Polish Journal of Microbiology
2014, Vol. 63, No 2, 147–156

ORIGINAL PAPER

Modular Genetic Architecture of the Toxigenic Plasmid pIS56-63 Harboring *cry1Ab21* in *Bacillus thuringiensis* subsp. *thuringiensis* Strain IS5056

EMILIA MURAWSKA[1], KRZYSZTOF FIEDORUK[2] and IZABELA SWIECICKA[1]*

[1] Department of Microbiology, University of Białystok, Białystok, Poland
[2] Department of Microbiology, Medical University of Białystok, Białystok, Poland

Submitted 9 February 2014, revised 15 February 2014, accepted 15 February 2014

Abstract

Bacillus thuringiensis subsp. *thuringiensis* IS5056, a strain highly toxic to *Trichoplusia ni* larvae, produces the newly described Cry1Ab21 δ-endotoxin encoded by a gene located in the 63.8 kb pIS56-63 plasmid. In this report we present the structure and functional similarity of this plasmid to other *B. thuringiensis* large toxigenic plasmids with particular interest focused on its modular architecture. The 61 open reading frames (ORFs) of the plasmid made four functional modules: (i) M1-*mic*, the mobile insertion cassette harboring *cry1Ab21*; (ii) M2-*tra*, the putative conjugative element; (iii) M3-*reg*, regulation sequence; and (iv) M4-*rep*, the *ori44* replicon. These modules display similarity to corresponding sequences in distinct *B. thuringiensis* plasmids, but, in general, not to plasmid of other *Bacillus cereus sensu lato.* The nucleotide sequence and organization of genes in pIS56-63 were highly similar (80–100%) to those in pHT73 of *B. thuringiensis* HT73, and in p03 of *B. thuringiensis* HD771, particularly within the M3-*reg* and M4-*rep* modules, and slightly less in M2-*tra*, the latter of which is composed of two segments exhibiting homology to sequences in pBMB28, pAH187_45, pCT83, and pIS56-85 or to pCT72, pBMB67, p04, and pIS56-68. The tetrapartite structure of the toxigenic pIS56-63 plasmid strongly suggests that its hybrid nature is a result of recombination of various genetic elements originating from different extrachromosomal and chromosomal sources in *B. thuringiensis*. The presence of *cry1Ab21* in the mobile cassette suggests that its occurrence on pIS56-63 resulted from recombination and transposition events during the evolution of the plasmid.

Key words: *Bacillus thuringiensis* IS5056, pIS56-63, genetic organization, *cry1Ab21*

Introduction

Bacillus thuringiensis, together with six other species of Gram-positive, endospore-forming bacilli, belongs to the *Bacillus cereus* group. In particular the most intriguing are toxigenic strains of *B. thuringiensis* that are entomopathogens (Sanahuja *et al.*, 2011), *Bacillus cereus sensu stricto* involved in food poisoning (Stenfors Arnes *et al.*, 2008), and *Bacillus anthracis*, the etiologic agent of anthrax (Mock and Fouet, 2001). Although these bacilli differ significantly with regard to virulence, target specificities, and ecology, they share a common chromosomal ancestry (Helgason *et al.*, 2000; Swiecicka, 2008). However, they differ markedly in their extrachromosomal profiles. Pathogenic *B. anthracis* and *B. cereus* strains contain virulence determinants on plasmids, respectively, pXO1 and pXO2 (Mock and Fouet, 2001), and pCER01 (Hoton *et al.*, 2005). Likewise, genes encoding the entomotoxins of *B. thuringiensis* reside on large extrachromosomal elements (Barry *et al.*, 2002; Murawska *et al.*, 2013). Therefore the acquired specific phenotypic features, peculiar pathogenicity, and capabilities in horizontal gene transfer (HGT) are largely a result of genes harbored on plasmids (Yuan *et al.*, 2010).

With regard to various strains of *B. thuringiensis*, many of which are used effectively as natural insecticides, much focus has been placed on the molecular biology and mode of action of its proteinaceous entomotoxins (Sanahuja *et al.*, 2011). In contrast, by far, less attention has been allotted to the genetic structure and evolution of its toxigenic plasmids. Only a few, such as pBtoxis of *B. thuringiensis* subsp. *israelensis* (Barry *et al.*, 2002) and pAW63 of *B. thuringiensis* HD73 (Van der Auwera *et al.*, 2005), have been described in detail regarding their nucleotide sequence and putative functions of encoded proteins. Therefore, data on comparative analyses regarding the evolution and modular arrangement of large toxigenic plasmids of entomopathogenic *B. thuringiensis* are generally lacking

* Corresponding author: I. Swiecicka, Department of Microbiology, Institute of Microbiology, University of Bialystok, 15-950 Białystok, 20B Świerkowa Street, Poland; phone: +48 857 457 332; fax: +48 857 457 301; **e-mail: izabelas@uwb.edu.pl**

in the literature. Undoubtedly, with ongoing *B. thuringiensis* genome sequencing projects, it is anticipated that much knowledge will be gained in the near future regarding their rich diversity of plasmids, the evolutionary history and biological functions as they relate to intra- and inter-species genetic exchanges and virulence. In this regard, we have sequenced the genome of a recently described strain (IS5056) of *B. thuringiensis* subsp. *thuringiensis* (Murawska *et al.*, 2013) that is highly toxic to the cabbage looper, *Trichoplusia ni*, and have identified a virulence plasmid (pIS56-63) that harbors the unique gene that encodes the principal entomotoxin, Cry1Ab21 (Swiecicka *et al.*, 2008). To gain more insights into the nature of virulence plasmids in *B. thuringiensis*, we undertook a comprehensive analysis of the structural organization of pIS56-63, its putative genes, and relatedness to other plasmids found in other strains of *B. thuringiensis* and *B. cereus*, with particular interest focused on modular architecture of the plasmid.

Experimental

Materials and Methods

***B. thuringiensis* subsp. *thuringiensis* strain IS5056 and its plasmids.** *B. thuringiensis* subsp. *thuringiensis* strain IS5056 was isolated in 2005 from soil collected in Biebrza National Park in northeastern Poland. The genotypic and phenotypic properties of IS5056 were described previously (Swiecicka *et al.*, 2008). Briefly, IS5056 synthesizes an unusual quasi-cuboidal/bipyramidal crystal composed of the Cry1Ab21 δ-endotoxin demonstrated to be highly toxic to *Trichoplusia ni* larvae. The whole genome of IS5056 is composed of 6.8 Mbp: the 5.5 Mb chromosome and 14 plasmids that vary in size from 6.9 to 328.2 kb (Murawska *et al.*, 2013). The pIS56-63 plasmid harbors the *cry1Ab21* gene (Swiecicka *et al.*, 2008).

Bioinformatics analyses of pIS56-63. Potential coding sequences (CDSs) were predicted using Glimmer3 (Delcher *et al.*, 2007), and each CDS was verified by comparison with non-redundant nucleotide database in GenBank. Sequence annotation was performed using Artemis ver. 14.0.0 software (Rutherford *et al.*, 2000). Protein sequence similarities were determined by BLAST analysis (*blastp* algorithm). In order to identify protein families and conserved domains the Conserved Protein Domain database (Marchler-Bauer *et al.*, 2013) was utilized. Potential transmembrane domains were identified with TMHMM server ver. 2.0 (Krogh *et al.*, 2001). The figures were generated using the DNA-Plotter ver. 1.10 (Carver *et al.*, 2009) and Easyfig ver. 2.1 (Sullivan *et al.*, 2011).

Results and Discussion

Molecular architecture of the pIS56-63 plasmid. Large *B. thuringiensis* plasmids are mostly known for the entomotoxin genes they harbor, while much less attention has been focused on other genetic determinants and their general organization that could shed light on their evolutionary history, particularly as it relates to the spread of entomopathogenic factors among *Bacillus* spp. species. To this end, we report on the genetic organization of the toxigenic plasmid pIS56-63 present in *B. thuringiensis* subsp. *thuringiensis* IS5056 (Swiecicka *et al.*, 2008). Based on whole genome sequencing (Murawska *et al.*, 2013), pIS56-63 was determined to be a 63,864 bp circular molecule [GenBank: CP004131] with a G + C content of 34.7% that did not deviate significantly from that of the host chromosome (35.4%) [GenBank: CP004123]. The plasmid contained 61 coding sequences (CDSs), with 12 of these defined to be in a clockwise orientation (Fig. 1 and Table I). According to Clusters of Orthologous Groups (COG), the largest fractions of CDSs constituted proteins associated with DNA replication and recombination (a total of 14 CDSs), and proteins that putatively play various roles in conjugation (11 CDSs). Other CDSs included those involved in transcription (three CDSs), two-component signal transduction (two CDSs), and outer membrane proteins that potentially mediate cell envelope biogenesis (two CDSs). Finally, pIS56-63 contained single copies of genes encoding a replication protein (CDS 54), the Cry1Ab21 insecticidal toxin (CDS 4), S-layer protein (CDS 42), and a globulin that putatively function in inorganic ion transport (CDS 6). Biological functions could not be assigned to 26 CDSs.

With regard to ascertaining possible genesis of pIS56-63, we determined whether it contained various putative functional cassettes that could have independent origins. Indeed, pIS56-63 contained four distinct functional modules: M1-*mic*, M2-*tra*, M3-*reg*, and M4-*rep*. The M1-*mic* module (mobile insertion cassette) consisted of 16 CDSs (CDS 1 to 9, and CDS 55 to 61); M2-*tra* (horizontal gene transfer), 34 CDSs (CDS 10 to 43); M3-*reg* (regulation module), CDS 44 to 53; and CDS 54, designated M4-*rep*, was assessed as the pIS56-63 replicon (Fig. 1). The modular arrangement attests to the hybrid nature of the plasmid that apparently originated as a result of dynamic exchanges among various genetic entities, including plasmids, transposable elements and host chromosome. The acquisition of the combination of genes, evidently from different genetic reservoirs, and their regrouping in pIS56-63 is not unusual among large plasmids and is likely a result of intra- and interspecies HGT, which is the main mechanism of adaptation to distinct environments,

Table I
CDSs of pIS56-63 and potential functions

pIS56--63 CDS	S	Coordinates		Size [aa]	Predicted product (protein/domain/family)[a]	Cellular localization	Functional category
		Start	Stop				
The mobile insertion cassette (M1-*mic*)							
#01	–	2460	63361	987	TnpA transposase (DDE_Tnp_Tn3 Tn3 transposase DDE domain)	Cytoplasm	Replication and recombination
#02	–	2479	3333	284	TnpI resolvase (Pfam00589: Phage integrase; Pfam02899: Phage integrase, N-terminal SAM-like domain; COG0582: Integrase)	Cytoplasm	Replication and recombination
#03	–	3706	5142	478	IS231A transposase (Pfam01609: DDE_Tnp_1 Transposase DDE domain; COG3385: Predicted transposase)	Cytoplasm	Replication and recombination
#04	–	5426	8893	1155	Delta-endotoxin Cry1Ab21 (Pfam00555: Endotoxin_M, delta endotoxin domain; Pfam03945: Endotoxin_N, delta endotoxin; Pfam03944: Endotoxin_C, delta endotoxin)	Cytoplasm	Insecticidal crystal protein
#05	–	9183	10139	318	N-acetylmuramoyl-L-alanine amidase (PF01510: Amidase_2 N-acetylmuramoyl-L-alanine amidase; COG3023: Negative regulator of beta-lactamase expression)	Secreted	Cell envelopes biogenesis
#06	–	10924	11784	286	Na+/H+ antiporter (Pfam00999: Na_H_Exchanger Sodium/hydrogen exchanger family; COG0475: Kef-type K+ transport system, membrane component)	Membrane	Inorganic ion transport
#07	+	11931	13367	478	IS*231A* transposase (Pfam01609 :DDE_Tnp_1 Transposase DDE domain; COG3385: Predicted transposase)	Cytoplasm	Replication and recombination
#08	+	13575	14870	431	IstA transposase (Pfam00665: Integrase core domain; COG4584: Transposase and inactivated derivatives)	Cytoplasm	Replication and recombination
#09	+	14860	15612	250	ATP-binding protein for IstB (COG1484: DNA replication protein)	Cytoplasm	Replication and recombination
The conjugative module, part I (M2-*tra*I)							
#10	–	15748	15945	65	Hypothetical protein	Membrane	Unknown
#11	–	15929	16105	58	Hypothetical protein	Cytoplasm	Unknown
#12	–	16134	16715	193	Hypothetical protein	Cytoplasm	Unknown
#13	+	17080	18309	409	Transposase (Pfam01548: DEDD_Tnp_IS110 Transposase Family; Pfam02371: Transposase_20; COG3547: Transposase)	Cytoplasm	Replication and recombination
#14	–	18458	19567	369	Glucosamidase-like protein (Pfam01832: Glucosaminidase, Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase; Pfam01551: Peptidase_M23; OG0739: membrane proteins related to metalloendopeptidases)	Membrane	Cell envelope biogenesis
#15	–	19639	23232	1197	TrbL-like protein (Pfam04610: TrbL/VirB6 plasmid conjugal transfer protein[a])	Membrane	Conjugation (T4SS-like system)
#16	–	23229	23744	171	Hypothetical protein	Membrane	Unknown
#17	–	23773	26361	862	ATPase-like protein (Pfam: 12846 AAA_10, AAA-like domain; COG0433: Predicted ATPase)	Cytoplasm	Conjugation (T4SS-like system)
#18	–	26426	26959	177	TcpE-like protein (Pfam12648: TcpE family)	Membrane	Conjugation (T4SS-like system)
#19	–	26984	27229	81	Hypothetical protein	Membrane	Unknown
#20	–	27245	28183	312	Transposase (Pfam12642: Conjugative transposon protein TcpC)	Membrane	Conjugation (T4SS-like system)
#21	–	28196	28420	74	Hypothetical protein	Membrane	Unknown
#22	–	28425	29258	277	Replication-relaxation-like protein (Pfam13814: Replic_Relax, Replication-relaxation)	Cytoplasm	Conjugation (T4SS-like system)

Table I continued

pIS56--63 CDS	S	Coordinates		Size [aa]	Predicted product (protein/domain/family)[a]	C ellular localization	Functional category
		Start	Stop				
#23	–	29279	32452	1057	ATPase-like protein (Pfam12846: AAA_10, AAA-like domain; COG053: InfB Translation initiation factor 2, GTPase; COG1418: Predicted HD superfamily hydrolase)	Membrane	Conjugation (T4SS-like system)
#24	–	32472	32651	59	Hypothetical protein[b] (No homology)	Cytoplasm	Unknown
#25	–	32639	32950	103	Hypothetical protein	Cytoplasm	Unknown
#26	–	32990	33181	63	Hypothetical protein	Membrane	Unknown
#27	–	33271	33462	63	Hypothetical protein	Cytoplasm	Unknown
#28	–	33532	33789	85	Hypothetical protein[b] (No homology)	Cytoplasm	Unknown
#29	–	33782	34099	105	Hypothetical protein[b] (No homology)	Cytoplasm	Unknown
#30	–	34341	34514	57	Hypothetical protein	Membrane	Unknown
#31	–	34519	34887	122	Hypothetical protein	Cytoplasm	Unknown
#32	–	34918	35049	43	Hypothetical protein	Membrane	Unknown
#33	–	35051	35248	65	Hypothetical protein	Membrane	Unknown
The conjugative module, part II (M2-*tra*II)							
#34	–	35567	35953	128	Hypothetical protein	Membrane	Unknown
#35	–	36020	36286	88	Hypoth etical protein	Membrane	Unknown
#36	+	36723	37916	397	Transposase (Pfam 01548: DEDD_Tnp_IS110; Pfam02371: Transposase_20 COG3547: Transposase and inactivated derivatives)	Cytoplasm	Replication and recombination
#37	–	38039	38932	297	Hypothetical protein	Membrane	Unknown
#38	–	38948	39895	315	TadB-like protein (COG4965: TadB, Flp pilus assembly protein TadB)	Membrane	Conjugation (T4SS-like system)
#39	–	39882	41339	485	T2SE Type II/IV secretion system protein (VirB11-like conjugation protein; COG0630: Type IV secretory pathway, VirB11 components; COG2805: PilT Predicted ATPases involved in pili biogenesis; COG2804: GspE Predicted ATPases involved in pili biogenesis; COG4962 : CpaF, Flp pilus assembly protein, ATPase CpaF)	Cytoplasm	Conjugation (T4SS-like system)
#40	–	41369	42184	271	ATPase-like protein (Pfam13614: AAA_31, ATPases associated with diverse cellular activities; COG1192: ATPases involved in chromosome partitioning; COG0455: ATPases involved in chromosome partitioning; COG0208: NrdF Ribonucleotide reductase beta subunit; COG2894: MinD Septum formation inhibitor-activating ATPase)	Cytoplasm	Conjugation (T4SS-like system)
#41	–	42201	42992	263	FlgA-like protein (cd11614: SAF_CpaB_FlgA_like, SAF domains of the flagella basal body P-ring formation protein FlgA and the flp pilus assembly CpaB)	Cytoplasm	Conjugation (T4SS-like system)
#42	–	43008	43502	164	Hypothetical protein	Membrane	Unknown
#43	–	43517	46066	849	Hypothetical conjugation protein (COG1192: ATPases involved in chromosome partitioning)	Cytoplasm	Cell division and chromosome partitioning
The regulatory module (M3-*reg*)							
#44	–	46152	47360	402	Surface layer protein (Pfam00395: SLH, S-layer homology domain)	Cytoplasm/ Membrane	
#45	–	47415	47963	182	Hypothetical protein	Membrane	Unknown
#46	–	48493	48747	84	Hypothetical protein	Cytoplasm	Unknown

Table I continued

pIS56--63 CDS	S	Coordinates		Size [aa]	Predicted product (protein/domain/family)[a]	C ellular localization	Functional category
		Start	Stop				
#47	–	48912	49151	79	Transcriptional regulator (COG1396: Predicted transcriptional regulators; COG1426: Uncharacterized BCR; cd00093: HTH_XRE, Helix-turn-helix XRE-family like proteins)	Cytoplasm	Transcription
#48	+	49289	49681	130	Transcriptional regulator MerR (COG1396: Predicted transcriptional regulators)	Cytoplasm	Transcription
#49	–	49710	49862	50	Hypothetical protein	Cytoplasm	Unknown
#50	–	50027	50218	63	Transcriptional regulator (COG1396: HTH_3 Helix-turn-helix Domain)	Cytoplasm	Transcription
#51	–	50370	50957	195	Putative two-component response regulator YhcZ (cd00156: REC, Signal receiver domain, originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB); COG2197: CitB, Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain)	Cytoplasm	Signal transduction mechanisms
#52	–	50961	52319	452	Multi-sensor signal transduction histidine kinase (COG0642: Signal transduction histidine kinase)	Membrane (multi-pass)	Signal transduction mechanism
#53	–	52335	53312	325	Hypothetical protein	Cytoplasm	Unknown
The replication module (M4-*rep*)							
#54	+	53861	54799	312	Replication associated protein (HTH_24 Winged helix-turn-helix DNA-binding[c])	Cytoplasm	Plasmid replication protein
The mobile insertion cassette (M1-*mic*)							
#55	–	55287	56039	250	ATP-binding protein IstB (pf01695: IstB_IS21 IstB-like ATP binding protein; COG1484: DNA replication protein)	Cytoplasm	Replication and recombination
#56	–	56029	57324	431	IstA Transposase (Pfam00665: Integrase core domain; COG4584: Transposase and inactivated derivatives)	Cytoplasm	Replication and recombination
#57	+	57949	59142	397	ParM-like protein (cd10227: ParM-like, plasmid segregation protein)	Cytoplasm	Replication and recombination
#58	+	59157	59585	142	Hypothetical protein	Cytoplasm	Unknown
#59	+	59847	60056	69	Hypothetical protein	Cytoplasm	Unknown
#60	+	60560	61495	311	Phage integrase domain protein (SAM domain) (COG0582: Integrase; COG4974: XerD, Site-specific recombinase XerD; cd01182: INT_REC_C, DNA breaking-rejoining enzymes, intergrase/ recombinases, C-terminal catalytic domain)	Cytoplasm	Replication and recombination
#61	+	61731	63167	487	IS*231* Transposase (COG3385: predicted transposase)	Cytoplasm	Replication and recombination

[a] Pfam, database of protein families and domains; COG, Clusters of Orthologous Groups; cd, conserved domains database,
[b] no homology in BLAST,
[c] insignificant Pfam hit,

differentiation and evolution among bacterial genera (Treangen and Rocha, 2011).

Characteristics of M1-*mic*, the large mobile insertion cassette. The CDSs of the 24,190-bp M1-*mic* had homologues in the database with the exception of CDS 58 and 59 (Fig. 2A). In this module, CDS 1 and 2 were homologues of transposase TnpA and resolvase TnpI of the Tn*4430*-like transposon (4,159 bp) flanked on both sides by IS*231* (CDS 3 and CDS 61). An additional copy of IS*231* (CDS 7) which lacked its inverted repeat, and thus probably incapable to transposition, was found downstream of CDS 6. Importantly, the entire M1-*mic* module was flanked by IS*232* and consisted of genes for transposase IstA (CDS 8 and 56) and an ATP-binding protein, IstB (CDS 9 and 55). Altogether, seven potential mobile fragments were recognized in M1-*mic* based on the presence of insertion sequences and the transposon-like element (for details see Fig. 2A).

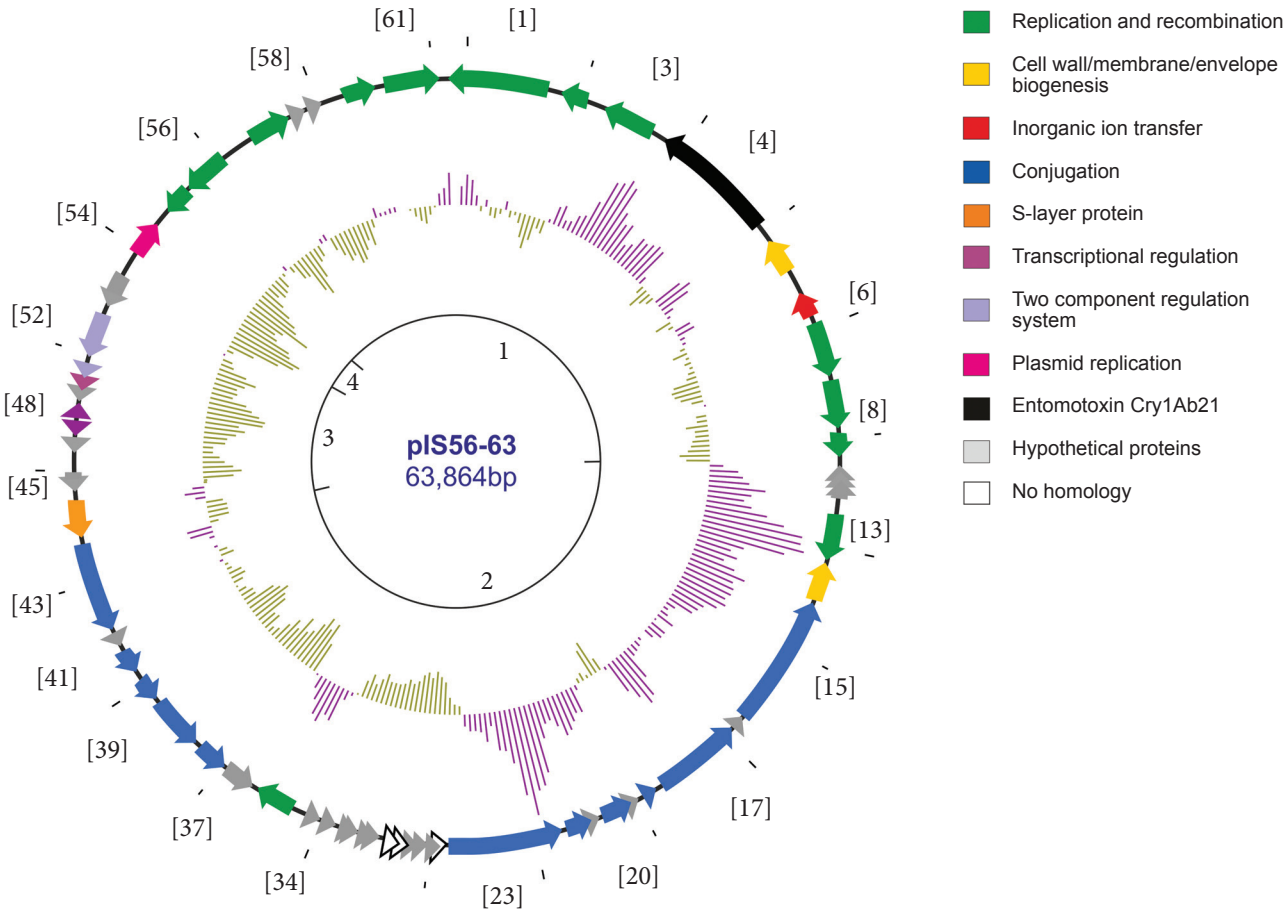

Fig. 1. Genetic map of pIS56-63. The circle in the center represents the functional modules found on the plasmid: **1**, the M1-*mic* module (the insertion mobile cassette); **2**, the M2-*tra* module (the conjugation region); **3**, the M3-*reg* module (the regulation region); **4**, the M4-*rep* module (replicon). The second circle shows a G + C content, with the mean value of 34.7% as baseline. The third circle illustrates coding sequences [CDS] represented as block arrows in line of the direction of transcription. The predicted function of each CDS is indicated by the color key as given in the legend. Numbers in brackets correspond to the main CDSs in Table I. The external circle is the scale marked with ticks every 6 kb.

In general, the genetic structure of M1-*mic* was similar to that found in pHT73 [GenBank: CP004070] and p03 [GenBank: CP003755] (Fig. 2B). However, in pHT73 the genes encoding Cry1Ac, amidase, Na^+/H^+ antiporter flanked by IS*232*, and Tn*4430* surrounded by IS*231* were harbored in an inverse orientation when compare with pIS56-63. The phage element, with one IS*232* in pHT73 and pIS56-63, had the same orientation. Additionally, Tn*4430* in pHT73 was flanked by a region encoding a putative transposase, and three small hypothetical proteins on one side, a region encoding a recombinase, transposase for transposon Tn*3*, and one small protein of unknown function on the opposite side; these regions were absent in pIS65-63. While in p03 of *B. thuringiensis* HD771, Tn*4430* was flanked on both sides by IS*231* in the same arrangement, but the phage elements were absent, and the *cry1Ac*, amidase and the sodium/hydrogen exchanger genes, called also Na^+/H^+ antiporter, were integrated in an inverse orientation.

Interestingly, the *cry1Ab21* δ-endotoxin gene (CDS 4) is also a component of M1-*mic*. The presence of the toxigenic *cry1Ab21* gene in the mobile element, together with IS*231*, IS*232*, and Tn*4430*, implies that its occurrence on the plasmid was mediated by a transposition event. In addition the cassette consisted a putative N-acetylmuramoyl-L-alanine amidase (CDS 5) involved in peptidoglycan biosynthesis, and Na^+/H^+ antiporter (CDS 6). The amidase homologs in IS5056, were also located on the chromosome (11 copies) and on pIS56-285 [GenBank: CP004136] of IS5056. The multiple copies of the gene in IS5056 could result in a significant increase in the level of N-acetylmuramoyl-L-alanine amidase, an enzyme that functions in peptidoglycan biosynthesis and its recycling (Johnson *et al.*, 2013). As a potential consequence, the rate of IS5056 sporulation and/or germination could be significantly enhanced, similar to that described for *B. cereus* E33L (Bourguet *et al.*, 2012). Indeed, IS5056 has been shown to proliferate rapidly in *Trichoplusia ni* larvae as it becomes moribund as a result of Cry1Ab21's entomocidal activity, and the propagation of this bacterium supersedes that of normal larval gut flora (Swiecicka

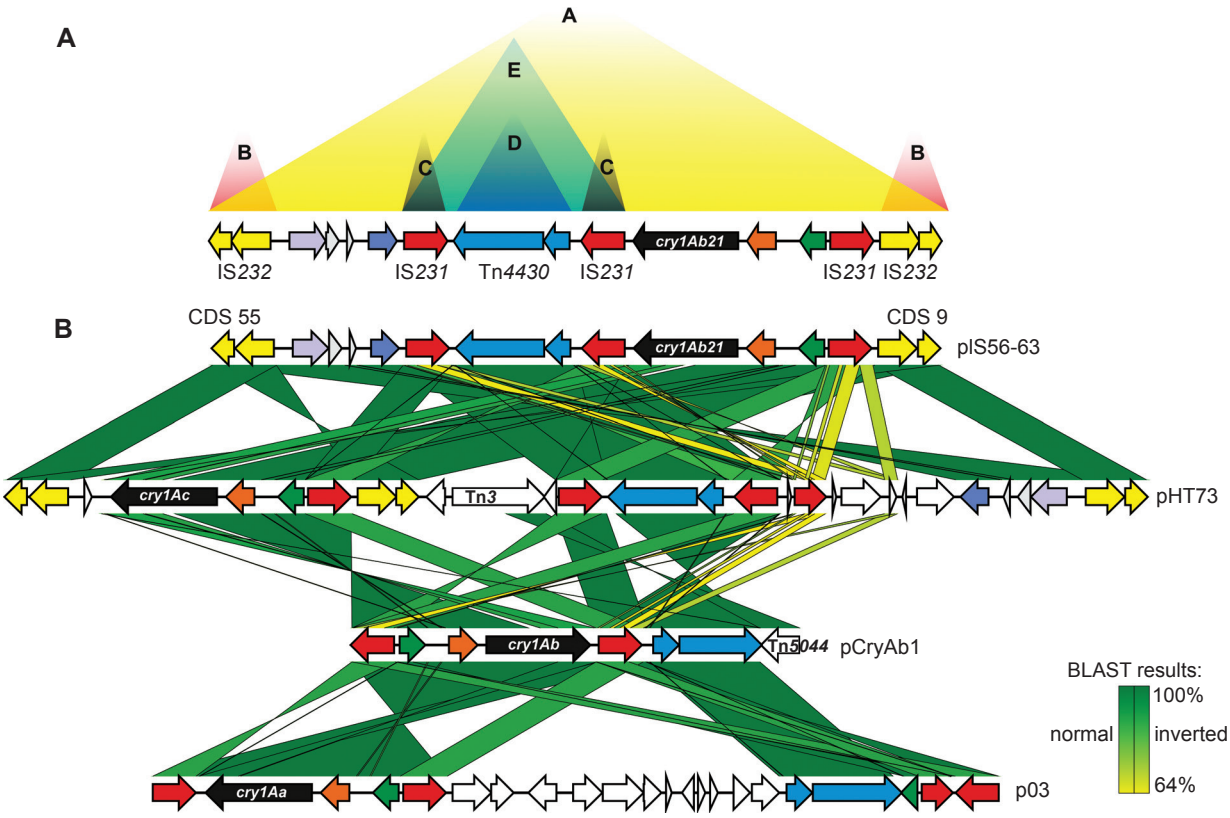

Fig. 2. Genetic organization of the 24-kb mobile insertion cassette (M1-*mic*) in pIS56-63 (A) and its homology to other *B. thuringiensis* plasmids (B). CDSs are represented by arrows in line with orientation of transcription. The potential seven mobile elements are indicated with triangles A-E in Fig. 2A. The color scale indicating the level of homology is presented in the legend.

et al., 2008). Thus, although Cry1Ab21 is the primary insect virulence factor, the amidase may significantly contribute to the virulence of IS5056 by optimizing propagation of the bacterium.

The M1-*mic* cassette contained CDS 57, present in similar form in pBMB28 of *B. thuringiensis* YBT-020 that encodes a ParM homologue, an actin-like protein that forms the cytomotive filament of the Type II plasmid segregation system, as described in *Escherichia coli* (Gayathri *et al.*, 2012). Moreover, CDS 60 encoded a site-specific recombinase XerD (phage integrase domain protein). Phage-specific integrase genes were also present in pBMB28, p03 of *B. thuringiensis* HD771, as well as pCT83 of *B. thuringiensis* CT-43, and with much lower homology in pAH187_45 of *B. cereus* AH187. In IS5056, the presence of these phage-specific CDSs is limited to pIS56-63.

Characteristics of M2-*tra*, the conjugative module. The M2-*tra* module, spanning two distinct segments within 30 kb, encompassed CDSs that encoded proteins with putative conjugation function (Fig. 3). The 36.8% G + C content of M2-*tra*I (CDS 10 to 33; 19,501 bp) deviated by ~2.3% from the average G + C content of its parental plasmid. The M2-*tra*II segment (CDSs 34 to 43; 10,500 bp) has a G + C content of 33.6%, which was only 1.1% different when compared to that of the pIS56-63. The genetic organization of the whole conjugative element in pIS56-63 is almost identical as in pHT73 of *B. thuringiensis* HD73. Moreover, while the M2-*tra*I shared significantly high homology with corresponding sequences in pBMB28 of *B. thuringiensis* YBT20 [GenBank: CP002510], lower homology with similar gene arrangement occurred with p03 [GenBank: CP003766] of *B. thuringiensis* HD789, pCT83 of *B. thuringiensis* CT-43 [GenBank: CP001915], pIS56-85 of *B. thuringiensis* IS5056 [GenBank: CP004133], and pAH187_45 of *B. cereus* AH187 [GenBank: CP001180]. The M2-*tra*II homologues were found in pBMB67 of *B. thuringiensis* YBT-1520 [GenBank: DQ363750], p04 of *B. thuringiensis* HD771 [GenBank: CP003756], pCT72 of *B. thuringiensis* CT-43 [GenBank: CP001913], and pIS56-68 of *B. thuringiensis* IS5056 [GenBank: CP004132].

The M2-*tra* module encodes proteins related to Type IV Secretion System (T4SS), the best described conjugative model among Gram-negative bacteria. A similar system is also known in Gram-positive bacilli (Van der Auvera *et al.*, 2005). The DNA VirB/D4 transfer system, described first in *Agrobacterium tumefaciens*, is regarded as the prototype of T4SS and includes genes coding for proteins involved in DNA transfer and replication (Dtr), mating-pair formation (Mpf), and

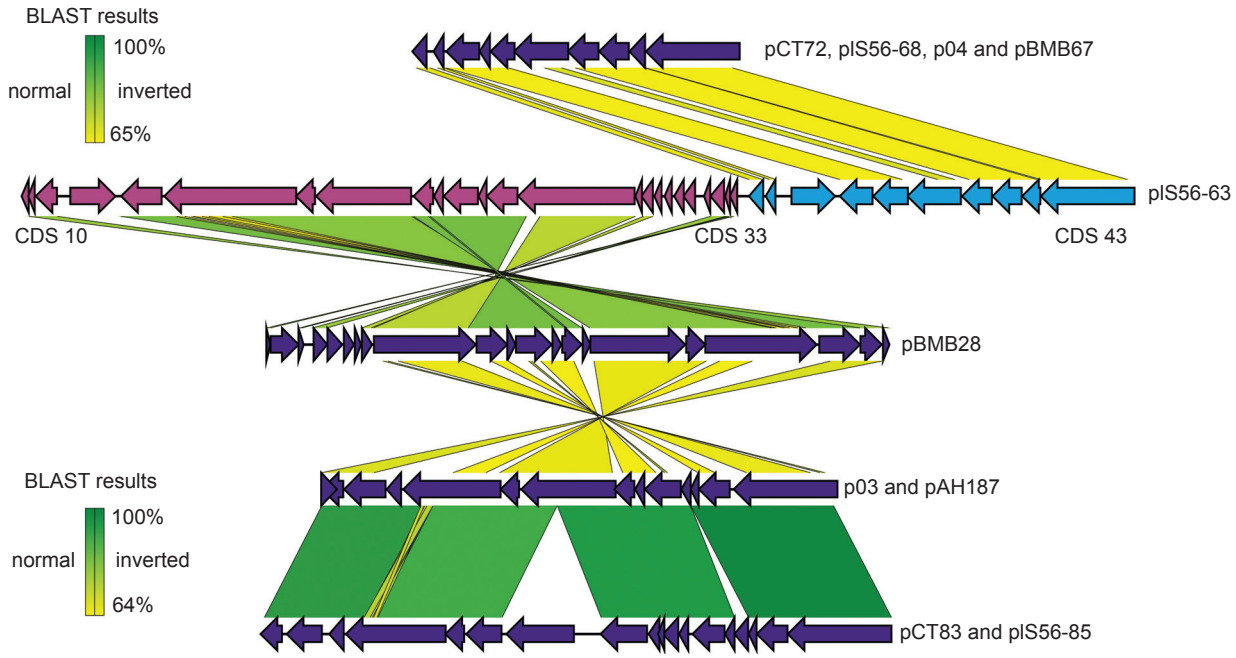

Fig. 3. Comparison of the 30-kb conjugative module (M2-*tra*) of pIS56-63, composed of M2-*tra*I (violet arrows) and M2-*tra*II (blue arrows) to corresponding segments in *B. thuringiensis* and *B. cereus* plasmids. The color scale indicating the level of homology is presented in the legend.

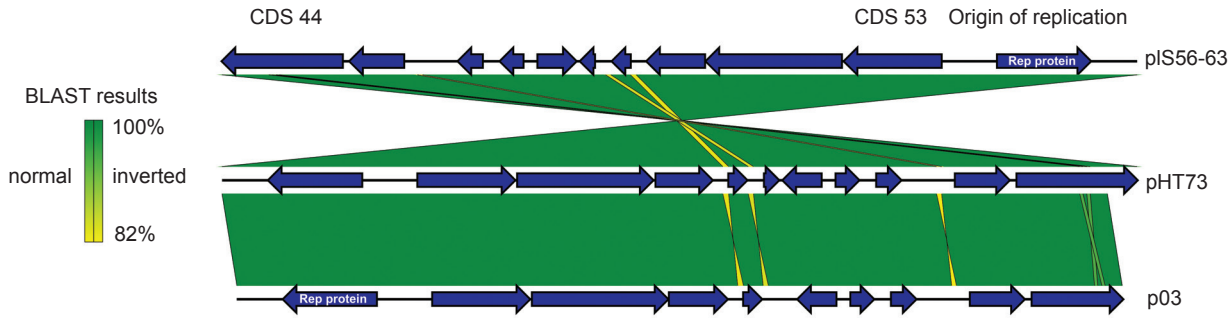

Fig. 4. Comparison of the 6.2-kb regulation (M3-*reg*) and the 1.7 replication (M4-*rep*) modules to corresponding segments in *B. thuringiensis* and *B. cereus* plasmids. The color scale indicating the level of homology is presented in the legend.

coupling proteins (CP or T4CP) (Grohmann *et al.*, 2003). In pIS56-63 the Dtr proteins are represented by CDS 22 for replicase/relaxase, and domains of CDSs 40, 41 and 43, responsible for plasmid replication, its transformation to ssDNA and intercellular transport. Functional domains in CDSs 15, 18, 20, 38 show similarity to the Mpf domains that mediate channels for nucleic acids transfer, while the domain within CDS 23 corresponds to VirD4, a protein known to interact with the relaxosome as a receptor (CP) during transfer of ssDNA through the conjugative channel. CDS 39 displayed 31% amino acid similarity to VirB11 (Souza *et al.*, 2012), a protein implicated in both Type II and Type IV secretion systems in linking the transmembrane domain with the inner membrane and stabilizing the transferred ssDNA. In fact, the *virB11* homolog is not known to be widespread among large plasmids in Gram-positive bacteria, though it has been described only among pXO2-like plasmids, such as pAW63 and p9727 (Van der Auwera *et al.*, 2005), which like pIS56-63, are toxigenic extrachromosomal elements. Moreover, CDS 14 was identified as a glucosaminidase that potentially has similar functions as the transglycosylase VirB1 that catalyzes murein degradation to create local pores, and CDS 17 coding for a cytoplasmic protein with ATPase activity which can be considered as a potential energy supplier.

The presence of the M2-*traI* (CDSs associated with membrane function) and M2-*traII* (CDSs associated with cytoplasmic functions) that differed in a G + C content in the conjugative region of pIS56-63 have not yet been observed among the *B. cereus s.l.* plasmids, and thus appear to be a novel feature among toxigenic plasmids in *Bacillus* species. Plasmids are among the

largest genetic elements capable of being mobilized among bacterial populations, and in many instances their acquisition can confer selective advantages to recipients (Toussaint *et al.*, 2002). In *B. cereus sensu lato* conjugation plays a key role in genes dispersion among members of the group (Van der Auwera *et al.*, 2005; Yuan *et al.*, 2010).

Characteristics of M3-*reg*, the regulatory module. M3-*reg* is composed of 6,168 bp and contains genes that putatively mediate regulatory functions (Table I, Fig. 4). This module harbors 11 CDSs (CDS 44 to 53) and has a G + C content of 32.4% that deviates by 2.3% from the average G + C content of its parental plasmid. CDS 47, 48, and 50 were predicted to be transcriptional regulators. CDS 47 and 48, contain a helix-turn-helix motif known in the XRE-like protein family and a MerR family transcriptional regulators, respectively, and showed similarity to corresponding sequences in pBMB28 of *B. thuringiensis* YBT-020, and pCT281 of *B. thuringiensis* [GenBank: CP001910]. CDS 50 which is related to the HTH-type transcriptional regulator SinR, was identical to the corresponding element in p03 and showed some similarity to pG9842_209 of *B. cereus* G9842 (coverage/max. identity: 46/75; [GenBank: CP001187]). CDS 51 and 52 formed a two-component signal transduction system. The HTH proteins together with sigma factors participate in a wide range of signaling pathways, for example, as repressors that inhibits sporulation (Cervin *et al.*, 1998), biofilm formation (Colledge *et al.*, 2011), and proteases secretion (Pflughoeft *et al.*, 2011). The two-component signal transduction system, encoded by CDS 51 and 52, allows cross-regulated gene expression, which in regard to pIS56-63 may be involved in regulating the T4SS system and conjugation (Martínez-Núñez *et al.*, 2010) as well as sporulation and accompanying Cry protein synthesis (Malvar *et al.*, 1994), CDS 45, 46, 49, and 53 encoded putative proteins, but their functions cannot be predicted at present (Table I). BLAST analysis (Fig. 4) showed 100% identity of these CDSs to corresponding fragment in pHT73, and p03, with the exception of CDS 44, a S-layer protein (coverage/max. identity: 75/100).

Characteristics of M4-*rep*, the replication module. The M4-*rep* module (1,721 bp), responsible for pIS56-63 replication, is composed of CDS 54 and its A + T rich regions (Fig. 4). A Shine-Dalgarno sequence was detected upstream of CDS 54. Also, direct and inverted repeats, called iterons, were found in the region preceding the start codon of CDS 54. Iterons are known to play a significant role in regulating plasmid copy number in bacterial cells, and to form complexes with replication (Rep) initiator proteins to achieve the regulatory effect. CDS 54 was identified as a putative replication associated protein, characterized by the presence of a winged helix-turn-helix DNA-binding domain. This protein was identical in sequence to the replication protein of p03 in *B. thuringiensis* HD771, pHT73 of *B. thuringiensis* HD73, and pCT83 of *B. thuringiensis* CT-43, and displayed strong similarity with replication proteins encoded by pBMB28 (coverage/max. identity: 99/94) of *B. thuringiensis* YBT-020, pFR55 in *B. thuringiensis* INTA-FR7-4 (coverage/max. identity: 46/65; [GenBank: EU362919.1]), and pAH187_45 (coverage/max. identity: 46/65) of *B. cereus* AH187. The *ori44* replicon was not found in the remaining thirteen plasmids in strain IS5056. The high conservation among the origin of replication and replication protein of pIS56-63 and other *B. cereus s.l.* plasmids, including pXO1-like plasmids, suggests that these toxigenic elements replicate by a conserved theta mechanism (Yuan *et al.*, 2010).

In summary, we described the complete nucleotide sequence and annotation of the toxigenic pIS56-63 plasmid that harbors the *cry1Ab21* entomotoxin, one of the 14 plasmids harbored by *B. thuringiensis* IS5056 (Murawska *et al.*, 2013; Swiecicka *et al.*, 2008). Importantly, our analyses provide further detailed insights into the gene composition of pIS56-63 and, in particular, its modular genetic architecture (M1-*mic*, M2-*tra*, M3-*reg*, M4-*rep*). The latter strongly suggests that the hybrid nature of pIS56-63 evolved through processes that involved recombination, in part mediated by transposition events. The flanking of the mobile insertion cassette (MI-*mic*) by the IS*232* insertion sequences, known to occur among *B. thuringiensis* strains only in opposite to IS*231* and widely propagated in the *B. cereus* group (Leonard *et al.*, 1997), and the high homology of the pIS56-63 modules to corresponding sequences in *B. thuringiensis*, indicate that the arrangement is typical of certain plasmids in *B. thuringiensis*. Our study, together with the few that have been previously described (Barry *et al.*, 2002; Van der Auwera *et al.*, 2005; Yuan *et al.*, 2010) lays a foundation for future analyses and characterization of large toxigenic plasmids among *Bacillus* species, as it relates to their replication, horizontal transfer, and virulence determinants.

Finally, various strains of *B. thuringiensis* are species-specific entomopathogens, and differ from one another primarily by the plasmid cassettes they harbor. Very little is known about (i) the mechanisms that maintain, or (ii) exclude, the plethora of plasmids in a given strain, (iii) the co-evolution of various modules, such as those described here, and (iv) the functional significance of genes, other than those that code for crystal toxins, in the pathobiology of *B. thuringiensis*. Regarding the latter, for example, it is tempting to speculate that the N-acetylmuramoyl-L-alanine amidase (CDS 5) encoded by pIS56-63, does indeed play a role in increasing biogenesis of vegetetive cells and spores of the bacterium following invasion of its target insect host. Further analyses, including genes deletion studies,

are required to establish whether such is the case, and are part of our ongoing work. In this regard, although complete genome sequences are accumulating in nucleotide databases that are publically available, detailed published descriptions of extrachromosomal elements of *B. thuringiensis* are generally lacking for these useful entomopathogens. It is hoped that detailed future comprehensive *in silico* analyses of these elements, combined with functional analyses, would shed light on the evolution, maintenance, and functional biology that allow various *B. thuringiensis* strains to thrive in their specific niches.

Acknowledgements

We thank Dennis K. Bideshi (California Baptist University and University of California Riverside, USA) for his helpful comments. This work was supported by grant N N302 656640 of the Ministry of Science and Higher Education in Poland (I. Swiecicka).

Literature

Berry C., S. O'Neil, E. Ben-Dov, A.F. Jones, L. Murphy, M.A. Quail, M.T. Holden, D. Harris, A. Zaritsky and J. Parkhill. 2002. Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. *israelensis*. *Appl. Environ. Microbiol.* 68: 5082–5095.

Bourguet F.A., B.E. Souza, A.K. Hinz, M.A. Coleman and P.J. Jackson. 2012. Characterization of a novel lytic protein encoded by the *Bacillus cereus* E33L gene ampD as a *Bacillus anthracis* antimicrobial protein. *Appl. Environ. Microbiol.* 78: 3025–3027.

Carver T., N. Thomson, A. Bleasby, M. Berriman and J. Parkhill. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25: 119–120.

Cervin M.A., R.J. Lewis, J.A. Brannigan and G.B. Spiegelman. 1998. The *Bacillus subtilis* regulator SinR inhibits spoIIG promoter transcription *in vitro* without displacing RNA polymerase. *Nucleic Acids Res.* 16: 3806–3812.

Colledge V.L., M.J. Fogg, V.M. Levdikov, A Leech, E.J. Dodson and A.J. Wilkinson. 2011. Structure and organisation of SinR, the master regulator of biofilm formation in *Bacillus subtilis. J. Mol. Biol.* 411: 597–613.

Delcher A.L., K.A. Bratke, E.C. Powers and S.L. Salzberg. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23: 673–679.

Gayathri P., T. Fujii, J. Møller-Jensen, F. van den Ent, K. Namba and J. Löwe. 2012. A bipolar spindle of antiparallel ParM filaments drives bacterial plasmid segregation. *Science* 338: 1334–1337.

Grohmann E., G. Muth and M. Espinosa. 2003. Conjugative plasmid transfer in Gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* 67: 277–301.

Helgason E., O.A. Okstad, D.A. Caugant, H.A. Johansen, A. Fouet, M. Mock, I. Hegna and A.-B. Kolstø. 2000. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* – one species on the basis of genetic evidence. *Appl. Environ. Microbiol.* 66: 2627–2630.

Hoton F., L. Andrup, I. Swiecicka and J. Mahillon. 2005. The cereulide genetic determinants of emetic *Bacillus cereus* are plasmid-borne. *Microbiology* 151: 2121–2124.

Johnson J.W., J.F. Fisher and S. Mobashery. 2013. Bacterial cell-wall recycling. *Ann. N. Y. Acad. Sci.* 1277: 54–75.

Krogh A., B. Larsson, G. von Heijne and E.L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305: 567–580.

Leonard C., Y. Chen and J. Mahillon. 1997. Diversity and differential distribution of IS*231*, IS*232* and IS*240* among *Bacillus cereus*, *Bacillus thuringiensis* and *Bacillus mycoides*. *Microbiology* 143: 2537–2547.

Malvar T., C. Gawron-Burke and J.A. Baum. 1994. Overexpression of *Bacillus thuringiensis* HknA, a histidine protein kinase homology, bypasses early Spo mutations that result in CryIIIA overproduction. *J. Bacteriol.* 176: 4742–4729.

Marchler-Bauer A., C. Zheng, F. Chitsaz, M.K. Derbyshire, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki and others. 2013. CDD: conserved domains and protein three-dimensional structure. *Nucleic. Acids. Res.* 41: D348–D352.

Martínez-Núñez C., P. Altamirano-Silva, F. Alvarado-Guillén, E. Moreno, C. Guzmán-Verri and E. Chaves-Olarte. 2010. The Two-Component System BvrR/BvrS Regulates the Expression of the Type IV Secretion System VirB in Brucella abortus. *J. Bacteriol.* 192: 5603–5608.

Mock M. and A. Fouet. 2001. Anthrax. *Ann. Rev. Microbiol.* 55: 647–671.

Murawska E., K. Fiedoruk, D.K. Bideshi and I. Swiecicka. 2013. Complete genome sequence of *Bacillus thuringiensis* subsp. *thuringiensis* IS5056, an isolate highly toxic to *Trichoplusia ni. Genome Announcem*. 2: e0010813.

Pflughoeft K.J., P. Sumby and T.M. Koehler. 2011. *Bacillus anthracis* sin locus and regulation of secreted proteases. *J. Bacteriol.* 193: 631–639.

Rutherford K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.A. Rajandream and B. Barrell. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16: 944–945.

Sanahuja G., R. Banakar, R.M. Twyman, T. Capell and P. Christou P. 2011. *Bacillus thuringiensis*: a century of research, development and commercial applications. *Plant Biotechnol. J.* 9: 283–300.

Souza R.C., G.D. Quispe Saji, M.O. Costa, D.S. Netto, N.C. Lima, C.C. Klein, A.T. Vasconcelos and M.F Nicolas. 2012. AtlasT4SS: A curated database for type IV secretion systems. *BMC Microbiol.* 12: 172.

Stenfors Arnes L.P., A. Fagerlund and P.E. Granum. 2008. From soil to gut: *Bacillus cereus* and its food poisoning toxins. *FEMS Microbiol. Rev.* 32: 579–606.

Sullivan M.J., N.K. Petty and S.A. Beatson. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics* 27: 1009–1010.

Swiecicka I., D.K. Bideshi and B.A. Federici. 2008. Novel isolate of *Bacillus thuringiensis* subsp. *thuringiensis* that produces a quasi-cuboidal crystal of Cry1Ab21 toxic to larvae of *Trichoplusia ni. Appl. Environ. Microbiol.* 74: 923–930.

Swiecicka I. 2008. Natural occurrence of *Bacillus thuringiensis* and *Bacillus cereus* in eukaryotic organisms: a case for symbiosis. *Biocontrol Sci. Technol.* 18: 221–239.

Toussaint A. and C. Merlin. 2002. Mobile elements as a combination of functional modules. *Plasmid* 47: 26–35.

Treangen T.J. and E.P.C Rocha. 2011. Horizontal transfer, not duplication, drives the expansion of protein families in Prokaryotes. *PLoS Genet*. 7: e1001284.

Van der Auwera G.A., L. Andrup and J. Mahillon. 2005. Conjugative plasmid pAW63 brings new insights into the genesis of the *Bacillus anthracis* virulence plasmid pXO2 and the *Bacillus thuringiensis* plasmid pBT9727. *BMC Genomics* 6: 103.

Yuan Y., D. Zheng, X. Hu, Q. Cai and Z. Yuan. 2010. Conjugative transfer of insecticidal plasmid pHT73 from *Bacillus thuringiensis* to *B. anthracis* and compatibility of this plasmid with pXO1 and pXO2. *Appl. Environ. Microbiol.* 76: 468–473.