Completed Genomic Sequence of *Bacillus thuringiensis* HER1410 Reveals a Cry-Containing Chromosome, Two Megaplasmids, and an Integrative Plasmidial Prophage

Ana Lechuga,*,1 Cédric Lood,†,‡,§ Margarita Salas,*,2 Vera van Noort,‡,§ Rob Lavigne,† and Modesto Redrejo-Rodríguez*,**,3

*Centro de Biología Molecular Severo Ochoa, **Departamento de Bioquímica, Universidad Autónoma de Madrid (UAM) and Instituto de Investigaciones Biomédicas Alberto Sols (CSIC-UAM) Madrid, Spain, †Department of Biosystems, Laboratory of Gene Technology, ‡Department of Microbial and Molecular Systems, Centre of Microbial and Plant Genetics, Laboratory of Computational Systems Biology, KU Leuven, Belgium, and §Institute of Biology, Leiden University, The Netherlands

ORCID IDs: 0000-0002-1387-3041 (A.L.); 0000-0001-7826-3378 (C.L.); 0000-0002-8436-6602 (V.N.); 0000-0001-7377-1314 (R.L.); 0000-0003-0014-4162 (M.R.-R.)

**ABSTRACT** *Bacillus thuringiensis* is the most used biopesticide in agriculture. Its entomopathogenic capacity stems from the possession of plasmid-borne insecticidal crystal genes (*cry*), traditionally used as discriminant taxonomic feature for that species. As such, crystal and plasmid identification are key to the characterization of this species. To date, about 600 *B. thuringiensis* genomes have been reported, but less than 5% have been completed, while the other draft genomes are incomplete, hindering full plasmid delineation. Here we present the complete genome of *Bacillus thuringiensis* HER1410, a strain closely related to *B. thuringiensis entomocidus* and a known host for a variety of *Bacillus* phages. The combination of short and long-read techniques allowed fully resolving the genome and delineation of three plasmids. This enabled the accurate detection of an unusual location of a unique *cry* gene, *cry1Ba4*, located in a genomic island near the chromosome replication origin. Two megaplasmids, pLUSID1 and pLUSID2 could be delineated: pLUSID1 (368 kb), a likely conjugative plasmid involved in virulence, and pLUSID2 (156 kb) potentially related to the sporulation process. A smaller plasmidial prophage pLUSID3, with a dual lifestyle whose integration within the chromosome causes the disruption of a flagellar key component. Finally, phylogenetic analysis placed this strain within a clade comprising members from the *B. thuringiensis* serovar *thuringiensis* and other serovars and with *B. cereus* s. s. in agreement with the intermingled taxonomy of *B. cereus* sensu lato group.

**KEYWORDS**

* Bacillus thuringiensis  
* Bacillus cereus* s. s.  
* tectivirus  
* complete genome  
* plasmids  

The *Bacillus cereus* group encompasses 21 published species with a common monophyletic origin that display highly differentiated phenotypes in terms of ecological niches and virulence spectra (Liu et al., 2017, Carroll et al., 2020). Among them, some have been extensively studied due to their medical and economic importance. This group contains well-known pathogenic species including *Bacillus cereus* sensu stricto, linked to food poisoning, emesis and diarrhea, and *Bacillus anthracis*, the etiological agent of anthrax. There are also non-pathogenic industrially relevant species like *Bacillus toyonensis* and *Bacillus thuringiensis*. The former is used as probiotic in veterinary medicine while the latter produces insecticidal toxins and is often used as an industrial biopesticide control agent (Ehling-Schulz et al., 2019). These phenotypes often find their genetic origin on large plasmids where toxins such as anthrax, bioinsecticidal crystal proteins, or emetic toxins are typically encoded (Adams et al., 2014). Toxin-bearing (mega)plasmids, as well as a variety of smaller plasmids, comprise the vast diversity of accessory elements characteristic of the *B. cereus* group. This accessory genome, as well as other mobile genetic elements (MGEs) like prophages and genomic islands, favor a rapid adaptive response to ecological change (Raymond and Bonsall, 2013, Fu et al., 2019).
Traditionally, B. cereus group members have been classified based on phenotypic criteria associated with their clinical and industrial relevance. These defined phenotypes do not necessarily agree with a genomic-based taxonomy as the plasmids causing the phenotypes can be transferred between less related species. A clear example of this problem is the extensive sequence similarity between B. cereus s. s. and B. thuringiensis where the latter is commonly discriminated based on the occurrence of parasporal crystal inclusion bodies. Two types of proteins, Cry (crystal) and Cyt (cytolytic) proteins, are the principal components of these crystal protein inclusions and have shown high specific activity against different insects (Ehling-Schulz et al., 2019). Cry and cyt genes are located in plasmids such as the well-studied B. thuringiensis sv. israelensis pBToxix which harbors up to five different cry genes and three cyt genes (Gillis et al., 2018). These genes are usually flanked by mobile elements (insertion sequences and transposons) and can grouped with other insecticidal proteins to form pathogenicity islands (PAI). This organization and the surrounding genetic context of these genes have been described to facilitate their mobility among plasmids and strains (Friederuk et al., 2017).

Numerous tailed bacteriophages preying Bacillus cereus sensu lato have been previously identified and, among them, a recently accepted genus of the Tectiviridae family, Betatectivirus was established (Gillis and Mahillon, 2014a). The recently renewed interest in the study of tectiviruses stems from the narrow host specificity of some members for dangerous human pathogens, their complex diversity and variability patterns, and a possible phylogenetic relationship with some groups of eukaryotic viruses and mobile elements (Krupovic and Koonin 2015). In the study of betatectiviruses, B. thuringiensis GB002, HER1410, and B. cereus HER1047 strains have been used as hosts for tectival infections. These three strains represent different serotypes and have been especially useful to study tectiviral infections features such as virus entry and host cell physiology (Gaidelyte et al., 2015). In the study of betatectiviruses, B. thuringiensis HER1410 strain was from Laboratory Stock. The original strain was obtained from culture collection Félix d’Herelle Reference Center for Bacterial Viruses of the Université de Laval (https://www.phage.ulaval.ca) and can be retrieved with HoststrainHERnumber1410.

DNA extraction and genome sequencing

B. thuringiensis HER1410 was inoculated in LB medium and incubated overnight at 37° to an OD600 of 1.14. Genomic DNA (gDNA) was isolated using the DNeasy Blood and Tissue kit (Qiagen) and purified by ethanol precipitation. The quality of the gDNA was checked using a ThermoFisher Scientific NanoDrop spectrophotometer (OD280/260 and OD230/260), and its integrity was verified by gel electrophoresis (1% agarose w/v).

The gDNA was sequenced using Illumina and Nanopore technologies. The first set of reads was obtained on an Illumina MiniSeq using a paired-end 2x150 bp approach with a library obtained using Nextera Flex (Illumina, inc.). A second set of reads was obtained on a MinION nanopore sequencer (Oxford Nanopore Technology) equipped with a flowcell of type R9.4.1 and a library prepared using the 1D ligation approach with native barcoding (Oxford Nanopore Technology).

Genome hybrid assembly and functional annotation

The Illumina reads were controlled for quality using FastQC v0.11.8 and trimomatic v0.38 for adapter clipping, quality trimming (LEADING:3 TRAILING:3 SLIDINGWINDOW:4,15), and minimum length exclusion (>50 bp) (Andrews, 2010, Bolger et al., 2014). The quality of the nanopore reads was evaluated visually using Nanoplot v1.27.0 (De Coster et al., 2018) and then further processed using Porechop v0.2.3 to clip barcodes and with Nanofilt v2.3.0 for quality exclusion (Q > 10) and length (>1,000 bp) (Wick 2017).

The genome was assembled de novo using the short-reads SPAdes assembler v3.13.1 with default options, and the hybrid Unicycler v0.4.8 assembler using bold mode (Bankevich et al., 2012, Wick et al., 2017). Unicycler was configured to use the Pilon polishing software (Walker et al., 2014), which performs repetitive cycles of mapping the raw short-reads onto the hybrid assemblies in order to correct possible mistakes. As shown in Supplemental Figure 1, short-reads overall mapped with a medium-high and homogeneous sequence depth (≥20-30) that assures the lowest possible error rate after the polishing step. The final assembly metrics reported in Supplemental Figure 2 were computed using QUAST v4 and the visualization of the assembly graphs was drawn with Bandage v0.8.1 (Gurevich et al., 2013, Wick et al., 2015).
The annotation of the genome was performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP), and the resulting annotated proteome was further annotated with COG categories and KEGG pathways using the eggNOG mapper (Tatusova et al., 2016, Huerta-Cepas et al., 2017). The annotation of the cry1Ba4-containing island was updated using ISFinder (Sigueri 2006) (Figure 2, Supplemental Table 3). The annotation of the extrachromosomal elements was updated using eggNOG mapper (Huerta-Cepas et al., 2017), PHASTER (Arndt et al., 2016) and PHMMER tools (Potter et al., 2018) (Figures 3 and 4, Supplemental tables 3 and 4).

Large structural variation analysis

The larger structural variations were detected using in combination the long-reads mapper Ngmlr v0.2.7 and the structural variation caller Sniffles v1.0.11 (Sedlazeck et al., 2018). The results were visualized using the software Ribbon v1.1 (Nattestad et al., 2016).

Genome analysis

The identification of genomic islands, i.e., regions of likely horizontal origin, of the annotated HER1410 genome was done using the Island Viewer 4 (Bertelli et al., 2017). The genomes were also mined for the presence of prophages using the web application PHASTER (Arndt et al., 2016). The identification of secondary metabolite regions was carried out using antiSMASH version 5.1.2 (Blin et al., 2019). Comparative analysis of the cry1Ba4-containing genomic island with plasmidial cry1Ba-containing cassettes was performed and visualized using EasyFig (Sullivan et al., 2011).

The virulence and antibiotic resistance genes were annotated using BTyper version 2.3.3 (default settings) (Carroll et al., 2017) and insecticidal toxin-encoding genes were annotated using BTyper version 3.0.2 (Carroll et al., 2020). CRISPR-Cas genes were detected using CRISPRCasFinder (Couvin et al., 2018) (Supplemental Table 3). Chromosomal and extrachromosomal features were visualized using Circos (Krzywinski et al., 2009)

Phylogenetic analysis

To establish the position of B. thuringiensis HER1410 within the population structure of B. cereus-thuringiensis, two taxon sets were created. The first set (Supplemental Table 5) comprised up to five strains of B. cereus or B. thuringiensis selected from differentiated clades in BCSL_114 and B. thuringiensis phylogenetic trees from previous population structure analyses (Bazinet, 2017, Gillis et al., 2018). The second set comprised 1) the closest genomes to our strain identified using the phylogenetic reconstruction of the first dataset (underlined in Figure 5), 2) 121 strains closely related to those according to phylogeny analysis of B. cereus group strains from Carroll et al. (2020), and 3) B. thuringiensis YBT-1518 type strain as outgroup (Supplemental Table 6). To ensure the assembly quality of both datasets, only genomes with an N50 size over 20 kbp were selected. The panlogenomes of B. cereus s. s. and B. thuringiensis were inferred using Roary version 3.11.2 (Page et al., 2015). First, the GenBank files from selected strains were downloaded from NCBI’s and converted to GFF3 using the hp_genbank2gff script from the Bioperl library (Stajich et al., 2002). Second, the gff3 annotations were provided to Roary to calculate the pangenome of the dataset and produce a multiple sequence alignment (MSA) of the concatenated core genes (present in >99% strains) using MAFFT. Third, the MSA of the core genome was used to generate a best-fit maximum likelihood phylogeny using IQTREE version 1.6.12 using ModelFinder optimization (Nguyen et al., 2015, Kalyaanamoorthy et al., 2017). Finally, the trees were visualized in FigTree v.1.4.4 (Rambaut 2012) and iTOL version 5.5 (Letunic and Bork 2016). ANI differences between B. thuringiensis sv. israelensis 4Q7 (GCF_000585975.1) and HER1410 were calculated using the web application ANI Calculator (Yoon et al., 2017).

Data availability

The genome of Bacillus thuringiensis HER1410 was deposited under the NCBI GenBank accession numbers CP050183-CP050186S. The datasets of Illumina and Nanopore reads are available in the NCBI SRA database via the bioproject accession number PRJNA613019. Supplemental material available at figshare: https://doi.org/10.25387/g3.12563612.

RESULTS AND DISCUSSION

Chromosome features

The B. thuringiensis HER1410 genome comprises a 5,585,577 bp circular chromosome which represents 90.86% of the whole genome. While gene content in chromosomes within the B. cereus group is well conserved, variation among them mainly stems from MGEs such as prophages, insertion elements, transposons and plasmids (Rasko et al., 2005). Accordingly, the search for genomic islands and prophages in HER1410 chromosome yielded 24 predicted genomic islands and nine prophages (three intact and six incomplete) that would belong to the Caudovirales (Figure 1, Supplemental Table 3). These prophages were assigned by Phaster software (Arndt et al., 2016) to the families Siphoviridae and Myoviridae, in agreement with previous studies of prophages infecting the B. cereus group (Gillis and Mahillon, 2014a). No tectiviruses were detected on the chromosome or the extrachromosomal elements, in line with the results from Verheust et al. (2005), which showed that no linear molecules could be detected by PFGE analysis of genomic DNA and where the sensitivity to tectiviral infections was confirmed.

B. thuringiensis has been typically defined by its ability to produce crystal proteins with specific activity against insects. Crystal encoding genes in B. thuringiensis are harbored by plasmids (Ehling-Schulz et al., 2019). Only few chromosome-located exceptions have been reported in early works pre-dating the high-throughput sequencing era (Kronstad et al., 1983, Carlson and Kolsto 1993). Strikingly, we found that HER1410 genome contains a single crystal protein-coding gene, cry1Ba4, that is not located in a plasmid but within the chromosome, within a few kbp from the predicted origin of replication (Figure 2). This unusual location could be explained by the surrounding genetic context, a genomic island that contains multiple transposase genes from the ISs, IS6, IS66, IS605, and Tn3 families is extensively described and highlighted the dynamics of horizontal transfer of toxins (Zheng et al., 2017). Within this island, several different transposases were annotated, suggesting multiple transposition events and pointing this region as a putative insertion hotspot (Figure 2). Interestingly, the HER1410 cry1Ba4-containing island organization resembles that of other B. thuringiensis plasmids where a N-acetylmuramoyl-l-alanine amidase coding sequence (CDS), coding for a germination key enzyme, precedes the insecticidal gene (Fiedoruk et al., 2017). When the HER1410 chromosomal cry1Ba4-containing island was compared with other plasmidial cryBa1-containing cassettes described in Fiedoruk et al. (2017), a similar genomic organization was
observed (Figure 2). Also, as in cry1Ba-containing plasmids, the HER1410 island only possesses one pesticidal gene (not within a PAI) and it is not preceded by a K\(^+\)/(Na\(^+)\)/H\(^+\) antiporter. These similarities suggest a plasmidial origin of the HER1410 cry1Ba\(^+\) -coding island. The insertion of this whole cassette into the chromosome close to the origin of replication may have been favored by a higher copy number of genes in this region during exponential growth, potentially allowing the loss of the plasmid (Bergstrom et al., 2000).

Since no additional crystal genes were detected, cry1Ba\(^+\) can be related to the observation that a single type of parasporal crystal exists in this strain, a feature that allowed its classification as B. thuringiensis (Verheust et al., 2005). More than 700 genes encoding Cry proteins have been identified and characterized in B. thuringiensis (Crickmore and Feitelson 2016) and they show highly specific insecticidal activity to target insects. In this case, Cry1Ba\(^+\) delta-toxin, purified from B. thuringiensis sv. entomocidus HD9, demonstrated mortality greater than 85% against Plutella xylostella, one of the more prevalent pests in Malaysia (Nathan et al., 2006). These results suggest that HER1410 may also possess insecticidal activity against some Lepidoptera insects.

Additional B. cereus s. l. toxins in this strain were found using the BTyper software. No emetic toxin cereulide, responsible of emetic syndrome, was detected, but heat-labile diarrheal syndrome enterotoxins cytolyisin K (CytK), non-hemolytic enterotoxin (Nhe), and hemolysin BL (Hbl), commonly present among B. cereus group members, were found within the chromosome of HER1410 (Figure 1). Other virulence factors such as InhA2 and bceT are also present in the HER1410 genome. These results correlate with the hemolysis activity and cytotoxicity exhibited by this strain, opening up also the possibility that HER1410 might cause diarrheal syndrome. However, it is difficult to predict the enterotoxic potential of a given strain from genomics data only since it stems from a complex mechanism where transcription levels and other virulence factors are involved.
The presence of antibiotic resistance genes and CRISPR-Cas systems was also searched, using BTyper and CRISPRCasFinder tools. Several vancomycin, beta-lactam and fosfomicin resistance genes but no CRISPR systems were detected along the chromosome (Figure 1, Supplemental Table 3).

Three novel extrachromosomal elements in Bacillus thuringiensis

Extrachromosomal elements in B. cereus s. l. are numerous, variable, mobile, and essential to the ecology and classification of its members. In particular, some plasmids are extensively studied since they carry virulence genes like anthrax containing pXO1 (182 kb) and pXO2 (95 kb) in B. anthracis or crystal toxins containing megaplasmids characteristic of B. thuringiensis (Adams et al., 2014). Despite the importance of plasmids in this group, the common use of short-reads sequencing techniques has often resulted in draft genome assemblies lacking described extrachromosomal elements (Méric et al., 2018). In this case, nanopore long-reads sequencing provided an efficient solution to identify two megaplasmids, denominated pLUSID1 (368 kb), pLUSID2 (155 kb), and a smaller plasmid, pLUSID3 (38 kb), comprising 9.14% of the genomic information. We used these three plasmids as queries in nucleotide-nucleotide BLASTn against the database of nucleotide collection (nr/nt) on NCBI. Based on BLASTn, pLUSID1 could be linked to some B. thuringiensis and B. cereus s. s. plasmids, being the closer homolog B. thuringiensis strain YC-10 plasmid pYC1 (70% coverage and 98.78% identity; Genbank CP011350.1). On the other hand, pLUSID2 and pLUSID3 hits corresponded mainly with B. thuringiensis and B. cereus s. s. chromosomes and, in the case of pLUSID3, BLASTn also highlighted a similarity to some Bacillus phages (see below).

The identification of pLUSID1, a megaplasmid of 368,179 bp, is in agreement with the previous report of direct observation by gel electrophoresis of a megaplasmid in HER1410 in Gillis et al. (2017). This element presents a lower GC-content than the chromosome and is detected in a low copy number of 1.0 ± 0.14 per chromosome, as is generally observed for large plasmids (Supplemental Table 1) (Bolotin et al., 2017). Within the B. cereus group, genes on plasmids are generally more similar to the chromosomal variable genes than to the chromosomal core genes. Also, these plasmids show some differences from chromosomes in the functions of the genes they harbor (Zheng et al., 2015). Comparing with the chromosome, the pLUSID1 functional annotation revealed an enrichment in defense mechanism genes and in secondary metabolite biosynthesis, transport and catabolism genes which are located in a particular region of the plasmid (Figure 3; Supplemental Figure 3). This region was identified using antiSMASH as a zwittermicin A cluster (96% similarity), an aminopolyol antibiotic produced by B. cereus (Supplemental Table 3). This antibiotic, among other functions, amplifies the insecticidal activity of the toxin protein from B. thuringiensis (Broderick et al., 2003). Besides, 20.54% of the annotated genes are related to replication, recombination, and repair representing the largest proportion of plasmid genes. Among them, several transposases and integrases were identified along with the plasmid. Two genomic islands were identified, one of them coinciding with an incomplete prophage. Also, within this category, a region located in one of the GC-skew changes contains some putative DnaA and cell division proteins (Ftsz-like protein and TadZ) (Figure 3, Supplemental Table 3). Similarly to the toxin-encoding plasmid B. thuringiensis pBtoxis, this region may contain the replication origin of pLUSID1 and these proteins could be responsible for a self-replication system and possibly the segregation and partitioning...
of this plasmid to daughter cells (Tang et al., 2007). Alternatively, this region might also contain a conjugation origin, as several conjugal transfer proteins were annotated in this location. Moreover, we could predict an S-layer protein, involved in plasmid mobilization (Gillis et al., 2018). Overall, these results suggest that pLUSID1 could be transferred by conjugation.

Unlike other B. thuringiensis megaplasmids, this element does not contain insecticidal toxin genes, but it harbors the three Hbl enterotoxin components, which means an additional copy to the chromosomal Hbl. Some antimicrobial toxins are also located in this plasmid. Thus, CDS coding for a bacitracin, a zeta toxin and four copies of a plantaricin C family lantibiotic as well as several ABC transporters, reportedly involved in bacteriocin secretion, were found (Berry et al., 2002). Interestingly, the region comprising the plantaricin C family lantibiotic genes also contains a lantipeptide synthetase gene and was identified as a secondary metabolite region with no similarity to described clusters (Supplemental Table 3). Also, several ABC transporter CDS are located in a predicted secondary metabolite region similar to thuricin H cluster, a bacteriocin that inhibits B. cereus F4552 growth (Lee et al., 2009). pLUSID1 also encodes spore-related proteins that contribute to the germination process. Holomogous proteins have been found in Clostridium botulinum where they act as germination receptors that have been shown to be dispensable (Clauwers et al., 2017). Well studied toxin-encoding plasmids such as B. anthracis pXO1 and pBtoxi also encode genes involved in sporulation and germination and they have shown to be important for the virulence of the host (Guidi-Rontani et al., 1999, Berry et al., 2002). Overall, although pLUSID1 does not carry pesticidal proteins, it seems to have a role in HER1410 virulence and defense.

A second megaplasmid, pLUSID2 (155,569 bp), was found in HER1410 (Figure 3). This plasmid has not been experimentally observed for this strain and BLASTn analysis yielded high similarity to other strains’ chromosomes, where the closer homolog corresponded to the B. thuringiensis serovar tolworthi chromosome (94% coverage and 99.09% identity; Genbank AP014864.1). Nevertheless, as a result of the structural variation analysis of pLUSID2, no variants that could indicate this region is part of the chromosome, could be detected. Furthermore, the only structural variant observed is a result from the linearization of this plasmid (Supplemental Figure 4). Thus, the nanopore data supports that pLUSID2 is a circular element present 2.00 ± 0.37 times per chromosome (Supplemental Table 1). This could mean that pLUSID2 is a cryptic plasmid in HER1410 and its high similarity to other B. thuringiensis chromosomes is consistent with the high mobility of MGE between plasmids and chromosomes within the B. cereus group (Méric et al., 2018). In agreement with that, several transposases integrases and recombinases were detected. However, this plasmid contains only one mobile island that coincides with the location of an incomplete prophage (Figure 3, Supplemental Table 3). ByTyper results did not yield any B. cereus-related toxins or antibiotic resistance genes suggesting pLUSID2 is not directly associated with HER1410 virulence. Thus, pLUSID2 seems to play some role in sporulation, since it does possess four sporulation genes as well as six tellurite resistance proteins which have been shown in B. cereus s. l. spores (Delvecchio et al., 2006). Gene functional analysis also suggests a possible role related to basal metabolism, such as those involved in carbohydrate transport and metabolism, and inorganic ion transport and metabolism and energy production and conversion whose proportion seems to be higher than on the chromosome (Supplemental Figure 3, Supplemental Table 2). This contrasts with the general characteristics of B. cereus plasmids where the proportion of gene families involved in basal metabolism was significantly lower than on chromosomes (Zheng et al., 2015).

In line with this, as accessory elements are usually under less stringent evolutionary forces than the chromosome, those genes may have a higher rate of neo-functionalization (Chevreton et al., 2020).

The smallest extrachromosomal element, pLUSID3, is a plasmid of 38,150 bp and coverage of 2.33x compared to the chromosome (Figure 4A, Supplemental Table 1). Annotation of pLUSID3 resulted in 40% of hypothetical proteins and a high number of phase related proteins (Supplemental Table 2). This annotation was updated with results from COG functional analysis and a high number of phase related proteins (Supplemental Table 2). Results were updated with conservation of genes with the highest identity with other strains detected in the same region of the chromosome, could be detected. Furthermore, the only structural variant observed is a result from the linearization of this plasmid (Supplemental Figure 4). Thus, the nanopore data supports that pLUSID3 is a circular element present 2.00 ± 0.37 times per chromosome (Supplemental Table 1). This could mean that pLUSID3 is a cryptic plasmid in HER1410 and its high similarity to other B. thuringiensis chromosomes is consistent with the high mobility of MGE between plasmids and chromosomes within the B. cereus group (Méric et al., 2018). In agreement with that, several transposases integrases and recombinases were detected. However, this plasmid contains only one mobile island that coincides with the location of an incomplete prophage (Figure 3, Supplemental Table 3). BByTyper results did not yield any B. cereus-related toxins or antibiotic resistance genes suggesting pLUSID2 is not directly associated with HER1410 virulence. Thus, pLUSID2 seems to play some role in sporulation, since it does possess four sporulation genes as well as six tellurite resistance proteins which have been shown in B. cereus s. l. spores (Delvecchio et al., 2006). Gene functional analysis also suggests a possible role related to basal metabolism, such as those involved in carbohydrate transport and metabolism, and inorganic ion transport and metabolism and energy production and conversion whose proportion seems to be higher than on the chromosome (Supplemental Figure 3, Supplemental Table 2). This contrasts with the general characteristics of B. cereus plasmids where the proportion of gene families involved in basal metabolism was significantly lower than on chromosomes (Zheng et al., 2015). In line with this, as accessory elements are usually under less stringent evolutionary forces than the chromosome, those genes may have a higher rate of neo-functionalization (Chevreton et al., 2020).

Phylogenetic positioning of HER1410
HER1410 strain has been classified either as serovar israelensis (Daugelavicius et al. 2007, Gaïdelyte et al. 2005, Gaïdelyte et al. 2009).
2006), or thuringiensis (Gillis, A., Mahillon, J., 2014b and Verheust et al., 2005), being the latter fundamented on plasmidic profile, type of crystal and insecticidal activity. Several studies have proposed to resolve the conflict between genomics and phenotypical classification (Bazinet, 2017, Baek et al., 2019, Carroll et al., 2020).

While B. thuringiensis has always been characterized by their ability to produce entomopathogenic crystals, some crystal producing strains are more genetically similar to B. cereus s. s. (Carroll et al., 2020). Here we aimed to characterize HER1410, not only by its capacity of producing entomotoxins, but also by its genomic features. Multiple different approaches have been used to establish the population structure of B. cereus s. l. such as multi-locus sequence typing, ANI or core genes-based phylogenies (Bolotin et al., 2017, Baek et al., 2019, Carroll et al., 2020). Here, in order to provide a high resolution on the population structure, we performed a phylogenomics analysis based on a MSA of the core genes present in the pangenome (Page et al., 2015). To accomplish this using a comprehensive dataset, we applied a two-stage phylogenetic analysis. First, we selected 45 genomically differentiated B. cereus s. s. and B. thuringiensis strains to situate this HER1410 among defined clusters of B. cereus (see Methods). The pangenome of the selected B. cereus-thuringiensis strains was determined with Roary and resulted in a total of 1,838 core genes (i.e., present in 99% of the strains), which were subsequently aligned to build a ML phylogenetic tree using IQ-TREE (Figure 5). We could observe that there are some clusters including both B. cereus and thuringiensis, in agreement with the reported high genomic similarities between both species. Interestingly, although HER1410 has been classified as B. thuringiensis sv. thuringiensis, our results locate this strain in a cluster that comprises both B. cereus s. s. and thuringiensis strains. Among these strains, we could also find the type strain for B. cereus s. s., ATCC 14579.

Furthermore, to find the closest genomes to our strain, we performed a focused analysis using 126 strains related to HER1410 (see Methods), with the addition of HER140 as outgroup and obtained the pangenome of this dataset, with a total of 2,886 core genes. This deeper analysis accurately anchored HER140 within a clade which contains both B. cereus and thuringiensis strains (Figure 6). These results agree with the taxonomic classification using BTyper3 of HER1410 as B. cereus s. s. biovar thuringiensis. This software attempts to reconcile genomic definitions of species with clinically and industrially relevant phenotypes (Carroll et al., 2020). In this case, HER1410 could be classified as Bacillus cereus s. s. because of its similarity to this genomospecies (ANI = 98.08%) and biovar thuringiensis due to the presence of the insecticidal gene cry1Ba4. For practical purposes and, in line with the currently accepted
nomenclature, we propose maintaining HER1410 as a *B. thuringiensis* serovar *thuringiensis* strain although the separation between these two species does not fully hold upon genomic inspection.

Interestingly, *B. thuringiensis* HER1410 has a high level of similarity with *B. thuringiensis* sv. *subtoxicus* and sv. *entomocidus*. The latter has shown an antimicrobial activity against Gram-positive bacteria including *Listeria monocytogenes*, one of four pathogenic *Pseudomonas aeruginosa* and several fungi. Despite the presence of a range of virulence-related genes, including Hbl, sv. *entomocidus* is not toxic against mammalian Vero cells (Cherif et al. 2003). This would not be the case for HER1410 which does present cytotoxic activity. Also, the *entomocidus* genome contains more than one cry gene besides some other entomocidal toxins not present in HER1410 suggesting both strains, although very similar, show different virulence phenotypes (Méric et al., 2018). Interestingly, as reported here for HER1410, *entomocidus* has been suggested to harbor cry genes within the chromosome (Carlson and Kolsto 1993). On the other hand, cry genes also seemed to be present in plasmids for *entomocidus*. This could explain the different number of cry genes for each strain by the presence of a cry-encoding plasmid in *entomocidus* but not in HER1410. Because the two closer strains (*entomocidus* and *subtoxicus*) have been sequenced using Illumina, the quality of the resulting assemblies did not allow the identification of extrachromosomal elements. However, BLASTn analysis indicate pPLUSID1, pPLUSID2 and pPLUSID3 are also present in these strains.

As mentioned above, besides HER1410, two other strains, HER1047 and GBJ002, are broadly used to the study of tectiviruses because of their sensitivity to multiple Betatectivirus members. This is in conflict with the hypothesis that these phages are probably able to infect only specific bacterial strains from closely related hosts, unlike the PRD1-like phages that have a broad host range (Gillis and Mahillon, 2014b). On one hand, HER1047 is classified as *B. cereus*, as it does not produce crystals but contains the *B. cereus* gene *clo* (Verheust et al., 2005) while, on the other hand, GBJ002 is derived from *B. thuringiensis* sv. *israelensis* 4Q7, which is cured of its plasmids. Interestingly, HER1410 and *israelensis* 4Q7 seems to be a distant relative in our phylogenetical analysis. This correlates with the ANI differences between them.

**Figure 4** pPLUSID3 features analysis (A) Circular representation of plasmid pPLUSID3. The outer two circles represent the coding regions on the positive and negative strands colored by their functional annotation. The next circle displays prophages predicted by PHASTER webserver where purple = questionable. The inner two rings indicate the changes in %GC content and the GC skew respectively. (B) Structural variation analysis of pPLUSID3. Long-reads from Nanopore sequencing show evidence for pPLUSID3 variants in HER1410 genome where it can be integrated in the chromosome or form a closed circular element (upper panel). One of the reads showing the chromosomal integration is indicated by an arrow and represented in the lower panel. Variants are visualized with Ribbon (Nattestad et al., 2016). (C) CDS map provided and colored by the PHASTER tool. CDSs annotations were updated using PHASTER, PHMMER and COG functional annotations (Supplemental Table 4). Pale and dark green colors correspond to hypothetical and phage-related functions. Other colors indicate identified and annotated phage functions.
near the limit of genomospecies definition (Jain et al., 2018). These results are fully consistent with results obtained in Gillis and Mahillon (2014b), where it was not possible to pinpoint a clear association between the phage infection pattern and the Bacillus species. The sequencing of both GBJ002 and HER1047 would allow a comparative analysis to obtain some clues about their common features that could establish a pattern for tectiviral susceptibility.

In conclusion, using a combination of short and long-read sequencing methods has proven to be particularly useful to unravel important features in a bacterial species where delineation of plasmids and location of genes is key. Our high-quality assembly of B. thuringiensis HER1410 genome led to the identification of an uncommon location of a cry gene, close to the replication origin of the chromosome, which should be studied further as it may have an effect on the entomopathogenicity of this strain against lepidoptera insects and, therefore, could make HER1410 a promising candidate as a specific biopesticide. Also, the identification of a small integrative plasmidial prophage related to a Bacillus virus found in the human bladder and its role in flagellar gene disruption has a great interest in the Bacillus phages research field and their interaction with the human microbiome. Finally, the genomic characterization of HER1410 will be valuable for studies on Bacillus phages. In particular, projects focused on phages

Figure 5 Overall position of B. thuringiensis HER1410 in the Bacillus cereus-thuringiensis phylogeny. An unrooted tree, representing the phylogenetic position of HER1410 (dark blue) among differentiated genomic sequences of the B. thuringiensis (Bt) and B.cereus (Bc) strains selected (Supplemental Table 5), was visualized with FigTree. A maximum likelihood tree was generated upon multiple sequence alignment of the core genome from selected strains. Labels correspond to their name in NCBI genomic database plus the assembly accession number. The colors represent clearly distinct clades. Strains in the same clade as HER1410 are underlined.
Acknowledgments

This research was funded by Fundación Ramón Areces (XIX Concurso Nacional de Investigación en Ciencias de la Vida y la Materia) and institutional grants from Fundación Ramón Areces and Banco de Santander to the Centro de Biología Molecular Severo Ochoa. A. L. is holder of a PhD fellowship [FPU15/05797] from the Spanish Ministry of Science, Innovation and Universities.


Zheng, J., Q. Gao, L. Liu, H. Liu, Y. Wang et al., 2017 Comparative genomics of Bacillus thuringiensis reveals a path to specialized exploitation of multiple invertebrate hosts. MBio 8: e00822-17. https://doi.org/10.1128/mBio.00822-17


Communicating editor: D. Baltrus