Genetic Environment of *cry1* Genes Indicates Their Common Origin

Krzysztof Fiedoruk[1,*], Tamara Daniluk[1], Jacques Mahillon[2], Katarzyna Leszczynska[1], and Izabela Swiecicka[3,4]

[1]Department of Microbiology, Medical University of Bialystok, Poland

[2]Laboratory of Food and Environmental Microbiology, Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

[3]Department of Microbiology, University of Bialystok, Poland

[4]Laboratory of Applied Microbiology, University of Bialystok, Poland

*Corresponding author: E-mail: krzysztof.fiedoruk@umb.edu.pl.

Accepted: August 22, 2017

Abstract

Although in *Bacillus thuringiensis* the *cry* genes coding for the insecticidal crystal proteins are plasmid-borne and are usually associated with mobile genetic elements, several aspects related to their genomic organization, diversification, and transmission remain to be elucidated. Plasmids of *B. thuringiensis* and other members of the *Bacillus cereus* group ($n = 364$) deposited in GenBank were screened for the presence of *cry1* genes, and their genetic environment was analyzed using a comparative bioinformatic approach. The *cry1* genes were identified in 27 *B. thuringiensis* plasmids ranging from 64 to 761 kb, and were predominantly associated with the *ori44*, *ori60*, or double *orf156/orf157* and *pXO1-16/pXO1-14* replication systems. In general, the *cry1* genes occur individually or as a part of an insecticidal pathogenicity island (PAI), and are preceded by genes coding for an *N*-acetylmuramoyl-L-alanine amidase and a putative $K^+(Na^+)/H^+$ antiporter. However, except in the case of the PAI, the latter gene is disrupted by the insertion of IS*231*B. Similarly, numerous mobile elements were recognized in the region downstream of *cry1*, except for *cry1I* that follows *cry1A* in the PAI. Therefore, the cassette involving *cry1* and these two genes, flanked by transposable elements, named as the *cry1* cassette, was the smallest *cry1*-carrying genetic unit recognized in the plasmids. Conservation of the genomic environment of the *cry1* genes carried by various plasmids strongly suggests a common origin, possibly from an insecticidal PAI carried by *B. thuringiensis* megaplasmids.

Key words: *Bacillus thuringiensis*, *cry1* genes, IS231, IS232, pathogenicity island, insertion sequences.

Introduction

Bacillus thuringiensis is a Gram-positive, spore-forming bacterium, known for the production of parasporal crystal inclusions composed of insecticidal Cry proteins. These toxins create a heterogeneous family of 74 different types of proteins (Cry1–Cry74) (Crickmore et al. 2016) harmful for various insect genera, including Lepidoptera, Coleoptera, Diptera, Hemiptera, and even some nematodes and snails (Palma et al. 2014). Within these toxins, Cry1 represent the most abundant group, accounting for ~17% of all Cry toxins (Crickmore et al. 2016), and exhibit activity against pests causing the highest damages in crops and forests. For example, Cry1A are toxic to Lepidoptera, whereas Cry1B and Cry1I have dual activity against Lepidoptera and Coleoptera. The Cry–host specificity is partly a consequence of prerequisite conditions for the Cry activation. For instance, the Cry1A crystals must be first solubilized in the alkaline midgut of Lepidoptera and subsequently processed into active forms by the proteolytic digestion of specific serine proteases (Palma et al. 2014). However, other factors associated with toxins processing or stability in the insects midgut, aside from the receptor binding, can influence Cry specificity and activity (reviewed by Jurat-Fuentes and Crickmore 2017).

In general, most *cry* genes are plasmid-borne and are usually flanked by various mobile elements such as insertion sequences (IS*231*, IS*232*, IS*240*, IS*Bt1*, and IS*Bt2*) and transposons (Tn*4430* or Tn*5401*) (Mahillon et al. 1994; Léonard et al. 1997; Mahillon and Chandler 1998; Schnepf et al. 1998). For example, *cry1A* genes are frequently found within a composite transposon structure flanked by IS*232* (Menou et al. 1990; Murawska et al. 2014). In addition, the *cry* genes can be organized in operons and/or co-localized with other toxin genes, for example those of the vegetative insecticidal proteins (*vip*), forming insecticidal pathogenicity islands (PAI)

(Aronson 1993; Zhu et al. 2015). This particular genetic environment of the *cry* genes is believed to be responsible for their amplification in bacterial cells, and to facilitate their recombination and exchange among plasmids, providing a source of novel specificities in crystal-producing strains (Aronson 1993; Léonard et al. 1997; Schnepf et al. 1998; Palma et al. 2014). However, the actual transfer mechanisms and the role of the transposable elements in the evolution of *cry*-carrying plasmids remain to be elucidated. Similarly, little attention has been paid to non-toxin genes located in the immediate *cry* genetic environment, while, as a part of insecticidal PAIs, they may contribute to the virulence or be implicated with certain *B. thuringiensis* specific traits, for example activation of spore germination in an alkaline pH (Abdoarrahem et al. 2009).

The aim of the study was to investigate the genetic and genomic environments of the *cry1* genes. To this end, a comparative analysis of the *cry1* loci was performed using the various *B. thuringiensis* plasmids deposited in GenBank.

Materials and Methods

In total, a set of 406 complete plasmids from *Bacillus thuringiensis* and other *Bacillus cereus* group members deposited in GenBank were screened for the presence of *cry1* genes using the *B. thuringiensis* Toxin_scanner (Ye et al. 2012) (supplementary table S1, Supplementary Material online). A genetic environment of *cry1* was visualized with Easyfig tool (Sullivan et al. 2011). In addition, ISfinder (Siguier et al. 2006) and PHASTER (Arndt et al. 2016) website tools were used in order to identify mobile elements and prophage regions, respectively. DNA sequence alignment and analysis were performed with CLC Genomic Workbench software (CLC Bio), and Blast2GO software was used for functional annotation of proteins (Conesa et al. 2005).

Results

Strains and Plasmids Characteristics

The *cry1*-carrying plasmids were detected in 15 *B. thuringiensis* strains (table 1). However, it should be noted that in the case of *B. thuringiensis* serovar (sv.) *kurstaki* strains HD-1 and YBT-1520 the same plasmids from different sequencing projects were included into analyses, since notable discrepancies were observed in their sequences (see below).

Eight *B. thuringiensis* strains have only one *cry1* bearing plasmid, whereas two or even three plasmids with *cry1* were noted in three (CT-43, HD-12, ATCC 10792) and four (HD-1, HD-29, YBT-1520, and IS5056) strains, respectively (table 1 and fig. 1). In total, we identified 14 plasmids carrying only a single *cry1* gene, *cry1A* ($n = 9$), *cry1B* ($n = 4$), or *cry1F* ($n = 1$), and 12 plasmids with more than one *cry1* or other *cry* genes (table 1). The latter group involved six megaplasmids carrying the PAI, as described by Zhu et al. (2015), which contains the *cry1Aa*, *cry1Ia*, *cry2Aa*, *cry2Ab*, and *vip3Aa* genes (hereinafter referred to as the "insecticidal PAI"), and additional seven megaplasmids containing its variants or derivatives (supplementary fig. S1, Supplementary Material online).

Replication System and Size of *B. thuringiensis* Plasmids Harbouring *cry1*

As shown in table 1, the majority of plasmids with a single *cry1A* gene possess the replication systems *ori44* ($n = 5$), *ori60* ($n = 5$), *ori43* ($n = 1$), or both *ori43/ori60* ($n = 1$) (Baum and Gilbert 1991) and are of a size below 100 kb, except for plasmids pBMB126 (126 kb and two *ori43/ori60* replication systems) and pHD120161 (161 kb and *repA* replication type). Similarly, plasmids ($n = 4$) with a single *cry1B* have the *ori60* replication system and size of 107 kb (pIS56–107), 109 kb (pBMB0558), 113 kb (poh2), and 127 kb (pCT127 and another replication system, *repA_N*). In contrast, plasmids with several *cry* genes are larger (from 250 to 761 kb) and are characterized by double *orf156/orf157* and *pXO1-16/pXO1-14* replication systems (Tang et al. 2007; Pomerantsev et al. 2009; Zheng et al. 2013).

Distribution and Variation of the *cry1* Genes

The *cry1* genes were represented by 12 types (*A–J*, *M*, and *N-like*) (fig. 1). The *cry1A* (*Aa*, *Ab*, *Ac*, and *Ae*) genes constitute the largest group among *cry1* (49%) and are present either in the insecticidal PAIs or separately. In addition, all *cry1A* located in the insecticidal PAI (fig. 2 and supplementary fig. S1, Supplementary Material online) of plasmids with the double *orf156/orf157* and *pXO1-16/pXO1-14* replication systems, belong to *cry1Aa* subtype, whereas *cry1Ac* and *cry1Aa*-c subtypes are associated with plasmids of the *ori60* and *ori44* replication systems, respectively (figs. 1 and 2). In contrast, *cry1Ia* or *cry1Id*, which represent 21% of the *cry1* genes, were found only within the insecticidal PAI, where they are located downstream of *cry1A* and *cry1E* (*cry1Ia)* or *cry1D* (*cry1Id)*. The *cry1N-like*, *cry1C*, *cry1E*, *cry1F*, *cry1G*, *cry1H*, *cry1J*, and *cry1M* genes occur only in individual plasmids, mostly as a part of the insecticidal PAI variants (supplementary fig. S1, Supplementary Material online).

Genetic Environment of the *cry1* Genes

In general, in the insecticidal PAI *cry1A* is followed by *cry1I* (fig. 3 and supplementary fig. S1, Supplementary Material online). In contrast, various mobile elements were found downstream of the *cry1* genes located outside the PAI or in the remaining plasmids, namely (i) IS*231*C followed by Tn*4430* (pIS56–63, pBMB65, and p03), (ii) IS*232* (pHT73, pBMB95 from strain HD1, pBMB126, and pYC1), (iii) a putative transposon (Tn*3* family) followed by Tn*4652* (pBMB95 from strain YBT-1520) and Tn*4430* (pBMB69), (iv) an IS*110* family member

Table 1

Characteristics of *B. thuringiensis* Strains and Their *cry1*-Carrying Plasmids

B. thuringiensis serovar and strain	Isolation			ST[a]	Plasmid						
	Country	Year	Source		Name	Accession Number	Size (bp)	Replication type	Insecticidal PAI[b]	*cry1* gene[c,d]	Other insecticidal toxin genes
kurstaki YBT-1520	China	1990	Soil	8	pBMB69	CP007613	69,416	*ori44*	No	*cry1Ac*	No
					pBMB293[e]	CP004861	293,574	*orf156/orf157 pXO1-16/pXO1-14*	Yes	(*cry1Aa, cry1Ia*)[c]	*cry2Ab, cry2Aa, vip3Aa*
					pBMB95	CP007614	94,637	*ori60*	No	*cry1Ac*	No
					pBMB94	CP004868	94,568	*ori60*	No	*cry1Ac*	No
kurstaki HD-1	United States	1967	*Pectinophora gossypiella*	8	pBMB95	CP004875	95,983	*ori60*	No	*cry1Ac*	No
					pBMB65	CP004873	65,873	*ori44*	No	*cry1Ab3*	No
					unnamed6	CP010003	69,317	*ori44*	No	*cry1Ab3*	No
					pBMB299	CP004876	299,843	*orf156/orf157 pXO1-16/pXO1-14*	Yes	(*cry1Aa, cry1Ia*)[c]	*cry2Ab, cry2Aa, vip3Aa*
					unnamed2	CP009999	317,336	*orf156/orf157 pXO1-16/pXO1-14*	Yes	*cry1Ac*, (*cry1Aa, cry1Ia*)[c]	*cry2Ab, cry2Aa, vip3Aa*
thuringiensis IS5056	Poland	2005	Soil	10	pIS56-63	CP004131	63,864	*ori44*	No	*cry1Ab21*	No
					pIS56-107	CP004134	107,431	*ori60*	No	*cry1Ba*	No
					pIS56-285	CP004136	285,459	*orf156/orf157 pXO1-16/pXO1-14*	Yes	(*cry1Aa, cry1Ia*)[c]	*cry2Ab, cry2Aa, vip3Aa*
chinensis CT-43	China	ND[f]	ND[f]	10	pCT127	CP001908	127,885	*ori60, repA_N*	No	*cry1Ba*	No
					pCT281	CP001910	281,231	*orf156/orf157 pXO1-16/pXO1-14*	Yes	(*cry1Aa, cry1Ia*)[c]	*cry2Ab, cry2Aa, vip3Aa*
galleriae HD-29	Czechoslovakia	1970	*Dendrolimus sibericus*	15	pBMB267	CP010091	267,359	*orf156/orf157 pXO1-16/pXO1-14*	Yes	(*cry1Aa, cry1Ia*)[c]	*cry2Ab, vip3Aa*
					pBMB426	CP010090	426,282	*orf156/orf157-like pXO1-16/pXO1-14*	Yes	(*cry1Ca, cryDa*)[c]	No
					pBMB126	CP010092	126,898	*ori60, ori43*	No	*cry1Ac*	No
HD-12	United States	2012	Soil	23	pHD120345	CP014853	345,196	*orf156/orf157 pXO1-16/pXO1-14*	Yes	(*cry1Hb, cry1Bb, cry1Ab24, cry1I-like, cry1Ja, cry1Id2, cry1Da3*)[c]	*cry2Ad, vip2Af, vip1Ca, vip2Ac, vip1Ba-like*
					pHD120161	CP014852	161,353	*repA*	No	*cry1Fb*	No
YC-10	China	2010	*Nicotiana tabacum* roots	8	pYC1	CP011350	761,374	*orf156/orf157 pXO1-16/pXO1-14*	Yes	*cry1Ac*, (*cry1Aa, cry1Ia36*)[c]	*cry2Aa, cry2Ab35, vip3Aa*
YWC2-8	China	2007	Soil	8	pYWC2-8-1	CP013056	250,706	*orf156/orf157 pXO1-16/pXO1-14*	Yes	(*cry1Ac*)[c]	*cry2Aa*

(continued)

Table 1 Continued

B. thuringiensis serovar and strain	Isolation			ST[a]	Plasmid						
	Country	Year	Source		Name	Accession Number	Size (bp)	Replication type	Insecticidal PAI[b]	*cry1* gene[c,d]	Other insecticidal toxin genes
HD-771	ND[f]	ND[f]	ND[f]	12	p03	CP003755	69,876	*ori44*	No	*cry1Aa*	No
kurstaki HD-73	ND[f]	ND[f]	ND[f]	8	pHT73	CP004070	77,351	*ori44*	No	*cry1Ac1*	No
alesti BGSC 4C1	Czechoslovakia	1987	*Bombyx mori*	12	pBMB267	CP015177	267,609	*pXO1-16/pXO1-14*	*Yes*	(*cry1Ae*, *cry1Ae*[d], *cry1Gb*[d], *cry1Ma*[d])[c]	*cry2Ab*, *vip3Aa*
tolworthi Pasteur Institute Standard	ND[f]	ND[f]	ND[f]		pKK2	NZ_AP014866		*orf156/orf157* *pXO1-16/pXO1-14*		(*cry1Ea*, *cry1Ia*[d])[c]	*cry2Aa*
CT-43	ND[f]	ND[f]	ND[f]	ND[f]	pBMB0558	NC_014937	109,464	*ori60*	No	*cry1Ba*[d]	No
L-7601	China	2015	ND[f]	197	unnamed3	CP020005	408,071	*orf156/orf157* *pXO1-16/pXO1-14*	Yes	(*cry1Ab*, *cry1Ia*, *cry1Ab*[d], *cry1Ia*[d], *cry1Bd*)[c]	*vip3Ah*
berliner ATCC 10792	Israel	ND[f]	*Ephestia kuhniella*	10	poh4	CP021065	86,488	*ori43*	No	*cry1Ab*	No
					poh2	CP021063	113,294	*ori60*	No	*cry1Ba*	No

[a]ST, sequence type determined by MLST (http://mlstoslo.uio.no/, last accessed April 3, 2017).
[b]PAI, Pathogenicity Island.
[c]*cry1* genes in brackets are located in the insecticidal PAI as showed in supplementary fig. S1, Supplementary Material online.
[d]Pseudogene.
[e]Identical plasmid sequence is deposited in GenBank under Acc. number CP007615.
[f]ND, not determined.

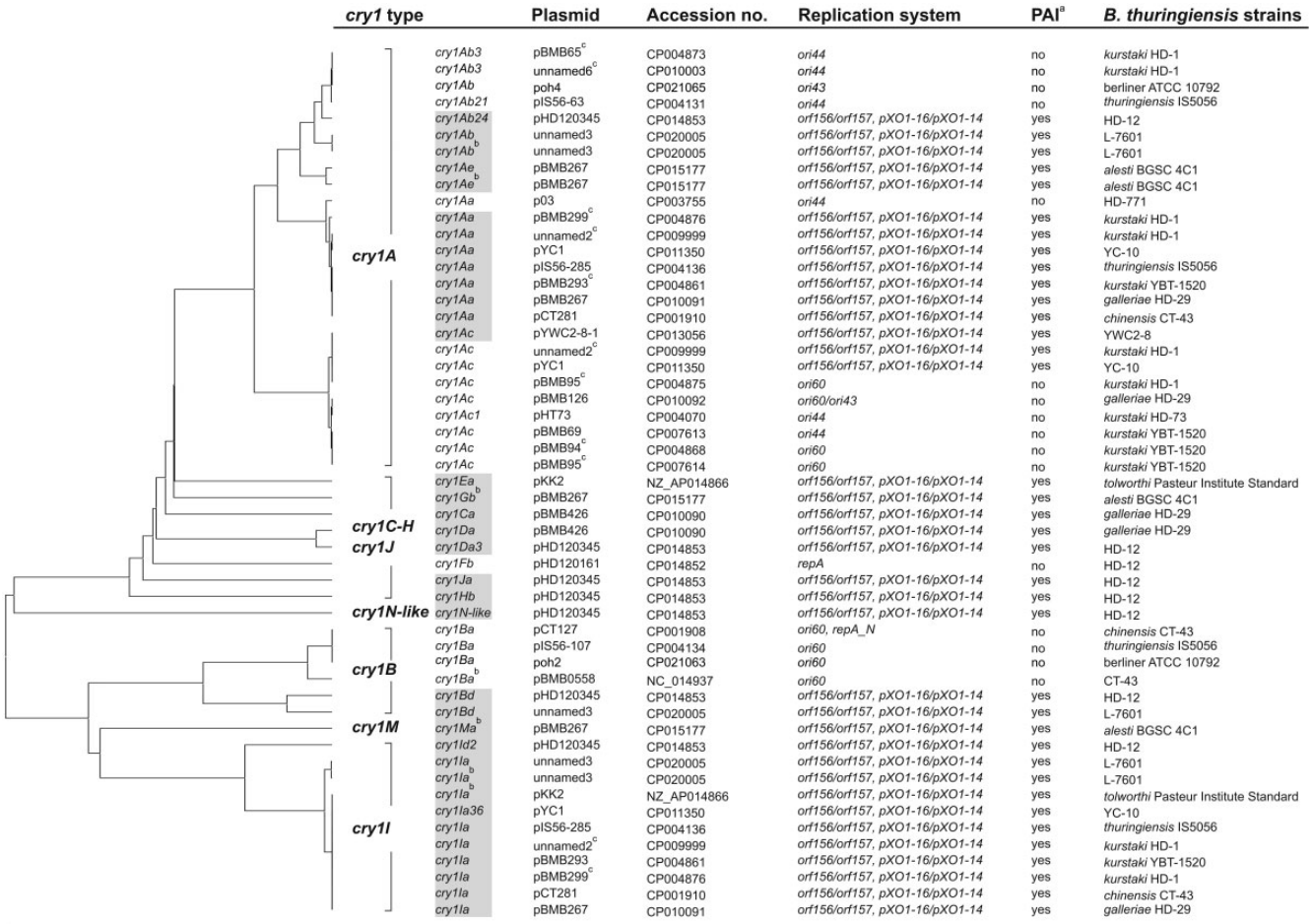

Fig. 1.—Comparison of the *cry1* genes from *B. thuringiensis* plasmids using UPGMA clustering. The *cry1* genes located in the insecticidal PAI, as showed in supplementary fig. S1, Supplementary Material online, are highlighted as grey boxes.

(pHD120161) (fig. 3 and supplementary fig. S2, Supplementary Material online). Overall, the mobile elements in the direct or more distal *cry1* environment include insertion sequences, transposons, and elements associated with DNA integration/recombination (integrases, resolvases) as well as group II intron reverse transcriptase/maturase genes (supplementary fig. S2, Supplementary Material online). More specifically, ISs are represented by four families: (i) IS*4* (IS*231*B, IS*231*C, IS*232*S, and IS*Bth4*), (ii) IS*21* (IS*232*), (iii) IS*110* (IS*Bth166*), and (iv) IS*200*/IS*605* (IS*Bth16*), while all the Tn transposase genes belong to the Tn*3* family. An accumulation of various transposable elements in *cry1* proximity is particularly apparent in the *ori44*-replication plasmids, where they represent up to 46% of the plasmid size (supplementary fig. S2, Supplementary Material online).

Within the downstream *cry1* sequence (except for *cry1Ia*) we noted the inverted repeat motif described by Wong and Chang (1986) as a positive retroregulator that stabilizes *cry1* mRNA. Interestingly, beside the original motif 5′-AAAACGGACATCACCTCC(N_8)GGAGTGATGTCCGTTTT-3′, variants characterized by different secondary mRNA structures were also observed (fig. 4 and supplementary fig. S3, Supplementary Material online). In addition, 45 bp downstream of the retroregulator sequence and 127 bp upstream of the *cry1Ia* promoter sequence, a second inverted repeat structure, 5′-AAGCAGAGATATTTTCA(N_8)TGAAAATATCTCTGCTT-3′ was also noted (supplementary fig. S4, Supplementary Material online).

In contrast, an upstream region of *cry1* is occupied by a gene (or its fragment in the case of *cry1B*) encoding for an *N*-acetylmuramoyl-L-alanine amidase, that is preceded, except for *cry1B*, by gene(s) encoding component(s) of a putative $K^+(Na^+)/H^+$ antiporter (fig. 3). However, in virtually all cases (but not in the PAI), the latter gene is disrupted by the insertion of IS*231*B and is usually followed by IS*232*. Therefore, the gene cassette involving *cry1* and these two genes or their

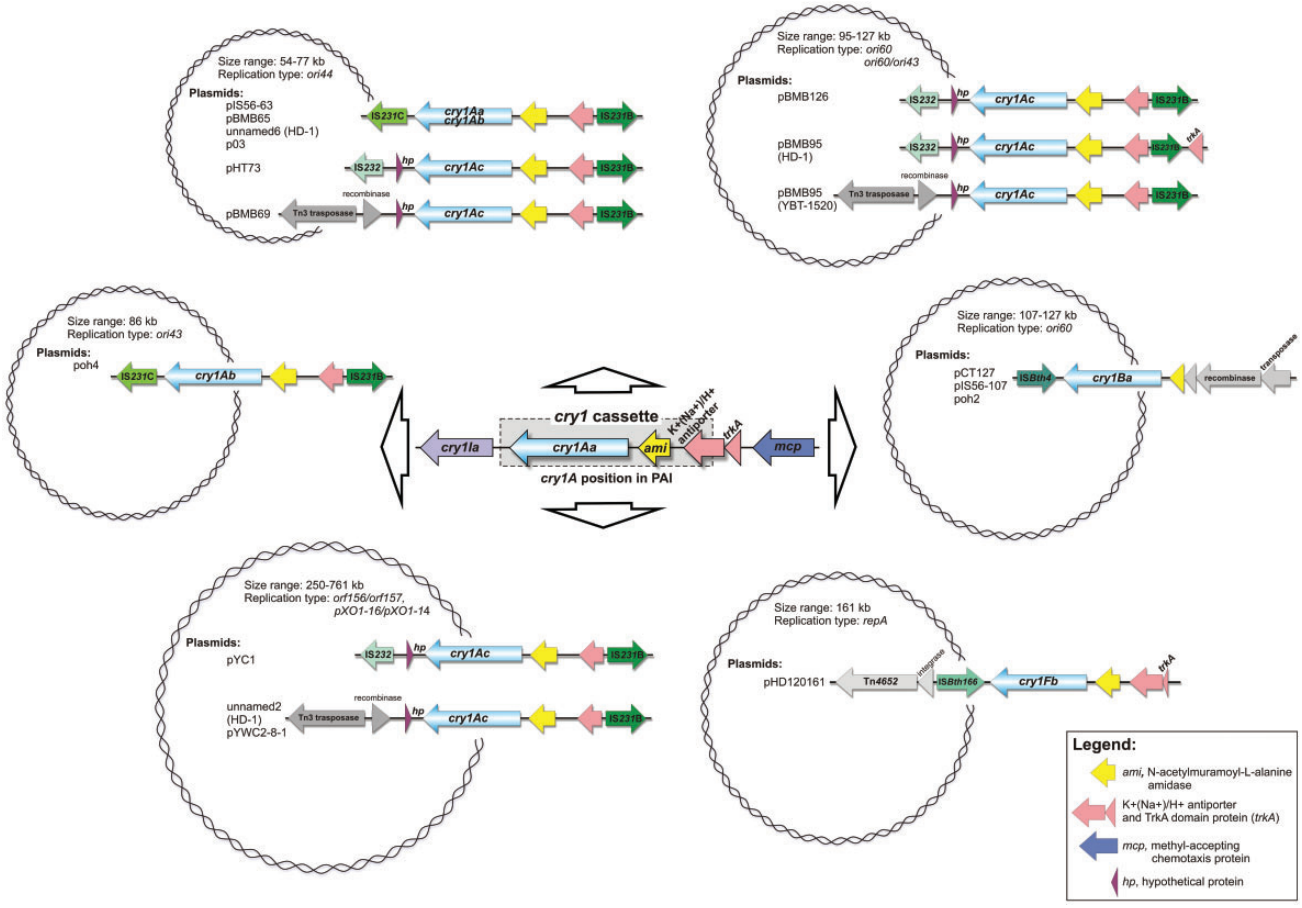

Fig. 2.—Distribution and variability of the *cry1* cassette in *B. thuringiensis* plasmids.

fragments (hereinafter named as the *cry1* cassette) is the smallest *cry1*-carrying (*cry1A*, *cry1F*, or *cry1B*) genetic unit recognized in different plasmids that possibly originated from the PAI (fig. 2). Interestingly, plasmids pYC1 (strain YC-10) and unnamed 2 (strain *kurstaki* HD-1) both carry the PAI and the *cry1* cassette (fig. 3).

Relatedness of *cry1*-Carrying Plasmids to Other *B. cereus* Group Plasmids

For all *cry1*-carrying plasmids, variants missing the *cry1* cassette (supplementary figs. S5–S8, Supplementary Material online) were identified. For instance, plasmids pKK2 (54 kb) and pBMB55 (55 kb) are in fact ~10 kb smaller *cry1*-negative variants of the *ori44*-type *cry1*-carrying plasmids pIS56-63, pBMB65, pHT73, pBMB69, and p03 (supplementary fig. S5, Supplementary Material online). Similarly, the lack of the *cry1* cassette is the only substantial difference between 7 plasmids (72–90 kb) and 12 plasmids (59–89 kb) displaying, respectively, the *ori60* or the *ori43* replication type, as compared with the *cry1A*- or *cry1B*-positive ones (supplementary figs. S6 and S7, Supplementary Material online). Finally, the lack of the large fragment (~120 kb) containing the insecticidal PAI differentiates pBTBC2 (171 kb) from the PAI-positive megaplasmids pBMB293, pIS56-285, pCT281, or pBMB299 (supplementary fig. S8, Supplementary Material online). Such *cry1*-negative plasmids were found exclusively in *B. thuringiensis* strains, except for the plasmid pH308197_73 from the emetic *B. cereus* strain H3081.97 (supplementary fig. S6, Supplementary Material online). However, it must be stated that certain inconsistencies for the presence of *cry1* genes in the same plasmids from different sequencing projects were noted (supplementary fig. S9, Supplementary Material online). For example, the *cry1Ac* cassette from the 95 kb plasmid pBMB95 (CP004875) of *B. thuringiensis* sv. *kurstaki* HD-1 is present in the 317 kb unnamed plasmid (CP009999).

Discussion

The occurrence of various *cry1* genes in the proximity of an *N*-acetylmuramoyl-L-alanine amidase gene (or its fragment), along with $K^+(Na^+)/H^+$ antiporter pseudogene (i.e., disrupted by IS*231*B), strongly suggests their common origin. The insecticidal PAI where both genes are complete and where *cry1A* is adjacent to *cry1I* instead of a mobile genetic element, seems

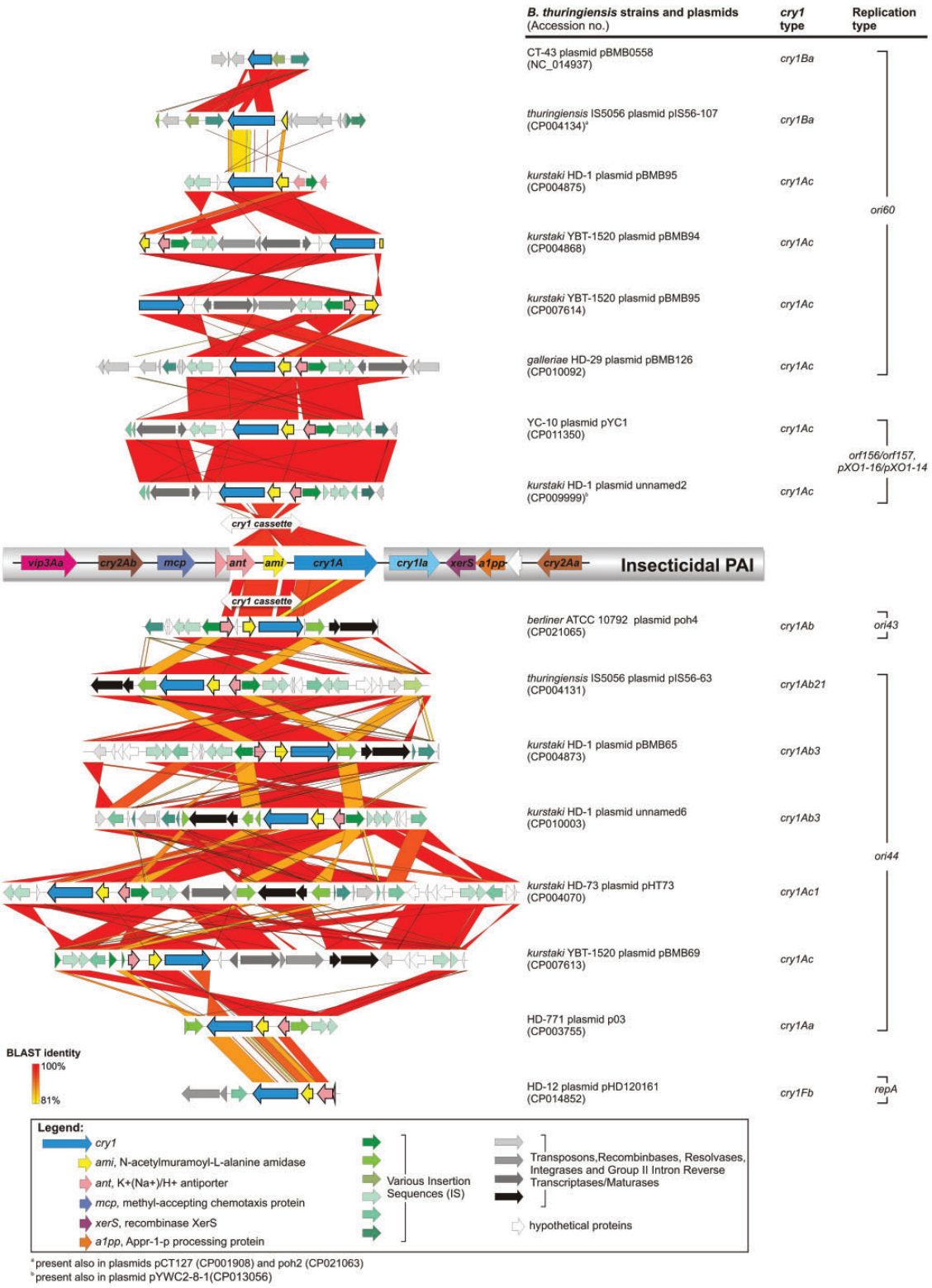

Fig. 3.—The *cry1* gene cassette and its genetic environment in *B. thuringiensis* plasmids. The insecticidal PAI was used as reference. The entirely annotated version is detailed in supplementary fig. S2, Supplementary Material online.

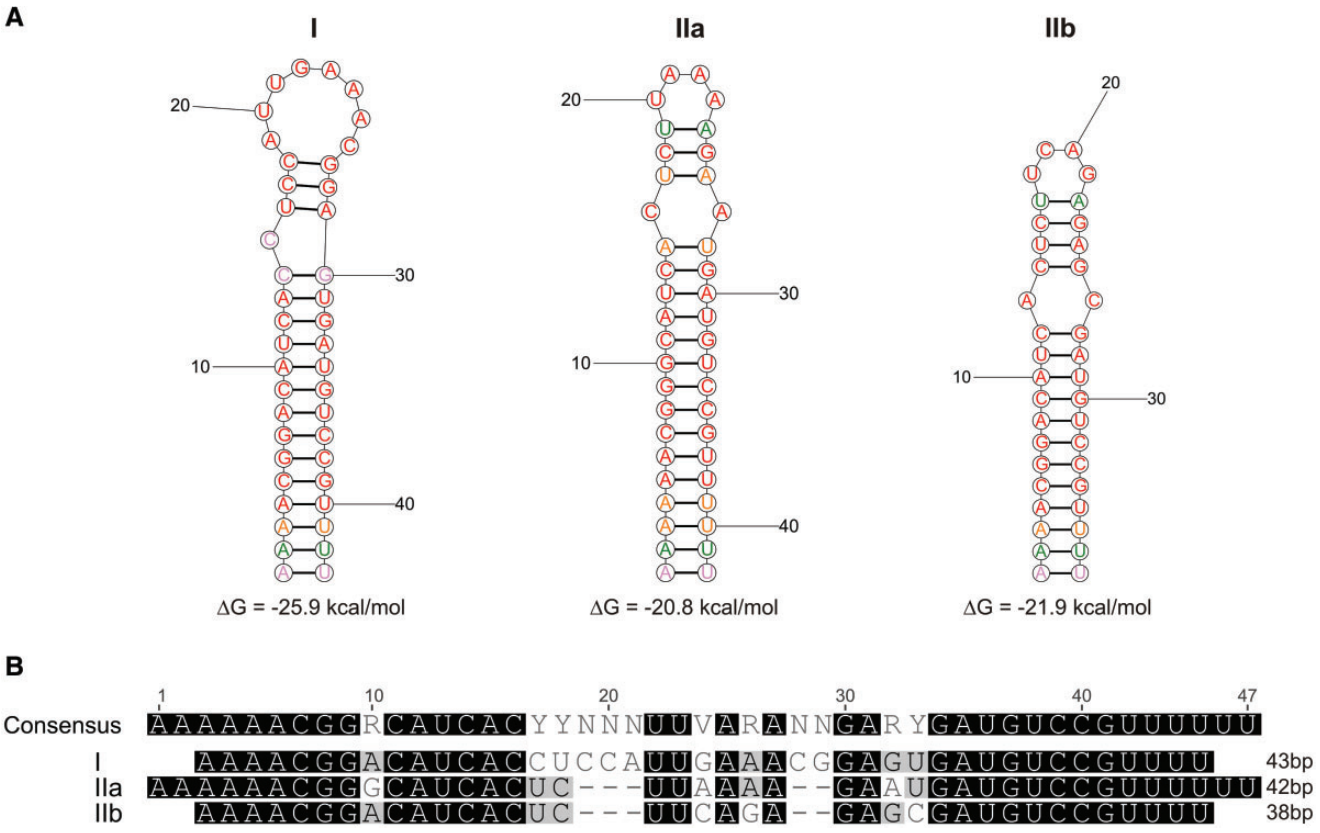

Fig. 4.—(*a*) Prediction of the lowest free energy (ΔG in kcal/mol) structure for mRNA of various variants (I, IIa, and IIb) of the *cry1* positive retroregulator; the analysis was done using RNAstructure web server (http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html, last accessed July 20, 2017). (*b*) Alignment of the I, IIa, and IIb positive retroregulator mRNA variants.

to be the natural candidate for such ancestor (fig. 2 and supplementary fig. S1, Supplementary Material online). Following this idea, we suggest two independent genetic events that "released" the *cry1* cassette from the PAI, (i) a disruption of $K^+(Na^+)/H^+$ antiporter operon by IS*231*B, and subsequently (ii) the insertion of an IS or a transposon into the *cry1* downstream region. The former event is visible in pKK2, where *cry1I* is preserved (supplementary fig. S1, Supplementary Material online). In fact, disruption of various genes by IS*231* elements is not an unusual phenomenon in *B. thuringiensis* (Qiu et al. 2010). However, it will be interesting to examine whether the presence of ISs downstream of the *cry1* genes is related to the presence of two types of inverted repeat sequences, including one acting as positive retroregulators of the *cry1* mRNA (Wong and Chang 1986), since such DNA motifs are common targets for ISs (Tobes and Pareja 2006). Interestingly, the second motif (supplementary fig. S4, Supplementary Material online) is not present in the *cry1* cassettes associated with ISs belonging to IS*4* family, namely IS*231*C and IS*Bth4* (fig. 2), which are flanked by the same direct repeat (5′-TGGCGGTACCC-3′). Further accumulation of numerous transposable elements in this region could be attributed to either homologous recombination or insertion of one transposable element into another. For example, Mahillon et al. (1987) revealed that IS*231* transposed into the IR of Tn*4430* without affecting the transposon structural integrity. Thus, the formation of such transposon-like structures (Menou et al. 1990; Mahillon et al. 1994) may be an important mechanism for the *cry1* cassette transposition among various plasmids. This mobility is supported by the observation of *cry1*-negative plasmids that are otherwise identical to those containing the *cry1* cassette (supplementary figs. S5–S7, Supplementary Material online). Therefore, duplication of the *cry1* cassette followed by further sequence divergence, may have led to small-scale diversification of Cry1 toxins active against the same insect targets, and explain the presence of different members of the same Cry toxin family in individual bacterial isolates (e.g., Cry1Aa, Cry1Ab, and Cry1Ac in *B. thuringiensis* sv. *kurstaki* HD1 or Cry1Aa and Cry1Ab in *B. thuringiensis* IS5056) (Murawska et al. 2013; Palma et al. 2014). In addition, diversification of the *cry1* genes also involves their positive retroregulator sequence that improves the *cry1* mRNA stability, which may be associated with insecticidal efficacy of certain *B. thuringiensis* strains (fig. 4). Furthermore, the presence of *cry1* cassettes, predominantly located in a limited number of plasmid types (i.e., *ori44* and *ori60*), lends further support to the mobilization of *cry*-carrying plasmids between *B. thuringiensis* strains as the

mechanism which explains why identical copies of *cry* genes are distributed among geographically separated isolates (Palma et al. 2014). This is the case for *B. thuringiensis* strain Na205-3 isolated in Spain (Palma et al. 2014a) and *B. thuringiensis* strain CT-43 isolated in China (He et al. 2011) that share some, but not all, *cry*-carrying plasmids with *B. thuringiensis* strain IS5056 from Poland (Swiecicka et al. 2008; Murawska et al. 2013) (table 1). In addition, *cry1I*, *cry1M*, and *cry1N-like* encoding smaller toxins, that is, devoid of C-terminal part responsible for crystallization, appeared to be associated with a gene encoding a "XerS" tyrosine recombinase and limited to plasmids with the PAI (supplementary fig. S6, Supplementary Material online). Although, those proteins are technically naturally truncated versions (65–70 kDa) of the 130–140 kDa toxins, we did not observe genetic events that might support their evolution neither by (i) a fragmentation or disruption of the most related *cry1B* genes nor (ii) in a manner characteristic for *cry10A*, *cry39A*, and *cry40A*, where ORFs encoding N-terminal and C-terminal parts of toxin are separated by short non-coding region (de Maagd et al. 2003). Nevertheless, the presence of *xerS* also in proximity of genes encoding Cry3A, toxic for Coleoptera, might be connected with dual Lepidoptera/Coleoptera activity of Cry1I, that is, as the result of domain swapping, since domains I and II of those toxins are phylogenetically related (de Maagd et al. 2001).

Considering the above observations, the genetic environment of *cry1* should be perceived from the perspective of the insecticidal PAI, where the *N*-acetylmuramoyl-L-alanine amidase and $K^+(Na^+)/H^+$ antiporter genes are preceded by an ORF encoding for a putative methyl-accepting chemotaxis protein (MCP). Consequently, this genetic context qualifies those genes as potential virulence factors or at least as part of *B. thuringiensis* adaptation machinery to its pathogenic lifestyle (Jensen et al. 2003; Raymond et al. 2010). However, an *ad hoc* explanation of this thesis is challenging, since these companion genes are associated with fundamental physiological processes, such as cell wall maintenance, ion transport and chemotaxis. Their contribution to the development of *B. thuringiensis* under the alkaline pH of insect midgut, may nevertheless be connected to potassium transport and net accumulation of K_2CO_3 (Dow 1984). For instance, the involvement of K^+/H^+ antiporter in alkaline pH homeostasis has been reported in several bacteria, and the presence of the proper antiporter is sufficient to enable a non-alkaliphilic bacterium to grow at extremely high pH (Padan et al. 2005). In addition, certain ion antiporters participate to spore germination in *Bacillus* spp. This is the case for *B. cereus* for which GerT significantly contributes to spore outgrowth from the germinated state during alkaline or Na^+ stress (Senior and Moir 2008). Similarly, it has been demonstrated that loss of the K^+ or $NH4^+$ transporter may affect endospore formation and germination in an alkaliphilic *Bacillus pseudofirmus* (Wei et al. 2003). A stimulation of spore germination in *B. thuringiensis* by an alkaline pH has also been reported in several studies (Benoit et al. 1995; Du and Nickerson 1996; Bhattacharya 1999; Abdoarrahem et al. 2009), and could be considered as an adaptation that enable to germinate at the appropriate time in the insect guts (Du and Nickerson 1996; Jensen et al. 2003).

The PAI $K^+(Na^+)/H^+$ antiporter proteins showed 80 and 62% identity with the YhaU and YhaT from *Bacillus subtilis*, respectively (Fujisawa et al. 2004). Interestingly, *yhaU* is strongly induced by alkaline pH and salt-induced stress and this antiporter may extrude K^+ and NH_4^+. Similarly, one MCP has been recognized as necessary for optimal pH homeostasis and for normal chemotaxis responses in the alkaliphilic *B. pseudofirmus* OF4 (Fujinami et al. 2007). However, it should be noted that the $K^+(Na^+)/H^+$ antiporter and MCP from the PAI are not plasmid-specific proteins, and that homologs are present in the *B. thuringiensis* chromosome (supplementary figs. S11 and S12, Supplementary Material online).

Certain *N*-acetylmuramoyl-L-alanine amidases have been shown as important enzymes for germination in *Bacillus* spp. (Makino and Moriyama 2002; Wu et al. 2015). It has also been revealed that the peptidoglycan hydrolases of *B. thuringiensis* are more active at high pH (Raddadi et al. 2004). These enzymes may contribute to *B. thuringiensis* virulence as an "evasin" which would ensure successful colonization of *B. thuringiensis* in the insect hemocoel before the host develops an immunological response, as it has been proposed for the AmiA amidase of *Bacillus anthracis* (Mesnage and Fouet 2002).

The *N*-acetylmuramoyl-L-alanine amidase from the insecticidal PAI has putatively bacteriophage origin, since it shows similarity with the amidase XlyA encoded by the defective prophage PBSX from *B. subtilis* (Krogh et al. 1998), and because its closest homolog on *B. thuringiensis* chromosome was identified in a prophage region (supplementary fig. S13, Supplementary Material online). A relationship of the PAI with prophages is also visible in the non-*cry* genes located between *cry1I* and *cry2Aa*. They code for an ADP-ribose 1-phosphate phosphatase and a potential ADP-ribosylase (protein family: pfam14487) (supplementary fig. S1, Supplementary Material online), that share homology with putative gene products of a *Geobacillus subterraneus* prophage (supplementary fig. S14, Supplementary Material online). Since ADP-ribosylation is used by various bacterial toxins, including the *B. thuringiensis* Vip1/Vip2 (Palma et al. 2014), they might also represent virulence attributes in this bacterium. As a whole, the insecticidal PAI appears to be a conglomerate of various genetic determinants of chromosomal and prophage origin, intermingled with insecticidal genes. In addition, the PAI might also be a part of a larger (~120 kb) genomic unit (supplementary fig. S8, Supplementary Material online), containing among others a novel haemolysin operon (Zhu et al. 2015) and an operon encoding spore delaying

proteins whose homologs where shown to contribute to cannibalism behaviour in *B. subtilis* (González-Pastor 2011) (supplementary table S2, Supplementary Material online).

In conclusion, the similarity of genetic environment of various *cry1* genes implies their common origin, likely the insecticidal PAI located in ~300 kb *B. thuringiensis* megaplasmids. We suggest that two independent insertion events "released" *cry1* from the PAI in the form of a *cry1* cassette, involving *N*-acetylmuramoyl-L-alanine amidase and fragment of $K^+(Na^+)/H^+$ antiporter genes. Hence, *cry1* sequences divergence and/or homologous recombination between *cry1* genes, including their positive retroregulator, occurring in this shared genetic environment appear to play a central role in the evolution and spread of the *cry1* genes.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Funding

This study was supported by grant No. N N302 656640 of Ministry of Science and Higher Education of Poland (I. Swiecicka), and by project numbers N/ST/ZB/16/001/1122 (K. Fiedoruk) and N/ST/ZB/16/004/1122 (T. Daniluk) from the Medical University of Bialystok.

Literature Cited

Abdoarrahem MM, Gammon K, Dancer BN, Berry C. 2009. Genetic basis for alkaline activation of germination in *Bacillus thuringiensis* subsp. *israelensis*. Appl Environ Microbiol. 75(19):6410–6413.

Arndt D, et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44(W1):W16–W21.

Aronson AI. 1993. The two faces of *Bacillus thuringiensis*: insecticidal proteins and post-exponential survival. Mol Microbiol. 7(4):489–496.

Baum JA, Gilbert MP. 1991. Characterization and comparative sequence analysis of replication origins from three large *Bacillus thuringiensis* plasmids. J Bacteriol. 173:5280–5289.

Benoit TG, Newnam KA, Wilson GR. 1995. Correlation between alkaline activation of *Bacillus thuringiensis* var. *kurstaki* spores and crystal production. Curr Microbiol. 31(5):301–303.

Bhattacharya PR. 1999. Activation and germination of spores of *Bacillus thuringiensis* var. *israelensis* by alkaline pH and larval (*Aedes aegypti*) gut fluid. Southeast Asian J Trop Med Public Health 30(2):338–342.

Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676.

Crickmore N, et al. 2016. *Bacillus thuringiensis* toxin nomenclature. Availabe from http://www.btnomenclature.info/. Cited 2017 Apr 3.

de Maagd RA, Bravo A, Crickmore N. 2001. How *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. Trends Genet. 17:193–199.

de Maagd RA, Bravo A, Berry C, Crickmore N, Schnepf HE. 2003. Structure, diversity, and evolution of protein toxins from spore-forming entomopathogenic bacteria. Annu Rev Genet. 37:409–433.

Dow JA. 1984. Extremely high pH in biological systems: a model for carbonate transport. Am J Physiol. 246:R633–R636.

Du C, Nickerson KW. 1996. *Bacillus thuringiensis* HD-73 spores have surface-localized Cry1Ac toxin: physiological and pathogenic consequences. Appl Environ Microbiol. 62(10):3722–3726.

Fujisawa M, Wada Y, Ito M. 2004. Modulation of the K+ efflux activity of *Bacillus subtilis* YhaU by YhaT and the C-terminal region of YhaS. FEMS Microbiol Lett. 231(2):211–217.

Fujinami S, et al. 2007. The voltage-gated Na+ channel NaVBP co-localizes with methyl-accepting chemotaxis protein at cell poles of alkaliphilic *Bacillus pseudofirmus* OF4. Microbiology 153(Pt 12):4027–4038.

González-Pastor JE. 2011. Cannibalism: a social behavior in sporulating *Bacillus subtilis*. FEMS Microbiol Rev. 35(3):415–424.

He J, et al. 2011. Complete genome sequence of *Bacillus thuringiensis* subsp. *chinensis* strain CT-43. J Bacteriol. 193(13):3407–3408.

Jensen GB, Hansen BM, Eilenberg J, Mahillon J. 2003. The hidden lifestyles of *Bacillus cereus* and relatives. Environ Microbiol. 5(8):631–640.

Jurat-Fuentes JL, Crickmore N. 2017. Specificity determinants for Cry insecticidal proteins: insights from their mode of action. J Invertebr Pathol. 142:5–10.

Krogh S, Jørgensen ST, Devine KM. 1998. Lysis genes of the *Bacillus subtilis* defective prophage PBSX. J Bacteriol. 180(8):2110–2117.

Léonard C, Chen Y, Mahillon J. 1997. Diversity and differential distribution of IS*231*, IS*232* and IS*240* among *Bacillus cereus*, *Bacillus thuringiensis* and *Bacillus mycoides*. Microbiology 143(8):2537–2547.

Mahillon J, Seurinck J, Delcour J, Zabeau M. 1987. Cloning and nucleotide sequence of different iso-IS*231* elements and their structural association with the Tn*4430* transposon in *Bacillus thuringiensis*. Gene 51(2–3):187–196.

Mahillon J, Rezsöhazy R, Hallet B, Delcour J. 1994. IS*231* and other *Bacillus thuringiensis* transposable elements: a review. Genetica 93(1–3):13–26.

Mahillon J, Chandler M. 1998. Insertion sequences. Microbiol Mol Biol Rev. 62(3):725–774.

Makino S, Moriyama R. 2002. Hydrolysis of cortex peptidoglycan during bacterial spore germination. Med Sci Monit. 8(6):RA119–RA127.

Menou G, Mahillon J, Lecadet MM, Lereclus D. 1990. Structural and genetic organization of IS*232*, a new insertion sequence of *Bacillus thuringiensis*. J Bacteriol. 172(12):6689–6696.

Mesnage S, Fouet A. 2002. Plasmid-encoded autolysin in *Bacillus anthracis*: modular structure and catalytic properties. J Bacteriol. 184(1):331–334.

Murawska E, Fiedoruk K, Bideshi DK, Swiecicka I. 2013. Complete genome sequence of *Bacillus thuringiensis* subsp. *thuringiensis* strain IS5056, an isolate highly toxic to *Trichoplusia ni*. Genome Announc. 1(2):e00108–13.

Murawska E, Fiedoruk K, Swiecicka I. 2014. Modular genetic architecture of the toxigenic plasmid pIS56-63 harboring *cry1Ab21* in *Bacillus thuringiensis* subsp. *thuringiensis* strain IS5056. Pol J Microbiol. 63:147–156.

Padan E, Bibi E, Ito M, Krulwich TA. 2005. Alkaline pH homeostasis in bacteria: new insights. Biochim Biophys Acta 1717(2):67–88.

Palma L, Muñoz D, Berry C, Murillo J, Caballero P. 2014. *Bacillus thuringiensis* toxins: an overview of their biocidal activity. Toxins (Basel) 6(12):3296–3325.

Palma L, Muñoz D, Murillo J, Caballero P. 2014a. Draft Genome Sequence of *Bacillus thuringiensis* serovar *tolworthi* strain Na205-3, an isolate toxic for *Helicoverpa armigera*. Genome Announc. 2:e00187–e00114.

Pomerantsev AP, Camp A, Leppla SH. 2009. A new minimal replicon of *Bacillus anthracis* plasmid pXO1. J Bacteriol. 191(16):5134–5146.

Qiu N, et al. 2010. Prevalence and diversity of insertion sequences in the genome of *Bacillus thuringiensis* YBT-1520 and comparison with other *Bacillus cereus* group members. FEMS Microbiol Lett. 310(1):9–16.

Raymond B, Johnston PR, Nielsen-LeRoux C, Lereclus D, Crickmore N. 2010. *Bacillus thuringiensis*: an impotent pathogen? Trends Microbiol. 18(5):189–194.

Raddadi N, et al. 2004. The autolytic phenotype of *Bacillus thuringiensis*. J Appl Microbiol. 97(1):158–168.

Schnepf E, et al. 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol Mol Biol Rev. 62:775–806.
Senior A, Moir A. 2008. The *Bacillus cereus* GerN and GerT protein homologs have distinct roles in spore germination and outgrowth, respectively. J Bacteriol. 190(18):6148–6152.
Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34(90001):D32–D36.
Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27:1009–1010.
Swiecicka I, Bideshi DK, Federici BA. 2008. Novel isolate of *Bacillus thuringiensis* subsp. *thuringiensis* that produces a quasicuboidal crystal of Cry1Ab21 toxic to larvae of *Trichoplusia ni*. Appl Environ Microbiol. 74(4):923–930.
Tang M, Bideshi DK, Park H-W, Federici BA. 2007. Iteron-binding ORF157 and FtsZ-like ORF156 proteins encoded by pBtoxis play a role in its replication in *Bacillus thuringiensis* subsp. *israelensis*. J Bacteriol. 189(22):8053–8058.
Tobes R, Pareja E. 2006. Bacterial repetitive extragenic palindromic sequences are DNA targets for Insertion Sequence elements. BMC Genomics 7:62.
Wei Y, et al. 2003. Mutational loss of a K+ and NH4+ transporter affects the growth and endospore formation of alkaliphilic *Bacillus pseudofirmus* OF4. J Bacteriol. 185(17):5133–5147.
Wong HC, Chang S. 1986. Identification of a positive retroregulator that stabilizes mRNAs in bacteria. Proc Natl Acad Sci U S A. 83(10):3233–3237.
Wu X, et al. 2015. Characterization of the activity of the spore cortex lytic enzyme CwlJ1. Biotechnol Bioeng. 112(7):1365–1375.
Ye W, et al. 2012. Mining new crystal protein genes from *Bacillus thuringiensis* on the basis of mixed plasmid-enriched genome sequencing and a computational pipeline. Appl Environ Microbiol. 78(14):4795–4801.
Zheng J, Peng D, Ruan L, Sun M. 2013. Evolution and dynamics of megaplasmids with genome sizes larger than 100 kb in the *Bacillus cereus* group. BMC Evol Biol. 13:262.
Zhu L, et al. 2015. Genomic and transcriptomic insights into the efficient entomopathogenicity of *Bacillus thuringiensis*. Sci Rep. 5:14129.

Associate editor: Bill Martin