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

Ana Lechuga,*,[1] Cédric Lood,[†,‡,1] 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.

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 adaptative 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. thuringiensi*s 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* pBtoxis 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 (Fiedoruk *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* GBJ002, HER1410, and *B. cereus* HER1047 strains have been used as hosts for tectiviral infections. These three strains represent different serotypes and have been especially useful to study tectiviruses features such as virus entry and host cell physiology (Gaidelyte *et al.*, 2006, Daugelavicius *et al.*, 2007). However, despite their importance to this field of research, none of them have been sequenced to date. In particular, *Bacillus thuringiensis* HER1410 is extensively used as a host of Bam35, the model species of *Betatectivirus*, for the molecular and structural studies on this virus (Ravantti *et al.*, 2003, Laurinmäki *et al.*, 2005, Berjon-Otero *et al.*, 2016). Further, HER1410 has shown high sensitivity to different *Bacillus* phages families. This has allowed the identification of several novel tectiviruses (Gillis and Mahillon, 2014b). Also, this strain can be infected by other tectiviruses and myoviruses (A. Gillis *et al.*, unpublished data). HER1410 mutants resistant to GIL01cp[25], a lytic variant of tectivirus GIL01, are susceptible to myovirus VP4 showing, in this case, the specificity of bacterial resistance (Gillis and Mahillon, 2014c). Besides, HER1410 has been also included in *B. cereus* group studies on hemolytic activity

and temperature dependent growth rates (Francis *et al.*, 1998, Prüss *et al.*, 1999a,b).

This study aims to characterize *Bacillus thuringiensis* HER1410, known as a sensitive host of a large variety of *Bacillus* phages and a reference strain to study tectiviruses. A combination of short and long-read sequencing methods was applied, known to allow the full resolution of bacterial genomes and delineation of plasmids (Albers *et al.*, 2018, T'Syen *et al.*, 2018, Botelho *et al.*, 2019). We successfully resolved the chromosome and identified and characterized three extrachromosomal elements comprising the HER1410 genome and present a detailed overview of these components, focusing on the key elements in *B. cereus* group such as the description of MGEs, virulence factors and phylogeny.

MATERIALS AND METHODS

Bacterial strain

Bacillus 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 of Laval (https://www.phage.ulaval.ca) andcanberetrievedwithHoststrainHERnumber1410.

DNA extraction and genome sequencing

B. thuringiensis HER1410 was inoculated in LB medium and incubated overnight at 37° to an OD_{600} 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 trimmomatic 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).

[1]These authors contributed equally to this work
[2]Deceased.
[3]Corresponding author: Departamento de Bioquímica, Universidad Autónoma de Madrid (UAM) and Instituto de Investigaciones Biomédicas Alberto Sols (CSIC-UAM), School of Medicine, Arzobispo Morzillo, 4 (ES28029), Madrid, Spain. E-mail: modesto.redrejo@uam.es.

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 (Siguier 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 BTyper3 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 phylogeny 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 pangenomes 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 bp_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 IS*4* family. The association of transposases and *B. thuringiensis* toxins, in particular, transposases in the IS*4*, IS*6*, IS*66*, IS*605*, and Tn*3* families is extensively described and highlights 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
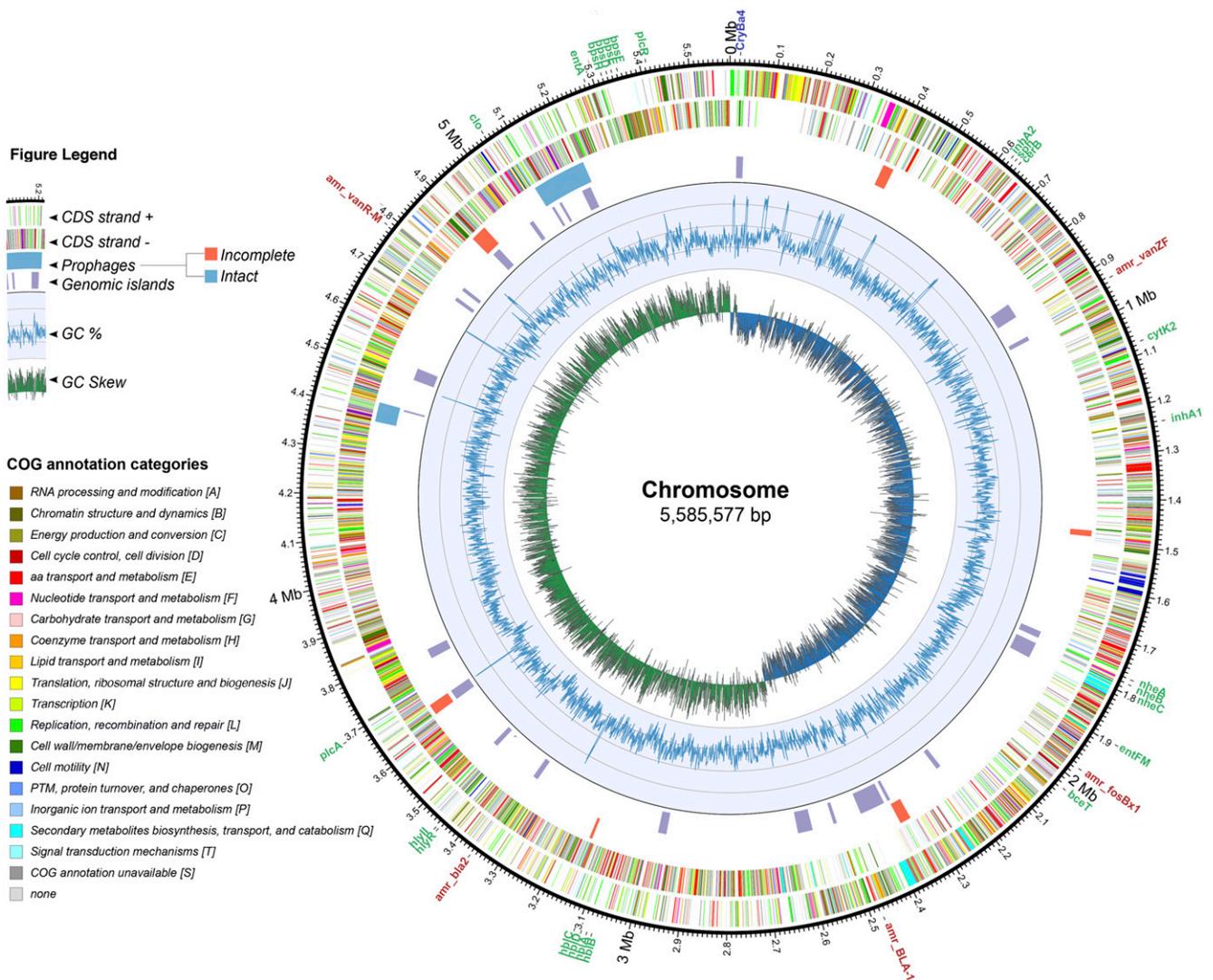
Figure 1 Circular representation of *B. thuringiensis* HER1410 chromosome. The outer labels refer to toxin (green) and antibiotic resistance genes (red), found using the bTyper and bTyper3 tools. The next two circles represent the coding regions on the positive and negative strands colored by their functional annotation. The next two circles indicate the genomic islands predicted by the IslandViewer package (purple) and prophages predicted by PHASTER webserver where orange = incomplete, blue = intact. The inner two rings indicate the changes in %GC content and the GC skew respectively. Coordinates of the highlighted features can be found in Supplemental Table 3. Chromosomal integration of pLUSID3 was excluded from the chromosome sequence and, therefore, this representation.

observed (Figure 2). Also, as in *cry1Ba*-containing plasmids, the HER1410 island only posseses one pesticidal gene (not within a PAI) and it is not preceded by a K$^{\text{p}}$(Na$^{\text{p}}$)/H$^{\text{p}}$ antiporter. These similarities suggest a plasmidial origin of the HER1410 *cry1Ba4*-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, *cry1Ba4* 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* Toxin Nomenclature (Crickmore and Feitelson 2016) and they show highly specific insecticidal activity to target insects. In this case, Cry1Ba4 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 cytolysin 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
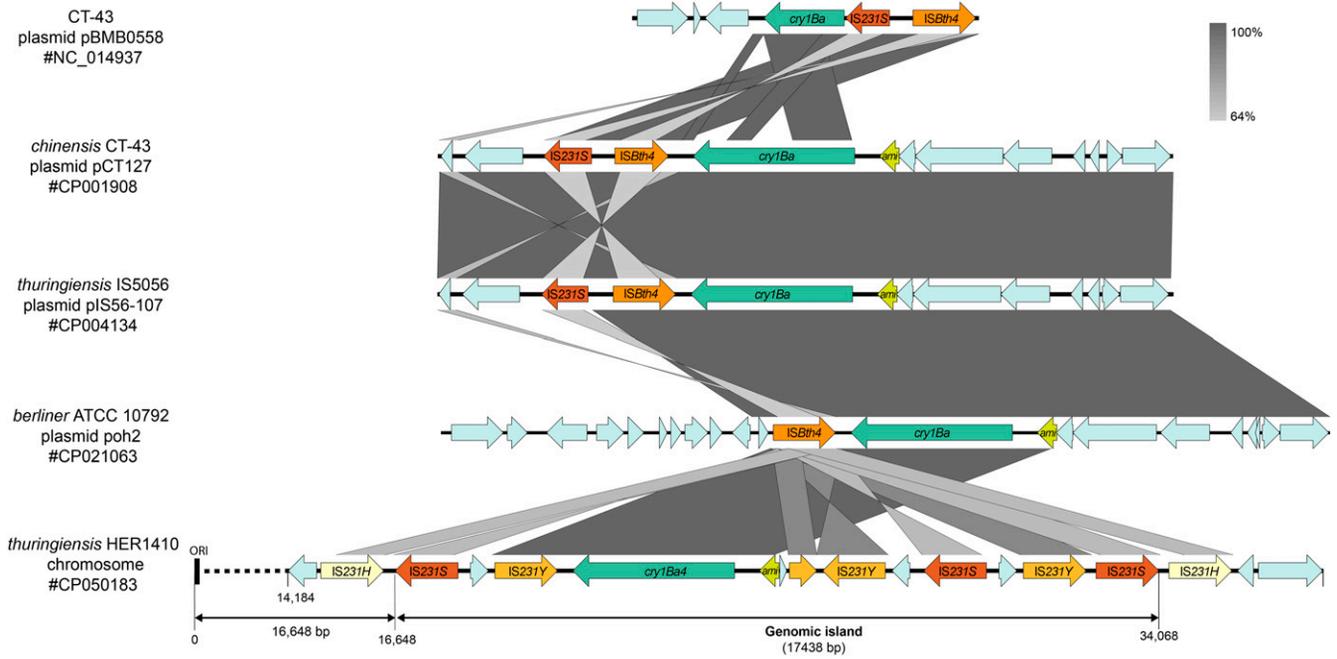
Figure 2 HER1410 chromosomal *cryBa4*-containing genomic island analysis. Comparison of *cry1Ba4*-containing genomic island with described *cry1Ba*-containing plasmid cassettes from Fiedoruk *et al.* (2017) was rendered with EasyFig and further details were added with Adobe Illustrator. Predicted protein-coding genes are indicated with arrows, indicating the direction of transcription. 'ami' stands for N-actylmuramoyl-L-alanine amidase. CDS coding for transposases were annotated using ISFinder (Siguier 2006). The length of the genomic island predicted by the IslandViewer package is indicated as well as the chromosomal origin of replication (ORI). Greyscale for similiarity levels by BLASTn is shown in the top right area of the figure.

(Prüss *et al.*, 1999a, Ehling-Schulz *et al.*, 2019). The presence of antibiotic resistance genes and CRISPR-Cas systems was also searched, using BTyper and CRISPRCasFinder tools. Several vancomycin, beta-lactam and fosfomycin 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 $\pm$ 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 thurincin 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. Homologous 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 pBtoxis 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 20.0 ± 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). BTyper 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 neofunctionalization (Chevrette *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 phage related proteins (Supplemental Table 2). This annotation was updated with results from COG functional analysis (Huerta-Cepas *et al.*, 2017), PHASTER (Arndt *et al.*, 2016) and PHMMER tools (Potter *et al.*, 2018) (Figure 4C, Supplemental Table 4). PHASTER results identified nearly the entire length of the plasmid as a questionable prophage, related to *Bacillus* prophage phBC6A52 (Genbank NC_004821.1). However, only 27% of pLUSID3 displays similarity to this prophage. The whole sequence of pLUSID3, as well as the annotated terminase and major capsid protein, were used as queries for BLASTn and BLASTp respectively against tailed phages NCBI database (taxid:28883). Results yielded similarity of pLUSID3 with phage Wrath (40% coverage, 83.63% identity) a *Bacillus* siphovirus found in the human bladder (Garretto *et al.*, 2019). These results suggest that pLUSID3 is a circular plasmidial prophage, not previously reported, showing partial similarity to characterized *B. cereus* prophages. This is not surprising given the high diversity exhibited by prophages in *B. thuringiensis* (Fu *et al.*, 2019). Strikingly, the structural variation analysis based on the nanopore data allowed us to show that pLUSID3 has a dual lifestyle, as it can be fully integrated into the chromosome but also detected as an excised, extrachromosomal element (Figure 4B). Functional annotation revealed the presence of three different recombinases located in the same region of the plasmid. These recombinases may participate in the integration of this putative prophage into the chromosome and may serve to resolve replication intermediates. It is remarkable that, in the limited clonal community of bacteria used for sequencing, we could detect several integration events of the whole plasmidial prophage that suggest a high mobility rate of this element. Interestingly, the integration was always detected in the same region of the chromosome, corresponding to the location of a flagellar hook associated protein CDS (*FlgK*). It has already been shown that the induction of the integration/excision of prophages in *B. thuringiensis* can modify some bacterial features. This is the case, for example, of phIS3501 that disrupts the production of hlyII toxin when it integrates into the chromosome (Moumen *et al.*, 2012). FlgK is an essential element of the hook-filament junction and its absence causes flagellar misassembly and accumulation in the extracellular milieu of *Bacillus subtilis* (Cairns *et al.*, 2014). The intermittent disruption of this gene by pLUSID3 integration could have an impact on the motility of *B. thuringiensis* HER1410 and biofilm formation (Houry *et al.*, 2010). Together with the presence of putative cro/cI type transcriptional repressors, these results point to a hypothetical superinfection exclusion mechanism. Also, this set of cro/cI type transcriptional regulator genes could be responsible for the induction of this lysogenic phage during the activation of the SOS response.

Phylogenetic positioning of HER1410

HER1410 strain has been classified either as serovar *israelensis* (Daugelavicius *et al.* 2007, Gaidelyte *et al.* 2005, Gaidelyte *et al.*
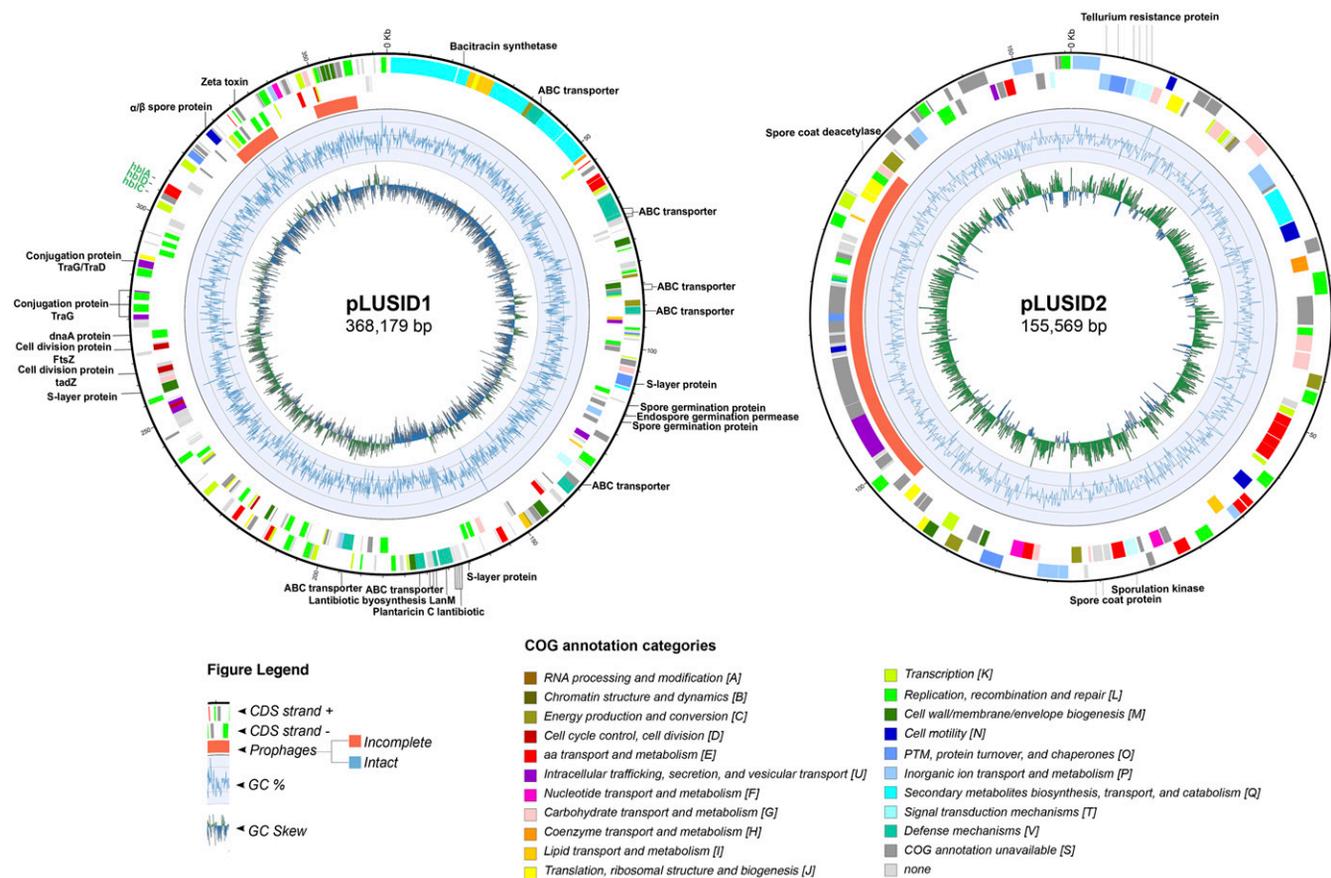
Figure 3 Circular representation of megaplasmids pLUSID1 and pLUSID2. The outer labels refer to toxin genes found using bTyper (green) and loci that are mentioned throughout the paper (black). The next two circles represent the coding regions on the positive and negative strands colored by their functional annotation. The next circle displays prophages predicted by PHASTER webserver where orange = incomplete, blue = intact. The inner two rings indicate the changes in %GC content and the GC skew respectively. Coordinates of the highlighted features can be found in Supplemental Table 3.

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 stablish 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 HER1410, and *B. thuringiensis* YBT-1518 type strain as outgroup and obtained the pangenome of this dataset, with a total of 2,886 core genes. This deeper analysis accurately anchored HER1410 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
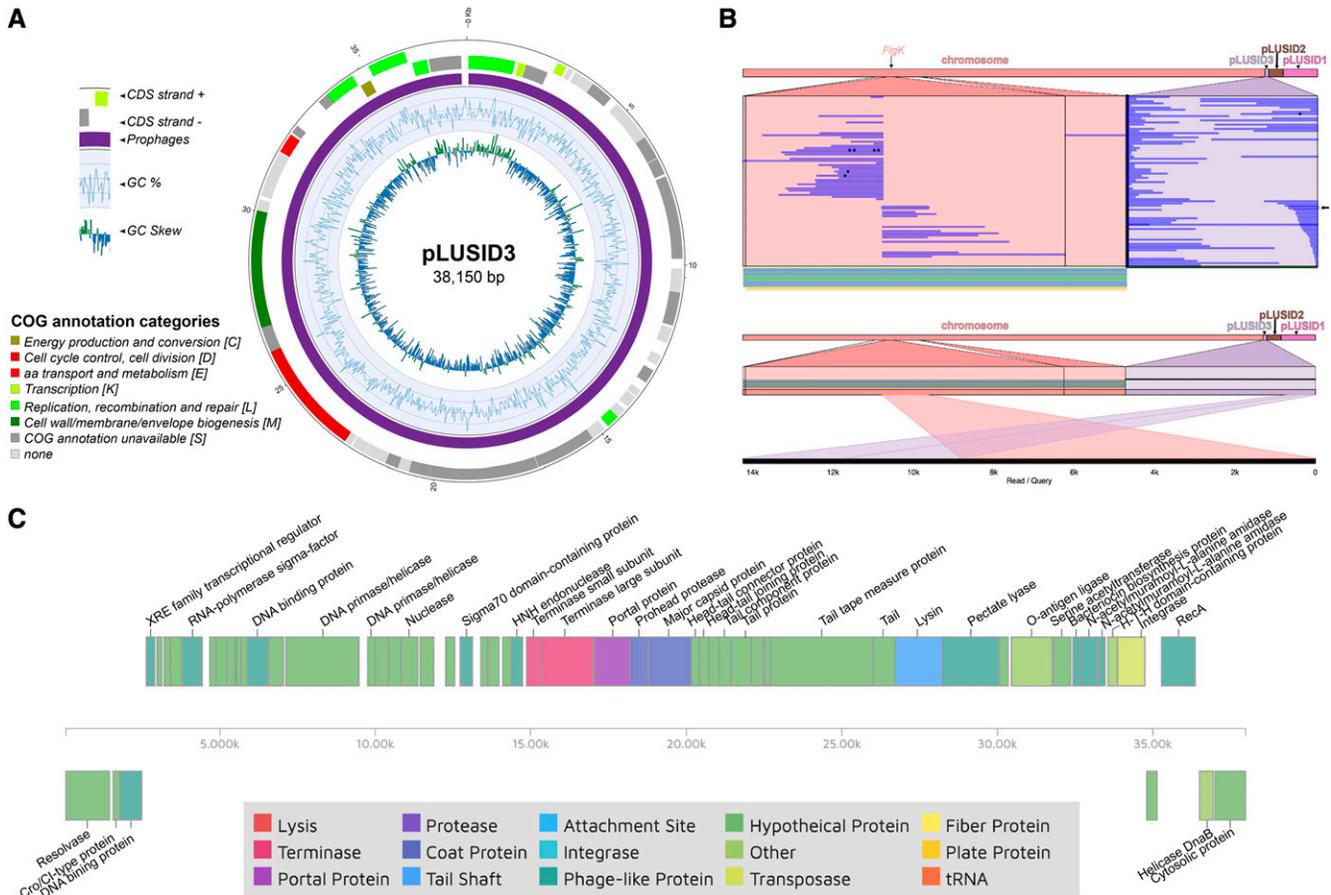
Figure 4 pLUSID3 features analysis (A) Circular representation of plasmid pLUSID3. The outer two circles represent the coding regions on the positive and negative strands colored by their functional annotation. The next circle displays prophages predicted by PHASTER webserver where purple = questionable. The inner two rings indicate the changes in %GC content and the GC skew respectively. (B) Structural variation analysis of pLUSID3. Long-reads from Nanopore sequencing show evidence for pLUSID3 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.

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 pLUSID1, pLUSID2 and pLUSID3 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
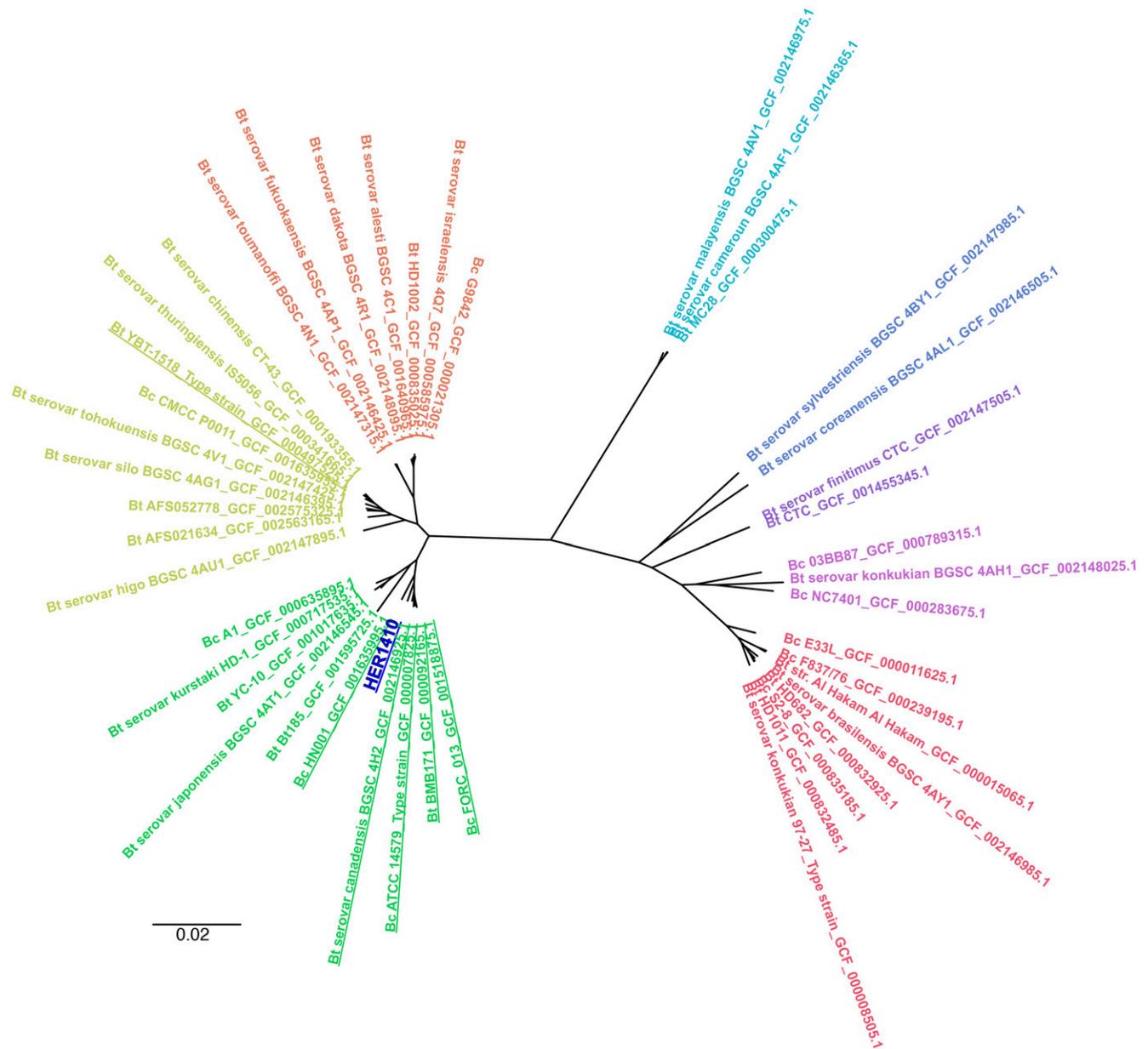
Figure 5 Overall position of *B. thuringiensis* HER1410 in the *Bacillus cereus-thuringiensis* phylogeny. An unrooted tree, representing the phylogenetic position of HER1410 (dark blue) among differentiated genomic sequences of the *B. thuringiensis* (Bt) and *B.cereus* (Bc) strains selected (Supplemental 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.

(95.92%), 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
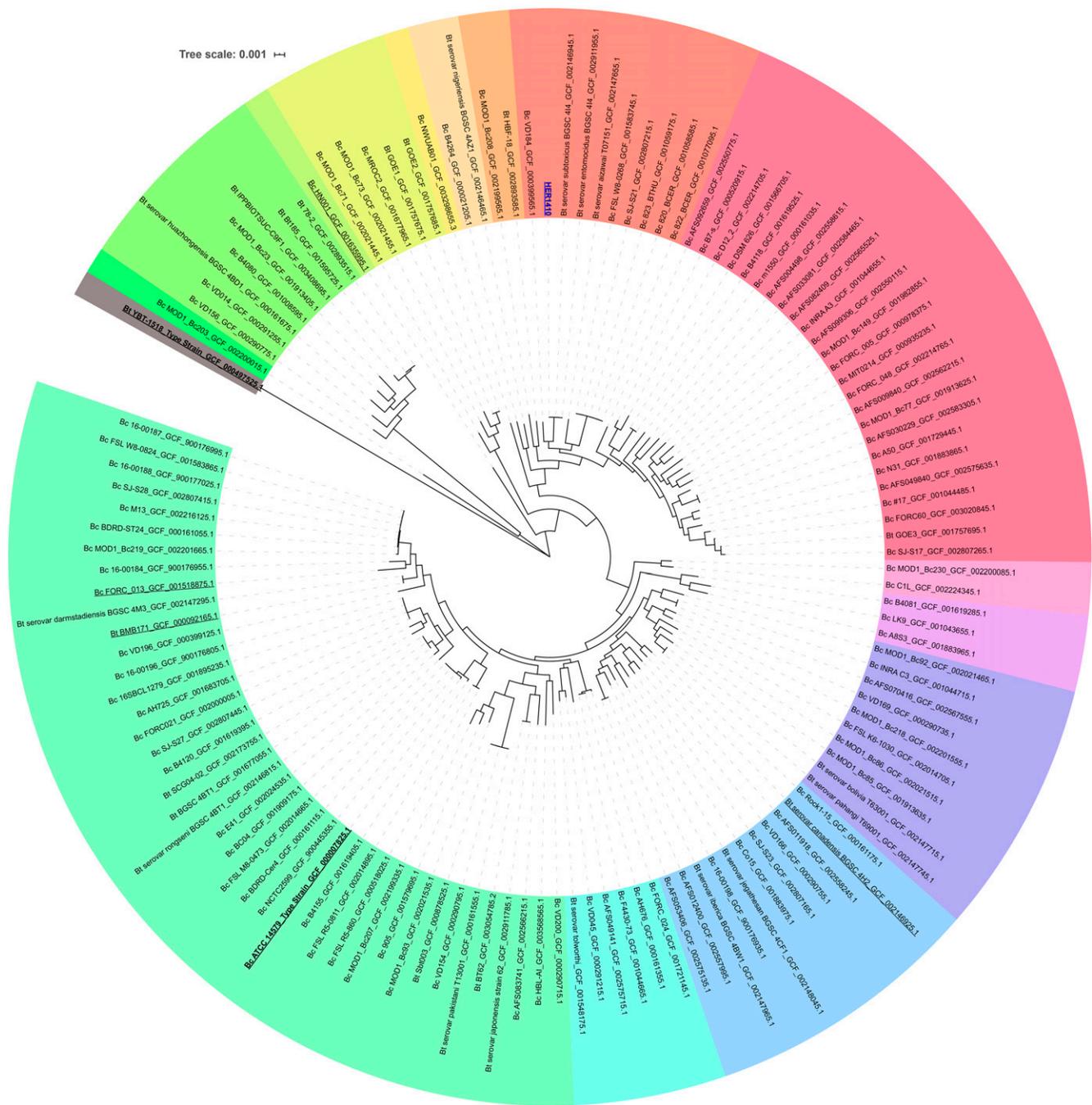
Figure 6 Accurate positioning of *B. thuringiensis* HER1410 in the *Bacillus cereus-thuringiensis* phylogeny. Maximum likelihood rooted tree representing the phylogeny of differentiated genomic sequences of the *B. thuringiensis* (Bt) and *B.cereus* (Bc) strains selected (Supplemental Tables 6) was visualized with iTOL. *Bt* YBT-1518 was used as an outgroup for pangenome analysis and core genome alignment and, subsequentially, to root the obtained tree. Labels correspond to their name in NCBI genomic database plus the assembly accession number. The colors represent clearly distinct clades. Common strains in the two phylogenetical analysis (Figures 5 and 6) are underlined and the strain of interest, HER1410 is colored in blue.

sensitivity of *Bacillus* will benefit from this work and comparative analysis of this strain with the strains HER1047 and GBJ002, which are also very sensitive to tectiviral infections (Gillis and Mahillon, 2014b). Furthermore, a better understanding of this highly sensitive strain to phages, especially to tectiviruses, may improve further investigations on virus-host interactions and *Bacillus* phages characterization.

G3·Genes | Genomes | Genetics

LITERATURE CITED

Adams, V., Li, J., Wisniewski, J. A., Uzal, F. A., Moore, R. J., et al., 2014. Virulence plasmids of spore-forming bacteria. Microbiol. Spectrum. 2: PLAS-0024–2014. https://doi.org/10.1128/microbiolspec.PLAS-0024-2014https://doi.org/10.1128/microbiolspec.PLAS-0024-2014

Albers, P., C. Lood, B. Ozturk, B. Horemans, R. Lavigne et al., 2018 Catabolic task division between two near-isogenic subpopulations co-existing in a herbicide-degrading bacterial consortium: Consequences for the inter-species consortium metabolic model. Environ. Microbiol. 20: 85–96. https://doi.org/10.1111/1462-2920.13994

Andrews, S., 2010. Fastqc: A quality control tool for high throughput sequence data. Available: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arndt, D., J. R. Grant, A. Marcu, T. Sajed, A. Pon et al., 2016 Phaster: A better, faster version of the phast phage search tool. Nucleic Acids Res. 44: W16–W21. https://doi.org/10.1093/nar/gkw387

Baek, I., K. Lee, M. Goodfellow, and J. Chun, 2019 Comparative genomic and phylogenomic analyses clarify relationships within and between Bacillus cereus and Bacillus thuringiensis: Proposal for the recognition of two Bacillus thuringiensis genomovars. Front. Microbiol. 10: 1978. https://doi.org/10.3389/fmicb.2019.01978

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin et al., 2012 Spades: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19: 455–477. https://doi.org/10.1089/cmb.2012.0021

Bazinet, A. L., 2017 Pan-genome and phylogeny of Bacillus cereus sensu lato. BMC Evol. Biol. 17: 176. https://doi.org/10.1186/s12862-017-1020-1

Bergstrom, C. T., M. Lipsitch, and B. R. Levin, 2000 Natural selection, infectious transfer and the existence conditions for bacterial plasmids. Genetics 155: 1505–1519.

Berjon-Otero, M., L. Villar, M. Salas, and M. Redrejo-Rodríguez, 2016 Disclosing early steps of protein-primed genome replication of the gram-positive tectivirus Bam35. Nucleic Acids Res. 44: 9733–9744. https://doi.org/10.1093/nar/gkw673

Berry, C., S. O'Neil, E. Ben-Dov, A. F. Jones, L. Murphy et al., 2002 Complete sequence and organization of pbtoxis, the toxin-coding plasmid of Bacillus thuringiensis subsp. israelensis. Appl. Environ. Microbiol. 68: 5082–5095. https://doi.org/10.1128/AEM.68.10.5082-5095.2002

Bertelli, C., M. R. Laird, K. P. Williams, B. Y. Lau, G. Hoad et al., Simon Fraser University Research Computing Group, 2017 Islandviewer 4: Expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 45: W30–W35. https://doi.org/10.1093/nar/gkx343

Blin, K., S. Shaw, K. Steinke, R. Villebro, N. Ziemert et al., 2019 antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 47: W81–W87. https://doi.org/10.1093/nar/gkz310

Bolger, A. M., M. Lohse, and B. Usadel, 2014 Trimmomatic: A flexible trimmer for illumina sequence data. Bioinformatics 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolotin, A., A. Gillis, V. Sanchis, C. Nielsen-LeRoux, J. Mahillon et al., 2017 Comparative genomics of extrachromosomal elements in Bacillus thuringiensis subsp. israelensis. Res. Microbiol. 168: 331–344. https://doi.org/10.1016/j.resmic.2016.10.008

Botelho, J., C. Lood, S. R. Partridge, V. van Noort, R. Lavigne et al., 2019 Combining sequencing approaches to fully resolve a carbapenemase-encoding megaplasmid in a pseudomonas shirazica clinical strain. Emerg. Microbes Infect. 8: 1186–1194. https://doi.org/10.1080/22221751.2019.1648182

Broderick, N. A., R. M. Goodman, J. Handelsman, and K. F. Raffa, 2003 Effect of Host Diet and Insect Source on Synergy of Gypsy Moth (Lepidoptera: Lymantriidae) Mortality to Bacillus thuringiensis subsp.

kurstaki by Zwittermicin A. Environ. Entomol. 32: 387–391. https://doi.org/10.1603/0046-225X-32.2.387

Cairns, L. S., V. L. Marlow, T. B. Kiley, C. Birchall, A. Ostrowski et al., 2014 Flgn is required for flagellum-based motility by Bacillus subtilis. J. Bacteriol. 196: 2216–2226. https://doi.org/10.1128/JB.01599-14

Carlson, C. R., and A. B. Kolsto, 1993 A complete physical map of a Bacillus thuringiensis chromosome. J. Bacteriol. 175: 1053–1060. https://doi.org/10.1128/JB.175.4.1053-1060.1993

Carroll, L. M., J. Kovac, R. A. Miller, and M. Wiedmann, 2017 Rapid, high-throughput identification of anthrax-causing and emetic Bacillus cereus group genome assemblies via btyper, a computational tool for virulence-based classification of Bacillus cereus group isolates by using nucleotide sequencing data. Appl. Environ. Microbiol. 83: e01096-17. https://doi.org/10.1128/AEM.01096-17

Carroll, L. M., M. Wiedmann, and J. Kovac, 2020 Proposal of a taxonomic nomenclature for the Bacillus cereus group which reconciles genomic definitions of bacterial species with clinical and industrial phenotypes. MBio 11: e00034-20. https://doi.org/10.1128/mBio.00034-20

Cherif, A., S. Chehimi, F. Limem, B. M. Hansen, N. B. Hendriksen et al., 2003 Detection and characterization of the novel bacteriocin entomocin 9, and safety evaluation of its producer, Bacillus thuringiensis ssp. entomocidus HD9. J. Appl. Microbiol., 95: 990–1000. https://doi.org/10.1046/j.1365-2672.2003.02089.x

Chevrette, M. G., K. Gutiérrez-García, N. Selem-Mojica, C. Aguilar-Martínez, A. Yañez-Olvera et al., 2020 Evolutionary dynamics of natural product biosynthesis in bacteria. Nat. Prod. Rep. 37: 566–599. https://doi.org/10.1039/C9NP00048H

Clauwers, C., C. Lood, B. Van den Bergh, V. van Noort, and C. W. Michiels, 2017 Canonical germinant receptor is dispensable for spore germination in Clostridium botulinum group ii strain nctc 11219. Sci. Rep. 7: 15426. https://doi.org/10.1038/s41598-017-15839-y

Couvin, D., A. Bernheim, C. Toffano-Nioche, M. Touchon, J. Michalik et al., 2018 Crisprcasfinder, an update of crisrfinder, includes a portable version, enhanced performance and integrates search for cas proteins. Nucleic Acids Res. 46: W246–W251. https://doi.org/10.1093/nar/gky425

Crickmore, N., and J. Feitelson, 2016. Bacillus thuringiensis toxin nomenclature. Available: http://www.lifesci.sussex.ac.uk/Home/Neil_Crickmore/Bt/

Daugelavicius, R., A. Gaidelyte, V. Cvirkaite-Krupovic, and D. H. Bamford, 2007 On-line monitoring of changes in host cell physiology during the one-step growth cycle of Bacillus phage Bam35. J. Microbiol. Methods 69: 174–179. https://doi.org/10.1016/j.mimet.2006.12.023

De Coster, W., S. D'Hert, D. T. Schultz, M. Cruts, and C. Van Broeckhoven, 2018 Nanopack: Visualizing and processing long-read sequencing data. Bioinformatics 34: 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

Delvecchio, V. G., J. P. Connolly, T. G. Alefantis, A. Walz, M. A. Quan et al., 2006 Proteomic profiling and identification of immunodominant spore antigens of Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis. Appl. Environ. Microbiol. 72: 6355–6363. https://doi.org/10.1128/AEM.00455-06

Ehling-Schulz, M., D. Lereclus, and T. M. Koehler, 2019 The Bacillus cereus group: Bacillus species with pathogenic potential. Microbiol. Spectr. 7. https://doi.org/10.1128/microbiolspec.GPP3-0032-2018

Fiedoruk, K., T. Daniluk, J. Mahillon, K. Leszczynska, and I. Swiecicka, 2017 Genetic environment of cry1 genes indicates their common origin. Genome Biol. Evol. 9: 2265–2275. https://doi.org/10.1093/gbe/evx165

Francis, K. P., R. Mayr, F. von Stetten, G. S. Stewart, and S. Scherer, 1998 Discrimination of psychrotrophic and mesophilic strains of the Bacillus cereus group by pcr targeting of major cold shock protein genes. Appl. Environ. Microbiol. 64: 3525–3529. https://doi.org/10.1128/AEM.64.9.3525-3529.1998

Fu, Y., Y. Wu, Y. Yuan, and M. Gao, 2019 Prevalence and diversity analysis of candidate prophages to provide an understanding on their roles in Bacillus thuringiensis. Viruses 11: 388. https://doi.org/10.3390/v11040388

Gaidelytė, A., S. T. Jaatinen, R. Daugelavicius, J. K. Bamford, and D. H. Bamford, 2005 The linear double-stranded DNA of phage Bam35 enters

lysogenic host cells, but the late phage functions are suppressed. J. Bacteriol. 187: 3521–3527. https://doi.org/10.1128/JB.187.10.3521-3527.2005

Gaidelyte, A., V. Cvirkaite-Krupovic, R. Daugelavicius, J. K. Bamford, and D. H. Bamford, 2006   The entry mechanism of membrane-containing phage Bam35 infecting *Bacillus thuringiensis*. J. Bacteriol. 188: 5925–5934. https://doi.org/10.1128/JB.00107-06

Garretto, A., T. Miller-Ensminger, A. J. Wolfe, and C. Putonti, 2019   Bacteriophages of the lower urinary tract. Nat. Rev. Urol. 16: 422–432. https://doi.org/10.1038/s41585-019-0192-4

Gillis, A., N. Fayad, L. Makart, A. Bolotin, A. Sorokin et al., 2018   Role of plasmid plasticity and mobile genetic elements in the entomopathogen *Bacillus thuringiensis* serovar *israelensis*. FEMS Microbiol. Rev. 42: 829–856. https://doi.org/10.1093/femsre/fuy034

Gillis, A., S. Guo, A. Bolotin, L. Makart, A. Sorokin et al., 2017   Detection of the cryptic prophage-like molecule pbtic235 in *Bacillus thuringiensis* subsp. *israelensis*. Res. Microbiol. 168: 319–330. https://doi.org/10.1016/j.resmic.2016.10.004

Gillis, A., and J. Mahillon, 2014a   Phages preying on *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*: Past, present and future. Viruses 6: 2623–2672. https://doi.org/10.3390/v6072623

Gillis, A., and J. Mahillon, 2014b   Prevalence, genetic diversity, and host range of tectiviruses among members of the *Bacillus cereus* group. Appl. Environ. Microbiol. 80: 4138–4152. https://doi.org/10.1128/AEM.00912-14

Gillis, A., and J. Mahillon, 2014c   An improved method for rapid generation and screening of Bacillus thuringiensis phage-resistant mutants. J. Microbiol. Methods 106: 101–103. https://doi.org/10.1016/j.mimet.2014.08.012

Guidi-Rontani, C., Y. Pereira, S. Ruffie, J.-C. Sirard, M. Weber-Levy et al., 1999   Identification and characterization of a germination operon on the virulence plasmid pxol of *Bacillus anthracis*. Mol. Microbiol. 33: 407–414. https://doi.org/10.1046/j.1365-2958.1999.01485.x

Gurevich, A., V. Saveliev, N. Vyahhi, and G. Tesler, 2013   Quast: Quality assessment tool for genome assemblies. Bioinformatics 29: 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Houry, A., R. Briandet, S. Aymerich, and M. Gohar, 2010   Involvement of motility and flagella in *Bacillus cereus* biofilm formation. Microbiology 156: 1009–1018. https://doi.org/10.1099/mic.0.034827-0

Huerta-Cepas, J., K. Forslund, L. P. Coelho, D. Szklarczyk, L. J. Jensen et al., 2017   Fast genome-wide functional annotation through orthology assignment by eggnog-mapper. Mol. Biol. Evol. 34: 2115–2122. https://doi.org/10.1093/molbev/msx148

Jain, C., L. M. Rodriguez-R, A. M. Phillippy, K. T. Konstantinidis, and S. Aluru, 2018   High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9: 5114. https://doi.org/10.1038/s41467-018-07641-9

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. Von Haeseler, and L. S. Jermiin, 2017   Modelfinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14: 587–589. https://doi.org/10.1038/nmeth.4285

Kronstad, J. W., H. E. Schnepf, and H. R. Whiteley, 1983   Diversity of locations for *Bacillus thuringiensis* crystal protein genes. J. Bacteriol. 154: 419–428. https://doi.org/10.1128/JB.154.1.419-428.1983

Krupovic, M., and E. V. Koonin, 2015   Polintons: A hotbed of eukaryotic virus, transposon and plasmid evolution. Nat. Rev. Microbiol. 13: 105–115. https://doi.org/10.1038/nrmicro3389

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne et al., 2009   Circos: An information aesthetic for comparative genomics. Genome Res. 19: 1639–1645. https://doi.org/10.1101/gr.092759.109

Laurinmäki, P. A., J. T. Huiskonen, D. H. Bamford, and S. J. Butcher, 2005   Membrane proteins modulate the bilayer curvature in the bacterial virus Bam35. Structure 13: 1819–1828. https://doi.org/10.1016/j.str.2005.08.020

Lee, H., J. J. Churey, and R. W. Worobo, 2009   Biosynthesis and transcriptional analysis of thurincin H, a tandem repeated bacteriocin genetic locus, produced by *Bacillus thuringiensis* SF361. FEMS

Microbiol. Lett. 299: 205–213. https://doi.org/10.1111/j.1574-6968.2009.01749.x

Letunic, I., and P. Bork, 2016   Interactive tree of life (itol) v3: An online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44: W242–W245. https://doi.org/10.1093/nar/gkw290

Liu, Y., J. Du, Q. Lai, R. Zeng, D. Ye et al., 2017   Proposal of nine novel species of the *Bacillus cereus* group. Int. J. Syst. Evol. Microbiol. 67: 2499–2508. https://doi.org/10.1099/ijsem.0.001821

Méric, G., L. Mageiros, B. Pascoe, D. J. Woodcock, E. Mourkas et al., 2018   Lineage-specific plasmid acquisition and the evolution of specialized pathogens in *Bacillus thuringiensis* and the *Bacillus cereus* group. Mol. Ecol. 27: 1524–1540. https://doi.org/10.1111/mec.14546

Moumen, B., C. Nguen-The, and A. Sorokin, 2012   Sequence analysis of inducible prophage phis3501 integrated into the haemolysin ii gene of *Bacillus thuringiensis* var *israelensis* atcc35646. Genet. Res. Int. 2012:543286. https://doi.org/10.1155/2012/543286

Nathan, S., D. H. Aziz, and N. M. Mahadi, 2006   Phage displayed *Bacillus thuringiensis* cry1Ba4 toxin is toxic to plutella xylostella. Curr. Microbiol. 53: 412–415. https://doi.org/10.1007/s00284-006-0164-9

Nattestad, M., C.-S. Chin, and M. C. Schatz, 2016   Ribbon: Visualizing complex genome alignments and structural variation. bioRxiv. (Preprint posted October 20, 2016).https://doi.org/10.1101/082123

Nguyen, L. T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh, 2015   Iq-tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32: 268–274. https://doi.org/10.1093/molbev/msu300

Page, A. J., C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter et al., 2015   Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics 31: 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Potter, S. C., A. Luciani, S. R. Eddy, Y. Park, R. Lopez et al., 2018   Hmmer web server: 2018 update. Nucleic Acids Res. 46: W200–W204. https://doi.org/10.1093/nar/gky448

Prüss, B. M., R. Dietrich, B. Nibler, E. Martlbauer, and S. Scherer, 1999a   The hemolytic enterotoxin hbl is broadly distributed among species of the *Bacillus cereus* group. Appl. Environ. Microbiol. 65: 5436–5442. https://doi.org/10.1128/AEM.65.12.5436-5442.1999

Prüss, B. M., K. P. Francis, F. von Stetten, and S. Scherer, 1999b   Correlation of 16s ribosomal DNA signature sequences with temperature-dependent growth rates of mesophilic and psychrotolerant strains of the *Bacillus cereus* group. J. Bacteriol. 181: 2624–2630. https://doi.org/10.1128/JB.181.8.2624-2630.1999

Rambaut, A., 2012. Figtree (version 1.4.0). Available: http://tree.bio.ed.ac.uk/software/figtree/.

Rasko, D. A., M. R. Altherr, C. S. Han, and J. Ravel, 2005   Genomics of the *Bacillus cereus* group of organisms. FEMS Microbiol. Rev. 29: 303–329. https://doi.org/10.1016/j.fmrre.2004.12.005

Ravantti, J. J., A. Gaidelyte, D. H. Bamford, and J. K. Bamford, 2003   Comparative analysis of bacterial viruses Bam35, infecting a gram-positive host, and PRD1, infecting gram-negative hosts, demonstrates a viral lineage. Virology 313: 401–414. https://doi.org/10.1016/S0042-6822(03)00295-2

Raymond, B., and M. B. Bonsall, 2013   Cooperation and the evolutionary ecology of bacterial virulence: The *Bacillus cereus* group as a novel study system. BioEssays 35: 706–716. https://doi.org/10.1002/bies.201300028

Sedlazeck, F. J., P. Rescheneder, M. Smolka, H. Fang, M. Nattestad et al., 2018   Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods 15: 461–468. https://doi.org/10.1038/s41592-018-0001-7

Siguier, P., 2006   Isfinder: The reference centre for bacterial insertion sequences. Nucleic Acids Res. 34: D32–D36. https://doi.org/10.1093/nar/gkj014

Stajich, J. E., D. Block, K. Boulez, S. E. Brenner, S. A. Chervitz et al., 2002   The bioperl toolkit: Perl modules for the life sciences. Genome Res. 12: 1611–1618. https://doi.org/10.1101/gr.361602

Sullivan, M. J., N. K. Petty, and S. A. Beatson, 2011   Easyfig: A genome comparison visualizer. Bioinformatics 27: 1009–1010. https://doi.org/10.1093/bioinformatics/btr039

G3 Genes | Genomes | Genetics

T'Syen, J., B. Raes, B. Horemans, R. Tassoni, B. Leroy *et al.*, 2018 Catabolism of the groundwater micropollutant 2,6-dichlorobenzamide beyond 2,6-dichlorobenzoate is plasmid encoded in aminobacter sp. Msh1. Appl. Microbiol. Biotechnol. 102: 7963–7979. https://doi.org/10.1007/s00253-018-9189-9

Tang, M., D. K. Bideshi, H, W, Park, B. A. Federici, B. A., 2007 Iteron-binding orf157 and ftsz-like orf156 proteins encoded by pbtoxis play a role in its replication in *Bacillus thuringiensis* subsp. *israelensis*. 189: 8053–8058. https://doi.org/10.1128/jb.00908–07https://doi.org/10.1128/jb.00908–07

Tatusova, T., M. DiCuccio, A. Badretdin, V. Chetvernin, E. P. Nawrocki *et al.*, 2016 Ncbi prokaryotic genome annotation pipeline. Nucleic Acids Res. 44: 6614–6624. https://doi.org/10.1093/nar/gkw569

Verheust, C., N. Fornelos, and J. Mahillon, 2005 GIL16, a new gram-positive tectiviral phage related to the *Bacillus thuringiensis* GIL01 and the *Bacillus cereus* pbclin15 elements. J. Bacteriol. 187: 1966–1973. https://doi.org/10.1128/JB.187.6.1966-1973.2005

Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel *et al.*, 2014 Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS One 9: e112963. https://doi.org/10.1371/journal.pone.0112963

Wick, R., 2017. Porechop. Available: https://github.com/rrwick/Porechop.

Wick, R. R., L. M. Judd, C. L. Gorrie, and K. E. Holt, 2017 Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Comput. Biol. 13: e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Wick, R. R., M. B. Schultz, J. Zobel, and K. E. Holt, 2015 Bandage: Interactive visualization of de novo genome assemblies. Bioinformatics 31: 3350–3352. https://doi.org/10.1093/bioinformatics/btv383

Yoon, S.-H., S.-M. Ha, J. Lim, S. Kwon, and J. Chun, 2017 A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie van Leeuwenhoek 110: 1281–1286. https://doi.org/10.1007/s10482-017-0844-4

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

Zheng, J., Z. Guan, S. Cao, D. Peng, L. Ruan *et al.*, 2015 Plasmids are vectors for redundant chromosomal genes in the *Bacillus cereus* group. BMC Genomics 16: 6. https://doi.org/10.1186/s12864-014-1206-5

Communicating editor: D. Baltrus