Bacillus wiedmannii biovar *thuringiensis*: A Specialized Mosquitocidal Pathogen with Plasmids from Diverse Origins

J. Nicolás Lazarte[1],[†], Rocio P. Lopez[1],[†], P. Daniel Ghiringhelli[2], and Corina M. Berón[1],*

[1]Instituto de Investigaciones en Biodiversidad y Biotecnología (INBIOTEC) – CONICET, FIBA, Mar del Plata, Argentina

[2]Laboratorio de Ingeniería Genética y Biología Celular y Molecular (LIGBCM), Area Virosis de Insectos (AVI), Departamento Ciencia y Tecnología, Universidad Nacional de Quilmes and CONICET, Bernal, Argentina

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: corina.beron@inbiotec-conicet.gob.ar.

Abstract

Bacillus cereus sensu lato also known as *B. cereus* group is composed of an ecologically diverse bacterial group with an increasing number of related species, some of which are medically or agriculturally important. Numerous efforts have been undertaken to allow presumptive differentiation of *B. cereus* group species from one another. FCC41 is a *Bacillus* sp. strain toxic against mosquito species like *Aedes aegypti*, *Aedes* (*Ochlerotatus*) *albifasciatus*, *Culex pipiens*, *Culex quinquefasciatus*, and *Culex apicinus*, some of them responsible for the transmission of vector-borne diseases. Here, we report the complete genome sequence of FCC41 strain, which consists of one circular chromosome and eight circular plasmids ranging in size from 8 to 490 kb. This strain harbors six crystal protein genes, including *cry24Ca*, two *cry4*-like and two *cry52*-like, a *cry41*-like parasporin gene and multiple virulence factors. The phylogenetic analysis of the whole-genome sequence of this strain with molecular approaches places this strain into the *Bacillus wiedmannii* cluster. However, according with phenotypical characteristics such as the mosquitocidal activity due to the presence of Cry proteins found in the parasporal body and *cry* genes encoded in plasmids of different sizes, indicate that this strain could be renamed as *B. wiedmannii* biovar *thuringiensis* strain FCC41.

Key words: *Bacillus cereus* group, whole-genome, *cry* genes, mosquitocidal strain.

Introduction

The current description of the genus *Bacillus* integrates a great number of species genetically extremely heterogeneous. In the last century, the species in the genus were redefined many times resulting in the reclassification of its constituent species (Bhandari et al. 2013). In particular, *Bacillus cereus sensu lato*, also known as the *B. cereus* group, was initially formed by few species and in the last years several new species had been included in this classification (Bazinet 2017). Among them, *Bacillus anthracis* is the etiological agent of anthrax and its major virulence factors are the anthrax toxins and an antiphagocytic polyglutamic capsule, both encoded by two large plasmids pXO1 and pXO2 (Moayeri et al. 2015). *Bacillus cereus sensu stricto*, or *B. cereus* as it is usually called, is ubiquitous in nature and an emerging human pathogen that causes gastroenteritis, considered the third most important cause of food intoxication incidents in Europe and an opportunistic human pathogen associated with other local and systemic infections. *Bacillus cereus* strains encode a range of toxins and other extracellular virulence factors on the chromosome, and it has been shown that the same genes are present and actively expressed in *Bacillus thuringiensis* strains (Maughan and Van der Auwera 2011).

Bacillus thuringiensis has been studied for its properties as an insect pathogen useful as microbial agent for insect pests in agriculture and insect vectors control. This bacterium produces a parasporal inclusion body composed primarily by the insecticidal proteins Cry and Cyt that are mainly encoded in megaplasmids. These kinds of inclusions or crystals must be

ingested by a susceptible insect and after some biochemical processes, disrupt the cell osmotic balance, then the midgut integrity, resulting in gut paralysis and the insects' death (Adang et al. 2014).

One of the most recently identified species of this bacterial group, *Bacillus wiedmannii*, produces hemolysin BL toxins and nonhemolytic enterotoxins and has been described as hydrolyze arginine negative being unable to produce acid from fermentation of sucrose and can be differentiated from the other *B. cereus* species by sequence comparison analysis (Miller et al. 2016).

In the past, species of *B. cereus* group were classified based on their habitats, morphological or physiological characteristics, *16S rRNA* gene sequences and characteristics such as the presence or absence of virulence plasmids that confer pathogenicity against mammals or insects. Their classification according with comparative genomics is controversial and shows incongruent phylogenies (Rasko et al. 2005; Liu et al. 2015). In order to demonstrate diverse phylogenetic hypothesis of *B. cereus* group, different molecular approaches have been applied (Guinebretière et al. 2008; Kim et al. 2014; Miller et al. 2016; Bazinet 2017). The use of multiple molecular methods has shown that the species boundaries between members of this bacterial group are difficult to define, forcing to rethink the current descriptions of some *Bacillus* species (Maughan and Van der Auwera 2011).

In a previous work, we characterized an Argentinian strain named as *B. thuringiensis* FCC41 that exhibited higher insecticidal activity against some mosquito species than *B. thuringiensis* subsp. *israelensis*, the reference strain used worldwide for vector control. This isolate produces a rounded crystal harboring two major proteins of ∼70–80 kDa and until recently, we could only sequence the complete encoding gene of two of the crystal proteins (Cry), Cry24Ca (Berón and Salerno 2007) and one Cry52-like protein, even though we suspected the presence of more Cry proteins. To determine other toxins and virulence factors that may play a role in the activity of this strain and to understand fully this highly important toxic activity against mosquitos, we have therefore determined its complete genome sequence obtained by next-generation sequencing technologies as presented here. Additionally, we discuss the reclassification of FCC41 strain according with phenotypical and genotypical characteristics as well as the phylogenetic analysis with molecular approaches.

Materials and Methods

Strain and Culture Condition

The strain FCC41 was originally isolated from soil samples (Berón and Salerno 2006) and cultured in nutrient broth medium (Difco) at 28 °C for 16 h under constant shaking at 180 rpm. In order to evaluate the psychrophilic nature of this strain, 100 μl of an overnight growth pre-inoculum were plated on nutrient broth agar plates and incubated at 5, 7, 10, and 12 °C during 72 h.

Genome Sequencing and Assembly

Bacterial DNA was isolated using the GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific Inc.). Whole-genome sequencing was performed using Illumina 1500 HiSeq (INDEAR, Rosario, Argentina) and PacBio RSII (Macrogen Inc., Korea) platforms. The Illumina sequencing was conducted with 100 bp paired-end reads from a Nextera XT library (broad size between 150 and 1,500 bp), yielding 2.2 G of sequenced bases. The PacBio sequencing was run in one single-molecule real-time cell (SMRT) with a 20 kb library, which generated 80,337 reads (average read length 10,771 bp) and 865,518,541 read bases. The complete genome was obtained by assembling short and long reads through a hybrid assembly approach using Unicycler software (Wick et al. 2017). Sequencing coverage was calculated with SAMtools (Li et al. 2009) by aligning all reads to the final assembly using Bowtie 2 (Langmead and Salzberg 2012) and minimap (Li 2016). In order to manually verify the final assembly these alignments were subsequently visualized with Integrative Genomics Viewer (Thorvaldsdóttir et al. 2013).

Genome Annotation

Genome annotation was performed with the NCBI Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/) and Prokka (Seemann 2014). Clusters of Orthologous Groups (COG) annotations were computed using WebMGA server (Wu et al. 2011).

Phylogenetic Analysis

The sequences of *panC*, *rpoB* and the seven housekeeping genes of the MLST scheme proposed by Priest et al. (2004; http://www.pubmlst.org/bcereus; Last accessed April 2018), were extracted of the whole-genome sequence of FCC41 strain and analyzed with BTyper software, which compares the sequences of the genes extracted to alleles present in specific online databases (Carroll et al. 2017). Single genes *panC* and *rpoB* sequence comparisons were performed using BLAST with NCBI nonredundant database and the Food Microbe Tracker database (http://www.foodmicrobetracker.com; Last accessed April 2018). For the MLST based phylogeny, as well as for the core genome and Average Nucleotide Identity (ANI) based phylogeny, we selected 35 *B. cereus sensu lato* genomes according to their distribution and representation along the phylogenetic tree previously reported by Bazinet (2017; see supplementary table S1, Supplementary Material online). These genomes were downloaded from NCBI RefSeq, and the sequence of the seven housekeeping genes of all genomes, including FCC41 strain, were extracted and concatenated, and the resulting sequences were aligned

Table 1

Genome Features of Strain FCC41

Features	Size (bp)	G + C Content (%)	No. of Genes	No. of tRNAs	No. of rRNAs	No. of *cry*-like Protein Genes	Accesion Number	Illumina Sequencing Coverage	PacBio Sequencing Coverage
Chromosome	55,13,449	36.09	5,893	108	39	–	CP024684	209.8	76
pFCC41-1-490k	4,90,693	32.65	467	1	–	–	CP024685	306.9	100.5
pFCC41-2-313k	3,13,432	32.46	287	–	–	1	CP024686	307.4	100.8
pFCC41-3-257k	2,57,021	34.49	259	–	–	4	CP024687	607.3	199.1
pFCC41-4-144k	1,44,434	33.82	143	–	–	–	CP024688	533.3	168.4
pFCC41-5-125k	1,24,955	34.10	134	–	–	1	CP024689	727.3	188.4
pFCC41-6-12k	11,912	31.51	18	–	–	–	CP024690	309.5	5.6
pFCC41-7-9k	9,217	35.20	13	–	–	–	CP024691	947.6	46.8
pFCC41-8-8k	8,368	29.36	8	–	–	–	CP024692	732.1	19.2
Total	68,73,481	35.05	7,222	109	39	6			

using ClustalO (Sievers and Higgins 2014). With this alignment we calculated the phylogenetic tree with RAxML performing 500 bootstraps (Stamatakis 2014).

For the core genome-based phylogeny, the bcgTree pipeline automatically extracts 107 essential housekeeping genes found in most bacteria using hidden Markov models, and then performs a partitioned maximum-likelihood analysis using the multiple alignment of the concatenated sequences (Ankenbrand and Keller 2016). The phylogenetic tree was constructed with RAxML (Stamatakis 2014) under a maximum likelihood model. To evaluate node support 100 bootstrap replicates were computed and the consensus tree visualized by FigTree 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

The selected genomes were analyzed by ANI using PyANI (https://github.com/widdowquinn/pyani) in ANIm mode, which uses MUMmer aligner (Kurtz et al. 2004).

Sequences of Interest Overview

Virulence factors and antibiotic resistance genes were predicted using BLASTP against the data set of Virulence Factor Database (VFDB; http://www.mgc.ac.cn/VFs/; Last accessed April 2018) and the Comprehensive Antibiotic Resistance Database (CARD; https://card.mcmaster.ca/; Last accessed April 2018), respectively.

Cry protein coding DNA sequences (CDS) were detected using Hidden Markov Modeler (HMMER, Eddy 1998). For that purpose, we constructed a HMMER profile with an alignment of all known deduced amino acid sequences of the Cry haplotype proteins (https://www.btnomenclature.info; Last accessed April 2018) and used it to search through the translated CDS obtained in the automated annotation step. To confirm the results, the positive matches were compared with known Cry proteins with a BLASTP search (Altschul et al. 1990).

Plasmid Analysis

Nucleotide sequences of plasmids of FCC41 strain were used to search for similar and complete plasmids DNA sequences in NCBI nonredundant nucleotide database, these plasmids were downloaded and used to analyze regions of shared homology between FCC41 strain plasmids and then were compared with BLAST Ring Image Generator (BRIG) and EasyFig software (Sullivan et al. 2011). On the other hand, in order to identify plasmid minireplicons from FCC41 strain BLASTP analysis was performed against known replication proteins as previously described (Zheng et al. 2013).

Results and Discussion

Whole-Genome Sequencing and Assembly

To characterize the mosquitocidal FCC41 strain and to find all putative Cry protein coding genes, its whole-genome was sequenced and analyzed. The complete genome was obtained by two different next-generation sequencing methodologies: Illumina HiSeq providing high fidelity, accurate short reads and higher coverage, resulting in an assembly with multiple gaps; and, PacBio SMRT providing longer but lesser precise reads. By combing both reads with the Unicycler assembly pipeline (Wick et al. 2017) the complete ungapped genome was obtained. The final assembly is comprised of one circular chromosome and eight circular plasmids, with a total genome size of 6,873,481 bp as described in table 1 and visualized in figure 1. This complete genome has been deposited in INSDC (International Nucleotide Sequence Database Collaboration: www.insdc.org) under the GenBank/EMBL/DDBJ accession numbers CP024684 to CP024692.

Phylogenetic Analysis

It has been discussed that the *B. cereus* group should be considered to form a single species with different ecotypes and pathotypes and according with that, the main phenotypical properties originally used to distinguish *B. cereus sensu stricto*, *B. thuringiensis* and *B. anthracis* were related to the presence or absence of large plasmids, which carry the genetic
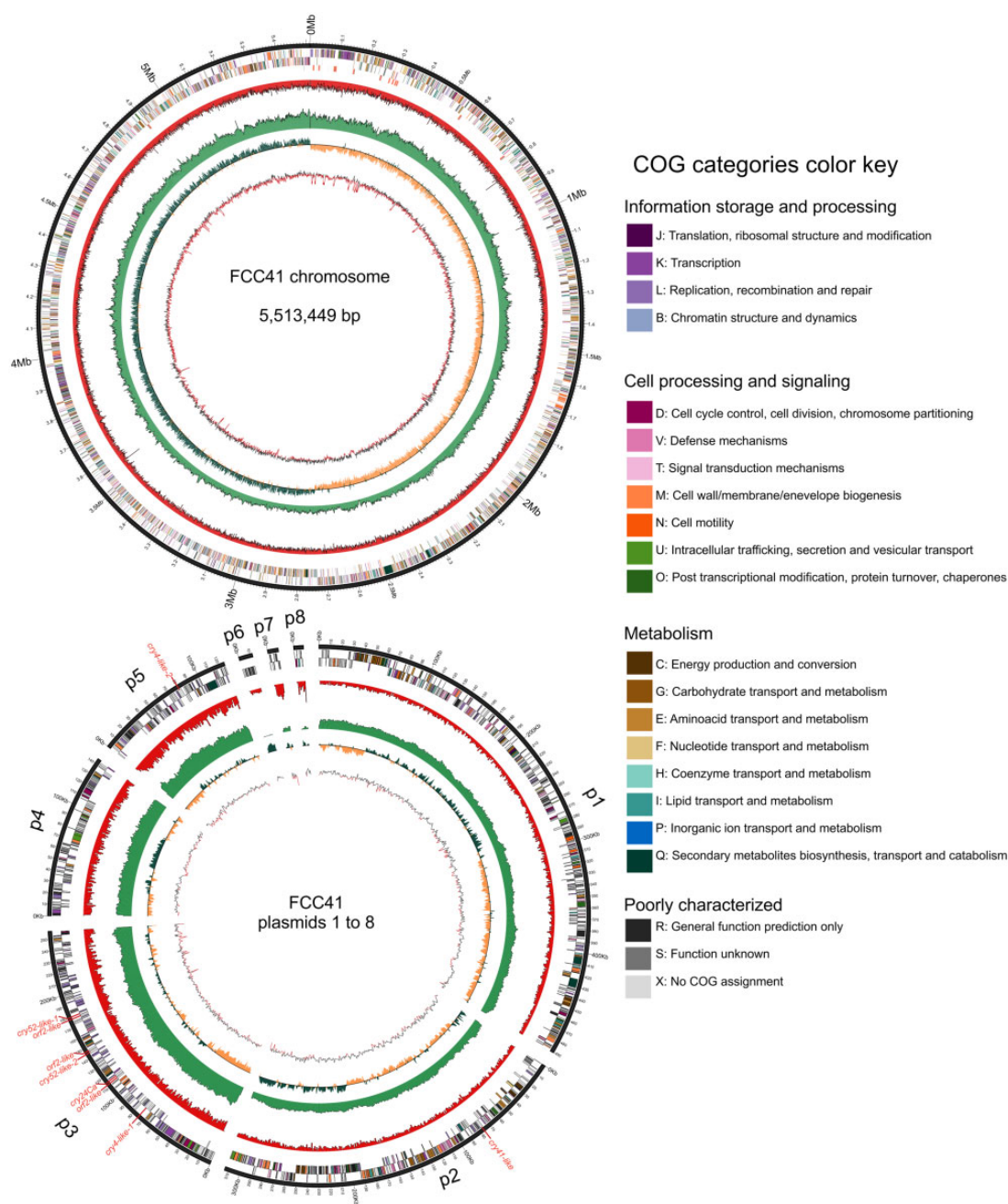
COG categories color key

Information storage and processing

- J: Translation, ribosomal structure and modification
- K: Transcription
- L: Replication, recombination and repair
- B: Chromatin structure and dynamics

Cell processing and signaling

- D: Cell cycle control, cell division, chromosome partitioning
- V: Defense mechanisms
- T: Signal transduction mechanisms
- M: Cell wall/membrane/enevelope biogenesis
- N: Cell motility
- U: Intracellular trafficking, secretion and vesicular transport
- O: Post transcriptional modification, protein turnover, chaperones

Metabolism

- C: Energy production and conversion
- G: Carbohydrate transport and metabolism
- E: Aminoacid transport and metabolism
- F: Nucleotide transport and metabolism
- H: Coenzyme transport and metabolism
- I: Lipid transport and metabolism
- P: Inorganic ion transport and metabolism
- Q: Secondary metabolites biosynthesis, transport and catabolism

Poorly characterized

- R: General function prediction only
- S: Function unknown
- X: No COG assignment

FIG. 1.—Genome map of FCC41 strain chromosome and plasmids. Rings from the outermost to the center: 1) scale marks, 2) protein-coding genes on the forward strand, 3) protein-coding genes on the reverse strand, 4) tRNA (black) and rRNA (red), 5) Illumina sequencing coverage, 6) PacBio sequencing coverage, 7) GC skew, 8) GC content. Protein-coding genes are color coded according to their COG categories. *cry*-like coding genes are indicated by red lines. The genome maps were drawn using the software Circos v0.69 (Krzywinski et al. 2009).

determinants that are responsible for those phenotypes. Despite these differences, the *16S rRNA* sequences of most members of *B. cereus* group have a homology >98%, providing insufficient resolution at the species level (Cohan 2002; Liu et al. 2015; Wang and Ash 2015). In addition to ribosomal genes, *panC* and *rpoB* genes are currently being used to differentiate among the most important *Bacillus* spp. as a more

accurate system for phylogenetic analysis (Caamaño-Antelo et al. 2015; Kovac et al. 2016). In particular, the *panC* sequence comparison analysis performed with BTyper reveals that the FCC41 strain belongs to Group II of the phylogeny proposed by Guinebretière et al. (2008) since it has 99% similarity to several *B. wiedmannii* strains, including strain FSL W8-0169. This group is comprised by *B. wiedmannii*,
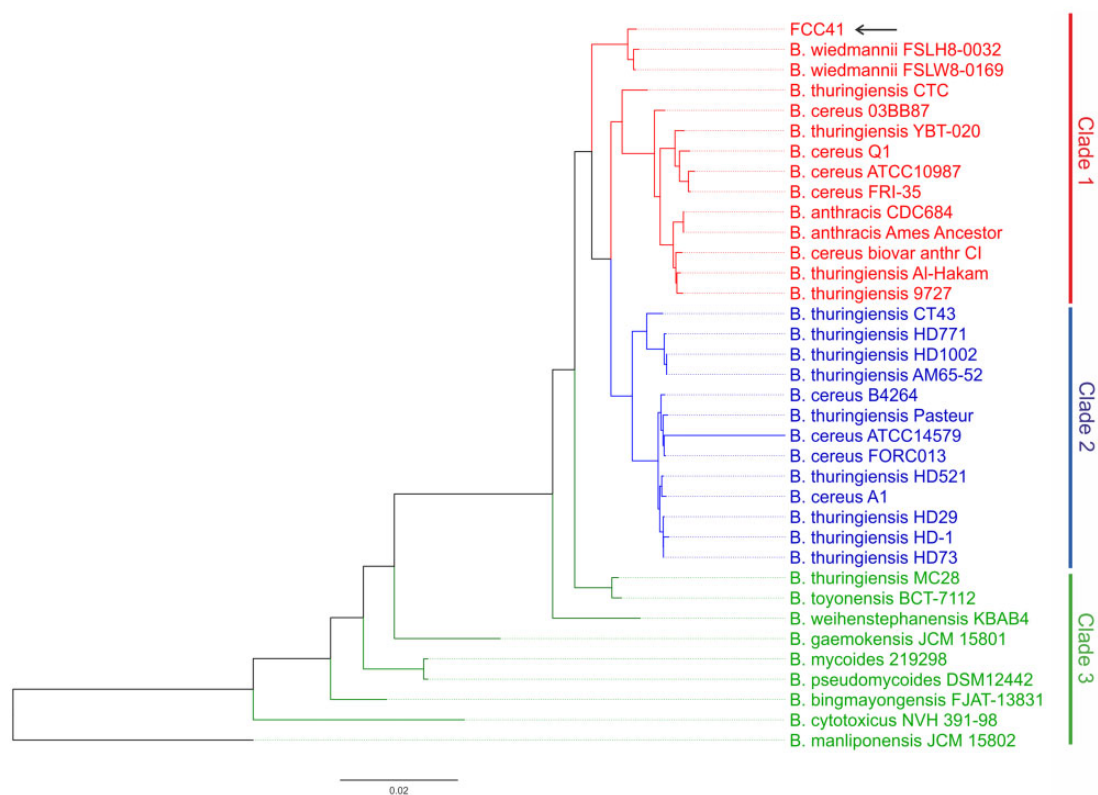
Fig. 2.—Phylogenetic analysis of *B. cereus sensu lato* strains. The phylogenetic tree was obtained using bacterial core genome tree (bcgTree) pipeline. *Bacillus manliponensis* JCM 15802 was used to root the tree and FCC41 is indicated by an arrow.

and some strains of *B. cereus* and *B. thuringiensis*; and most of its members are psychrotolerant and able to grow at 7 °C, since they carry one or more copies of the cold shock protein A gene (*cspA*). FCC41 strain has seven copies of *cspA* and the culture's growth was detected at temperatures of 10 °C, or higher. On the other hand, BTyper tool showed that *rpoB* allelic type (AT) of this strain has 100% identity to AT 410, which is associated to 12 *B. cereus sensu lato* strains deposited in the Food Microbe Tracker database (http://www.foodmicrobetracker.com; Last accessed April 2018.); additionally, there are 115 identical matches of the *rpoB* FCC41 sequence in NCBI nonredundant database, 94 of them corresponding to recently sequenced *B. wiedmannii* strains.

Moreover, different MLST schemes of housekeeping genes have been used in several works in order to make a greater effort in the clarification of *B. cereus* group phylogeny (Ko et al. 2004; Sorokin et al. 2006; Tourasse and Kolstø 2008). To compare FCC41 strain using a MLST approach, seven housekeeping genes from the *B. cereus* PubMLST typing scheme were used. Although the AT of the seven housekeeping genes already existed in the database, the combination of this seven AT represents a novel sequence type (ST). The resulting phylogenetic tree indicated that strain FCC41 is placed in the same group as other *B. wiedmannii* strains (see supplementary fig. S1, Supplementary Material online).

Additionally, the core genome phylogenetic analysis was performed with bcgTree which shows that strain FCC41 and the others *B. wiedmannii* strains are grouped together and close to Clade I in a similar way as described in Bazinet (2017; fig. 2).

ANI has been used to compare prokaryotic genome sequences and is recognized as one of the most robust measurements of relatedness between strains. An ANI value of 95% is accepted as the cutoff value for the delimitation of bacterial species. The ANI analysis value of 96.4% between FCC41 and *B. wiedmannii* FSL W8-0169 indicates that this strain belongs to the recently described species (fig. 3; see supplementary table S2, Supplementary Material online), and the results were highly related to the grouping obtained with bcgTree and previously reported phylogenetic trees (Bazinet 2017).

Virulence Factors and Antibiotic Resistance Genes

The virulence factors predicted by the BLASTP analysis of FCC41 strain reveals proteins that are widely distributed in *B. cereus sensu lato* such as phospholipases, proteases (*inhA*), enhancins, chitinases, among others, which have been suggested as virulence assistance factors, to aid *B. thuringiensis* in the invertebrate host and tissue invasion (Raymond et al. 2010; see supplementary table S3,
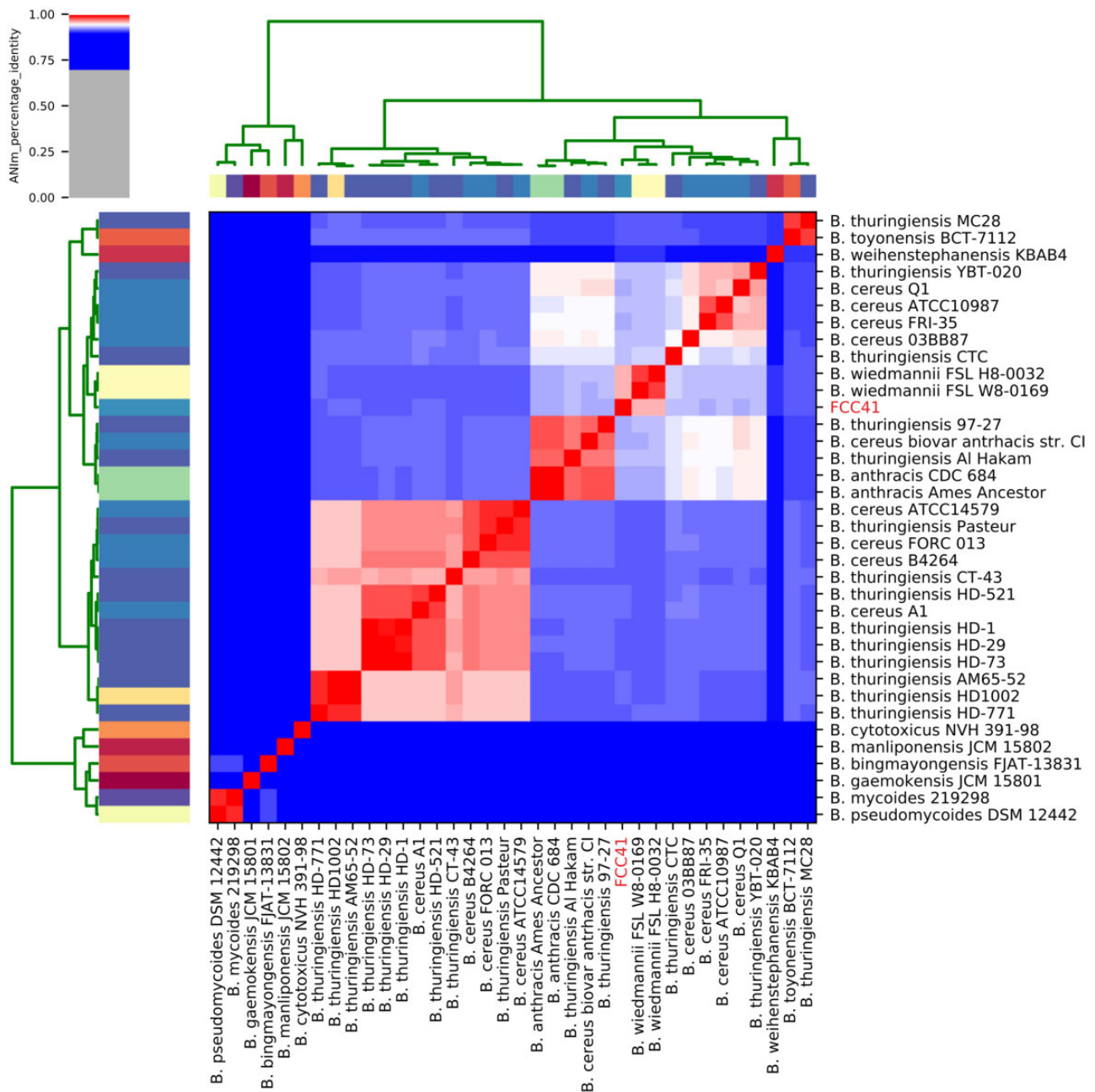
FIG. 3.—Comparison of the average nucleotide identity values between each genome of the 36 strains of *B. cereus sensu lato* strains: ANI analysis based on MUMmer alignment of the genome sequences was performed using PYANI. Red squares represent ANI values of 96% or higher.

Supplementary Material online). This strain carries coding genes of hemolysin BL, enterotoxin, cereolysin and other common chromosomal toxin genes mainly related with *B. cereus*. Additionally, FCC41 harbors two copies of the complete operon of the nonhemolytic enterotoxin, *nheABC*, one of them encoded in the chromosome and the other in plasmid pFCC41-5-125k, probably because of a horizontal gene transfer event. In particular, operon *nheABC* is abundant within *B. cereus* strains isolated from infections and from the environment, and it is usually found

in the chromosome. It has been described that the duplication of *nheABC* is extremely infrequent and appears to be phylogenetically unstable (Böhm et al. 2015). The operon present in plasmid pFCC41-5-125k is highly divergent from the chromosome copy and does not have the PlcR recognition sequence, which is commonly found in *nheABC* operon and other virulence factors coding genes. Moreover, the enterotoxin C encoded by *nheC* has an additional C-terminal domain with homology to ricin domain.

Table 2

Homologous Sequences to *cry* Genes Present in FCC41 Strain

Gene ID	Location	From	To	Length (bp)	Closest Cry Protein Homology	Identity (%)	Coverage (%)
cry41-like	Plasmid 2	75,380	77,956	2,577	Cry41Ba2 (ZP_04099652)	50	99
					Cry41Aa1 (BAD35157)	42	99
cry4-like1	Plasmid 3	72,700	76,215	3,516	Cry4Ca2 (AJW76682)	51	99
cry24Ca	Plasmid 3	1,13,304	1,15,364	2076	Cry24Ca1 (CAJ43600)	99	95
cry24Ca orf2	Plasmid 3	1,11,767	1,13,236	1,470	Cry39Aa1 orf2 (BAB72017)	85	99
cry52-like1	Plasmid 3	1,78,658	1,80,751	2094	Cry52Aa1 (ABU96490)	82	100
cry52-like1 *orf2*-like	Plasmid 3	1,76,911	1,78,590	1,680	Cry39Aa1 orf2 (BAB72017)	87	100
cry52-like2	Plasmid 3	1,43,701	1,45,767	2067	Cry52Ca1 (AJW76683)	63	100
cry52-like2 *orf2*-like	Plasmid 3	1,45,835	1,47,508	1,674	Cry39Aa1 orf2 (BAB72017)	82	99
cry4-like2	Plasmid 5	77,853	81,524	3,572	Cry4Ca2 (AJW76682)	51	99

Although this strain has pXO1-like and pXO2-like plasmids it does not carry any of the virulence factors associated with anthrax-like symptoms (protective antigen, and edema and lethal factors), nor capsule coding genes (*capABCD*), typical characteristics of *B. anthracis*.

Furthermore, we detected antibiotic resistance genes which may help this bacterium to survive and compete with other gut bacteria. Most of these virulence factors might contribute to insect pathogenicity (Raymond et al. 2010) of this strain and are shown in supplementary table S3, Supplementary Material online.

Cry-like Protein Coding Genes

We identified nine CDS with conserved Cry-like domains using hmmrsearch and further BLASTP analysis. The translated CDS were identified as two Cry4-like (both have 51% identity with Cry4Ca2, and 48% identity between them); two Cry52-like (82% identity with Cry52Aa1 and 63% identity with Cry52Ca1, respectively) with their corresponding downstream Orf2-like regions; the previously described Cry24Ca with its Orf2-like; and three contiguous pseudogenes within the same operon with Cry proteins domains homology. Cry4-like and Cry52-like proteins may contribute to the mosquitocidal activity since Cry4 and Cry52 proteins previously described have insecticidal activity against some species of *Aedes* sp., *Anopheles* sp., and *Culex* sp. (Ben-Dov 2014; Zheng et al. 2017). As well, a Cry41-like sequence with 50% identity to Cry41Ba2 and 42% identity to Cry41Aa1 was detected. In particular, this Cry41-like sequence contains a conserved ricin domain also detected in Cry41Aa. It has been previously reported that similar Cry proteins, called parasporins, cause cytotoxicity *in vitro* towards cancerous cells (Krishnan et al. 2017). All these sequences are encoded in three plasmids (fig. 1; table 2). In contrast with *B. thuringiensis* mosquitocidal strains like the subsp. *israelensis*, FCC41 does not have coding sequence genes for the previously described Cyt proteins (Soberón et al. 2013; Ben-Dov 2014).

Plasmid Analysis

Some of the most important *B. cereus* group phenotypic determinants are encoded in plasmids >100 kb. In most cases, these megaplasmids contain two or more different putative minireplicons, which are the minimum regions necessary to permit plasmid replication and comprise the origin of replication and the replication proteins coding genes (Zheng et al. 2013). In megaplasmids of the *B. cereus* group at least six types of minireplicons have been described, and the diversity and wide distribution are related with the evolution of those plasmids. FCC41 strain has eight plasmids, and five of them are megaplasmids (table 1) that harbor minireplicons previously found in *B. cereus* group (Zheng et al. 2013; Wang et al. 2018). In particular, plasmids pFCC41-1-490k and pFCC41-4-144k have a copy of a *repX*-like gene (see supplementary figs. S2 and S4, Supplementary Material online), defined as replication initiator proteins involved in the replication of megaplasmids. pFCC41-4-144k has additionally the pXO1-14/pXO-16-like proteins genes, initially described in *B. anthracis* pXO1 (Tinsley and Khan 2006), also present in other *B. cereus* group strains plasmids but not found in those with only one minireplicon (Zheng et al. 2013). It has been suggested that the presence of multiple minireplicons of different types could avoid the incompatibility between plasmids that possess the same minireplicon (Hülter et al. 2017). The lack of incompatibility between plasmids suggests that there are other replications proteins involved. Additionally, plasmid pFCC41-3-257k has a *repA*-like gene, and plasmids pFCC41-2-313k and pFCC41-5-125k have also different variants of the replication associated protein gene *ori44*-like (RepFR55), as shown in supplementary figures S3 and S5, Supplementary Material online.

The presence of two or more minireplicons in *B. cereus* group megaplasmids suggests that some of them may have resulted from the integration of two or more smaller plasmids (Zheng et al. 2013). pFCC41-3-257k shows a clear confirmation of such an integration event: About one fourth of the plasmid sequence has similarity with plasmids pMC183 from
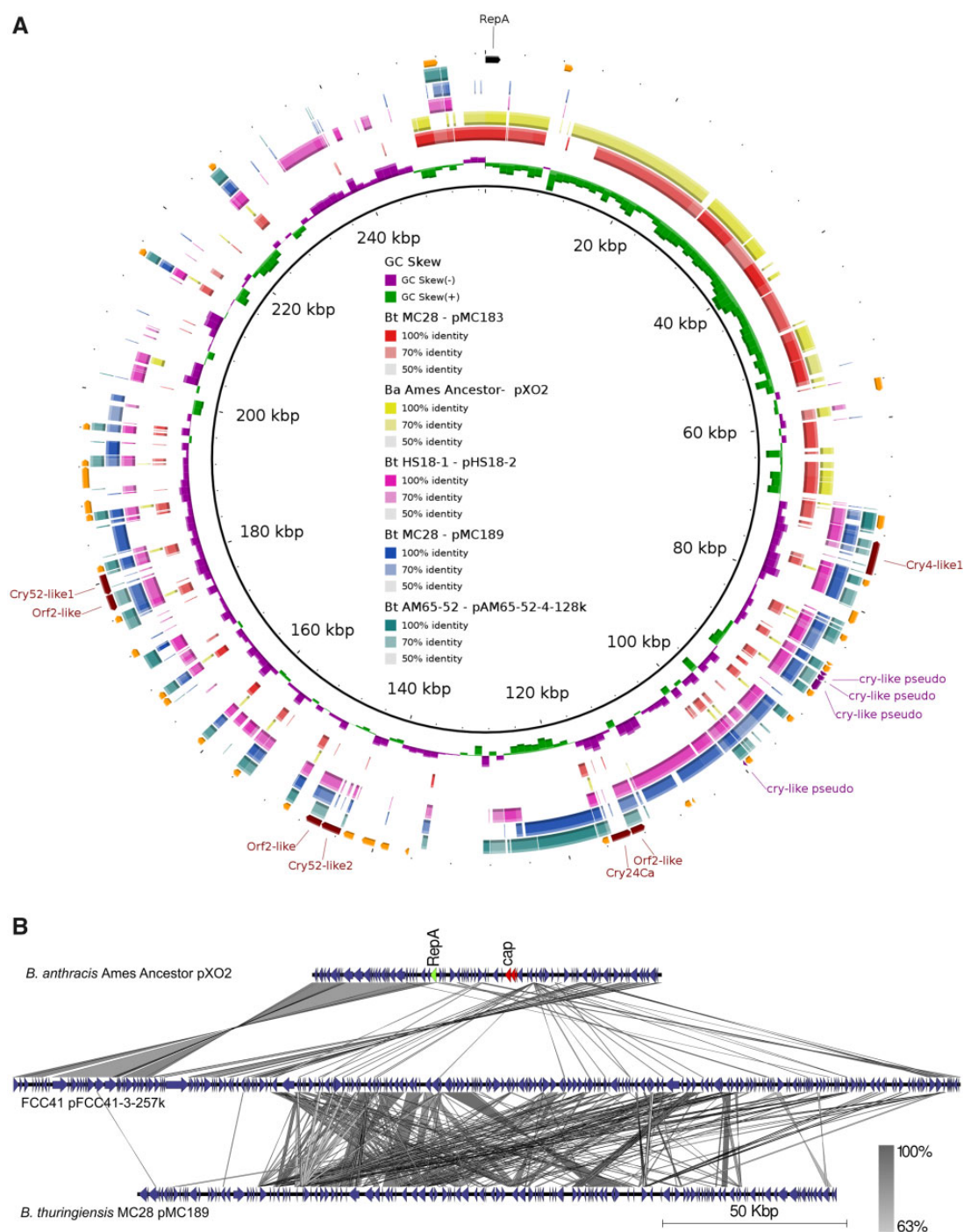
Fig. 4.—(A) Comparison of pFCC41-3-257K to homologous plasmids (*B. thuringiensis* MC28 pMC183, *Bacillus anthracis* Ames Ancestor pXO2, *B. thuringiensis* HS18-1 pHS18-2, *B. thuringiensis* MC28 pMC189, *B. thuringiensis* AM65-52 pAM65-52-4-128k) using Blast Ring Image Generator (BRIG). Rings from the outermost to the center: 1) minireplicon associated protein coding genes (in black), transposase coding genes (orange), Cry protein coding genes and *orf2*-like genes (brown), *cry*-like pseudo genes (violet), 2–6) sequence percentage identity to homologous plasmids, 7) GC skew, 8) scale marks. (B) Comparison of pFCC41-3-257k to *B. anthracis* Ames Ancestor pXO2 and *B. thuringiensis* MC28 pMC189 employing BlastN and EasyFig. Genes of operon *cap* are marked in red, and minireplicon associated gene *repA* in green.

B. thuringiensis MC28 and pXO2 from *B. anthracis*, and the remainder of the plasmid shares regions of partial homology with plasmids pAM65-52-4-128k from *B. thuringiensis*

israelensis AM65-52, pMC189 from *B. thuringiensis* MC28, and pHS18-2 from *B. thuringiensis* HS18-1. However, we only detected one minireplicon in this plasmid. The regions of

homology shift between related plasmids correspond with regions of GC skew sign change, which is known to be associated to DNA replication origins or terminus (Tillier and Collins 2000).

Plasmid pFCC41-1-490k has regions of shared homology with plasmids pMC429 and pHS18-1 from mosquitocidal *B. thuringiensis* MC28 and HS18 strains, respectively; and pAR460 from plant pathogen control agent and growth promoter *B. cereus* AR156 strain; and has homology with contigs 14 and 17 from *B. wiedmannii* strain AFS023094 (see supplementary fig. S2, Supplementary Material online). Plasmid pFCC41-2-313k is very similar to the previously described plasmid pAM65-52-1-360k present in *B. thuringiensis* subsp. *israelensis* strain AM65-52 and to contigs from strains ATCC35646, IBL4222 and IBL4200 of *B. thuringiensis* subsp. *israelensis*. This plasmid contains metabolite transporter genes, FtsK-like DNA transport protein, and type IV secretion system protein VirDIV. It has been previously mentioned that this type of plasmids appeared to be restricted to *B. thuringiensis* species (Bolotin et al. 2017). However, in recent times there has been an increasing number of published genomes of *B. cereus* and *B. wiedmannii* strains that have plasmids or contigs homologous to pFCC41-2-313k (see supplementary fig. S3, Supplementary Material online). Additionally, plasmid pFCC41-3-257k lacks the region of pXO2 that has the coding sequences for *B. anthracis* capsid proteins. This plasmid has the highest density of transposons and contains most of the Cry protein coding genes of this strain, which are all flanked by transposase coding genes (fig. 4). Nucleotide BLAST results indicate that plasmid pFCC41-4-144K shows high homology with pXO1 from *B. anthracis*; however, it lacks the coding genes of anthrax virulence factors (see supplementary fig. S4, Supplementary Material online). Plasmid pFCC41-5-125K has homology regions with plasmids from *Bacillus mycoides* and multiple strains of *B. thuringiensis* (see supplementary fig. S5, Supplementary Material online). Smaller plasmids, which range in size from 8 to 12 kb, have conserved regions presents in plasmids from other strains of *B. cereus* group (see supplementary figs. S6–S8, Supplementary Material online).

Horizontal gene transfer is a driving force in bacterial evolution, in which genetic exchange occurs, mainly caused by transposons, phages or integrons, and can take place between highly divergent strains (Zheng et al. 2015). FCC41 chromosome has high similarity to *B. wiedmannii* strains sequences; however, its plasmids appear to be from different origins. The presence of *cry* genes, a high number of transposons particularly in highly divergent plasmids (fig. 4A) and large plasmid regions similar to plasmid sequences present in other *Bacillus* species suggest strong evidence of horizontal gene transfer events, considering that the involved species belong to different clades (Bazinet 2017). Recent studies in comparative genomics suggest that these lateral processes

result in phylogenetic networks that explain more adequately the complexity of the evolutionary route (Zheng et al. 2015). Guinebretière et al. (2008) discussed that the global evolution of the *B. cereus* group seems to be strongly determined by ecological adaptations, being the temperature tolerance an important mechanism by which strains of the different groups adapted to novel environments and by which these groups have diverged. Most recently, Méric et al. (2018) reinforce this hypothesis proposing that the strains of different clades which share an ecological niche had increased opportunities for plasmid transfer between them. Most of the plasmids from strain FCC41 share homologous regions with plasmids of distant lineages from other clades, such as specialized insect pathogens from clade 2, which could indicate recent horizontal transfer. Strain FCC41 must have shared the ecological niche with other members of the *B. cereus* group. Our results support the hypothesis of cooperative infection, which predicts that conjugative plasmids carrying social genes such as Cry toxins could be widely distributed across clades (Méric et al. 2018).

The classification criteria for the *B. cereus* group division are widely discussed, the main problem has been to decide between evolutionary criteria or the use of phenotypical traits and traits of clinical and biotechnological importance, which are not clearly clustered in current phylogenetic trees (Raymond and Federici 2017; Zheng et al. 2017). The complete genome of FCC41 enriches the information about the genomic diversity and evolution among this group and proposes new questions about the position of *B. thuringiensis* strains that apparently belong to different lineages of evolution.

Conclusion

FCC41 was previously described as *B. thuringiensis* according to the phenotypic traits and the characterization of at least one Cry-toxic protein (Berón and Salerno 2007). This strain has mosquitocidal activity against *Aedes aegypti*, *Aedes (Ochlerotatus) albifasciatus*, *Culex pipiens*, *Culex quinquefasciatus*, and *Culex apicinus*, showing higher activity against *Ae. aegypti* than the *B. thuringiensis* subsp. *israelensis* reference strain. In this work we present the complete genome of FCC41 strain which includes one circular chromosome and eight circular plasmids. The genome-wide screening of this strain reveals new putative *cry* genes, its regulatory regions and some virulence factors potentially useful for the development of biotechnological applications for the control of vector-borne diseases important in human health. According with the bioinformatics analysis this strain is placed into the *B. wiedmannii* cluster, but due to the phenotypical characteristics it could be renamed as *B. wiedmannii* biovar *thuringiensis* strain FCC41, being the first reported *B. wiedmannii* specialized mosquitocidal pathogen strain that has Cry

proteins with insecticidal activity, a phenotypic characteristic that is commonly associated with *B. thuringiensis*.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Literature Cited

Adang MJ, Crickmore N, Jurat-Fuentes JL. 2014. Diversity of *Bacillus thuringiensis* crystal toxins and mechanism of action. In: Dhadialla TS, Gill SS, editors. Advances in insect physiology. Oxford: Academic Press. p. 39–87.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Ankenbrand MJ, Keller A. 2016. bcgTree: automatized phylogenetic tree building from bacterial core genomes. Genome 59(10):783–791.

Bazinet AL. 2017. Pan-genome and phylogeny of *Bacillus cereus sensu lato*. BMC Evol Biol. 17(1):176.

Ben-Dov E. 2014. *Bacillus thuringiensis* subsp. *israelensis* and its dipteran-specific toxins. Toxins 6(4):1222–1243.

Berón CM, Salerno GL. 2006. Characterization of *Bacillus thuringiensis* isolates from Argentina potentially useful in insect pest control. BioControl 51(6):779–794.

Berón CM, Salerno GL. 2007. Cloning and characterization of a novel crystal protein from a native *Bacillus thuringiensis* isolate highly active against *Aedes aegypti*. Curr Microbiol. 54(4):271–276.

Bhandari V, Ahmod NZ, Shah HN, Gupta RS. 2013. Molecular signatures for *Bacillus* species: demarcation of the *Bacillus subtilis* and *Bacillus cereus* clades in molecular terms and proposal to limit the placement of new species into the genus *Bacillus*. Int J Syst Evol Microbiol. 63(Pt 7):2712–2726.

Böhm ME, Huptas C, Krey VM, Scherer S. 2015. 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.

Bolotin A, et al. 2017. Comparative genomics of extrachromosomal elements in *Bacillus thuringiensis* subsp. *israelensis*. Res Microbiol. 168(4):331–344.

Caamaño-Antelo S, et al. 2015. Genetic discrimination of foodborne pathogenic and spoilage *Bacillus* spp. based on three housekeeping genes. Int J Food Microbiol. 46:288–298.

Carroll LM, Kovac J, Miller RA, Wiedmann M. 2017. Rapid, high-throughput identification of anthrax-causing and emetic *Bacillus cereus* group genome assemblies using BTyper, a computational tool for virulence-based classification of *Bacillus cereus* group isolates using nucleotide sequencing data. Appl Environ Microbiol. 83:e01096–e01017.

Cohan FM. 2002. What are bacterial species? Annu Rev Microbiol. 56:457–487.

Eddy SR. 1998. Profile hidden Markov models. Bioinformatics 14(9):755–763.

Guinebretière MH, et al. 2008. Ecological diversification in the *Bacillus cereus* group. Environ Microbiol. 10(4):851–865.

Hülter N, et al. 2017. An evolutionary perspective on plasmid lifestyle modes. Curr Opin Microbiol. 38:74–80.

Kim M, Oh HS, Park SC, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and *16S rRNA* gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol. 64(Pt 2):346–351.

Ko KS, et al. 2004. Population structure of the *Bacillus cereus* group as determined by sequence analysis of six housekeeping genes and the *plcR* gene. Infect Immun. 72(9):5253–5261.

Kovac J, et al. 2016. Production of hemolysin BL by *Bacillus cereus* group isolates of dairy origin is associated with whole-genome phylogenetic clade. BMC Genomics. 17:581.

Krishnan V, et al. 2017. The human cancer cell active toxin Cry41Aa from *Bacillus thuringiensis* acts like its insecticidal counterparts. Biochem J. 474(10):1591–1602.

Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19(9):1639–1645.

Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5(2):R12.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9(4):357.

Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.

Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32(14):2103–2110.

Liu Y, et al. 2015. Genomic insights into the taxonomic status of the *Bacillus cereus* group. Sci Rep. 5:14082.

Maughan H, Van der Auwera G. 2011. *Bacillus* taxonomy in the genomic era finds phenotypes to be essential though often misleading. Infect Genet Evol. 11(5):789–797.

Méric G, et al. 2018. Lineage-specific plasmid acquisition and the evolution of specialized pathogens in *Bacillus thuringiensis* and the *Bacillus cereus* group. Mol Ecol. 27(7):1524–1540.

Miller RA, et al. 2016. *Bacillus wiedmannii* sp. nov., a psychrotolerant and cytotoxic *Bacillus cereus* group species isolated from dairy foods and dairy environments. Int J Syst Evol Microbiol. 66(11):4744–4753.

Moayeri M, Leppla SH, Vrentas C, Pomerantsev AP, Liu S. 2015. Anthrax pathogenesis. Annu Rev Microbiol. 69:185–208.

Priest FG, Barker M, Baillie L, Holmes EC, Maiden MC. 2004. Population structure and evolution of the *Bacillus cereus* group. J Bacteriol. 186(23):7959–7970.

Rasko DA, Altherr MR, Han CS, Ravel J. 2005. Genomics of the *Bacillus cereus* group of organisms. FEMS Microbiol Rev. 29(2):303–329.

Raymond B, Federici B. 2017. In defense of *Bacillus thuringiensis*, the safest and most successful microbial insecticide available to humanity—a response to EFSA. FEMS Microbiol Ecol. 93:1–8.

Raymond B, Johnston PR, Nielsen-LeRoux C, Lereclus D, Crickmore N. 2010. *Bacillus thuringiensis*: an impotent pathogen? Trends Microbiol. 18(5):189–194.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14):2068–2069.

Sievers F, Higgins DG. 2014. Clustal Omega, accurate alignment of very large numbers of sequences. In: Russell DJ, editor. Multiple sequence alignment methods. Totowa (NJ): Humana Press. p. 105–116.

Soberón M, López-Díaz JA, Bravo A. 2013. Cyt toxins produced by *Bacillus thuringiensis*: a protein fold conserved in several pathogenic microorganisms. Peptides 41:87–93.

Sorokin A, et al. 2006. Multiple-locus sequence typing analysis of *Bacillus cereus* and *Bacillus thuringiensis* reveals separate clustering and a distinct population structure of psychotropic strains. Appl Environ Microbiol. 72(2):1569–1578.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.

Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27(7):1009–1010.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 14(2):178–192.

Tillier ER, Collins RA. 2000. Replication orientation affects the rate and direction of bacterial gene evolution. J Mol Evol. 51(5):459–463.

Tinsley E, Khan SA. 2006. A novel FtsZ-like protein is involved in replication of the anthrax toxin-encoding pXO1 plasmid in *Bacillus anthracis*. J Bacteriol. 188(8):2829–2835.

Tourasse NJ, Kolstø AB. 2008. SuperCAT: a supertree database for combined and integrative multilocus sequence typing analysis of the *Bacillus cereus* group of bacteria (including *B. cereus*, *B. anthracis* and *B. thuringiensis*). Nucleic Acids Res. 36:461–468.

Wang A, Ash GJ. 2015. Whole genome phylogeny of *Bacillus* by feature frequency profiles (FFP). Sci Rep. 5:13644.

Wang P, Zhu Y, Shang H, Deng Y, Sun M. 2018. A minireplicon of plasmid pBMB26 represents a new typical replicon in the megaplasmids of *Bacillus cereus* group. J Basic Microbiol. 58(3):263–272.

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 13(6):e1005595.

Wu S, Zhu Z, Fu L, Niu B, Li W. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics 12:444.

Zheng J, et al. 2015. Plasmids are vectors for redundant chromosomal genes in the *Bacillus cereus* group. BMC Genomics 16:6.

Zheng J, et al. 2017. Comparative genomics of *Bacillus thuringiensis* reveals a path to specialized exploitation of multiple invertebrate hosts. MBio 8:e00822–e00817.

Zheng J, Peng D, Ruan L, Sun M. 2013. Evolution and dynamics of megaplasmids with genome sizes larger than 100 kb in the *Bacillus cereus* group. BMC Evol Biol. 13:262.