For publication

Comparative genomics of extrachromosomal elements in *Bacillus*
thuringiensis subsp. *israelensis*

Alexandre Bolotin^a, Annika Gillis^b, Vincent Sanchis^a, Christina Nielsen-LeRoux^a,
Jacques Mahillon^b, Didier Lereclus^a, Alexei Sorokin^{a*}

^a Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

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

*Correspondence and reprints:

Alexei Sorokin
INRA MICALIS,
CRJ Jouy-en-Josas,
Domaine de Vilvert, bat. 440, niveau -1, pièce 325;
F-78352 Jouy-en-Josas cedex
France.
Phone: 33 1 34 65 27 24.
Fax: 33 1 34 65 25 21.
E-mail: alexei.sorokine@jouy.inra.fr

Co-author's e-mails:

alexandre.bolotine@jouy.inra.fr
annika.gillis@uclouvain.be
vincent.sanchis@jouy.inra.fr
christina.nielsen@jouy.inra.fr
jacques.mahillon@uclouvain.be
didier.lereclus@jouy.inra.fr
alexei.sorokine@jouy.inra.fr

Abstract

Bacillus thuringiensis subsp. *israelensis* is one of the most important microorganisms used against mosquitoes. It was intensively studied following its discovery and became a model bacterium of the *B. thuringiensis* species. Those studies focused on toxin genes, aggregation-associated conjugation, linear genome phages, etc. Recent announcements of genomic sequences of different strains have not been explicitly related to the biological properties studied. We report data on plasmid content analysis of four strains using ultra-high-throughput sequencing. The strains were commercial product isolates, with their putative ancestor and type *B. thuringiensis* subsp. *israelensis* strain sequenced earlier. The assembled contigs corresponding to published and novel data were assigned to plasmids described earlier in *B. thuringiensis* subsp. *israelensis* and other *B. thuringiensis* strains. A new 360 kb plasmid was identified, encoding multiple transporters, also found in most of the earlier sequenced strains. Our genomic data show the presence of two toxin-coding plasmids of 128 and 100 kb instead of the reported 225 kb plasmid, a co-integrate of the former two. In two of the sequenced strains, only a 100 kb plasmid was present. Some heterogeneity exists in the small plasmid content and structure between strains. These data support the perception of active plasmid exchange among *B. thuringiensis* subsp. *israelensis* strains in nature.

Keywords: *Israelensis* plasmids; *Israelensis* sequencing; pBtoxis; Conjugative plasmid pXO16.

1. Introduction

Despite great progress in development of insect-resistant plants carrying the toxin genes from *Bacillus thuringiensis*, this bacterium is still commonly used [1]. Presumably, the bacterium cannot be replaced by using genetically modified plants, particularly in applications involving control of mosquito propagation [2]. Other applications can also be envisaged, such as protection of forests and fields, seeded with unprotected plants, against sudden propagation of undesired insects. The biological agents carrying the required properties for the former application are strains of *B. thuringiensis* subsp. *israelensis* [3]. Due to its practical importance, this *B. thuringiensis* lineage has been one of the most intensively studied, in particular, targeting conjugation and retromobilization carried by an aggregation mechanism [4, 5]. Studies of conjugation provided detailed descriptions of electrophoretic plasmid profiles [6, 7]. *B. thuringiensis* subsp. *israelensis* is the only *B. thuringiensis* lineage for which a plasmidless strain, *B. thuringiensis* subsp. *israelensis* 4Q7 (known also as 4Q2-81), was constructed and intensively used [5, 8]. This strain, along with other acrySTALLIFEROUS mutants in which individual toxins were cloned, was very useful for discovery and detailed analysis of toxin synergies [9, 10], complementing in vitro studies [11]. Also, several studies on the phages well represented in *B. thuringiensis* subsp. *israelensis* were recently published [12-15]. It can be concluded that *B. thuringiensis* subsp. *israelensis* lineage is one of the most accountable experimental models for studies related to *B. thuringiensis*.

The strong interest in experimental studies using *B. thuringiensis* subsp. *israelensis* was not adequately reflected by extensive genomic undertakings. Only four genomic sequence

assemblies have been reported thus far, and only one of them completed. These include: strains *B. thuringiensis* subsp. *israelensis* 4Q7, a plasmid-less derivative of strain HD-500 (or 4Q2) [7, 16], acc. # JEOC00000000; *B. thuringiensis* subsp. *israelensis* HD-789, the only strain for which genome assembly, including six plasmids, was completed [17], acc.##CP003763-69; *B. thuringiensis* subsp. *israelensis* IBL4222 [18], acc # ACNL00000000; and strain *B. thuringiensis* subsp. *israelensis* ATCC35646, the first *B. thuringiensis* strain sequenced [19], acc ##AAJM1000001-AAJM1000866. In all cases, sequence analysis was rather formal and was not complemented with thorough biological reporting, limiting the use of such data. In particular, plasmids carrying insecticidal toxin genes were not analyzed in sufficient detail to explain inconsistency in determination of their sizes, which in one case appeared to be 128 kb and in another case 225 kb [17, 20]. Another example of ambiguity relates to identification of plasmid pXO16, responsible for aggregation and conjugation. The estimated earlier size of the plasmid was 135 MDa, corresponding to approximately 230 kb [7]. At the same time, this is the largest plasmid detected by electrophoresis in *B. thuringiensis* subsp. *israelensis* strains [6]. Sequencing revealed a candidate plasmid, designated pBTHD789-1, 350 kb in size [17]. The obvious difference in plasmid sizes was not explained, but was finally experimentally resolved in a recent paper that identifies the 350 kb pBTHD789-1 as pXO16 [21]. This identification was based on direct sequencing of pXO16 parts adjacent to the Tn5401 transposon, used for plasmid labeling, and also on sequencing of several PCR amplicons from the pXO16 transferred into a plasmidless strain. The authors also re-annotated the plasmid genes, and suggested that the conjugation system could be related to the FtsK/SpoIIIG family protein encoded by the plasmid.

In this work, we sequenced several representative strains of *B. thuringiensis* subsp. *israelensis* and performed PCR amplification experiments resolving ambiguities in

1 identification of correspondence between experimentally studied earlier plasmids and their
2 formalized presentation as sequencing contigs. Counting of sequencing reads covering
3 chromosomes and plasmids provided data for estimation of plasmid copy numbers as well as for
4 detection of several regions of local DNA over-synthesis. In some cases, this led to identification
5 of potential plasmid replication genes presumed to be situated near the replication origins. The
6 strains that we chose for sequencing, *B. thuringiensis* subsp. *israelensis* AM65-52 and *B.*
7 *thuringiensis* subsp. *israelensis* BMP144, were isolated from commercial products, Vectobac[®]
8 and Aquabac[®], respectively. We also sequenced strain *B. thuringiensis* subsp. *israelensis* IPS82,
9 a putative ancestor of one or both of these two strains. Another strain sequenced in the current
10 study was *B. thuringiensis* subsp. *israelensis* ATCC35646, sequenced earlier using Sanger
11 technology, but in draft mode [19].

2. Materials and methods

2.1. Bacterial strains

Strain *B. thuringiensis* subsp. *israelensis* IPS82 was from the Pasteur Institute Collection. Strain *B. thuringiensis* subsp. *israelensis* ATCC35646 (original name USDA HD522), originally obtained from the American Type Culture Collection (ATCC), was provided by Dr. A. Lapidus from Integrated Genomics (Chicago). It is the same strain sample used for sequencing by this company [19]. Strains *B. thuringiensis* subsp. *israelensis* BMP144 (Aquabac[®], company CERA) and *B. thuringiensis* subsp. *israelensis* AM65-52 (Vectobac[®], Sumitomo) were isolated from commercial samples. For this, 200 mg of each powder were resuspended in 1 ml of tryptone yeast (TY) broth. The suspensions were serial-diluted in TY broth and 100 µl of each dilution plated. The titers of the original product sample were 5x10⁹ cfu/g for Aquabac and 2x10⁹ cfu/g for Vectobac. Four separated colonies of the isolation were frozen for storage, designated WNF20#1, 2, 3 and 4 for Aquabac and WNF19#1, 2, 3 and 4 for Vectobac. Clones WNF20#1 and WNF19#1 were used for DNA preparation to be sequenced and designated *B. thuringiensis* subsp. *israelensis* BMP144 and *B. thuringiensis* subsp. *israelensis* AM65-52, respectively. Strain *B. thuringiensis* subsp. *israelensis* IBL 4222 was kindly provided by Dr. P. Martin (USDA-ARS) and strain *B. thuringiensis* subsp. *israelensis* BGSC 4Q5 was obtained from the Bacillus Genetics Stock Center (BGSC).

2.2. DNA preparation and sequencing.

The O/N cultures of each strain grown in TY broth were diluted 1:100 and grown in 20 ml at 37°C with shaking (200 rpm) up to middle exponential phase (OD₆₀₀~0.6). After centrifugation, they were resuspended in 6 ml of lysis solution (50 mM glucose, 25 mM

Tris/HCl, pH 8.0, 10 mM EDTA) each and treated by chicken egg lysozyme (5 mg/ml) at 37°C for 2 h. SDS was then added to final concentration of 0.5% and proteinase K to 1 mg/ml. The samples were incubated ON at 37°C. The mix was then treated twice under very soft shaking by water-saturated phenol, pH 8.0. The nucleic acids were then precipitated twice by adding sodium acetate, pH 4.8 (0.3 M final) and two volumes of ethanol. The pellets, washed with 70% ethanol, were slowly dried and resuspended in 0.8 ml of water. RNase A treatment was applied to remove the RNA.

Sequencing libraries and runs were done by the ImaGif sequencing platform (CNRS, Gif-sur-Yvette). Four standard libraries to be run in parallel on a HiSeq machine were prepared for each DNA sample, and one mate-paired library to be run on a MiSeq machine was prepared for the AM65-52 strain. Long-range PCR (LR PCR) was done as described earlier [22].

2.3. Large plasmid gel electrophoresis

Large plasmid profiles from *B. thuringiensis* subsp. *israelensis* strains BMP144, ATCC 35646, BGSC 4Q5 and IBL 4222 were obtained as described by Andrup et al. [23]. Bacteria were grown in 7 ml of brain heart infusion (BHI) broth for 15 h at 30 °C with shaking (120 rpm). A sample of 2 ml was centrifuged (8000 g for 7 min) and pellets were carefully resuspended in 100 µl of buffer E (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 addition of 20 µl of proteinase K (20 mg/ml). The solutions were gently inverted and incubated at 37°C for 60 min. One milliliter of phenol–chloroform–isoamyl alcohol (25:24:1) was added and the solutions were inverted several times. After centrifugation (8000 g for 7 min), the upper aqueous layer was subjected to 25 h of agarose gel 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 for 5 days (at 4°C).

2.4. Data analysis and release

The completed genome of *B. thuringiensis* subsp. *israelensis* HD-789 strain (acc.## CP003763÷9) was used as a template for alignment of the sequencing reads. The paired reads data were aligned using MAQ [24] or BOWTIE [25] software. Non-template assembling was gained using SPAdes or SOAPdenovo software [26, 27]. Gap4 of the Staden package [28] was used for final editing of the *B. thuringiensis* subsp. *israelensis* AM65-52 strain sequence. The most intense calculation part of data analysis was done using the computing resources of MetaGenoPolis (MGP) unit at INRA (Jouy-en-Josas). Plasmid genome comparison drawings were generated using Easyfig software [29]. The completed sequence of strain *B. thuringiensis* subsp. *israelensis* AM65-52 will appear in GenBank under the accession numbers CP013275 (chromosome) and CP013276-84 (plasmids). The most advanced assemblies of the other strains will be shortly released after final verification.

3. Results and discussion

3.1. Analysis of previously sequenced *B. thuringiensis* subsp. *israelensis* genomes

Multiple locus sequence typing (MLST) data indicate that the *B. thuringiensis* subsp. *israelensis* lineage represents a group of strains with very closely related genomic sequences. Such strains constitute about 10% of random forest soil isolates of *B. thuringiensis* [30]. Usually, the genes sequenced for MLST have 100% similarity. This property can be used to detect such genomes in databases in the cases in which the provided genome description does not relate it to *B. thuringiensis* subsp. *israelensis*. We performed this analysis using the Integrated Microbial Genome (IMG) database [31] and seven gene sequences from the SuperCAT MLST database [32]. Up until the date of analysis (March, 2015), five detected genomes, referred to as *B. thuringiensis* subsp. *israelensis*, with 100% similarity for these genes, were found and are listed in Table 1. However, during preparation of this manuscript, the genome of *B. thuringiensis* strain HD1002 became available for analysis [33] as a completely assembled sequence. The strain appeared to be highly similar to *B. thuringiensis* subsp. *israelensis* ATCC35646. Some features of interest in the plasmids found in this genome are discussed below.

Apart from *B. thuringiensis* HD1002, the only genome that was entirely assembled and finished using data generated with Illumina and 454 technologies is that of *B. thuringiensis* subsp. *israelensis* HD-789 [17]. Sequencing revealed the presence of six plasmids. The sequencing data assembly was the only experimental evidence of completeness of this plasmid set in strain *B. thuringiensis* subsp. *israelensis* HD-789.

The completed assembly of the *B. thuringiensis* subsp. *israelensis* HD-789 sequence provides an estimation of the chromosome size to be 5,495 kb. This figure is slightly different from the 5,491 kb reported for *B. thuringiensis* HD1002, but far different from the sum of contig lengths, 5,043 kb, in assembly for strain *B. thuringiensis* subsp. *israelensis* 4Q7 [16]. The latter was believed to be a plasmidless derivative of strain HD-500 (or 4Q2), and thus the assembly size of *B. thuringiensis* subsp. *israelensis* 4Q7 would also provide a minimal estimation of the chromosome length. Jeong et al. indicated two important differences between the *B. thuringiensis* subsp. *israelensis* 4Q7 and the *B. thuringiensis* subsp. *israelensis* HD-789 genome (Fig. 1A). These comprise the area between coordinates 1783 to 2275 kb in *B. thuringiensis* subsp. *israelensis* HD-789, almost completely absent in *B. thuringiensis* subsp. *israelensis* 4Q7, and a contig of 235 kb, mentioned to be unique in *B. thuringiensis* subsp. *israelensis* 4Q7 [16]. Surprisingly, the authors did not point out that this contig corresponds to plasmid pBTHD789-2 of 235,425 bp, present in all *B. thuringiensis* subsp. *israelensis* genomes sequenced to date (Fig.1B and RESULTS below). Thus, *B. thuringiensis* subsp. *israelensis* 4Q7 does not appear to be cured of all plasmids contained in its direct precursor *B. thuringiensis* subsp. *israelensis* 4Q2, as was expected. The accompanying paper of Gillis et al. describes the 235 kb plasmid, designated there as pB. *thuringiensis* subsp. *israelensis* c235, in detail.

B. thuringiensis subsp. *israelensis* ATCC35646 was the first strain sequenced for the cluster of *israelensis* [19]. The original strain was provided to ATCC from USDA under the name HD522, which was shown to be active against Dipteran larvae [34]. The total coverage of sequencing, done using standard shot-gun strategy based on Sanger technology and capillary machines, was 6x, and reads were assembled into 866 contigs. No further sequencing to reach better finishing was elaborated and the quality of the data remains unclear. No detailed analysis

was performed to separate contigs related to the chromosome from those encoding for the plasmids, although the contig corresponding to the linear DNA phage pBth35646 [12, 35], similar to pGIL01 family [15], can be easily detected. The total size of the assembled reads is 5,881 kb. Assuming that at least 600 kb should correspond to the plasmid DNA, this supposes that the size of the chromosome is smaller than 5.3 Mb. Data on re-sequencing of this strain and identification of plasmids are presented below.

Strains *B. thuringiensis* subsp. *israelensis* IBL4222 and *B. thuringiensis* subsp. *israelensis* IBL200 were sequenced in parallel with 43 other *B. cereus* group strains [18]. The reads generated with 454 technology were assembled into 383 contigs containing 6.6 Mb for the former and 244 contigs with 6.73 Mb in them for the latter. The plasmids were not identified and sizes of the chromosomes are therefore not known. A remarkable feature of these strains is that they were isolated from a domestic animal (cat) and a human and, respectively, were not used as insecticides.

Very recently, a novel fully assembled *B. thuringiensis* subsp. *israelensis* genome was reported for strain HD1002, although the authors did not mention that this could be *B. thuringiensis* subsp. *israelensis*, classifying it as *B. thuringiensis* [33]. Apparently, the genome, produced using three sequencing technologies (see Table 1, was meticulously assembled and contains an additional plasmid of 360 kb not reported in *B. thuringiensis* subsp. *israelensis* HD-789. Interestingly, the strain *B. thuringiensis* HD1002 also contains a 107 kb plasmid, designated pHD1002-4, similar to a counterpart in strain *B. thuringiensis* subsp. *israelensis* ATCC35646, but not pBtoxis that carries most insecticidal toxins of the *B. thuringiensis* subsp. *israelensis* cluster. The insect toxicity profile of these two strains as compared to others is therefore not clear.

In addition, several plasmids were independently sequenced earlier. The most important is plasmid pBtoxis carrying crystal toxin genes *cry* and *cyt* [6]. The plasmid was sequenced using strain *B. thuringiensis* subsp. *israelensis* 4Q5, a derivative of *B. thuringiensis* subsp. *israelensis* 4Q2 cured from all but pBtoxis plasmid [20]. It appeared, however, that this plasmid of 128 kb is rather different in size from the 225 kb plasmid pBTHD789-3 of strain *B. thuringiensis* subsp. *israelensis* HD-789, also carrying the toxin genes [17]. As mentioned above, the sequencing assemblies of *B. thuringiensis* subsp. *israelensis* ATCC35646 [19] and *B. thuringiensis* HD1002 [33] do not contain contigs corresponding to pBtoxis.

Another report concerns sequencing of three small multicopy plasmids, pTX14-1, -2 and -3, of 5.4, 6.8 and 7.6 kb, respectively [36]. All these plasmids originated from the *B. thuringiensis* subsp. *israelensis* 4Q2 strain, and their sequencing was done after subcloning in *E. coli*. Two of them, pTX14-2 and pTX14-3, have counterparts in the sequenced *B. thuringiensis* subsp. *israelensis* HD-789, *B. thuringiensis* HD1002 and *B. thuringiensis* subsp. *israelensis* ATCC35646 strains. pTX14-1 appears to be present only in assembly of the *B. thuringiensis* subsp. *israelensis* IBL4222 strain, where pTX14-2 is also present, but pTX14-3 was not found.

3.2. Previous experimental studies of plasmid content in *B. thuringiensis* subsp. *israelensis* strains

Plasmid content was experimentally studied for strain *B. thuringiensis* subsp. *israelensis* HD-567 using agarose gel electrophoresis [6]. The authors identified 9 plasmids with sizes of 135, 105, 75, 68, 10 (linear plasmid), 10.6, 4.9, 4.2 and 3.3 MD, that is, 230, 179, 128, 116, 17 (linear), 18, 8.3, 7.1 and 5.6 kb, respectively.

The entirely assembled genome of *B. thuringiensis* subsp. *israelensis* HD-789 [17] revealed the presence of six plasmids, of 349.6, 235.4, 224.9, 14.9, 7.7 and 6.8 kb, designated pBTHD789-1 to pBTHD789-6, respectively. No correspondence between the above mentioned plasmids, identified by electrophoresis, and this set was reported in the communication. The 224.9 kb pBTHD789-3 plasmid was reported to contain all *cry* and *cyt* toxin genes, which were earlier believed to all be in 128 kb pBtoxis [20]. Two sequenced strains, *B. thuringiensis* subsp. *israelensis* ATCC35646 and *B. thuringiensis* HD1002, appeared to contain only the part of the pBTHD789-3, not corresponding to pBtoxis. However, these plasmids of 104-107 kb carry two toxin genes and could therefore direct crystal protein synthesis and related toxicity to insects. It was recently shown that the electrophoresis band estimated to be 135 MD (230 kb) should in fact correspond to the 349.6 kb pBTHD789-1, and appeared to be identical to the conjugative pXO16 [21]. Also, a 360 kb plasmid was reported in strain *B. thuringiensis* HD1002, never detected earlier by electrophoresis. Apparently, there exist important discrepancies in plasmid size profiles between the two data sets. The analysis presented below resolves ambiguities in correspondence between plasmids functionally or electrophoretically identified, and the contigs resulted from sequence assemblies.

3.3. Insecticidal crystal protein genes containing plasmids

Among plasmids detected by electrophoresis, 75 MD pBtoxis, containing insecticidal crystal protein genes, was identified using conjugative transfer [6]. The plasmid appeared to have the exact size of 128 kb, was mapped with restriction enzymes and sequenced [20, 37]. For this purpose, strain 4Q2-72 (also known as 4Q5), a direct predecessor of plasmid-less 4Q7,

was constructed containing only this plasmid. More recently, genomic sequence assembly of the strain *B. thuringiensis* subsp. *israelensis* HD-789 identified a 225 kb plasmid, designated pBTHD789-3, which contained the toxin genes in the part highly similar to pBtoxis [17]. In contrast, assemblies of shot-gun reads of type strain *B. thuringiensis* subsp. *israelensis* ATCC35646 [19], as well as the recently reported *B. thuringiensis* HD1002 [33], do not contain contigs corresponding to pBtoxis, but each contains a contig corresponding to another part of pBTHD789-3 (Figs. 1C and 3).

Two possibilities might explain such a difference in sizes of these plasmids. One is that the reported assembly of plasmid pBTHD789-3 contains errors due to artificial joining of sequencing contigs, possibly because of multiple repeated sequences in this plasmid. If this is the case, only two types of such plasmids are contained in different *B. thuringiensis* subsp. *israelensis* strains. One of them is pBtoxis or its counterpart, encoding the majority of crystal protein genes. The second plasmid, corresponding to the rest of pBTHD789-3, is present alone in strains *B. thuringiensis* subsp. *israelensis* ATCC35646 and *B. thuringiensis* HD1002, devoid of pBtoxis. All other sequenced strains contain both of these plasmids. The known *B. thuringiensis* subsp. *israelensis* strains containing pBtoxis alone are only the ones artificially constructed.

The second possibility also supposes the existence of two plasmids, pBtoxis of 128 kb and another of approximately 100 kb, namely pATCC35646-3 or pHD1002-4, as a counterpart of the rest of pBTHD789-3. The smaller plasmid encodes two crystal protein genes, presumed to provide the capacity for bacteria to synthesize the toxins and be lethal to insects. If we accept that the assembly of pBTHD789-3 is correct, we might suppose that the two plasmids can co-integrate, resulting in pBTHD789-3 or its counterparts in other strains. In this model, it can also

be suggested that two different plasmid replication machineries are compatible in the same molecule. Although not studied in sufficient detail so as to be conclusive, such a recombination event was detected in experiments on curing of conjugative plasmid pXO16 by high-temperature treatment [7] (Fig. 2, lane f in this reference). Presumably, the hybrid plasmid suffered a deletion after the recombination event since, otherwise, the pure co-integrate of 225 kb would coincide with the 235 kb plasmid band on this gel.

Non-direct evidence that two plasmids are present in the strains studied here follows from sequencing read distribution data for the pBTHD789-3 area, shown in Fig. 1C and Supplementary Fig. S1A. In addition to the absence of the pBtoxis part in strain *B. thuringiensis* subsp. *israelensis* ATCC35646, three other sequenced strains have read distribution shapes that are significantly different in regions corresponding to pBtoxis and pATCC35646-3. Moreover, parts of distributions covering the area corresponding to the latter plasmid are very similar to that of strain *B. thuringiensis* subsp. *israelensis* ATCC35646. If a co-integrate of these two plasmids is present in any of these strains, the shape of read distribution in this complex plasmid would be homogenous over the whole of its length. This question was definitely resolved by LR PCR using total DNA of strain *B. thuringiensis* subsp. *israelensis* AM65-52 as template. These data, partially presented in Figs. 2 and 3 and, in more detail, in Suppl. Fig. S1 are explicitly compatible only if two plasmids are present in *B. thuringiensis* subsp. *israelensis* AM65-52. The counterparts of pBtoxis and pATCC35646-3 were designated pAM65-52-4-128K and pAM65-52-5-100K, respectively.

Fig. 2D shows that the most important difference between pAM65-52-4-128K and pBtoxis is symmetrical inversion of the region covering half of the plasmid centered at the origin or terminus of replication. Such a difference can be explained by only one recombination

event, and it is rather common in closely related plasmids. Positioning of a putative replication origin in these plasmids, based on identification of potential replication genes, corresponds to that based on distributions of sequencing reads, shown in Fig. 1C and S1A. These data indicate a feeble peak of over-coverage at a position around 95 kb on pBTHD789-3. This area is close to the BTF1_32191 gene encoding the FtsZ/tubulin-related protein (exact position is 80 kb in pBTHD789-3). The homolog of this protein, pBt156, together with pBt157, was shown to be able to maintain autonomous replication of plasmids as mini-replicons in the *B. thuringiensis* subsp. *israelensis* strains [38, 39]. Moreover, another replication system, designated ORF14-ORF16 according to gene numbering in pXO1 where it was first identified [40], is also present in pBTHD789-3 (BTF1_32141 and BTF1_32146, near 70 kb), pBtoxis (pBt007 and pBt009) (Figs. 2A, 2B and 2D) and pAM65-52-4-128K. It is noteworthy that the two replication systems are close to each other in all plasmids. The two systems were demonstrated to be able to maintain full-length replication of pXO1-like plasmids, although the ORF14-ORF16 replicon was found to be more efficient [41]. Similarity search using the sequence of protein BTF1_32146 reveals ORF14-ORF16-like genes encoding proteins with 75% identity existing in all *B. thuringiensis* subsp. *israelensis* assemblies, excluding only strains *B. thuringiensis* subsp. *israelensis* HD-789 and *B. thuringiensis* subsp. *israelensis* 4Q7. These genes are believed to be responsible for replication of another plasmid of 360 kb, described below.

The distribution of reads over the region covering the pATCC35646-3 part of pBTHD789-3 is rather unusual (Figs. 1C and S1A). First, the integral read coverage of this area is two to three times higher than that of the rest of the plasmid, strongly suggesting that corresponding contigs represent two independently replicating plasmids. Strain *B. thuringiensis* subsp. *israelensis* HD-789 was not, however, used in our experimental study, so we cannot

completely exclude that pBTHD789-3 is their co-integrate. Second, several relatively easily detectable over-coverage peaks can be detected at positions near 10, 154, 180 and 199 kb (Fig. S1A). The 180 kb peak is near the gene *repB* (BTF1_32576 in pBTHD789-3), presumably involved in initiation of plasmid replication (Fig. 3).

The complete sequence of plasmid pATCC35646-3 appeared to be available as one 104714 bp contig (acc. #NZ_AAJM01000002) of the 866 ones of draft genome assembly [19] (Fig. 3B). However, assembly of our shotgun data and LR PCR experiments (Fig. 3E) provided results that are more compatible with assembly of a similar plasmid, pHD1002-4, reported for another closely related strain *B. thuringiensis* HD1002 (acc. # NZ_CP009350) (Fig. 3C). It is worth noting that assembly problems were in the same region as in the case of pBTHD789-3 (Fig. 3 and S2 for more details). We found in this region a genetic element, consisting of six putative CDS around BTF1_32541 of pBTHD789-3, which is conserved in all strains and present in both plasmids, containing toxin genes, pBtoxis and pATCC35646-3 counterparts (Figs. 2, 3 and S3). BTF1_32541 encodes a transposase and the adjacent gene codes for recombinase similar to BinR of *S. aureus* transposon Tn552. Presumably, the element represents a transposon that may have played an important role in formation of co-integrate plasmid pBTHD789-3, if this was the case.

We did not detect any larvicidal activity of our *B. thuringiensis* subsp. *israelensis* ATCC35646 strain in a standard test against *Aedes aegypti* (not shown), although it was reported in the test against *Haematobia irritans* for the same strain, used under the name *B. thuringiensis* subsp. *israelensis* HD522 [34]. In our positive control test, 75% of mortality of larvae was seen after 30 min incubation with a similar preparation of the *B. thuringiensis* subsp. *israelensis*

AM65-52 culture diluted 1:100. Due to apparent ambiguity, we cannot, however, conclude as to whether the toxins encoded in pATCC35646-3 or counterparts are active or not.

3.4. Conjugative plasmid pXO16

The conjugative plasmid designated in the course of functional studies as pXO16, responsible for aggregation and conjugation phenotypes in its host strains [42], was recently shown to be the counterpart of pBTHD789-1 [21]. Although there are alternative mechanisms of transfer of crystal protein containing plasmids, pXO16 is presumably the main plasmid monitoring such transfer between *B. thuringiensis* subsp. *israelensis* strains in natural conditions. The specificity of this process is thus one of the factors responsible for the remoteness of such strains from others in the environment.

The contigs corresponding to the pBTHD789-1 were identified in all sequenced strains except 4Q7, including the four studied here (Fig.1C). Comparisons of read coverage of this plasmid with that of the chromosome indicated that the plasmid copy number varies between 2.2 to 3.6 in different strains (Table 2). A 20 kb area near 230 kb of the plasmid is 4-5 times over-covered compared to the average (Fig. 4A). We interpret this as over-synthesis of the plasmid replication origin at the end of exponential growth, where the bacteria were taken for total DNA extraction. This interpretation concords with the close positioning of multiple genes; those functions were annotated as related to plasmid replication initiation (Fig. 4B). It is also remarkable that the FtsK/SpoIIIE-family protein, believed to be involved in the segregation-conjugation phenotype [21], is also encoded in this proximity.

3.5. Prophage-like plasmid pBTHD789-2 counterparts

As mentioned above, the contig of 235 kb in sequence assembly of strain *B. thuringiensis* subsp. *israelensis* 4Q7 [16] corresponds to plasmid pBTHD789-2 of 235,425 bp, present in all *B. thuringiensis* subsp. *israelensis* genomes sequenced to date, including the four genomes studied in this work (Fig. 1). More detailed characterization of this element is presented in the accompanying paper by Gillis et al. The estimation of the copy number of this plasmid gave 0.7-0.8 per chromosome (Table 2). Some stringent mechanism of regulation of its replication could therefore be active.

3.6. A new large plasmid pAM65-52-1-360K in *B. thuringiensis* subsp. *israelensis* genomes

During de novo assembly of HiSeq reads for the four *B. thuringiensis* subsp. *israelensis* genomes using SOAPdenovo software, we detected multiple contigs that were not mapped into *B. thuringiensis* subsp. *israelensis* HD-789 contigs present in GenBank. We carefully scrutinized assemblies of these contigs and, finally, using high-redundancy mate-paired data for the *B. thuringiensis* subsp. *israelensis* AM65-52 strain, succeeded in producing a contiguous sequence of 360 kb, that we designated pAM65-52-1-360K (Table 2). This indicates that an additional low copy number plasmid, not detected earlier by electrophoresis or genomic sequencing, is present in *B. thuringiensis* subsp. *israelensis* strains. Generating the consensus sequence that was used for template-dependent assemblies we found that such a plasmid exists in all strains sequenced in this work (Fig. 1C). Also, corresponding contigs can be found in assemblies of earlier sequenced *B. thuringiensis* subsp. *israelensis* ATCC35646, IBL4222 and IBL200 strains (Fig. 5B), excluding *B. thuringiensis* subsp. *israelensis* HD-789 and plasmidless

B. thuringiensis subsp. *israelensis* 4Q7 (not shown). Extensive analysis of the most recent data for all sequenced *Bacillus cereus* group genomes revealed that a very similar contiguous sequence is present in strain *B. thuringiensis* HD1002 (Fig. 5), which was not assigned to the *B. thuringiensis* subsp. *israelensis* cluster, but was also isolated in Israel [33]. The plasmid contains multiple metabolite transporter genes (Fig. 5A). It also includes genes encoding the FtsK-like DNA transport protein and type IV secretion system protein VirDIV (Fig. 5A). We thus suppose that the plasmid was or still is conjugative, and its transfer to different strains was mediated by this process. Genome organization comparison revealed that one other, non-*B. thuringiensis* subsp. *israelensis* strain, from the 243 strains of the *B. cereus* group available for genome analysis, *B. thuringiensis* subsp. *monterrey* BGSC 4AJ1, also contains a similar plasmid (Fig. 5B). The plasmid presence is thus not limited to *B. thuringiensis* subsp. *israelensis*, but appears to be restricted to the *B. thuringiensis* species.

Agarose large plasmid electrophoresis of several *B. thuringiensis* subsp. *israelensis* isolates revealed two types of profiles, differing by the presence of a 128 kb band (Fig. 5D). It is, however, rather difficult to separate two slightly different molecules near 235 or 350 kb, and thus to confirm or reject whether the corresponding bands are single or double. This explains the difficulties in detecting, by this approach, 225 and 360 kb plasmids shaded by molecules of similar sizes. Surprisingly, in the same experiment, we detected the three-band profile for strain *B. thuringiensis* subsp. *israelensis* 4Q5, obtained from BGSC and believed to give only the 128 kb band. For the moment, we do not have a reasonable explanation for this observation.

3.7. Small plasmids in *B. thuringiensis* subsp. *israelensis* strains

The genomic sequencing of strain *B. thuringiensis* subsp. *israelensis* HD-789 revealed three plasmids smaller than 15 kb [17]. These are pBTHD789-4, -5 and -6, of 14935, 7697 and 6824 bp, respectively. Alignment of reads for the four strains sequenced in this work indicated that the first and third plasmid exist in all four strains, but pBTHD789-5 is present only in *B. thuringiensis* subsp. *israelensis* ATCC35646 (Fig. 6A and B). Analysis of plasmid databases showed that pBTHD789-4 corresponds to linear DNA phages like GIL01 [15] or similar linear DNA plasmids (Fig. 6C). pBTHD789-5 and -6 correspond to pTX14-3 and -2 [36], respectively, characterized earlier. The copy number of presented plasmids, estimated by sequence coverage, varied from 10 to 40 depending on the strain (Fig. 6 and Table 2).

The assembly of *B. thuringiensis* subsp. *israelensis* AM65-52, independently of template alignment, resulted in four contigs, corresponding to small multicopy plasmids. They were designated pAM65-52-6-16K, pAM65-52-7-15K, pAM65-52-8-7K and pAM65-52-9-5K, with sizes 16.1, 15, 6.8 and 4.8 kb, respectively (Table 2). Homology analysis revealed that the 15 kb plasmid pAM65-52-7-15K corresponds to the linear phage mentioned above. The 6.8 kb plasmid pAM65-52-8-7K corresponds to plasmid pBTHD789-6. Both of these elements are omnipresent, and they have very similar sizes in all *B. thuringiensis* subsp. *israelensis* strains. Their coverage by reads for different strains is presented in Fig. 6A. The 4.8 kb plasmid pAM65-52-9-5K corresponds to previously sequenced pTX14-1, although the size of the latter, 5415 bp, is slightly different [36]. This plasmid was not found in strains *B. thuringiensis* subsp. *israelensis* ATCC35646, *B. thuringiensis* subsp. *israelensis* HD-789 or *B. thuringiensis* HD1002.

We did not find any reliable counterpart of the 16.1 kb plasmid pAM65-52-6-16K, although 90-96% similarity over up to 4 kb regions was detected in other assemblies of *B.*

thuringiensis subsp. *israelensis* and *B. thuringiensis* strains, also including contigs longer than 100 kb. The elements of this plasmid thus appear to be extremely mobile. In particular, almost 100% identity over two 7-8 kb areas corresponding to this plasmid were detected with the reads of the *B. thuringiensis* subsp. *israelensis* IPS82 strain. However, this region was not covered with reads for *B. thuringiensis* subsp. *israelensis* ATCC35646 and *B. thuringiensis* subsp. *israelensis* BMP144 (Fig. 6A). It is worth mentioning that the existence of a plasmid slightly bigger in size than the linear pro-phage of 15 kb concords with the electrophoretic profile of plasmids in strain *B. thuringiensis* subsp. *israelensis* HD-567 discussed above [6].

In total, plasmid DNA in *B. thuringiensis* subsp. *israelensis* strains represents about 15-20% (18% in the strain *B. thuringiensis* subsp. *israelensis* AM65-52, Table 2) of the chromosome by size, and 34 to 57% by amount (Table 2). This is far below the almost 1.6 genome equivalent found in a strain of *kurstaki* subspecies using a different method [43]. Plasmids carry the most essential properties distinguishing *B. thuringiensis* subsp. *israelensis* from other *B. cereus* group strains. In particular, they encode toxicity to particular insect groups, conjugation properties and, in the case of pAM65-52-1-360K, multiple transport functions. In fact, they determine distinction of *B. thuringiensis* subsp. *israelensis* from other closely related bacteria. Thus, since they are responsible for genetic and phenotypic isolation, they in fact determine formation of a novel bacterial species on the basis of alliance of strains that contain them.

In conclusion, we can sum our findings up as follows:

1. Four strains of *B. thuringiensis* subsp. *israelensis*: *B. thuringiensis* subsp. *israelensis* BMP144 (Aquabac[®], CERA), *B. thuringiensis* subsp. *israelensis* AM65-52 (Vectobac[®], Sumitomo), *B. thuringiensis* subsp. *israelensis* IPS82 (Pasteur Institute) and *B. thuringiensis* subsp. *israelensis* ATCC35646 (ATCC type strain) were sequenced by Illumina technology, with chromosome coverage of three to six hundred. The assembling of sequencing reads was done using completely assembled earlier strain *B. thuringiensis* subsp. *israelensis* HD-789 as a template or de novo.
2. Alignment of the sequencing reads over the *B. thuringiensis* subsp. *israelensis* HD-789 genome revealed complete coverage of its chromosome, without evident gaps, for all sequenced strains. This was also the case for four of its plasmids. Plasmid pBTHD789-3 revealed non-homogenous distribution of reads and half of the plasmid was not covered with the reads of *B. thuringiensis* subsp. *israelensis* ATCC35646. We showed that, in *B. thuringiensis* subsp. *israelensis* AM65-52, there are two smaller plasmids (100 and 128 kb), instead of one large one(225 kb), and in strain *B. thuringiensis* subsp. *israelensis* ATCC35646, the 128 kb plasmid is absent. Small multicopy plasmid pBTHD789-5 is present in strain *B. thuringiensis* subsp. *israelensis* ATCC35646 and absent in three other sequenced strains. The profiles of other small plasmids also vary among the studied strains.
3. De novo assembly of reads for strain *B. thuringiensis* subsp. *israelensis* AM65-52 revealed a new large plasmid of 360 kb, designated pAM65-52-1-360K, and not identified earlier in any *B. thuringiensis* subsp. *israelensis* strain. Its counterparts exist, among a total of 243 *B. cereus* group strains analyzed, in *B. thuringiensis*

HD1002 and *B. thuringiensis* subsp. *monterrey* BGSC 4AJ1 strain, the latter distant from *B. thuringiensis* subsp. *israelensis*. Plasmid profile similarity permits assigning *B. thuringiensis* HD1002 to the *B. thuringiensis* subsp. *israelensis* cluster, although its toxicity to insects may be different. pAM65-52-1-360K counterparts are present in all sequenced natural *B. thuringiensis* subsp. *israelensis* isolates, probably including *B. thuringiensis* subsp. *israelensis* HD-789, where it was not identified. The plasmid encodes multiple metabolite transporters.

4. This study brings the number of completed *B. thuringiensis* subsp. *israelensis* genomes to three: *B. thuringiensis* subsp. *israelensis* HD-789, *B. thuringiensis* subsp. *israelensis* HD-1002 and *B. thuringiensis* subsp. *israelensis* AM65-52 (NCBI references provided in Table 1). The *B. thuringiensis* subsp. *israelensis* HD-789 genome assembly lacks the 360 kb plasmid, and the accuracy of assembly of 225 kb plasmid requires experimental verification. The two other genomes can be used as references for the completed genomic *B. thuringiensis* subsp. *israelensis* assemblies.

Acknowledgements

We thank Nayla Aspirot (CERA) for providing us commercial samples of Aquabac and Vectobac, and Dr. A. Lapidus (IG) and BGSC for *B. thuringiensis* subsp. *israelensis* strains. Dr. P. Martin from USDA-ARS is credited with providing strain IBL 4222. We also thank the ImaGif sequencing platform for preparing libraries and providing excellent quality raw sequencing data. The MGP unit (INRA) is credited with giving access to computing facilities. AB, VS, CN, AS and DL are supported by the French Agence nationale de la Recherche (ANR) grant (project PathoBactEvol). AG holds a Chargé de Recherches fellowship from the National Fund for Scientific Research (FNRS). JM is supported by the FNRS and the Université catholique de Louvain.

References

- [1] Sansinenea E. Discovery and Description of *Bacillus thuringiensis*. in: E Sansinenea (Ed.), *Bacillus thuringiensis Biotechnology*, Springer, Dordrecht Heidelberg London New York, 2012, pp. 3-18.
- [2] Ben-Dov E. *Bacillus thuringiensis* subsp. *israelensis* and its dipteran-specific toxins. *Toxins* 2014;6:1222-43.
- [3] Geissbuhler Y, Kannady K, Chaki PP, Emidi B, Govella NJ, Mayagaya V, et al. Microbial larvicide application by a large-scale, community-based program reduces malaria infection prevalence in urban Dar es Salaam, Tanzania. *PLoS One* 2009;4:e5107.
- [4] Andrup L, Jorgensen O, Wilcks A, Smidt L, Jensen GB. Mobilization of "nonmobilizable" plasmids by the aggregation-mediated conjugation system of *Bacillus thuringiensis*. *Plasmid* 1996;36:75-85.
- [5] Timmerly S, Modrie P, Minet O, Mahillon J. Plasmid capture by the *Bacillus thuringiensis* conjugative plasmid pXO16. *J Bacteriol* 2009;191:2197-205.
- [6] Gonzalez JM, Jr., Carlton BC. A large transmissible plasmid is required for crystal toxin production in *Bacillus thuringiensis* variety *israelensis*. *Plasmid* 1984;11:28-38.
- [7] 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.
- [8] Jensen GB, Andrup L, Wilcks A, Smidt L, Poulsen OM. The aggregation-mediated conjugation system of *Bacillus thuringiensis* subsp. *israelensis*: host range and kinetics of transfer. *Current microbiology* 1996;33:228-36.
- [9] Monnerat R, Pereira E, Teles B, Martins E, Praca L, Queiroz P, et al. Synergistic activity of *Bacillus thuringiensis* toxins against *Simulium* spp. larvae. *Journal of invertebrate pathology* 2014;121:70-3.
- [10] Crickmore N, Bone EJ, Williams AJ, Ellar DJ. Contribution of the individual components of the δ -endotoxin crystal to the mosquitocidal activity of *Bacillus thuringiensis* subsp. *israelensis*. *FEMS Microbiol Lett* 1995;131:249-54.
- [11] Perez C, Fernandez LE, Sun J, Folch JL, Gill SS, Soberon M, et al. *Bacillus thuringiensis* subsp. *israelensis* Cyt1Aa synergizes Cry11Aa toxin by functioning as a membrane-bound receptor. *Proc Natl Acad Sci U S A* 2005;102:18303-8.
- [12] 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.
- [13] Hardies SC, Thomas JA, Serwer P. Comparative genomics of *Bacillus thuringiensis* phage 0305phi8-36: defining patterns of descent in a novel ancient phage lineage. *Virology* 2007;4:97.
- [14] Moumen B, Nguen-The C, Sorokin A. Sequence Analysis of Inducible Prophage phIS3501 Integrated into the Haemolysin II Gene of *Bacillus thuringiensis* var *israelensis* ATCC35646. *Genetics Research International* 2012;2012:1-9.
- [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] Jeong H, Park S-H, Chol S-K. Genome Sequence of the AcrySTALLiferous *Bacillus thuringiensis* Serovar *Israelensis* Strain 4Q7, Widely Used as a Recombination Host. *Genome Announcements* 2014;2:e00231-14.
- [17] 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.

- [18] 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.
- [19] 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.
- [20] 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.
- [21] Makart L, Gillis A, Mahillon J. pXO16 from *Bacillus thuringiensis* serovar *israelensis*: Almost 350 kb of terra incognita. *Plasmid* 2015.
- [22] 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.
- [23] Andrup L, Barfod KK, Jensen GB, Smidt L. Detection of large plasmids from the *Bacillus cereus* group. *Plasmid* 2008;59:139-43.
- [24] Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 2008;18:1851-8.
- [25] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357-9.
- [26] Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 2012;1:18.
- [27] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology* 2012;19:455-77.
- [28] Staden R, Judge DP, Bonfield JK. Sequence assembly and finishing methods. *Methods of biochemical analysis* 2001;43:303-22.
- [29] Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics* 2011;27:1009-10.
- [30] Sorokin A, Candelon B, Guilloux K, Galleron N, Wackerow-Kouzova N, Ehrlich SD, et al. Multiple-locus sequence typing analysis of *Bacillus cereus* and *Bacillus thuringiensis* reveals separate clustering and a distinct population structure of psychrotrophic strains. *Appl Environ Microbiol* 2006;72:1569-78.
- [31] Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic Acids Res* 2010;38:D382-90.
- [32] Tourasse NJ, Kolsto AB. SuperCAT: a supertree database for combined and integrative multilocus sequence typing analysis of the *Bacillus cereus* group of bacteria (including *B. cereus*, *B. anthracis* and *B. thuringiensis*). *Nucleic Acids Res* 2008;36:D461-8.
- [33] 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.
- [34] Temeyer KB. Larvicidal activity of *Bacillus thuringiensis* subsp. *israelensis* in the dipteran *Haematobia irritans*. *Appl Environ Microbiol* 1984;47:952-5.
- [35] Sozhamannan S, McKinstry M, Lentz SM, Jalasvuori M, McAfee F, Smith A, et al. Molecular characterization of a variant of *Bacillus anthracis*-specific phage AP50 with improved bacteriolytic activity. *Appl Environ Microbiol* 2008;74:6792-6.
- [36] 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.

- [37] Ben-Dov E, Nissan G, Pelleg N, Manasherob R, Boussiba S, Zaritsky A. Refined, circular restriction map of the *Bacillus thuringiensis* subsp. *israelensis* plasmid carrying the mosquito larvicidal genes. *Plasmid* 1999;42:186-91.
- [38] Tang M, Bideshi DK, Park HW, Federici BA. Minireplicon from pBtoxis of *Bacillus thuringiensis* subsp. *israelensis*. *Appl Environ Microbiol* 2006;72:6948-54.
- [39] Tang M, Bideshi DK, Park HW, Federici BA. Iteron-binding ORF157 and FtsZ-like ORF156 proteins encoded by pBtoxis play a role in its replication in *Bacillus thuringiensis* subsp. *israelensis*. *J Bacteriol* 2007;189:8053-8.
- [40] Pomerantsev AP, Camp A, Leppla SH. A new minimal replicon of *Bacillus anthracis* plasmid pXO1. *J Bacteriol* 2009;191:5134-46.
- [41] Akhtar P, Khan SA. Two independent replicons can support replication of the anthrax toxin-encoding plasmid pXO1 of *Bacillus anthracis*. *Plasmid* 2012;67:111-7.
- [42] Andrup L, Damgaard J, Wassermann K. Mobilization of small plasmids in *Bacillus thuringiensis* subsp. *israelensis* is accompanied by specific aggregation. *J Bacteriol* 1993;175:6530-6.
- [43] Zhong C, Peng D, Ye W, Chai L, Qi J, Yu Z, et al. Determination of plasmid copy number reveals the total plasmid DNA amount is greater than the chromosomal DNA amount in *Bacillus thuringiensis* YBT-1520. *PLoS One* 2011;6:e16025.
- [44] Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394-403.

Table 1. Statistics of *B. thuringiensis* subsp. *israelensis* genome sequencing^a.

Strain	Submitter ^b	Method	Contigs	Total length,		rRNA	Proteins	NCBI ref id ^a	Publication
				Mb					
ATCC 35646	IG	Sanger	866	5.88		15	6132	29401	[19]
HD-789	LANL	Illum + 454	7	6.33		42	6261	167057	[17]
IBL 4222	NMRC	454	383	6.61		6	6428	167077	[18]
IBL 200	NMRC	454	244	6.73		5	6193	167076	[18]
HD 1002	LANL	Illum+454+PacBio	8	6.57		39	6471	219528	[33]
4Q7	KRIBB	Illum	51	5.04		9	4982	167087	[16]
AM65-52	this work	Illum	10	6.72		42	6641	nd	This study
BMP144	this work	Illum	51	6.62		nd	nd	nd	This study
IPS 82	this work	Illum	58	6.63		nd	nd	nd	This study
ATCC 35646	this work	Illum	33	6.52		nd	nd	nd	This study

^a All data apart from this work are from NCBI:

[http://www.ncbi.nlm.nih.gov/genome/genomes/486?genome_assembly_id=NCBI ref id](http://www.ncbi.nlm.nih.gov/genome/genomes/486?genome_assembly_id=NCBI%20ref%20id)

^b IG: Integrated Genomics; LANL: Los Alamos National Laboratory; NMRC: Naval Medical Research Center; KRIBB: Korea Research Institute of Bioscience and Biotechnology; nd = not yet determined.

1

2 **Table 2. Statistics of template genome coverage with HiSeq reads for different strains^a.**

Strain or plasmid	Size, kb	AM65-52	ATCC35646	BMP144	IPS82	AM65-52	ATCC35646	BMP144	IPS82
<i>B. thuringiensis</i> subsp. <i>israelensis</i> HD-789									
chromosome	5495	311	477	457	605	1.0	1.0	1.0	1.0
pBTHD789-1	349.6	1005	1054	1196	1452	3.2	2.2	2.6	2.4
pBTHD789-2	235.4	205	324	336	477	0.7	0.7	0.7	0.8
pBTHD789-3	224.9	586	443	864	1202	1.9	0.9	1.9	2.0
pBTHD789-4	14.9	3873	6934	4914	6737	12.4	14.5	10.8	11.1
pBTHD789-5	7.7	135	6480	1	111	0.36	13.6	0.0	0.2
pBTHD789-6	6.8	12036	11774	11252	11787	38.6	24.7	24.6	19.5
<i>B. thuringiensis</i> subsp. <i>israelensis</i> AM65-52									
chromosome	5500.0	326	489	465	612	1.0	1.0	1.0	1.0
pAM65-52-1-360K	359.6	332	346	478	678	1.0	0.7	1.0	1.1
pAM65-52-2-350K	349.6	1164	1187	1314	1555	3.6	2.4	2.8	2.5
pAM65-52-3-235K	235.4	228	348	356	497	0.7	0.7	0.8	0.8
pAM65-52-4-128K	127.9	364	49	516	763	1.1	0.1	1.1	1.2
pAM65-52-5-100K	100.0	883	959	1260	1660	2.7	2.0	2.7	2.7
pAM65-52-6-16K	16.1	4664	191	1	3977	14.3	0.4	0.0	6.4
pAM65-52-7-15K	15.0	4155	7240	5115	6952	12.7	14.8	11.0	11.4
pAM65-52-8-7K	6.8	12613	12183	11755	12197	38.7	24.7	25.3	19.9
pAM65-52-9-5K	4.8	15786	3	12184	15154	48.4	0.0	26.2	24.7
total DNA, kb^b	6715.2					8608	7377	7913	7935
plasmid DNA, kb	1215.2					3108	1877	2406	2436
% plasmid/chromos. DNA	18%					57%	34%	44%	44%

3

4 ^a Mean coverage of each nucleotide and ratio of this coverage to chromosomal is indicated for
5 two assemblies of each studied strain, using *B. thuringiensis* subsp. *israelensis* HD-789 (upper
6 part) and *B. thuringiensis* subsp. *israelensis* AM65-52 (bottom part) completed assemblies as
7 templates.

1 ^b Integrated and relative amount of plasmid DNA in each strain is calculated accounting for each
2 plasmid coverage using *B. thuringiensis* subsp. *israelensis* AM65-52 strain sequence as reference
3 template.

4

Figure legends

Fig. 1. Identification of plasmids in sequenced *B. thuringiensis* subsp. *israelensis* strains.

- A. Mauve [44] alignment of *B. thuringiensis* subsp. *israelensis* HD-789, *B. thuringiensis* subsp. *israelensis* 4Q7 and *B. cereus* ATCC14579 genomic sequences. Red square and arrows indicate chromosomal regions absent in the *B. thuringiensis* subsp. *israelensis* 4Q7 strain. Red ovals show contigs corresponding to plasmid pBTHD789-2, thus also present in the *B. thuringiensis* subsp. *israelensis* 4Q7 strain. The bottom panel shows alignment with higher resolution of presentation of similarity areas, allowing better comparison and rendering the identity of contigs in *B. thuringiensis* subsp. *israelensis* 4Q7 genome assembly and pBTHD789-2 more evident.
- B. CDS organization in an area corresponding to the plasmid pBTHD789-2 contig. Identical CDS colors indicate homologous contigs also present in the *B. thuringiensis* subsp. *israelensis* 4Q7, *B. thuringiensis* subsp. *israelensis* ATCC35646 and *B. thuringiensis* subsp. *israelensis* IBL4222 strains. Source of information and figure design: IMG database of DOE Joint Genome Institute [31].
- C. Distribution of sequencing reads for different strains over the area of a consensus genome, corresponding to plasmids. The positions corresponding to pBTHD789-1, -2 and -3 plasmids of the *B. thuringiensis* subsp. *israelensis* HD-789 strain sequence are indicated by their names and black arrows. The position corresponding to the 360 kb plasmid absent in the reported *B. thuringiensis* subsp. *israelensis* HD-789 assembly is indicated by dashed arrow. Copy numbers of plasmids, estimated by integral coverage, are presented in Table 2.

Fig. 2. Maps of the crystal protein genes containing plasmids - the pBtoxis part.

- A. CDS map of pBTHD789-3. Genes for toxins *cry* and *cyt* of different types, and also replication genes *ftsZ*, *orf14-16* and *repB*, are indicated. Colored CDS, distinct of light yellow, indicate identified functions. Semi-transparent green and red bars correspond to homologous regions with pBtoxis, in direct and inverted orientation around the pBtoxis replication origin, respectively. Brown semi-transparent bars highlight putative replication genes. Blue semi-transparent bar indicates inverted homologous region with plasmid pATCC35646-3, shown in Fig. 3, and blue rectangle indicates the region identical in all three plasmids. The primers (small filled red arrows) and PCR amplification links (continuous and dashed red lines), most essential for plasmid pAM65-52-4-128K structure verification, are also indicated. Dashed red lines indicate unexpected but experimentally obtained primer PCR links. For simplicity, only two expected and confirmed PCR links are indicated by continuous red lines. Source of map information and figure design: IMG database of DOE Joint Genome Institute [31].
- B. CDS map of plasmid pBtoxis using the same symbols as in A.
- C. The most important PCR amplification results for verification of pAM65-52-4-128K structure. The product sizes are indicated as those expected from the reported pBTHD789-3 sequence (acc. # CP003766) and actually obtained, the latter in parentheses. Marker (λ DNA/*Hind*III) sizes are indicated in kb on the right. PCR experiments were done using total DNA preparations of strain *B. thuringiensis* subsp. *israelensis* AM65-52 as template.
- D. Comparison of genomes of toxin-encoding plasmids from different *B. thuringiensis* subsp. *israelensis* strains. Green and red colors refer to direct and inverted homologous

regions, respectively. Orange arrows in pATCC35646-3, pBTHD789-3 and pBtoxis maps correspond to toxin genes; pink arrows are transposition and IS-related genes, as annotated in the GenBank entries. Starting nucleotide of plasmids pAM65-52-4-128K and pBTHD789-3 are made close to putative origins of replication in concordance with the pBtoxis EMBL entry (AL731825), which is presented here as the inverted complement of the EMBL entry. Starting nucleotide of pATCC35646-3 and pAM65-52-5-100K is near its putative replication gene; *repB*. pATCC35646-3 sequence corresponds to the GenBank entry contig (acc#AAJM01000002) before correcting due to LR PCR data (Fig.3). Light green arrows, indicated only in pATCC35646, pBTHD789-3 and pBtoxis maps, are the replication genes. The BLASTn identity scale is shown in the bottom right-hand corner. The drawing was generated using the Easyfig software [29].

Fig. 3. Maps of crystal protein genes containing plasmids - the pATCC35646-3 part.

CDS maps of pBTHD789-3 (A), pATCC35646-3 (B) and pHD1002-4 (C). Colored regions and indicated genes are as in Fig. 2A. The primers (small filled violet arrows) and PCR amplification links (continuous and dashed violet lines), most essential for plasmid pAM65-52-5-100K and pATCC35646-3 structure verification, are also indicated. The dashed violet lines indicate unexpected but experimentally obtained primer PCR links. For simplicity, a minimal number of expected and confirmed PCR links are indicated by continuous violet lines. Source of maps and figure design: IMG database of DOE Joint Genome Institute [31].

D and E: LR PCR amplification experiments clarifying the structure of pAM65-52-5-100K (D) and pATCC35646-3 (E). Only the most important reactions are shown. Product sizes are indicated as those expected from the reported sequences (acc. ## CP003766 and NZ_AAJM1000002) and actually obtained, with the latter in parentheses. Marker (λ DNA/*Hind*III) sizes are indicated in kb on the right. PCR experiments were done using total DNA preparation of strains *B. thuringiensis* subsp. *israelensis* AM65-52 (D) and *B. thuringiensis* subsp. *israelensis* ATCC35646 (E) as templates.

Fig. 4. Identification of the replication initiation area in pBTHD789-1 and counterparts of other strains.

- A. Distribution of *B. thuringiensis* subsp. *israelensis* AM65-52 reads over the area corresponding to pBTHD789-1 in the *B. thuringiensis* subsp. *israelensis* HD-789 sequence. The over-covered region peak corresponds to a putative replication origin.
- B. CDS map of the pBTHD789-1 plasmid as presented in the IMG database [31]. Replication and DNA transfer (*ftsK*)-related gene function annotations are pinned to the corresponding CDS by blue arrows. The part of the map over-covered with sequencing reads is indicated by a blue rectangle. pBTHD789-1 corresponds to conjugative plasmid pXO16 [21].

Fig. 5. Counterparts of pAM65-52-1-360K in different strains of *B. thuringiensis*.

- A. CDS map of the Ga0077856_15 contig from strain *B. thuringiensis* HD1002 as presented in the IMG database [31]. Metabolite-transport-related gene annotations,

replication (*orf14-16*) and DNA transfer (*ftsK*) genes are pinned to the corresponding CDS by blue arrows. The part of the map over-covered with sequencing reads and thus presumably containing the replication origin is indicated by a blue rectangle. Starting nucleotide and map orientation are as in the IMG database, different from the corresponding GenBank entry (acc. # CP009349).

- B. Comparison of CDS maps of different *B. thuringiensis* strains in the area corresponding to the *orf14* replication gene (locus tag Ga0077856_15200) of Ga0077856_15 contig from genomic sequence of *B. thuringiensis* HD1002. All except *B. anthracis* and *B. cereus* group strains (243 in total) from the IMG database were used to derive the alignment. Only alignments for strains containing a region with similar gene organization are shown.
- C. Comparison of genome of plasmid pAM65-52-1-360K and contig Ga0077856_15 of *B. thuringiensis* HD1002. Green colors indicate homologous regions. Pink arrows of Ga0077856_15 map correspond to transposition and IS-related genes, dark blue to transporter genes, light green to replication genes. Starting nucleotides are made near the replication genes. The BLASTn identity scale is shown in the bottom right-hand corner. The drawing was generated using Easyfig software [29].
- D. Electrophoresis profiles of large plasmids in different *B. thuringiensis* subsp. *israelensis* strains: 1: BMP144; 2: ATCC 35646; 3: BGSC 4Q5; 4: IBL4222. The proposed sizes of plasmids in each band are indicated on the right. chr: chromosomal DNA.

Fig. 6. Small, high-copy number plasmids in *B. thuringiensis* subsp. *israelensis* strains.

- A. Distributions of sequencing reads for different strains over the area corresponding to small plasmids detected in strain *B. thuringiensis* subsp. *israelensis* AM65-52. For clarity, the

graphs that are parts of those shown in Fig.1A are extended. Double arrows indicate plasmids whose names are shown above. Dashed arrows indicate two 7-8 kb areas, separated in this preliminary assembly version, common between plasmid pAM65-52-6-16K and related elements in strain *B. thuringiensis* subsp. *israelensis* IPS82. The names of strains are indicated below the corresponding graphs. The copy numbers of plasmids, estimated by integral coverage, are presented in Table 2.

B. Distributions of sequencing reads for different strains over the area corresponding to pBTHD789-5.

C. Comparison of CDS maps of different strains in the area corresponding to pBTHD789-4 plasmid as presented in the IMG database [31]. Identical colors correspond to identical gene functions. Sequence identifiers are copied as they are in the IMG database. Shown on this figure, Bam35 and GIL16 (Bam35c and GIL16c for clear-plaques mutants) were demonstrated to be the active phages [12].

Figure 1

C

Figure 1

Figure 2

A

B

C

D

Figure 2

Figure 3

Figure 3

Figure 4

A

B

Figure 4

Figure 5

Figure 5

Figure 6

Figure 6