentage nucleotide sequence similarity		
	16S rRNA	<i>gyrB</i>	<i>aroE</i>
Median value	99	91	91
Average	99	92	91
Lowest value	98	87	77
Highest value	100	100	100

*gyrB* gene polymorphism among serovars from same H serotype and among strains from same serovar

The capability of the *gyrB* gene nucleotide sequences to distinguish *B. thuringiensis* serovars from the same H serotype and strains from the same serovar was further studied. Table 3 lists the 42 *B. thuringiensis* strains used in this analysis. This includes eight *B. thuringiensis* strains from Table 1 and 34 additional ones. Together, they encompass 12 H serotypes and 25 serovars. Amplifications were carried out as described before using both pairs of primers, *gyrB*-F1/-R1 and *gyrB*-F2/-R2. Amplicon sizes were as described previously.

*gyrB* nucleotide sequences-inferred phylogenetic analyses

Figure 6 shows the phylogenetic relationships between the *gyrB* gene nucleotide sequences for the 42 *B. thuringiensis* strains shown in Table 3. Four clusters, I–IV, were revealed at the 96% nucleotide sequence identities.

Although some serovars from the same H serotype are found in the same clusters, this is not necessarily true for all serovars from the same H serotype. Indeed, serovars from the same H serotype do not necessarily cluster together. For example, H5 serovars *galleriae* and *canadensis* are both found in cluster I. H8 serovars *ostrinae* and *morrisoni* are in cluster IV. H11 serovars *toumanoffi* and *kyushuensis* are in cluster IV. H18 serovars *yosoo* and *kumamotoensis* are in cluster I. For H3, however, whereas serovar *kurstaki* is found in cluster I, serovars *alesti*, *sumiyoshiensis* and *fukuokaensis* are found in cluster IV. Likewise, H4 serovars *kenyae* and *sotto* are found

in clusters I and IV, respectively. H10 serovars *londrina* and *darmstadiensis* are found in clusters I and IV. H20 serovars *yunnanensis* and *pondicheriensis* are found in clusters I and II. H24 serovars *novosibirsk* and *neoleonensis* are found in clusters III and IV. H28 serovars *jegathesan* and *monterrey* are found in clusters I and II. The *gyrB* gene nucleotide sequences could distinguish *B. thuringiensis* serovars from the same H serotype. No *gyrB* gene nucleotide sequences from *B. thuringiensis* serovars from the same H serotype were identical.

Intra-serovar strains, i.e. strains from the same serovar, are found in the same cluster in close proximity. This is true for strains in the *kurstaki*, *galleriae*, *kenyae* and *entomocidus* serovars in cluster I, *finitimus* in cluster II, and *alesti*, *sotto*, *morrisoni* in cluster IV. Although strains from the same serovar were found in the same cluster, none was undistinguishable. The *gyrB* gene nucleotide sequence was sufficient to distinguish each intra-serovar strain. This is also true for the pathovars and biovars studied here as exemplified by *B. thuringiensis* serovar *morrisoni* pathovars *tenebrionis* and *sandiego*, by *B. thuringiensis* serovar *entomocidus* and its biovar *subtoxicus*, and by *B. thuringiensis* serovar *sotto* and its biovar *dendrolimus*.

## Discussion

Our aim, at the onset of this work, was to investigate the capability of each of the three genes under study, 16S rRNA, *gyrB* and *aroE*, to discriminate among *B. thuringiensis* H serotypes, serovars from the same H serotype and strains from the same serovar. First, these genes were analyzed in 21 *B. thuringiensis* H serotypes and four other *B. cereus* sensu lato species (*B. cereus*, *B. anthracis*, *B. mycoides* and *B. weihenstephanensis*). Nucleotide sequences were pairwise compared and phylogenetic relationships were inferred. The groups were analyzed at the species level. Each of the three genes grouped all three *B. anthracis* together. All three genes also grouped all five *B. weihenstephanensis* strains together. Perhaps surprisingly, the four *B. mycoides* were found in different clusters suggesting a higher level of heterogeneity. The 16S rRNA, *gyrB* and *aroE* gene nucleotide sequences grouped the four *B. cereus* in a single or in three or two different clusters,

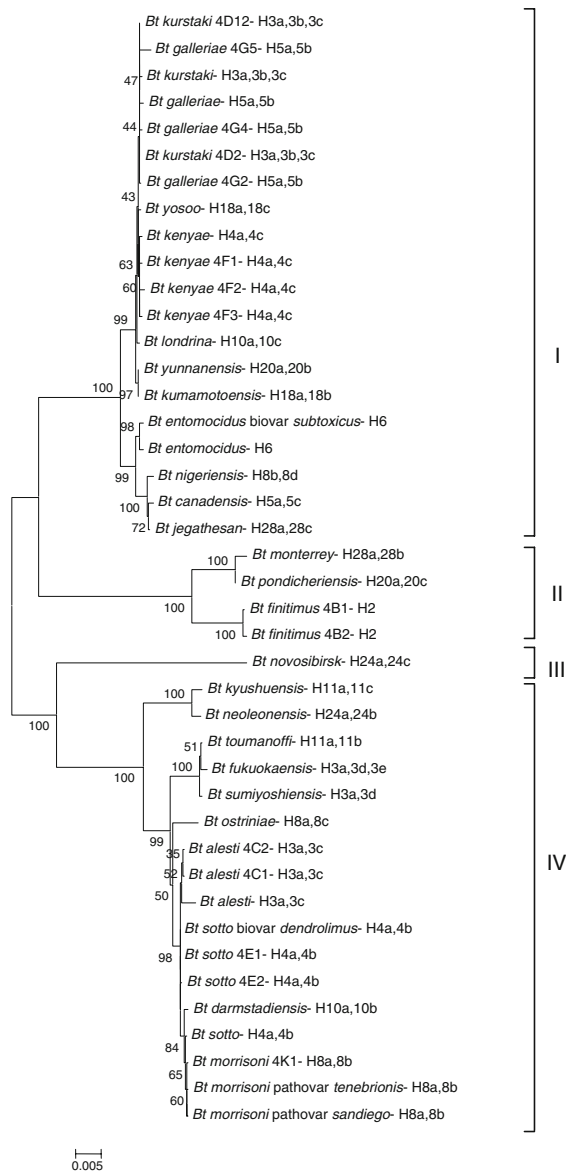
**Table 3** *Bacillus thuringiensis* serovars and strains used in Fig. 6

Bt serovars	H antigen <sup>b</sup>	Source-strain	Country of origin	GenBank accession no.
<i>finitimus</i>	H2	BGSC-4B1 <sup>c</sup>	USA	EF210248
<i>finitimus</i>	H2	BGSC-4B2 <sup>c</sup>	USA	EF210249
<i>kurstaki</i>	H3a,3b,3c	IEBC-T03A001	France	EU761187
<i>kurstaki</i>	H3a,3b,3c	BGSC-4D2	unknown	EU761176
<i>kurstaki</i>	H3a,3b,3c	BGSC-4D12	England	EU761177
<i>alesti</i>	H3a,3c	IEBC-T03 001 <sup>c</sup>	France	EF210264
<i>alesti</i>	H3a,3c	BGSC-4C1	Former Czechoslovakia	EU761175
<i>alesti</i>	H3a,3c	BGSC-4C2	France	EU761164
<i>sumiyoshiensis</i>	H3a,3d	IEBC-T03B001	Japan	EU761173
<i>fukuokaensis</i>	H3a,3d,3e	IEBC-T03C001 <sup>c</sup>	Japan	EF210279
<i>sotto</i>	H4a,4b	IEBC-T04 001 <sup>c</sup>	Canada	EF210259
<i>sotto</i>	H4a,4b	BGSC-4E1	USA	EU761166
<i>sotto</i>	H4a,4b	BGSC-4E2	France	EU761181
<i>sotto biovar dendrolimus</i>	H4a,4b	IEBC-T04A001	France	EU761186
<i>kenyae</i>	H4a,4c	IEBC-T04B001 <sup>c</sup>	Kenya	EF210257
<i>kenyae</i>	H4a,4c	BGSC-4F1	England	EU761165
<i>kenyae</i>	H4a,4c	BGSC-4F2	Kenya	EU761178
<i>kenyae</i>	H4a,4c	BGSC-4F3	USA	EU761179
<i>galleriae</i>	H5a,5b	IEBC-T05 001	Former USSR	EU761193
<i>galleriae</i>	H5a,5b	BGSC-4G2	USA	EU761192
<i>galleriae</i>	H5a,5b	BGSC-4G4	England	EU761197
<i>galleriae</i>	H5a,5b	BGSC-4G5	Former Czechoslovakia	EU761184
<i>canadensis</i>	H5a,5c	IEBC-T05A001	Canada	EU761180
<i>entomocidus</i>	H6	IEBC-T06 001 <sup>c</sup>	Canada	EF210255
<i>entomocidus biovar subtoxicus</i>	H6	IEBC-T06A001	Canada	EU761172
<i>morrisoni</i>	H8a,8b	BGSC-4K1	USA	EU761196
<i>morrisoni</i> pathovar <i>tenebrionis</i>	H8a,8b	IEBC-T08 017	Germany	EU761189
<i>morrisoni</i> pathovar <i>sandiego</i> <sup>a</sup>	H8a,8b			EU761195
<i>ostrinae</i>	H8a,8c	IEBC-T08A001 <sup>c</sup>	China	EF210275
<i>nigeriensis</i>	H8b,8d	IEBC-T08B001	Former Czechoslovakia	EU761170
<i>darmstadiensis</i>	H10a,10b	IEBC-T10 001	Germany	EU761185
<i>londrina</i>	H10a,10c	IEBC-T10A001	Brazil	EU761169
<i>toumanoffi</i>	H11a,11b	IEBC-T11 001	Germany	EU761174
<i>kyushuensis</i>	H11a,11c	IEBC-T11A001	Japan	EU761167
<i>kumamotoensis</i>	H18a,18b	IEBC-T18 001	Japan	EU761194
<i>yosoo</i>	H18a,18c	IEBC-T18A001	South Korea	EU761183
<i>yunnanensis</i>	H20a,20b	IEBC-T20 001	China	EU761190
<i>pondicheriensis</i>	H20a,20c	IEBC-T20A001	India	EU761188
<i>neoleonensis</i>	H24a,24b	IEBC-T24 001	Mexico	EU761191
<i>novosibirsk</i>	H24a,24c	IEBC-T24A001	Former USSR	EU761171
<i>monterrey</i>	H28a,28b	IEBC-T28 001	Mexico	EU761168
<i>jegathesan</i>	H28a,28c	IEBC-T28A001	Malaysia	EU761182

<sup>a</sup> *B. thuringiensis* serovar *morrisoni* pathovar *sandiego* was purified from M-One™, a commercial formulation developed by Mycogen Corp

<sup>b</sup> IEBC, International Entomopathogenic *Bacillus* Centre, Institut Pasteur, Paris, France; BGSC, *Bacillus* Genetic Stock Center, The Ohio State University, Columbus, OH, USA

<sup>c</sup> *Bacillus thuringiensis* strains also present in Table 1



**Fig. 6** Bootstrapped neighbor-joining tree of *Bacillus thuringiensis* serovars and strains generated from the alignment of *gyrB* nucleotide sequences. Bootstrap values higher than 30% are indicated. Roman numerals, I–IV, refer to the four clusters revealed at the 96% nucleotide sequence identities. The horizontal bar represents 0.005% differences in nucleotide identities

respectively. The 16S rRNA gene grouped all *B. thuringiensis* strains in cluster I with two exceptions: serovars *bolivia* and *vazensis*, both found in cluster II along with the *B. weihenstephanensis* strains suggesting a closer relationship. The *gyrB* gene saw the *B. thuringiensis* strains distributed in

five clusters with serovar *vazensis* in a sixth cluster along with the *B. weihenstephanensis* strains. The *aroE* gene saw the *B. thuringiensis* strains distributed in all five clusters. Interestingly enough, here also, serovars *bolivia* and *vazensis* were found in cluster IV along with the *B. weihenstephanensis* strains adding support to the closer relationship between these two *B. thuringiensis* serovar with this *B. cereus* sensu lato species. On a side note, the presence of a parasporal inclusion body is the single most discriminating criterion for the identification of *B. thuringiensis* among *Bacillus*. Whether *B. thuringiensis* serovar *vazensis*, or serovar *bolivia*, might share common additional features with *B. weihenstephanensis* remains to be confirmed. We are currently pursuing studies of this type. Each of the three genes under study could discriminate among *B. thuringiensis* H serotypes and the discriminatory power of each was calculated. Both house-keeping genes, *gyrB* and *aroE*, showed a median percentage of nucleotide sequence divergence among *B. thuringiensis* H serotypes up to nine fold greater than the one of the 16S rRNA gene. The *gyrB* gene was retained for subsequent analyses: the discrimination of *B. thuringiensis* serovars from the same H serotype, and the discrimination of strains from the same serovar. A total of 42 *B. thuringiensis* strains which encompassed 25 serovars from 12 H serotypes were analyzed. The *gyrB* gene nucleotide sequences could distinguish *B. thuringiensis* serovars from the same H serotype. In addition, the *gyrB* gene nucleotide sequence was sufficient to distinguish *B. thuringiensis* strains from the same serovar. Clearly, whereas all three genes under study, 16S rRNA, *gyrB* and *aroE* could discriminate among *B. thuringiensis* H serotypes, the *gyrB* gene has proven capable of discriminating serovars from the same H serotype and strains from the same serovar.

## References

- Bavykin SG, Lysov YP, Zakhariev V, Kelly JJ, Jackman J, Stahl DA, Cherni A (2004) Use of 16S rRNA, 23S rRNA, and *gyrB* gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. *J Clin Microbiol* 42:3711–3730. doi:10.1128/JCM.42.8.3711-3730.2004 Erratum in: *J Clin Microbiol* (2006) 44:2676
- Claus D, Berkeley RCW (1986) Genus *Bacillus* Cohn 1872. In: Sneath PHA, Mair NS, Sharpe ME, Holt JG (eds)

- Bergey's manual of systematic bacteriology, vol 2. Williams and Wilkins, Baltimore, pp 1105–1139
- de Barjac H, Bonnefoi A (1962) Essai de classification biochimique et sérologique de 24 souches de *Bacillus* du type *B. thuringiensis*. *Entomophaga* 7:5–31. doi:[10.1007/BF02375988](https://doi.org/10.1007/BF02375988)
- Glare TR, O'Callaghan M (2000) *Bacillus thuringiensis*: biology, ecology and safety. Wiley, Toronto
- Höfte H, Whiteley HR (1989) Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol Rev* 53:242–255
- Ko KS, Kim J-W, Kim J-M, Kim W, Chung S-I, Kim IJ, Kook Y-H (2004) Population structure of the *Bacillus cereus* group as determined by sequence analysis of six house-keeping genes and the *plcR* Gene. *Infect Immun* 72:5253–5261. doi:[10.1128/IAI.72.9.5253-5261.2004](https://doi.org/10.1128/IAI.72.9.5253-5261.2004)
- Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163. doi:[10.1093/bib/5.2.150](https://doi.org/10.1093/bib/5.2.150)
- La Duc MT, Satomi M, Agata N, Venkateswaran K (2004) *gyrB* as a phylogenetic discriminator for members of the *Bacillus anthracis-cereus-thuringiensis* group. *J Microbiol Methods* 56:383–394. doi:[10.1016/j.jmimet.2003.11.004](https://doi.org/10.1016/j.jmimet.2003.11.004)
- Lecadet M-M, Frachon E, Cosmao Dumanoir V, Ripouteau H, Hamon S, Laurent P, Thiéry I (1999) Updating the H-antigen classification of *Bacillus thuringiensis*. *J Appl Microbiol* 86:660–672. doi:[10.1046/j.1365-2672.1999.00710.x](https://doi.org/10.1046/j.1365-2672.1999.00710.x)
- Maiden MCJ, Bygraves JA, Feil EJ, Morelli G, Russell JE, Urwin R, Zhang Q, Zurth K, Caugant D, Feavers IM, Achtman M, Spratt BG (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 95:3140–3145. doi:[10.1073/pnas.95.6.3140](https://doi.org/10.1073/pnas.95.6.3140)
- Park S-H, Kim H-J, Kim J-H, Kim T-W, Kim H-Y (2007) Simultaneous detection and identification of *Bacillus cereus* group bacteria using multiplex PCR. *J Microbiol Biotechnol* 17:1177–1182
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sanchis V, Chafaux J, Lereclus D (1996) Amélioration biotechnologique de *Bacillus thuringiensis*: les enjeux et les risques. *Ann Inst Pasteur/Actual* 7:271–284
- Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680. doi:[10.1093/nar/22.22.4673](https://doi.org/10.1093/nar/22.22.4673)
- Watanabe K, Nelson J, Harayama S, Kasai H (2001) ICB database: the *gyrB* database for identification and classification of bacteria. *Nucleic Acids Res* 29:344–345. doi:[10.1093/nar/29.1.344](https://doi.org/10.1093/nar/29.1.344)
- Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51:221–271
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 87:4576–4579. doi:[10.1073/pnas.87.12.4576](https://doi.org/10.1073/pnas.87.12.4576)
- Xu D, Côté J-C (2003) Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 3' end 16S rDNA and 5' end 16S–23S ITS nucleotide sequences. *Int J Syst Evol Microbiol* 53:695–704. doi:[10.1099/ijs.0.02346-0](https://doi.org/10.1099/ijs.0.02346-0)
- Xu D, Côté J-C (2006) Sequence diversity of the *Bacillus thuringiensis* and *B. cereus* sensu lato flagellin (H antigen) protein: comparison with H serotype diversity. *Appl Environ Microbiol* 72:4653–4662. doi:[10.1128/AEM.00328-06](https://doi.org/10.1128/AEM.00328-06)
- Yamamoto S, Harayama S (1995) PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl Environ Microbiol* 61:1104–1109