For publication

Detection of the cryptic prophage-like molecule pBtic235 in *Bacillus thuringiensis* subsp. *israelensis*

Annika Gillis^{a§}, Suxia Guo^{a§‡}, Alexandre Bolotin^{b,c}, Lionel Makart^a,
Alexei Sorokin^{b,c}, Jacques Mahillon^{a*}

^a Laboratory of Food and Environmental Microbiology, Earth and Life Institute, UCL, Croix du Sud 2, L7.05.12, B-1348 Louvain-la-Neuve, Belgium

^b INRA, UMR1319 Micalis, F-78350 Jouy-en-Josas, France

^c AgroParisTech, UMR1319 Micalis, F-78350 Jouy-en-Josas, France

*Correspondence and reprints:

E-mail address: jacques.mahillon@uclouvain.be

Tel: +32 10 473370

Fax: +32 10 473440

[§]A. Gillis and S. Guo contributed equally to this work.

[‡]Present address: S. Guo. Haikou Experimental Station, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

Co-author's e-mails:

annika.gillis@uclouvain.be (A. Gillis)

34473172@qq.com (S. Guo)

alexandre.bolotine@jouy.inra.fr (A. Bolotin)

lionel.makart@uclouvain.be (L. Makart)

alexei.sorokine@jouy.inra.fr (A. Sorokine)

jacques.mahillon@uclouvain.be (J. Mahillon)

Abstract

Bacillus thuringiensis has long been recognized to carry numerous extrachromosomal molecules. Of particular interest are the strains belonging to the *B. thuringiensis* subsp. *israelensis* lineage, as they can harbor at least seven extrachromosomal molecules. One of these elements seems to be a cryptic molecule that may have been disregarded in strains considered plasmid-less. Therefore, this work focused on this cryptic molecule, named pBtic235. Using different approaches that included transposition-tagging, large plasmid gel electrophoresis and Southern blotting, conjugation and phage-induction experiments, in combination with bioinformatics analyses, it was found that pBtic235 is a hybrid molecule of 235,425 bp whose genome displays potential plasmid- and phage-like modules. The sequence of pBtic235 has been identified in all sequenced genomes of *B. thuringiensis* subsp. *israelensis* strains. Here, the pBtic235 sequence was considered identical to that of plasmid pBTHD789-2 from strain HD-789. Despite the fact that the pBtic235 genome possesses 240 putative CDSs, many of them have no homologs in the databases. However, CDSs coding for potential proteins involved in replication, genome packaging and virion structure, cell lysis, regulation of lytic-lysogenic cycles, metabolite transporters, stress and metal resistance, were identified. The candidate plasmidial prophage pBtic235 exemplifies the notable diversity of the extrachromosomal realm found in *B. thuringiensis*.

Keywords: *Israelensis* plasmids; pBtic235; Plasmidial prophage, Biopesticide

1. Introduction

Bacillus thuringiensis is a well-known biopesticide that has been used worldwide for almost 80 years. This rod-shaped Gram-positive bacterium belongs to the group of *Bacillus cereus* sensu lato that also includes, among others, emetic and diarrheic pathotypes of *B. cereus* and the mammal pathogen *Bacillus anthracis* [1, 2]. In many cases, the genetic determinants conferring specific characteristics to the different members of this bacterial group are plasmid-borne and have the capacity to be transferred horizontally [3]. *B. thuringiensis* is rather unique in this bacterial group, since it has adapted its lifestyle as an efficient pathogen of specific insect larvae [4]. One of the peculiarities of *B. thuringiensis* is the extent of its extrachromosomal pool, with strains harboring up to 17 distinct plasmid molecules, with sizes ranging from 2 to 600 kb [5-8]. This high degree of plasmid plasticity also plays a significant role in the insecticidal activity displayed by the distinct subspecies of *B. thuringiensis*. In fact, the main entomopathogenic properties that distinguish *B. thuringiensis* from the others members of the *B. cereus* group are due to large plasmids carrying different arrays of *cry* genes coding for entomopathogenic crystal proteins [9].

Among the numerous subspecies of *B. thuringiensis*, “*israelensis*” (serovar H14) is certainly emblematic since its host spectrum is apparently restricted to dipteran insects like mosquitoes and black flies, vectors of important human diseases such as malaria, dengue, Zika, yellow fever and river blindness. *B. thuringiensis*, subsp. *israelensis* possesses a complex arsenal of dipteran toxins (Cry4Aa, Cry4Ba, Cry10Aa, Cry11Aa, Cyt1Aa, Cyt1Ca and Cyt2Ba, also known as delta-endotoxins), encoded by genes located on the large pBtoxis plasmid (128 kb) [10, 11]. Despite the fact that several *B. thuringiensis* subsp. *israelensis* strains have been widely used as experimental models for toxicity on insect larvae, the contribution of the different plasmids to the bona fide life cycle of this bacterium has long

been overlooked, with the obvious exception of the pBtoxis plasmid. *B. thuringiensis*, subsp. *israelensis* contains at least six other dsDNA extrachromosomal molecules ranging from ~5 to 350 kb [6, 12-14], including one linear plasmidial prophage of ~15 kb [15, 16]. Lately, sequencing efforts have permitted obtaining the draft genome of *B. thuringiensis* subsp. *israelensis* plasmid-less strain 4Q7 (also known as 4Q2-81) [17] and completed the genome assemblies of strains HD-789 [13] and AM65-52. The latter bacterium was shown to harbor nine plasmids, some of them also present in *B. thuringiensis* strain HD 1002 [7], including a large plasmid of 360 kb. All these sequenced strains illustrate the plethora of plasmids that a single *B. thuringiensis* strain can shelter, and suggest how dynamics can be the transfer of these extrachromosomal elements in the *B. cereus* group. The accompanying paper by Bolotin et al. describes in detail the comparative genomics of extrachromosomal elements present in strain AM65-52 and in the *B. thuringiensis* subsp. *israelensis* lineage in general.

In this work, we addressed the question of whether or not a potential cryptic plasmid-like molecule is present in a plasmid-cured derivative strain of *B. thuringiensis* subsp. *israelensis* 4Q7, since a previous report suggested the presence of an extrachromosomal element in this ‘plasmid-less’ strain [18]. Using different approaches such as transposition-mediated plasmid tagging, large plasmid gel electrophoresis and Southern blotting, conjugation and phage-induction experiments as well as bioinformatics analyses, it was found that, indeed, a cryptic prophage-like molecule, namely pBtic235, is not only found in this previously considered ‘plasmid-less’ strain, but in all strains sequenced to date belonging to the *B. thuringiensis* subsp. *israelensis* lineage.

2. Materials and methods

2.1. Bacterial strains, plasmids and culture conditions

All *B. thuringiensis* strains and plasmids used in this study are listed in Table 1. Bacteria were routinely cultured in lysogeny broth (LB) medium (5 g/liter NaCl, 5 g/liter yeast extract, 10 g/liter tryptone), unless otherwise indicated. For agar plates, LB medium was solidified with 1.4% (wt/vol) agar (LB-agar). Antibiotics (Sigma) were added to the culture medium, when appropriate, at the following concentrations (µg/ml): chloramphenicol (Cm), 15; nalidixic acid (Nal), 15; streptomycin (Sm), 100 and tetracycline (Tet), 4.

2.2. Plasmid tagging by transposition and sequencing

Strain GBJ002 was transformed by electroporation with plasmid pEG922 containing the transposon Tn5401, according to the previously described protocol [19]. Transposition events were analyzed after removing the pEG922 temperature-sensitive shuttle vector as described in [20]. Run-off sequencing was performed in a collection of strains that were transposition-positive to retrieve the sequences flanking the transposon insertion sites, using primers Tn5401-1 and Tn5401-1 (Table 2) that hybridize to Tn5401 but are oriented outwards [14].

2.3. Large plasmid gel electrophoresis and Southern blot hybridization

Large plasmid profiles from *B. thuringiensis* subsp. *israelensis* strains were obtained following the procedure described by Andrup et al. [21] as follows: bacteria were grown in 7 ml of brain heart Infusion (BHI) broth for 15 h at 30 °C with shaking (120 rpm). A 2-ml volume of cells was centrifuged (8,000 g for 7 min) and pellets were carefully resuspended in 100 µl E buffer (15% sucrose, 40mM Tris-hydroxide, 2mM EDTA, pH 7.9). Then, 200 µl of lysing solution (3% SDS, 50mM Tris, pH 12.5) were added. The lysates were heated at 60 °C for 60 min, followed by the addition of 20 µl of proteinase K (20 mg/ml). The solutions were

gently inverted and incubated at 37 °C for 60 min. One ml of phenol–chloroform–isoamyl alcohol (25:24:1) was added and the solutions were inverted carefully several times. After centrifugation (8,000 g for 7 min), the upper aqueous layer was subjected to 25 h agarose electrophoresis (80 V, 4 °C). DNA was stained in 1 µg/ml of ethidium bromide in MQ water for 20 min and destained in MQ water up to 7 days (at 4 °C). Plasmid sizes were estimated by comparing with the reference plasmids of strain AND508 (Table 1) [21]. Probe labeling was performed using the pBtic235-repOF and pBtic235-repOF primers (Table 2) according to the kit manufacturer's protocol (PCR DIG probe synthesis kit, Roche). Hybridization was performed as described previously [22].

2.4. PCR screening and sequencing

Total bacterial DNA extractions were screened by PCR for the presence of plasmid pBtic235 using the pairs of primers pBtic235-135k, pBtic235-91k and pBtic235-LamF (Table 2). PCR amplicons were purified using a GenElute PCR cleanup kit (Sigma) and sequenced directly using both strands.

2.5. Plasmid curing

Strain GBJ002 was subjected to plasmid curing using heat. Briefly, 14 mL of liquid LB with nalidixic acid was inoculated with an overnight colony of GBJ002 growing in LB plates at 30 °C. The culture was incubated overnight at 30°C and 120 rpm. Then, 500 µl of the overnight culture were transferred to 50 ml of LB-Nal broth and incubated 4 h at the maximum temperature that allowed growth of strain GBJ002 (i.e. 43 °C). Again, 500 µl of the 4 h culture were transferred to 50 ml LB-Nal broth and incubated 4 h at 43 °C. Finally, 500 µl of the 4 h culture were transferred to 50 ml liquid LB-Nal broth and incubated overnight at 43 °C. These rounds of cycles were performed for 3 days. At each cycle, several LB-Nal plates were inoculated with 100 µl of the culture. All colonies were screened by PCR, using the pair

of primers pBtic235-135k, pBtic235-91k and pBtic235-LamF (Table 2), to evaluate the loss of plasmid pBtic235. With this procedure strain GSX002 was obtained (Table 1).

2.6. Mating tests

Plasmid pBtic235 was tagged using a chloramphenicol resistance (Cm^R) cassette by homologous recombination, and two derivatives strains from strain GBJ002 were obtained: GBJ002 (pBtic235:: Cm^{R1}) and GBJ002 (pBtic235:: Cm^{R2}) (Table 1). Mating experiments were then performed between GBJ002 (pBtic235:: Cm^{R1}) or GBJ002 (pBtic235:: Cm^{R2}) (Nal^R and Cm^R , donor strains) and GBJ001 (Sm^R , recipient strain). Conjugation assays were performed using the drop-on-drop method as previously described [23]. The transconjugants were confirmed by PCR using the pair of primers pBtic235-135k, pBtic235-91k and pBtic235-LamF (Table 2).

2.7. Prophage induction experiments

Prophage induction from strain GBJ002 was carried out by adding 100 ng/ml mitomycin C (AppliChem) to 20 ml of a mid-log-phase bacterial culture as previously described [16, 24]. Phage stocks were assayed on lawns of *B. thuringiensis* strain HER1410 (Table 1). Plaques were purified by at least three rounds of single-plaque purification using strain HER1410 as host. Final lysogens growing on LB-agar were screened by PCR to confirm lysogeny of pBtic235.

2.8. Bioinformatics analyses

Sequences obtained from transposition-based plasmid tagging experiments were compared to those present in the GenBank database using BLASTn searches [25, 26]. Retrieved genomes were aligned using MUSCLE [27], MAUVE [28] and progressive MAUVE [29]. A plasmid pBtic235 inferred sequence was also screened in all *B. cereus* sensu lato sequenced genomes using BLASTn searches. Additionally, predicted coding sequences

169 (CDSs) were analyzed using BLASTn searches [25, 26], and functional annotation and
170 relevant homologs were manually recovered. Prophinder [30] and PHAST [31] software were
171 used to predict the potential prophage region(s) in plasmid pBtic235. The genetic map of
172 plasmid pBtic235 was generated using the CGView program [32], and particular displayed
173 characteristics were manually edited.

174

3. Results and discussion

3.1. Detection of a cryptic molecule in *B. thuringiensis subsp. israelensis* strain GBJ002

To address the possibility that a cryptic extrachromosomal element could be present in the ‘plasmid-less’ strain GBJ002 (derivative of strain 4Q7, Table 1), a plasmid tagging approach by means of the Tet^R transposon Tn5401 was used. To this end, strain GBJ002(pXO16), containing the conjugative plasmid pXO16, was transformed with plasmid pEG922, a temperature-sensitive shuttle vector, containing transposon Tn5401. After a transposition assay in GBJ002(pXO16) [20], mating-out was performed between the “transposants” of GBJ002 (pXO16) (acting as donors) and GBJ001 (Sm^R, acting as recipient). Upon selection on Sm and Tet, the GBJ001 transconjugants were then tested by PCR for the absence of pEG922 using pUC18-derived primers [20]. From this collection of Sm^R-Tet^R transconjugants, the sequences flanking the Tn5401 insertion sites were analyzed by run-off sequencing using outward-oriented hybridizing primers for Tn5401 (Table 2). The sequences were compared to those present in the GenBank database using BLASTn searches. At the time that this work was performed, the only two draft *B. thuringiensis subsp. israelensis* genomes that were available in the GenBank database were those belonging to strains ATCC 35646 (Acc. Nos. AAJM01000001-AAJM01000866, distributed in 867 contigs) [33] and IBL 4222 (Acc. No. ACNL00000000; distributed in 383 contigs) [34]. Interestingly, none of the Tn5401 insertion sites were found in the pXO16 sequence, but rather, they all mapped to a specific contig of strain IBL 4222: contig number 38 with 235,355 bp (Acc. No. ACNL01000230). Transposon Tn5401 insertion site flanking sequences also matched mainly sequences present in five other contigs belonging to the *B. thuringiensis subsp. israelensis* strain ATCC 35646 (Acc. Nos. AAJM01000001, AAJM01000003, AAJM01000016, AAJM01000080 and AAJM01000603). When the five contigs of strain ATCC 35646 were

aligned with contig number 38 from strain IBL 4222, their nucleotide sequences were almost identical, except that there were some gaps among the five contigs of strain ATCC 35646 (Fig. S1a). In addition, some Tn5401 insertion sites flanking sequences showed high nucleotide similarity with regions of the mega-plasmid pBWB401 (~417 kb) from *Bacillus weihenstephanensis* KBAB4 (Acc. No. CP000904) [35] and the jumbo-phage 0305phi8-36 (~219 kb) infecting *B. thuringiensis* (Acc. No. EF583821) [36] (Fig. S1b). All together, these results indicated that the transposon Tn5401 might be inserted in a cryptic plasmid present in *B. thuringiensis* subsp. *israelensis* strain GBJ002.

3.2. Plasmid content in the strain GBJ002: habemus plasmid

To further assess the potential presence of a cryptic plasmid in strain GBJ002, large plasmid extractions were carefully prepared and electrophoresed in agarose gels. Remarkably, a ‘ghost’ band migrating at the same size of the second largest plasmid from *B. thuringiensis* subsp. *israelensis* strain AND508 (~110 MDa, [37]) was detected in strain GBJ002 (Fig. 1). This result supported a previous report where a faint band from the ‘plasmid-less’ strain 4Q7 was sporadically observed migrating in an agarose gel at the same position as the 110-MDa plasmid of strain AND508 [37]. Moreover, a band displaying the same migration as the ‘ghost’ molecule detected in strain GBJ002 was also present in large plasmid extractions of strain IBL 4222, along with three other large plasmids (Fig. S2). The ‘ghost’ plasmid-like molecule in strain GBJ002, and its equivalent in strain IBL 4222, were not observed when total genomic DNA extractions from these two strains were electrophoresed together with large plasmid preparations (Fig. S2). These results indicated that strain GBJ002 harbors a large cryptic plasmid.

3.3. Detection of a large cryptic plasmid by Southern blot hybridization

In order to determine whether contig 38 from strain IBL 4222 could correspond to the sequence of the ‘ghost’ plasmid detected in strain GBJ002, a probe (pBtic235-RepO) that

targets a candidate coding region for the origin of plasmid replication (replication initiator A family protein) was designed based on the sequence of this contig. The probe was used for Southern blot hybridization of large plasmid extractions. As shown in Fig. 2, this analysis confirmed that the sequence of contig 38 from strain IBL 4222 corresponds to the ‘ghost’ plasmid detected in strain GBJ002. Moreover, a derivative strain of GBJ002, designated here as GSX002, that had lost the ‘ghost’ plasmid via heat plasmid-curing procedures, did not show any band or probe hybridization in large plasmid extractions (Fig. 2). *B. thuringiensis* subsp. *israelensis* strains IBL 4222, GBJ002 and GSX002 were also checked by PCR using several primers that targeted various regions of contig 38, using total DNA preparations. For strains IBL 4222 and GBJ002 amplicons of expected sizes were consistently obtained, whereas, for strain GSX002, no PCR amplicons could be obtained (data not shown), confirming the curing of the ‘ghost’ plasmid.

To investigate whether the sequence of contig 38 from strain IBL 4222 corresponds to a circular molecule and whether the sequence was complete, two primers (pBtic235-EndF and pBtic235-EndR, Table 2) were designed based on the 5’- and 3’-ends of the contig 38 sequence. PCR amplicons from strains IBL 4222 and GBJ002 were sequenced and, to our surprise, we obtained a 70-bp sequence that was missing in the released sequence of contig 38 that permitted circularizing the molecule. With these additional 70 bp, the size of the circular ‘ghost’ molecule was calculated to be 235,425 bp. Therefore, this plasmid-like molecule was named pBtic235 (plasmid from *B. thuringiensis* subsp. *israelensis*, circular with 235 kb).

By confirming the circular nature of the ‘ghost’ plasmid, whole genome sequencing efforts completed the genome assembly of *B. thuringiensis* subsp. *israelensis* strain HD-789, where six plasmids were reported, including a circular molecule of 235,425 bp (pBTHD789-2, Acc. No. CP003765) [13] which is identical to the inferred sequence of pBtic235. Moreover, the draft genome of ‘plasmid-less’ strain 4Q7 distributed in 51 contigs has been

recently released [17]. The report for ‘plasmid-less’ strain 4Q7 indicated that contig number 5 (235 kb) (Acc. No. JEOC01000005), encoding mostly hypothetical proteins and phage-related proteins, was specific to 4Q7 [17]. In addition, the *B. thuringiensis* strain HD 1002 complete genome has also been released, revealing the presence of a 235-kb plasmid (HD 1002 plasmid 3; Acc. No. CP009347) [7]. When nucleotide sequences of pBTHD789-2 from strain HD-789, plasmid 3 from strain HD 1002 and contig 5 from strain 4Q7 were aligned against the sequence of pBtic235 (derived from strain IBL 4222 contig 38 plus the 70 bp), the alignments revealed that the sequences were very similar (Fig. S3), indicating that these sequences correspond to the same plasmid: pBtic235. Furthermore, when sequencing several PCR amplicons of pBtic235 from strain GBJ002, obtained with primers targeting the IBL 4222 contig 38 sequence, the sequences were 100% identical to their counterparts in pBTHD789-2.

Taken together, these data indicate that pBtic235 is found in all *B. thuringiensis* subsp. *israelensis* genomes sequenced to date, including those of strains ATCC 35646 and AM65-52 (Acc. No. CP013278). Moreover, pBtic235 is also found in *B. thuringiensis* strain HD 1002, although the authors stated that this strain belongs to the subsp. *thuringiensis*. Nevertheless, based on the information specified in the ‘Technical Bulletin of Cultures available from the U.S. Department of Agriculture’, *B. thuringiensis* strain HD 1002 belongs to subsp. *israelensis* [38]. Our bioinformatics analyses did not detect pBtic235 in other sequenced members of the *B. cereus* group, thus indicating that this molecule is specific to *B. thuringiensis* subsp. *israelensis*. The accompanying paper by Bolotin et al. provides further details about the copy number per chromosome of pBtic235 in strain AM65-52, and supports our findings concerning the presence of this plasmid in all strains belonging to the *B. thuringiensis* subsp. *israelensis* lineage.

3.4. Analysis of the conjugative transfer capabilities and induction of pBtic235

To assess whether pBtic235 is capable of self-transfer to a recipient bacteria by conjugation, pBtic235 was tagged by means of a Cm^R -cassette in two different regions of its genome (Fig. 3) to obtain two derivatives from strain GBJ002. Plasmid conjugation transfer was performed using strains GBJ002 (pBtic235:: Cm^{R1}) and GBJ002 (pBtic235:: Cm^{R2}) as donor strains and strain GBJ001 as recipient. No transconjugants were ever observed for pBtic235 using this experimental setup. However, when GBJ002 (pBtiC235:: Cm^{R1}) and GBJ002 (pBtiC235:: Cm^{R2}) carrying plasmid pXO16 [14, 39] were used as donor strains, and GBJ001 as recipient strain, the mobilization of plasmid pBtic235 was observed at frequencies of 10^{-5} transconjugants per recipient cells. Consequently, pBtic235 is apparently unable to sustain self-transfer, but can take advantage of co-resident plasmid pXO16 to promote its own transfer, a feature that may be essential in natural environments.

Since the sequence of pBtic235 shares similarities with several regions coding for phage structural proteins (see section 3.5), phage induction experiments were performed by means of the DNA-damaging agent mitomycin C and using *B. thuringiensis* HER1410 as host. Remarkably, some turbid plaques were occasionally obtained on lawns of strain HER1410. After single-plaque purification and infection of strain HER1410 with supernatants derived from the purified plaques, a lysogenic strain, namely HER1410/pBtic235, was obtained. Lysogeny was confirmed by PCR using primers targeting the IBL 4222 contig 38 sequence. Furthermore, host strain HER1410 and lysogenic strain HER1410/pBtic235 were subjected to Southern blot hybridization using the previously mentioned probe, confirming that the lysogenic strain harbors pBtic235 and that strain HER1410 was originally lacking this phage-like molecule (Fig. 2).

The fact that pBtic235 is induced by DNA-damaging treatments suggests that this prophage-like molecule might correspond to the previously reported prophage SU-11, induced from a *B. thuringiensis* subsp. *israelensis* strain [40]. Prophage SU-11 was also described as

an extrachromosomal element migrating at a position higher than that of the linear chromosomal fragments when visualized by agarose gel electrophoresis [40]. To answer the question as to whether pBtic235 and SU-11 are the same molecule, more studies are required, including sequencing the phage SU-11 genome.

3.5. General features of pBtic235 genome

Since the sequence of plasmid pBTHD789-2 from strain HD-789 is 100% identical to that inferred for pBtic235, from now on in this paper, the sequence of pBTHD789-2 will be considered as that of pBtic235. According to the NCBI annotation of pBTHD789-2 [13], pBtic235 has a total of 240 CDSs and 17 tRNAs, for a total of 257 genes representing almost 88% of coding capacity. The G+C percentage of pBtic235 is 36.59%, and does not vary significantly along its genome (Fig. 3).

As the sequence of pBTHD789-2 was annotated automatically, and many hypothetical proteins remained in the annotation, all predicted CDSs were functionally analyzed using BLASTp searches against the non-redundant NCBI database. Table 3 lists all CDSs for which functional annotation was available or for which additional relevant homologs were manually recovered. The genome of pBtic235 codes for proteins involved in DNA replication, genome packaging and virion structure, cell lysis, regulation of lytic-lysogenic cycles and plasmid replication, among others (Fig. 3, Table 3). Interestingly, the pBtic235 genome has two modules: one encoded by the sense DNA strand, mainly coding for plasmid replication-related proteins, including potential replication initiators and terminators and some metabolite transporters, stress and metal resistance proteins and a second module, composed of 54 CDSs coding mainly for virion (phage) structure, DNA packaging and lysis proteins, and encoded by the anti-sense DNA strand (Fig. 3). This module, that expands between positions 700-88517, was also predicted using the bioinformatics tool Prophinder [30]. Therefore, the first module is herein referred to as the ‘plasmid-like module’, whereas the second one is referred

to as the ‘phage-like module’. Despite the fact that there are two clearly demarcated regions in the genome of pBtic235, some phage-related proteins can also be found in the plasmid-like region, particularly a prophage antirepressor, which may be involved in regulation of lysogeny and lysogenic cycles of this cryptic prophage-like molecule.

Also, using PHAST, a prophage prediction tool [31], it was possible to distinguish two candidate overlapping regions in the plasmid-like module of pBtic235 that display prophage characteristics. The first region comprises 14.2 kb, including 19 CDSs (that expand between positions 163866-178084) and shares homology with regions present in many *Bacillus* phages, including the myoviruses 0305phi8-36 [36] and BCP8-2 (Acc. No. KJ081346) [41]. This region includes putative phage integrase (CDS 220) and prophage antirepressor (CDS 178). Moreover, it comprises two genes coding for a putative thymidylate synthase (TS) (CDS 143) and dihydrofolate reductase (DHFR) (CDS 144) (Fig. 3) that have been recently proposed as signature genes of the “bastille-like” group of phages in the subfamily *Spounavirinae* of the *Myoviridae* [42]. The second region of 55.3 kb (between positions 173658-228998), includes 70 CDSs. This region also has homology with many *Bacillus* phages, including the two myoviruses mentioned above and the giant temperate siphovirus vB_BanS-Tsamsa [43]. These two candidate regions, however, represent incomplete prophages based on the predictions of the bioinformatics tool used. Thus, it seems that the genome of pBtic235 has dual characteristics that are found both in plasmids and in phages. This type of hybrid molecule has been previously reported in the *B. cereus* group [40, 44-46]. For instance, the siphovirus vB_BceS-IEBH replicates as a circular plasmid in the prophage state and also has plasmid-like and phage-like regions [46]. For more information about phages with a plasmidial prophage state in the *B. cereus* group, readers are referred to a recent review [47].

As pointed out above, some regions of pBtic235 have similarity with the virulent jumbo-myovirus 0305phi8-36 (Fig. S1b). Many of the hypothetical phage and virion structural-related proteins found in pBtic235 (Table 3) display strong homology with their counterparts in 0305phi8-36. Comparative genomics of this jumbo-phage found that the closest homologues for structural or morphogenesis protein-encoding genes and some replication genes are located in the *B. thuringiensis* subsp. *israelensis* ATCC 35646 genome [48], in the same region in which homologies were found in GenBank using the IBL 4222 contig 38 sequence as a query (see section 3.1). In addition, the non-structural genes of 0305phi8-36 include remnants of two replicative systems, suggesting that this phage might have originated by fusion of two ancestral viruses [48], or perhaps, as the case of pBtic235 suggests, the fusion of a plasmid and a prophage. Therefore, it can be speculated that pBtic235 and 0305phi8-36 might share a common origin. Nevertheless, further analyses are required to evaluate this possibility.

3.6. Concluding remarks

Since the discovery of *B. thuringiensis* subsp. *israelensis* in 1976 [49], extensive data has proven its efficacy in controlling mosquitoes, many species of which are vectors of important human diseases. Although this bacterium harbors a plethora of extrachromosomal elements that confer a certain degree of plasticity upon its ‘core-genome’, many of these elements remain poorly characterized, with one exception: plasmid pBtoxis (128 kb) [11]. In the present work, a potential hybrid molecule, pBtic235 (235 kb), was identified, whose genome displays both plasmid- and phage-like modules. This large cryptic prophage-like molecule is present in all *B. thuringiensis* subsp. *israelensis* strains sequenced, and has not been detected in other members of the *B. cereus* group thus far. Although the genomic sequence of pBtic235 was uncovered and some clues about its potential nature were found, the question as to how pBtic235 deals with its dual plasmid-prophage nature persists. Future

373 experiments on pBtic235 will shed light on how the extrachromosomal elements in *B.*
374 *thuringiensis* subsp. *israelensis* participate in the lifestyle of its host bacterium.

375

376

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We are thankful to Dr. P. Martin from USDA-ARS for providing the *B. thuringiensis* subsp. *israelensis* strain IBL 4222. This work was supported by the National Fund for Scientific Research (FNRS) (grants to A.G. and J.M.); the Université catholique de Louvain (grants to S.G. and J.M.) and the Research Department of the *Communauté française de Belgique* (Concerted Research Action) (grants to L.M. and J.M.). A.B. and A.S. are supported by French Agence nationale de la Recherche (ANR) grant (project PathoBactEvol).

References

- [1] Jensen GB, Hansen BM, Eilenberg J, Mahillon J. The hidden lifestyles of *Bacillus cereus* and relatives. *Environ Microbiol* 2003;5:631-40.
- [2] Liu Y, Lai Q, Göker M, Meier-Kolthoff JP, Wang M, Sun Y, et al. Genomic insights into the taxonomic status of the *Bacillus cereus* group. *Sci Rep* 2015;5:14082.
- [3] Kolstø AB, Tourasse NJ, Økstad OA. What sets *Bacillus anthracis* apart from other *Bacillus* species? *Annu Rev Microbiol* 2009;63:451-76.
- [4] Raymond B, Johnston PR, Nielsen-LeRoux C, Lereclus D, Crickmore N. *Bacillus thuringiensis*: an impotent pathogen? *Trends Microbiol.* 2010;18:189-94.
- [5] Lereclus D, Lecadet MM, Ribier J, Dedonder R. Molecular relationships among plasmids of *Bacillus thuringiensis*: conserved sequences through 11 crystalliferous strains. *Mol Gen Genet* 1982;186:391-8.
- [6] Reyes-Ramirez A, Ibarra JE. Plasmid patterns of *Bacillus thuringiensis* type strains. *Appl Environ Microbiol* 2008;74:125-9.
- [7] Johnson SL, Daligault HE, Davenport KW, Jaissle J, Frey KG, Ladner JT, et al. Complete genome sequences for 35 biothreat assay-relevant *Bacillus* species. *Genome Announc* 2015;3.
- [8] Zheng J, Peng D, Ruan L, Sun M. Evolution and dynamics of megaplasmids with genome sizes larger than 100kb in the *Bacillus cereus* group. *BMC Evol Biol* 2013;13:262.
- [9] Aronson A. Sporulation and delta-endotoxin synthesis by *Bacillus thuringiensis*. *Cell Mol Life Sci* 2002;59:417-25.
- [10] González Jr J, Carlton BC. A large transmissible plasmid is required for crystal toxin production in *Bacillus thuringiensis* variety *israelensis*. *Plasmid* 1984;11:28-38.
- [11] Berry C, O'Neil S, Ben-Dov E, Jones AF, Murphy L, Quail MA, et al. Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. *israelensis*. *Appl Environ Microbiol* 2002;68:5082-95.
- [12] Andrup L, Jensen GB, Wilcks A, Smidt L, Hoflack L, Mahillon J. The patchwork nature of rolling-circle plasmids: comparison of six plasmids from two distinct *Bacillus thuringiensis* serotypes. *Plasmid* 2003;49:205-32.
- [13] Doggett NA, Stubben CJ, Chertkov O, Bruce DC, Detter JC, Johnson SL, et al. Complete genome sequence of *Bacillus thuringiensis* serovar *israelensis* strain HD-789. *Genome Announc* 2013;1:e01023-13
- [14] Makart L, Gillis A, Mahillon J. pXO16 from *Bacillus thuringiensis* serovar *israelensis*: Almost 350 kb of terra incognita. *Plasmid* 2015;80:8-15.
- [15] Verheust C, Jensen G, Mahillon J. pGIL01, a linear tectiviral plasmid prophage originating from *Bacillus thuringiensis* serovar *israelensis*. *Microbiology* 2003;149:2083-92.
- [16] Gillis A, Mahillon J. Prevalence, genetic diversity and host range of tectiviruses among members of the *Bacillus cereus* group. *Appl Environ Microbiol* 2014;80:4138-52.
- [17] Jeong H, Park S-H, Choi S-K. Genome sequence of the acrySTALLIFEROUS *Bacillus thuringiensis* serovar *israelensis* strain 4Q7, widely used as a recombination host. *Genome Announc* 2014;2:e00231-14.
- [18] Jensen G andrup L, Wilcks A, Smidt L, Poulsen O. The aggregation-mediated conjugation system of *Bacillus thuringiensis* subsp. *israelensis*: host range and kinetics of transfer. *Curr Microbiol* 1996;33:228-36.
- [19] Peng D, Luo Y, Guo S, Zeng H, Ju S, Yu Z, et al. Elaboration of an electroporation protocol for large plasmids and wild-type strains of *Bacillus thuringiensis*. *J Appl Microbiol* 2009;106:1849-58.
- [20] Baum JA. Tn5401, a new class II transposable element from *Bacillus thuringiensis*. *J Bacteriol* 1994;176:2835-45.
- [21] Andrup L, Barfod KK, Jensen GB, Smidt L. Detection of large plasmids from the *Bacillus cereus* group. *Plasmid* 2008;59:139-43.

- [22] Sambrook J, Russell DW. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY., 2001.
- [23] Guo S, Mahillon J. pGIAK1, a heavy metal resistant plasmid from an obligate alkaliphilic and halotolerant bacterium isolated from the Antarctic Concordia Station confined environment. *PloS one* 2013;8:e72461.
- [24] Gillis A, Mahillon J. Influence of lysogeny of tectiviruses GIL01 and GIL16 on *Bacillus thuringiensis* growth, biofilm formation and swarming motility. *Appl Environ Microbiol* 2014;80:7620-30.
- [25] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403-10.
- [26] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389-402.
- [27] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792-7.
- [28] Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394-403.
- [29] Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one* 2010;5:e11147.
- [30] Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics* 2008;24:863-5.
- [31] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res* 2011.
- [32] Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res* 2008;36:W181-W4.
- [33] Anderson I, Sorokin A, Kapatral V, Reznik G, Bhattacharya A, Mikhailova N, et al. Comparative genome analysis of *Bacillus cereus* group genomes with *Bacillus subtilis*. *FEMS Microbiol Lett* 2005;250:175-84.
- [34] Zwick ME, Joseph SJ, Didelot X, Chen PE, Bishop-Lilly KA, Stewart AC, et al. Genomic characterization of the *Bacillus cereus sensu lato* species: Backdrop to the evolution of *Bacillus anthracis*. *Genome Res* 2012;22:1512-24.
- [35] Lapidus A, Goltsman E, Auger S, Galleron N, Segurens B, Dossat C, et al. Extending the *Bacillus cereus* group genomics to putative food-borne pathogens of different toxicity. *Chem Biol Interact* 2008;171:236-49.
- [36] Thomas JA, Hardies SC, Rolando M, Hayes SJ, Lieman K, Carroll CA, et al. Complete genomic sequence and mass spectrometric analysis of highly diverse, atypical *Bacillus thuringiensis* phage 0305φ8–36. *Virology* 2007;368:405-21.
- [37] Jensen GB, Wilcks A, Petersen SS, Damgaard J, Baum JA andrup L. The genetic basis of the aggregation system in *Bacillus thuringiensis* subsp. *israelensis* is located on the large conjugative plasmid pXO16. *J Bacteriol* 1995;177:2914-7.
- [38] Nakamura LK, Dulmage HT. *Bacillus thuringiensis* cultures available from the U.S. Department of Agriculture. in: USDoAAR Service. (Ed.), Technical Bulletin, 1988, p. 38.
- [39] Timmer S, Modrie P, Minet O, Mahillon J. Plasmid capture by the *Bacillus thuringiensis* conjugative plasmid pXO16. *J Bacteriol* 2009;191:2197-205.
- [40] Kanda K, Ohderaotoshi T, Shimojyo A, Kato F, Murata A. An extrachromosomal prophage naturally associated with *Bacillus thuringiensis* serovar *israelensis*. *Lett Appl Microbiol* 1999;28:305-8.
- [41] Asare P, Bandara N, Jeong T-Y, Ryu S, Klumpp J, Kim K-P. Complete genome sequence analysis and identification of putative metallo-beta-lactamase and SpoIIIE homologs in *Bacillus cereus* group phage BCP8-2, a new member of the proposed Bastille-like group. *Arch Virol* 2015;1-4.
- [42] Asare P, Jeong T-Y, Ryu S, Klumpp J, Loessner M, Merrill B, et al. Putative type 1 thymidylate synthase and dihydrofolate reductase as signature genes of a novel bastille-like group of phages in the subfamily *Spounavirinae*. *BMC Genomics* 2015;16:582.
- [43] Ganz HH, Law C, Schmuki M, Eichenseher F, Calendar R, Loessner MJ, et al. Novel giant siphovirus from *Bacillus anthracis* features unusual genome characteristics. *PloS one* 2014;9:e85972.

- 492 [44] Kanda K, Tan Y, Aizawa K. A novel phage genome integrated into a plasmid in *Bacillus*
493 *thuringiensis* strain AF101. J Gen Microbiol 1989;135:3035-41.
- 494 [45] Inal JM, Karunakaran KV. phi 20, a temperate bacteriophage isolated from *Bacillus anthracis*
495 exists as a plasmidial prophage. Curr Microbiol 1996;32:171-5.
- 496 [46] Smeesters PR, Drèze PA, Bousbata S, Parikka KJ, Timmery S, Hu X, et al. Characterization of a
497 novel temperate phage originating from a cereulide-producing *Bacillus cereus* strain. Res Microbiol
498 2011;162:446-59.
- 499 [47] Gillis A, Mahillon J. Phages preying on *Bacillus anthracis*, *Bacillus cereus* and *Bacillus*
500 *thuringiensis*: past, present and future. Viruses 2014;6:2623-72.
- 501 [48] Hardies SC, Thomas JA, Serwer P. Comparative genomics of *Bacillus thuringiensis* phage
502 0305phi8-36: defining patterns of descent in a novel ancient phage lineage. Virol J 2007;4:97.
- 503 [49] Golberg LJ, Margalit J. A bacterial spore demonstrating rapid larvicidal activity against *Anopheles*
504 *sergentii*, *Uranotaenia unguiculata*, *Culex univittatus*, *Aedes aegypti* and *Culex pipiens*. Mosq News
505 1977;37:355-8.
- 506 [50] Andrup L, Damgaard J, Wassermann K. Mobilization of small plasmids in *Bacillus thuringiensis*
507 subsp. *israelensis* is accompanied by specific aggregation. J Bacteriol 1993;175:6530-6.

508

Table 1. *B. thuringiensis* strains used in this study.

<i>B. thuringiensis</i> strain/plasmid	Subsp.	Relevant features	Source/reference
<i>B. thuringiensis</i> strains			
AND508	<i>israelensis</i>	Derivative of NB31, a commercial strain of serotype H14, cured from its three small plasmids. Used as molecular marker, contains four large plasmids.	[50]
GBJ001	<i>israelensis</i>	Derivative of strain 4Q7. Chromosomal resistance to streptomycin (Sm ^R).	[18]
GBJ002	<i>israelensis</i>	Derivative of strain 4Q7. Chromosomal resistance to nalidixic acid (Nal ^R).	[18]
GBJ002 (pBtic235::Cm ^{R1})	<i>israelensis</i>	Derivative of strain GBJ002. Chloramphenicol resistance (Cm ^R) in plasmid pBtic235 (in phage-like module).	This study
GBJ002 (pBtic235::Cm ^{R2})	<i>israelensis</i>	Derivative of strain GBJ002. Cm ^R in plasmid pBtic235 (in plasmid-like module).	This study
GBJ002 (pXO16)	<i>israelensis</i>	Strain containing pXO16 introduced by conjugation.	This study
GSX002	<i>israelensis</i>	Plasmid-less derivative of strain GBJ002. Chromosomal resistance to Nal.	This study
IBL 4222	<i>israelensis</i>	Isolated from a cat.	USDA-ARS (USA)/ [34]
HER1410	<i>thuringiensis</i>	Used as a host strain for pBtic235.	Félix d'Hérelle Center for Bacterial Viruses (Canada)
HER1410/pBtic235	<i>thuringiensis</i>	Lysogen for pBtic235.	This study
Plasmids			
pEG922	-	Tn5401-based transposon vector	[20]
pBtic235	-	A 235-kb prophage-like plasmid from <i>B. thuringiensis</i> subsp. <i>israelensis</i>	This study
pXO16	-	A 350-kb conjugative plasmid from <i>B. thuringiensis</i> subsp. <i>israelensis</i> .	[37]

Table 2. Primer pairs used in this study.

Primer	Sequence (5' to 3')	Expected amplicon size (bp)
pBtic235-135kF	GGGCTATTTGCGTTTGAAGA	622
pBtic235-135kR	CAATATGCGTATGGGACCAA	
pBtic235-91kF	GAATTGTGCCCCGATACTCGT	572
pBtic235-91kR	CGGTCGCTTCCCTACTATCA	
pBtic235-LamF	GTCGTCACGGTCAGCTGTAA	526
pBtic235-LamR	CATCAGTAGCGACTGCAGGA	
pBtic235-EndF	ATGGCTATGCAACAACCACA	-
pBtic235-EndR	TGGTATTGACTGCAGCTTCG	
pBtic235-RepOF	ACCGGGGGTAGGAGAAAATA	-
pBtic235-RepOR	AGCGAAAAATGCTGGTTCAT	
Tn5401-1	CGAATAATGTCCGCTAATGC	-
Tn5401-2	GGACAAGAAGAGCAATTAGGT	

518

519 **Table 3.** Functional annotations available on NCBI and additional relevant homologs or
 520 domains manually recovered for pBtic235.

521

Locus tag (NCBI)	CDS No.	Coding strand	Start/end	Product name or identified conserved homologs when available from NCBI (acc. No. CP003765)	Additional relevant homologs or domains ^a
BTF1_30592	4	-	4552/5766	-	Virion structural protein
BTF1_30602	6	-	6744/7493	-	Virion structural protein
BTF1_30617	9	-	8271/8933	-	Hypothetical phage protein
BTF1_30622	10	-	11757/18029	Phage-related protein, prophage LambdaBa01, minor structural protein	Carbohydrate-binding protein CenC; tail fiber protein
BTF1_30627	11	-	18130/21648	-	Phage protein
BTF1_30637	13	-	21946/26460	-	Virion structural protein
BTF1_30647	15	-	29020/32319	Peptidoglycan-binding LysM	-
BTF1_30662	18	-	35699/37282	Low copy number virion structural protein	-
BTF1_30667	19	-	37299/38099	Baseplate protein	-
BTF1_30672	20	-	38099/38587	-	Hypothetical phage protein
BTF1_30677	21	-	38591/39160	-	Hypothetical phage protein
BTF1_30692	24	-	39773/41095	M23 peptidase domain-containing protein	-
BTF1_30697	25	-	41070/43121	Baseplate hub protein	<i>N</i> -acetylmuramoyl-L-alanine amidase
BTF1_30702	26	-	43198/50571	Nuclease	Baseplate hub protein
BTF1_30712	28	-	57966/64448	-	Virion structural protein
BTF1_30717	29	-	64463/65038	Tail chaperonin <i>B. cereus</i> group	-
BTF1_30727	31	-	65610/66326	-	Virion structural protein
BTF1_30732	32	-	66332/66901	Virion structural protein	-
BTF1_30737	33	-	66935/67537	Major structural protein	-
BTF1_30742	34	-	67609/69324	Phage tail sheath family protein	-
BTF1_30747	35	-	69371/69991	-	Virion structural protein
BTF1_30757	37	-	71484/72023	Virion structural protein	-
BTF1_30762	38	-	72035/73648	Endoglucanase-like protein	Bacterial Ig-like domain protein
BTF1_30772	40	-	74175/75278	Phage major capsid protein, HK97	-
BTF1_30777	41	-	75478/76734	-	Major virion structural protein
BTF1_30782	42	-	76766/78163	-	Protease/scaffold
BTF1_30787	43	-	78166/80271	Phage portal family protein	-
BTF1_30797	45	-	81705/84599	DNA helicase	Terminase large subunit
BTF1_30802	46	-	84615/85061	-	Hypothetical phage protein
BTF1_30822	50	-	87110/87814	-	Hypothetical phage protein
BTF1_30827	51	-	87855/88127	-	Hypothetical phage protein
BTF1_30832	52	-	88260/88517	gp48	Phage family protein
BTF1_30847	55	+	94051/95448	Recombinase D	ATP-dependent RecD-like DNA helicase
BTF1_30852	56	+	95908/96654	-	Putative repetitive glutamine-rich protein

BTF1_30892	64	+	100158/100403	-	Ubiquinone-binding protein
BTF1_30932	72	+	103835/104092	-	Hypothetical phage protein
BTF1_30942	74	+	104907/105458	-	Phage protein
BTF1_30957	77	+	105971/107494	tRNA nucleotidyltransferase	-
BTF1_30992	84	+	112590/112889	Regulatory protein SpoVG	-
BTF1_31012	88	-	115431/116582	Replication initiator A family protein	-
BTF1_31017	89	+	117285/117719	Regulatory protein spx	ArsC family protein
BTF1_31027	91	+	118147/118407	-	Putative reductase
BTF1_31052	96	+	123600/124475	UBA/THIF-type NAD/FAD binding protein	-
BTF1_31057	97	+	124459/125145	Ribonuclease HI	-
BTF1_31062	98	+	125136/125555	OB-fold tRNA/helicase-type nucleic acid binding protein	-
BTF1_31082	102	+	127994/129535	DNA polymerase III subunit alpha	-
BTF1_31087	103	+	129678/130445	-	Intein-containing protein
BTF1_31092	104	+	130745/132607	DNA polymerase III DnaE	-
BTF1_31097	105	+	132647/133114	-	DNA-binding protein
BTF1_31102	106	+	133164/133403	-	Hypothetical phage protein
BTF1_31107	107	+	133465/135435	NAD-dependent DNA ligase LigA	-
BTF1_31112	108	+	135824/136522	RNA-directed DNA polymerase	Reverse transcriptase family protein
BTF1_31117	109	+	137045/138307	RNA-directed DNA polymerase	Group II intron reverse transcriptase/maturase
BTF1_31122	110	+	138365/138754	Reverse transcriptase	-
BTF1_31127	111	+	138814/139800	-	Appr-1-p processing domain protein
BTF1_31132	112	+	139818/140969	GIY-YIG catalytic domain-containing protein	Group I intron endonuclease family protein; hypothetical phage protein
BTF1_31137	113	+	140973/141710	tRNA ^{His} guanylyltransferase family protein	Hypothetical phage protein; putative phage tRNA-His guanylyltransferase
BTF1_31142	114	+	141707/142027	-	ABC transporter substrate-binding protein
BTF1_31157	117	+	142870/143277	-	Peptide ABC transporter substrate-binding protein
BTF1_31172	120	+	144773/145870	Tetratricopeptide domain-containing protein	Response regulator aspartate phosphatase
BTF1_31182	122	+	147071/147346	AbrB family transcriptional regulator	-
BTF1_31187	123	+	147387/147785	-	DNA primase
BTF1_31192	124	+	147782/148429	-	Tetratricopeptide repeat protein
BTF1_31212	128	+	150202/151044	ThiF family protein	-
BTF1_31222	130	+	153966/154949	<i>N</i> -acetylmuramoyl-L-alanine amidase	Putative phage alanine amidase
BTF1_31242	133	+	158878/159462	Thymidine kinase	-
BTF1_31252	135	+	159997/160533	Single-strand binding protein family	Putative phage single-stranded DNA binding protein
BTF1_31262	137	+	160893/161981	DNA recombination protein RecA	-
BTF1_31277	140	+	163866/165482	-	Phage protein
BTF1_31282	141	+	165594/166673	-	ATPase associated with various cellular activities; putative phage AAA superfamily ATPase

BTF1_31292	143	+	167181/168041	Thymidylate synthase	-
BTF1_31297	144	+	168038/168556	Dihydrofolate reductase	Putative phage dihydrofolate reductase
BTF1_31302	145	+	168559/168768	-	Chaperone protein DnaJ
BTF1_31307	146	+	168862/169752	-	Phage-related protein
BTF1_31312	147	+	169864/170079	Holin	Putative phage holin
BTF1_31317	148	+	170229/172190	ATP-dependent DNA helicase PcrA	-
BTF1_31322	149	+	172221/172703	Deoxyuridine 5'-triphosphate nucleotidohydrolase Dut	dUTP diphosphatase
BTF1_31327	150	+	172700/173314	Thymidylate kinase	-
BTF1_31332	151	+	173318/173797	-	Hypothetical phage protein
BTF1_31342	153	+	174093/174842	-	Hypothetical phage protein
BTF1_31347	154	+	174965/175678	-	Membrane protein
BTF1_31352	155	+	176019/176210	-	LysR family transcriptional regulator
BTF1_31367	158	+	177299/178084	-	Phage 3D domain-containing protein
BTF1_31372	159	+	178109/179896	SNF2-like protein	Superfamily II DNA/RNA helicase
BTF1_31382	161	+	180973/181680	Two component transcriptional regulator	-
BTF1_31387	162	+	181655/183148	Histidine kinase	-
BTF1_31392	163	+	183148/184674	Sensor histidine kinase	-
BTF1_31397	164	+	184754/185050	-	Sensor histidine kinase
BTF1_31402	165	+	185047/185421	-	Mannose-1-phosphate guanylyltransferase
BTF1_31412	167	+	185771/186430	Deoxynucleoside kinase	-
BTF1_31427	170	+	187840/188571	Metal-dependent hydrolase	Metallo-beta-lactamase phage superfamily protein
BTF1_31432	171	+	188590/189267	GTP pyrophosphokinase/guanosine-3',5'-bis (diphosphate) 3-pyrophosphohydrolase	Hypothetical phage protein
BTF1_31447	174	-	190348/190830	Holliday junction resolvase	<i>Lactococcus</i> phage M3 family protein
BTF1_31467	178	+	192820/193635	Phage antirepressor KilAC domain protein	-
BTF1_31487	182	+	195512/195835	-	Hypothetical phage protein
BTF1_31492	183	+	195810/196043	-	Hypothetical phage protein
BTF1_31497	184	+	196043/196297	-	Hypothetical phage protein
BTF1_31507	186	+	197115/197306	-	Hypothetical phage protein
BTF1_31517	188	+	197766/198194	Nucleoside 2-deoxyribosyltransferase	-
BTF1_31527	190	+	198477/199304	-	Phage protein
BTF1_31537	192	-	199645/200070	-	Hypothetical phage protein
BTF1_31542	193	-	200096/200899	-	Phage RecB protein
BTF1_31557	196	+	201735/202175	-	DNA-binding response regulator
BTF1_31562	197	+	202296/203018	-	Hypothetical phage protein
BTF1_31567	198	+	202993/204108	-	Phage primase
BTF1_31572	199	+	204135/205340	DnaB, putative	Phage dnaB
BTF1_31582	201	+	205609/206046	-	Hypothetical phage protein
BTF1_31587	202	+	206068/206502	Glutamyl-tRNA amidotransferase	GatB/Yqey domain protein
BTF1_31592	203	+	206557/207522	Ribose-phosphate pyrophosphokinase	Putative phage ribose-phosphate pyrophosphokinase
BTF1_31597	204	+	207549/209084	Nicotinate phosphoribosyltransferase	Putative phage nicotinamide

					phosphoribosyl transferase
BTF1_31602	205	+	209192/210073	Excinuclease ABC subunit C	Hypothetical phage protein
BTF1_31607	206	+	210267/210665	-	Hypothetical phage protein
BTF1_31617	208	+	210887/211537	Tellurium resistance protein TerE	Tellurium phage resistance protein TerE
BTF1_31622	209	+	211589/212191	Tellurium resistance protein TerD	-
BTF1_31632	211	+	212557/213306	Tellurium resistance protein terC	Membrane protein
BTF1_31637	212	+	213326/214396	Phosphoesterase	-
BTF1_31642	213	+	214416/215141	NAD-dependent deacetylase	-
BTF1_31647	214	+	215351/215872	-	Hypothetical phage protein
BTF1_31652	215	+	215958/217142	CNT family transporter	Pyrimidine nucleoside transporter NupC
BTF1_31657	216	+	217376/218272	Peptidase S8	Thermitase; Subtilisin Carlsberg
BTF1_31662	217	+	218402/219040	-	Sigma-70, region 4 family; phage protein
BTF1_31667	218	+	219369/219755	Replication terminator protein	-
BTF1_31677	220	-	220090/221187	Site-specific tyrosine recombinase XerS	Integrase
BTF1_31682	221	+	221372/221734	Ribonucleotide reductase stimulatory protein	-
BTF1_31687	222	+	221716/223800	Ribonucleotide-diphosphate reductase subunit alpha	Hypothetical phage protein
BTF1_31692	223	+	223819/224847	Ribonucleotide-diphosphate reductase subunit beta	Hypothetical phage protein
BTF1_31697	224	+	224880/225131	Thiol reductase thioredoxin	SPBc2 prophage-derived thioredoxin-like protein yosR
BTF1_31712	227	+	225877/226167	-	Hypothetical phage protein
BTF1_31717	228	+	226169/226423	-	-
BTF1_31722	229	+	226420/227181	-	Putative phage nucleotidyltransferase
BTF1_31727	230	+	227301/228533	Type VII secretion protein	-
BTF1_31732	231	+	228546/228998	SMI1 / KNR4 family protein	1,3-beta-glucan synthase regulator
BTF1_31737	232	+	229721/230023	-	Hypothetical phage protein
BTF1_31777	240	+	234076/235323	-	Virion structural protein

522 ^a Similar hypothetical proteins found in phages are indicated.

523

Figure legends

Fig. 1. Large plasmid electrophoresis profiles of *B. thuringiensis* subsp. *israelensis* strains AND508 and GBJ002. Strain AND508, containing four large plasmids, is used as reference for plasmid sizes [50]. The sizes of plasmids pXO16 and pBtoxis are indicated. The cryptic plasmid in strain GBJ002 (aka pBtic235) is highlighted by the black arrow. chr, chromosomal DNA.

Fig. 2. Large plasmid electrophoresis profiles and Southern blot patterns of *B. thuringiensis* strains. **a)** Plasmid profiles; **b)** Southern blot hybridization patterns with a probe targeting a fragment of the sequence coding for the putative replication initiator protein. *B. thuringiensis* strains: lane **1** and508; lane **2**, GBJ002 (harboring pBtic235, 235kb); lane **3**, GSX002 (plasmid-less strain, derivative of GBJ002); lane **4**, HER1410 (*B. thuringiensis* subsp. *thuringiensis* wild type strain, harboring one single plasmid larger than pBtic235); lane **5**, HER1410/pBtic235 (lysogen containing pBtic235); lane **6**, IBL 4222 (*B. thuringiensis* subsp. *israelensis* wild type strain). The sizes of plasmids pXO16, pBtic235 and pBtoxis are indicated. chr, chromosomal DNA. The black triangles highlight the positive signal of probe hybridization against pBtic235.

Fig. 3. Circular genetic map of pBtic235. The block arrows in the outer circles indicate the predicted CDSs in their direction of transcription. The black circle represents the GC content plotted using a sliding window, as the deviation from the average GC content of the entire sequence. The green/magenta circles represent the GC-skew calculated using a sliding window, as $(G-C)/(G+C)$ and plotted as the deviation from the average GC skew of the entire sequence. The light and dark gray semi-circles indicate the phage- and plasmid-like modules,

respectively. CDSs in blue represent genes coding for virion structural-related proteins, whereas those in rose code for proteins found in phages. CDSs coding for proteins associated with lytic functions are indicated in green. tRNAs are presented as red block arrows. Remaining predicted CDSs (with or without functional annotation or relevant homologs) are indicated in light purple. Predicted genes coding for interesting proteins are indicated (for more information, see Table 3). Putative thymidylate synthase (TS) and dihydrofolate reductase (DHFR) are highlighted in orange. Sites used for plasmid tagging with chloramphenicol resistance (Cm^{R}) are highlighted in red. Map was generated by CGView [32] using the sequence of *B. thuringiensis* subsp. *israelensis* HD-789 plasmid pBTHD789-2 (Acc. No. CP003765) [13].

Supplementary figure legends

Fig. S1. Progressive MAUVE [29] alignments. **a)** Alignment of *B. thuringiensis* subsp. *israelensis* IBL 4222 contig 38 sequence (Acc. No. ACNL01000230) and five contigs belonging to *B. thuringiensis* subsp. *israelensis* ATCC 35646 (contigs were ordered as follows: Acc. Nos. AJM01000016 (reverse complement), AAJM01000003 (reverse complement), AAJM01000001, AAJM01000080 (reverse complement) and AAJM01000603). The order of the contigs from strain ATCC 35646 was previously determined by the function ‘ordering contigs’ available in MAUVE [28] and verified by MUSCLE alignments [27] using the sequence of IBL 4222 contig 38 as template. The black double square represents the same genomic region in both sequences. **b)** Alignment of IBL 4222 contig 38 sequence against the *B. weihenstephanensis* KBAB4 plasmid pBWB401 (Acc. No. CP000904) [35] and *B. thuringiensis* myovirus 0305phi8-36 (Acc. No. EF583821) [36].

Fig. S2. Large plasmid electrophoresis profiles of *B. thuringiensis* strains. Lane **1** and 508 (containing four large plasmids, used as reference for plasmid sizes [50]); lane **2**, GBJ002 (harboring pBtic235, 235kb); lane **3**, HER1410 (*B. thuringiensis* subsp. *thuringiensis* wild type strain, harboring one single plasmid larger than pBtic235); lane **4**, HER1410/pBtic235 (lysogen containing pBtic235); lane **5**, IBL 4222 (*B. thuringiensis* subsp. *israelensis* wild type strain); lane **6**, GBJ002 (pBtiC235::Cm^{R1}); lane **7**; IBL 4222 total genomic DNA preparation; lane **8**, GBJ002 total genomic DNA preparation. The plasmid sizes are indicated. chr, chromosomal DNA.

Fig. S3. Progressive MAUVE [29] alignment of different pBtic235 sequences retrieved from *B. thuringiensis* strains for which the genomes were available in GenBank. **1)** Plasmid

585 pBTHD789-2 from *B. thuringiensis* subsp. *israelensis* HD-789 (Acc. No. CP003765) [13]; **2)**
586 plasmid pBtic235 sequence derived from *B. thuringiensis* subsp. *israelensis* IBL 4222 contig
587 38 and completed in this study; **3)** plasmid pAM65-52-3-235K from *B. thuringiensis* subsp.
588 *israelensis* AM65-52 (Acc. No. CP013278, see accompanying paper by Bolotin *et al.*); and **4)**
589 plasmid 3 from *B. thuringiensis* HD 1002 (Acc. No. CP009347) [7]. Sequences of *B.*
590 *thuringiensis* subsp. *israelensis* IBL 4222 contig 38 (**5**) and the five contigs belonging to *B.*
591 *thuringiensis* subsp. *israelensis* ATCC 35646 (**6**) were added to facilitate comparisons with
592 Fig. S1. Colors highlight the same genomic regions across the different sequences.

Figure 1

Figure 2

Figure 3