Published online 28 June 2008 *Nucleic Acids Research, 2008, Vol. 36, No. 14* ***4529–4548***
doi:10.1093/nar/gkn372

Survey of group I and group II introns in 29 sequenced genomes of the *Bacillus cereus* group: insights into their spread and evolution

Nicolas J. Tourasse and Anne-Brit Kolstø*

Laboratory for Microbial Dynamics (LaMDa), Department of Pharmaceutical Biosciences, University of Oslo, Oslo, Norway

Received February 25, 2008; Revised April 28, 2008; Accepted May 28, 2008

ABSTRACT

Group I and group II introns are different catalytic self-splicing and mobile RNA elements that contribute to genome dynamics. In this study, we have analyzed their distribution and evolution in 29 sequenced genomes from the *Bacillus cereus* group of bacteria. Introns were of different structural classes and evolutionary origins, and a large number of nearly identical elements are shared between multiple strains of different sources, suggesting recent lateral transfers and/or that introns are under a strong selection pressure. Altogether, 73 group I introns were identified, inserted in essential genes from the chromosome or newly described prophages, including the first elements found within phages in bacterial plasmids. Notably, bacteriophages are an important source for spreading group I introns between strains. Furthermore, 77 group II introns were found within a diverse set of chromosomal and plasmidic genes. Unusual findings include elements located within conserved DNA metabolism and repair genes and one intron inserted within a novel retroelement. Group II introns are mainly disseminated via plasmids and can subsequently invade the host genome, in particular by coupling mobility with host cell replication. This study reveals a very high diversity and variability of mobile introns in *B. cereus* group strains.

INTRODUCTION

Group I and group II introns are well-known genetic elements that were discovered >20 years ago. They are catalytic RNAs (ribozymes) that are capable of self-splicing, i.e. excising themselves out of RNA transcripts and ligating their flanking RNA sequences (hereafter referred as exons). They are also mobile elements as they typically encode proteins that allow them to invade genomic sequences (1–10). Introns can spread into cognate (homologous) intron-less DNA sites, a process called homing, or insert into ectopic (novel) genomic locations, a process called transposition, which usually occurs at lower frequencies. Altogether, these elements are found in all three domains of life: group I introns are present in bacteria, bacteriophages and eukaryotes (organellar and nuclear genomes), while group II introns are present in bacteria, archaea and eukaryotic organelles. Due to their widespread occurrence and their dual functionality as splicing ribozymes and mobile elements, group I and group II introns have been extensively studied and the molecular details of the RNA catalysis and insertion events are well characterized (3–6,10–14). These introns are also of great interest as they play an important role in genome dynamics within and/or between bacteria, archaea, bacteriophages and eukaryotes (2,7,8,15–18). Another relevant, but less thoroughly examined aspect, is the regulatory connection between intron splicing and host gene or protein expression (19–21). Group I and group II introns have also been engineered as tools in biotechnology and molecular medicine, e.g. for targeted gene knock-out/knock-down, gene delivery or gene therapy systems (4,9,22–25), and may as well be the targets for therapeutics (26).

Group I and group II introns are unrelated in sequence, and their structures, splicing and mobility mechanisms are distinct. Their overall distribution patterns in bacterial genomes are also different and sporadic, reflecting their mobility pathways. Group I introns can be found in bacterial tRNA or rRNA genes (21,27–30) and are common in bacteriophages, where they are preferentially inserted in essential genes, notably those coding for proteins involved in DNA metabolism, such as ribonucleotide reductase (1,17,18,31–33). Group I introns fold into a secondary structure made up of usually 9 or 10 paired elements (called P1–P10, with optional peripheral structures, see Figures 2 and 5) which is further stabilized by tertiary interactions and splice via a two-step transesterification

*To whom correspondence should be addressed. Tel: +47 22 85 69 23; Fax: +47 22 84 49 44; Email: a.b.kolsto@farmasi.uio.no

pathway mediated by an external guanosine cofactor (3,34,35). During this process, the intron RNA makes specific base-pairings with sequence motifs in the target site at the 5′ exon (internal guide sequence, IGS) and 3′ exon sides (P10 interaction). Full group I introns contain an open reading frame (ORF) coding for a homing endonuclease, which is responsible for mobility and the size of group I introns range from 200 bp to >1 kb depending on whether an ORF is present. The endonuclease recognizes and generates a double-strand break at homologous intron-less DNA sites, which is then repaired following the double-strand break repair (DSBR) or the synthesis-dependent strand annealing (SDSA) pathway (1,13,36). These pathways are dependent on homologous recombination between exon sequences and lead to copy of the intron as well as conversion of part of the flanking exons from the intron donor DNA into the recipient DNA. Homing endonucleases are generally very site-specific allowing intron insertion into cognate target sites (37). Reverse splicing of the intron RNA directly into RNA or DNA, a reaction which only depends on formation of the IGS and thus has more limited target site requirements, has been suggested as a mechanism for transposition into ectopic sites and/or for mobility of ORF-less introns (38,39). In this case, no exon coconversion occurs.

Group II introns in prokaryotes are rarely found within highly conserved or housekeeping genes and are also often present in noncoding regions, and there is a significant amount of truncated or fragmented elements in bacterial genomes (4,7,8,16). Interestingly, bacterial group II introns tend to be located within mobile DNA elements such as plasmids, insertion sequence (IS) elements, transposons or pathogenicity islands, which could account for their spread among bacteria (16,40,41). The secondary structure of the group II RNA is made up of six domains (Figure 6) linked by tertiary interactions and splicing proceeds via two transesterifications, which typically releases the intron RNA in a lariat form (circle with tail) (4,5). Due to similarities in this splicing mechanism and shared structural features, group II introns are thought to be the ancestors of the nuclear spliceosomal introns of eukaryotes (42). Typical group II introns encode a multifunctional ORF, which has reverse transcriptase (RT), maturase (splicing) and optionally endonuclease activities. Group II introns are therefore genetic retroelements (and have been called retrointrons) and are much larger than group I elements: 700 nt to 3 kb whether or not the ORF is present. The general retrohoming mechanism in bacteria follows the target-primed reverse transcription target-primed reverse transcription (TPRT) pathway with the following steps: the intronic protein recognizes and binds the DNA target site, the intron RNA reverse splices into one strand of the target DNA, the intronic protein cleaves the second strand and uses it as a primer to reverse transcribe the intron, and finally host repair functions complete the intron integration and conversion into DNA (4,9,14). In contrast to group I introns, the repair mechanism is independent of homologous recombination between donor and recipient DNA and thus no conversion of flanking exon sequences occurs upon group II intron insertion in bacteria. During splicing and reverse splicing, specific base-pairing contacts are made between motifs in the intron RNA (exon-binding sequences, EBS) and the complementary sequences in the flanking exons (intron-binding sequences, IBS). Introns whose ORF does not contain the endonuclease domain insert into their target sites by using nascent strands at DNA replication forks to prime reverse transcription (43). The latter mechanism is also used by introns with nuclease activity for retrotransposition into ectopic sites that do not support second strand cleavage (44).

The *Bacillus cereus* group is a group of endospore-forming bacteria belonging to the *Firmicutes* phylum, i.e. low G + C% Gram-positive bacteria. The *B. cereus* group includes bacterial species that are of medical and/or economic importance, such as *B. anthracis*, an obligate mammalian pathogen causing the lethal disease anthrax; *B. cereus*, an opportunistic human pathogen involved in food-poisoning incidents and contaminations in hospitals; *B. thuringiensis*, an insect pathogen and one of the world's most widely used biopesticides and *B. weihenstephanensis*, a cold-tolerant species known for contaminating dairies. These species are very closely related at the genomic level (45,46). *Bacillus cereus* group genomes are usually 5.2–5.4 Mb in length including a single circular chromosome and most strains carry one or several extrachromosomal plasmids that are responsible for the main phenotypic differences between the species. *Bacillus cereus* group organisms have been shown to carry a number of group I and group II introns, some of them exhibiting unusual properties (46–52), and this group of bacteria is among those with the highest numbers of sequenced genomes. In this study, we have conducted an extended bioinformatics survey of the distribution and features of group I and group II introns in the 29 currently available genomes from the *B. cereus* group, which cover strains isolated from various origins and geographical locations. As the virtually complete set of introns in these isolates can then be identified, the goal of such an analysis was to gain insights into the spread and evolution of mobile intronic elements in a worldwide-distributed bacterial population.

IDENTIFICATION AND ANALYSIS OF INTRONS IN *B. CEREUS* GROUP GENOMES

Intron search and analysis procedure

The list of complete or nearly complete *B. cereus* group genome sequences examined in this study is given in Table 1. All sequence data were retrieved from the GenBank database and include chromosomes and plasmids. Eleven *B. anthracis* strains have been sequenced; however, *B. anthracis* isolates are highly monomorphic and virtually identical at the genomic level (53) and no difference was found between their intron sets (except for isolates lacking the group II intron-containing plasmid pXO1 and differences due to missing sequences in unfinished genomes) [(52), this study]. Therefore, data for the 'Ames Ancestor' strain only will be presented in this article. The procedures for group I and group II introns search and analysis, as well as *B. cereus* group phylogenomic reconstruction, are described in Supplementary Material. Individual intron information and sequence

Table 1. Strain information and numbers of full-length group I and group II introns in *B. cereus* group genomes

Species/strain	Origin/source	Genome status	GenBank accession number[a]	GroupI introns[b]	Group II introns[b]
B. anthracis Ames Ancestor A0581[c]	Cow (Texas, USA)	Finished	AE017334, AE017336 (pXO1), AE017335 (pXO2)	3	2
B. cereus ATCC 14579	Farm (USA, 1916)	Finished	AE016877, AE016878 (pBClin15)	0	1
B. cereus ATCC 10987	Dairy cheese (Canada, 1930)	Finished	AE017194, AE017195 (pBc10987)	2	7
B. cereus G9241	Human, sputum and blood (Louisiana, USA, 1994)	12X shotgun	AAEK00000000, DQ889680 (pBCXO1), DQ889679 (pBC210)	1	3
B. thuringiensis konkukian 97-27	Human, severe tissue necrosis (Yugoslavia, 1995)	Finished	AE017355, CP000047 (pBT9727)	2	0
B. thuringiensis israelensis ATCC 35646	Sewage (Israel)	8X shotgun	AAJM01000000	0	2
B. cereus E33L (formerly Zebra Killer)	Dead zebra (Namibia, 1996)	Finished	CP000001, CP000040 (pE33L466), CP000041 (pE33L5), CP000042 (pE33L54), CP000043 (pE33L8), CP000044 (pE33L9)	5	5
B. weihenstephanensis KBAB4	Forest soil (France, 2000)	Finished	CP000903, CP000904 (pBWB401), CP000905 (pBWB402), CP000906 (pBWB403), CP000907 (pBWB404)	5	0
B. thuringiensis Al Hakam	Suspected bioweapons facility (Iraq)	Finished	CP000485, CP000486 (pALH1)	1	1
B. cereus subsp. cytotoxis NVH391-98	Vegetable puree (France, 1998)	Finished[d]	CP000764, CP000765 (pBC9801)	1	2
B. cereus AH187 (F4810/72)	Human, vomit (UK, 1972)	8X shotgun[e]	AAUF00000000, DQ889676 (pCER270)	4	12
B. cereus AH820	Human, periodontitis (Norway, 1995)	8X shotgun[e]	AAUE00000000, DQ889677 (pPER272)	4	4
B. cereus AH1134	Human, eye (Oklahoma, USA)	8X shotgun[e]	ABDA00000000	2	1
B. cereus G9842	Human, stool (Nebraska, USA, 1996)	8X shotgun[e]	ABDJ00000000	3	4–5[f]
B. cereus B4264	Human, blood and pleural fluid (1969)	8X shotgun[e]	ABDI00000000	1	0
B. cereus NVH0597-99	Spice mix (Norway, 1999)	8X shotgun[e]	ABDK00000000	2	0
B. cereus 03BB108	Dust (Texas, USA, 2003)	8X shotgun[e]	ABDM00000000	2	4
B. cereus H3081.97	Food (USA, 1997)	8X shotgun[e]	ABDL00000000	3	13–14[f]
B. cereus W	Soil	8X shotgun[e]	ABCZ00000000	2	0

[a]The first accession number given is that of the chromosome for finished genomes or the full set of sequence contigs for unfinished genomes. Following numbers are for complete plasmids, when applicable (plasmid names are given in parentheses).
[b]Individual intron information and sequence data will be deposited in the GISSD (http://www.rna.whu.edu.cn/gissd/index.html) and the group II intron database (http://www.fp.ucalgary.ca/group2introns/), and are also available from the authors upon request.
[c]Eleven *B. anthracis* strains have been sequenced (see http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=1392). Since *B. anthracis* isolates are highly monomorphic and virtually identical, data for the 'Ames Ancestor' strain only are presented in this article.
[d]Strain NVH391-98 has a reduced size genome of 4.1 Mb, whereas the estimated genome size of all other strains is 5.2–5.9 Mb.
[e]No gene annotation available for these genomes at the time of study.
[f]Due to incomplete sequence data the assignment of a few intron fragments to the same or separate elements could not be confirmed.

data will be submitted for inclusion in the Group I intron database (GISSD), http://www.rna.whu.edu.cn/gissd/index.html (54) and the Group II intron database, http://www.fp.ucalgary.ca/group2introns/ (55), which are publicly available. Data may also be obtained from the authors on request.

DISTRIBUTION AND EVOLUTION OF GROUP I INTRONS IN THE *B. CEREUS* GROUP

Introns inserted in essential chromosomal genes

Bacteria from the *B. cereus* group do not carry group I introns in their tRNA or rRNA genes, as opposed to a few other bacterial groups including cyanobacteria, proteobacteria, Thermotogales and Chlamydiales (21,27–30). However, *B. cereus* group organisms are exceptions as they are the only known bacterial species in which stand-alone chromosomal protein-coding genes, encoding the large (α) and small (β) subunits of ribonucleotide reductase (nrdE and nrdF, respectively) and the recombinase A (recA), contain group I introns (49). Ribonucleotide reductase and recA are essential enzymes involved respectively in generating the ribonucleotides necessary for DNA synthesis and in DNA recombination and repair. Interestingly, introns are usually found in homologues of these genes in bacteriophages infecting mainly bacteria belonging, like the *B. cereus* group, to the *Firmicutes*

family, and it might therefore be that the introns and/or the genes in the *B. cereus* group have originally been acquired via phages (46,49). However, since no other phage-associated copies of nrdE, nrdF and recA are present in *B. cereus* group genomes (with the exception of an intron-less nrdEF operon in a *B. thuringiensis israelensis* ATCC 35 646 phage), other mechanisms may have been involved, and competence (i.e. uptake of foreign DNA), which is common among low GC Gram-positive bacteria, has been suggested (48,49,56).

In the 29 *B. cereus* group genomes four different group I introns are inserted in four positions throughout the nrdE gene (denoted IVS2-5). Remarkably, every strain contains only a single intron and the four insertion sites are occupied in different subsets of isolates (Table 2). In contrast, introns in the nrdF and recA genes are found in one location only. The observed intron location pattern in the nrdEF operon also holds for an additional 26 *B. thuringiensis* and *B. cereus* strains recently analyzed by Nord *et al*. (49).

Detailed examination of the intron insertion sites revealed two interesting patterns with respect to intron mobility. The first relates to target site specificity. Virtually all known group I introns are inserted after a uridine (U), which is involved in a U–G basepair with an internal guanosine (G) within the intron and is critical for 5′ splice-site selection (3,34). As expected, all seven *B. cereus* group strains that carry a cytosine (C) rather than U at the intron location site within the recA gene do not contain the intron and probably never did (Figure 1A). A second feature of group I intron insertion sites is the coconversion of exon sequences due to DNA repair of the double-strand break generated by the intron-encoded homing endonuclease. As can be seen in Figure 1, the 5′ exons flanking the introns located at the IVS2 and IVS5 sites within nrdE and at the unique site within nrdF (IVS6) exhibit clear sequence differences compared to the corresponding empty sites. Given the very high degree of conservation of the nrdEF genes within the *B. cereus* group (>99% nucleotide sequence identity), the observed differences at the intron sites are indicative of recombination events, most likely with non-*B. cereus* group donors. The longest and most conspicuous event occurred at the IVS5 site, shared by *B. thuringiensis* Al Hakam and *B. cereus* 03BB108, where the visible conversion track extends to 145 bp. Sequence searches of public databases could not identify the potential intron donors. For nrdF (IVS6 site), the recombined exon sequence matches the sequence of the *Bacillus sp*. BSG40 prophage (32). However, that particular phage does not contain a related intron at that particular site and was therefore not the donor, but it might have been a related bacteriophage. There is also similarity between the exon sequences at the IVS2 site and *Bacillus* prophages

Table 2. Full-length introns shared between *B. cereus* group genomes

Intron	Strains sharing the intron	Location[a]
Group I		
recA	*Anthracis* (11 strains), *konkukian* 97-27, E33L, ATCC 10987, 03BB108, AH187, AH820, W, H3081.97, NVH0597-99 and AH1134[b]	S
nrdE-IVS2	KBAB4 and NVH391-98	S
nrdE-IVS3	*Anthracis* (11 strains), *konkukian* 97-27, W and AH820[c]	S
nrdE-IVS4	ATCC 10987, E33L, G9241, AH187 and H3081.97	S
nrdE-IVS5	03BB108 and Al Hakam	S
nrdF-IVS6	NVH0597-99, AH1134, AH187 and H3081.97[d]	S
TMP-IVS1	E33L, AH820, AH187 and G9842 (2 copies)	S
TMP-IVS2	*Anthracis* (11 strains) and KBAB4	S
TMP-IVS3a	KBAB4 and E33L	S
TMP-IVS3b	KBAB4, G9842 and AH820	S
tail tube	KBAB4 and B4264	S
Group II		
B.a.I1	*Anthracis* (9 strains) and G9241[e]	S
B.a.I2	*Anthracis* (9 strains) and G9241[e]	S
B.c.I1	ATCC 14579, ATCC 10987 (2 copies), E33L, AH820, AH187 (3 copies) and H3081.97	S and D
B.c.I2	ATCC 10987, AH820, AH187 and H3081.97 (2 copies in each strain)[f]	S
B.c.I3	ATCC 10987, Al Hakam and 03BB108	S
B.c.I5	ATCC 10987, AH187 and H3081.97[e]	S
B.c.I7	E33L (4 copies) and AH1134[g]	D
B.c.I10	AH187 (6 copies), H3081.97 (7 copies) and 03BB108[h]	S and D
B.c.I11	AH820 and H3081.97[e]	S

[a]S, D, same or different host gene (or insertion site), respectively.
[b]The *B. cereus* AH1134 recA intron encodes a full homing endonuclease gene (HEG), while only the last 39 bp of the HEG remain in the other strains.
[c]The *B. thuringiensis konkukian* 97-27 intron lacks the HEG (only the first 72 bp and the last 66 bp remain), while the other strains carry a full HEG.
[d]The *B. cereus* AH187 intron lacks the HEG (only the first 10 bp and the last 104 bp remain), while the other strains carry a full HEG.
[e]The *B.a*.I1, *B.a*.I2, *B.c*.I5 and *B.c*.I11 introns are found exclusively on plasmids.
[f]One copy on the chromosome and one on a plasmid.
[g]*Bacillus thuringiensis konkukian* 97-27 and *B. cereus subsp. cytotoxis* NVH391-98 only carry a truncated copy of the intron. Three of the four copies in *B. cereus* E33L lack a full RT ORF (only the first 13 bp and the last 28 bp remain). *B.c*.I7 is related to the full-length intron (*B.th*.I1) in the pAW63 plasmid of *B. thuringiensis kurstaki* HD73 (60% nucleotide sequence identity).
[h]The *B.c*.I10 intron is highly similar to the ORF-less intron (*B.th*.I2) in the pAW63 plasmid of *B. thuringiensis kurstaki* HD73 (83% nucleotide sequence identity).

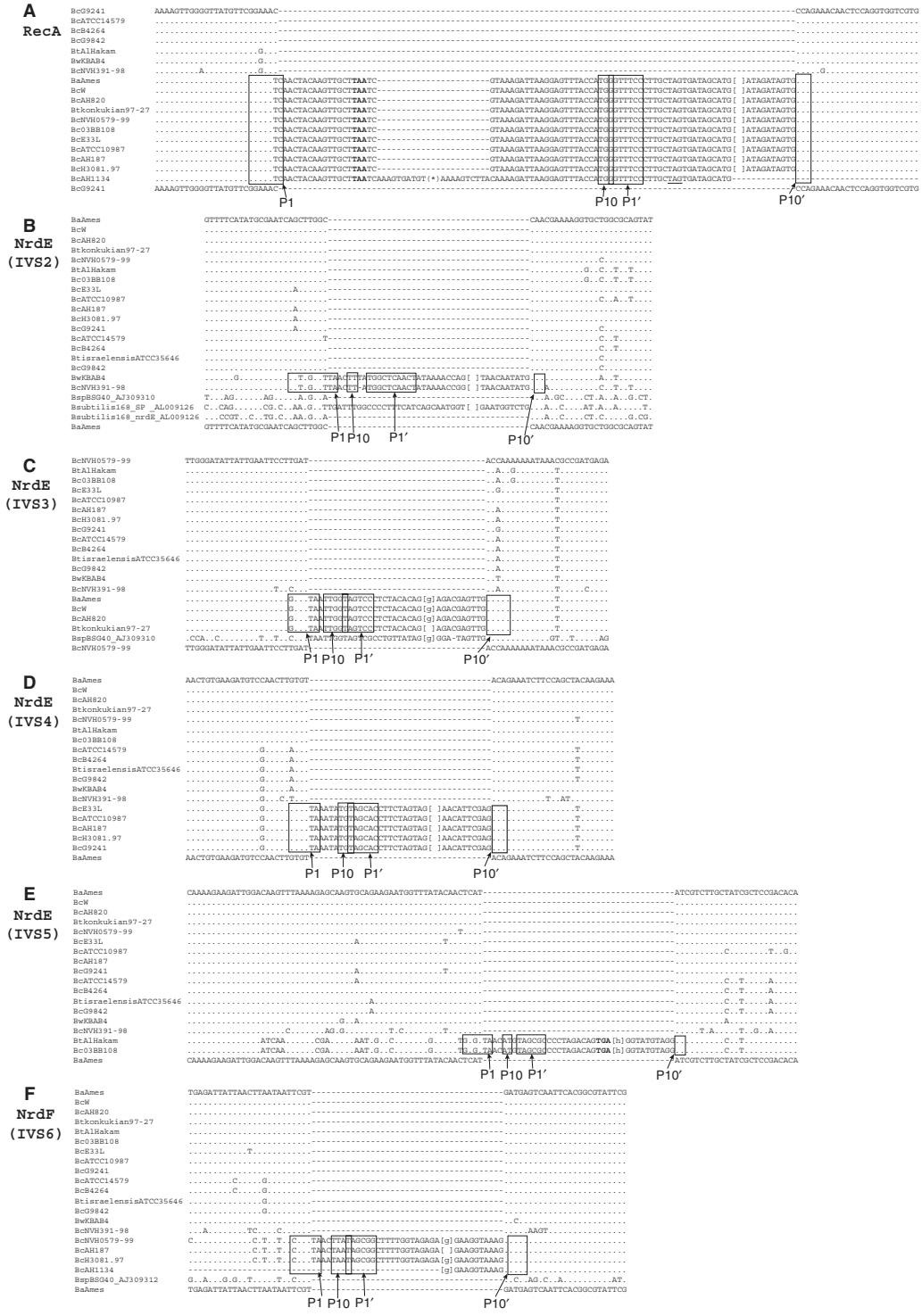

Figure 1. Multiple sequence alignments of the group I intron insertion sites within the recA, nrdE and nrdF genes of sequenced *B. cereus* group strains. When appropriate, the sequences from other organisms were included for comparison and their GenBank accession numbers are given in the last part of the strain identifiers. For the exons, only the differences to an arbitrary reference sequence are shown. Positions identical to the reference are displayed as dots. The reference sequence is shown on the top and bottom lines of each alignment. For the introns, only the first few nucleotides at the 5′- and 3′-ends are shown and are separated by square brackets. A 'g' or an 'h' in-between the brackets indicate that the intron encodes a HEG of the GIY-YIG or H-N-H family, respectively. Complementary nucleotide sets predicted to be involved in the formation of the P1 pairing (P1-P1′) and the P10 pairing (P10-P10′) for recognition of the 5′ and 3′ splice sites, respectively, are boxed and indicated by arrows. For introns that do not start with TAA or TAG or are not inserted in-frame within the host gene, the stop codon that would terminate translation coming from the 5′ exon is in bold. The *B. cereus* AH1134 recA intron encodes a putative HEG located within P1. For simplicity, the HEG is not shown in full and the central part is replaced with a parenthesized asterisk, indicating that the HEG family is unknown. The predicted stop codon of the HEG is underlined. Note that the coconversion tracts in the 5′ exon of nrdE-IVS5 in *B. thuringiensis* Al Hakam and *B. cereus* 03BB108 actually extends to 145 bp; however, only the last 60 bp are shown here due to space limitations.

(Figure 1). The fact that the three coconversion cases observed at IVS1, IVS5 and IVS6 are asymmetric, only affecting the 5′ exons, could suggest that the intron insertions were preferentially repaired through the SDSA pathway, which is more asymmetric than DSBR (13,36). For nrdE-IVS3, as well as for recA, the exon sequences in the intron-containing and intron-less strains are highly similar (Figure 1), implying that the introns in these sites were acquired from donors carrying nrdE or recA sequences virtually identical to those of the recipient strains, donors possibly from the *B. cereus* group. Finally, it should be noted that, as expected, for all three genes the 5′ exons in the strains harboring introns display the motifs that are involved in formation of the IGS with the P1 intron domain (usually the last 4–6 nt of the 5′ exon; boxed in Figure 1). This, in combination with motifs in the 3′ exon (P10 pairing), will ensure correct splicing of the introns. Indeed, *in vivo* splicing activity has been demonstrated for the *B. anthracis* recA and nrdE introns, as well as for the nrdF element in *B. thuringiensis* (47–49).

Although the six introns located within recA, nrdE and nrdF are different overall, they share some sequence and structural features in the 5′- and 3′-ends and the central intron core regions including the P3 and P7 domains [data not shown; see ref. (49) for an example]. The secondary structure is used to divide the group I introns into subclasses (34,35) and the six *B. cereus* group introns have features of different structural subclasses: IA1 (nrdE-IVS2), IA2 (nrdE-IVS4 and nrdF-IVS6), IA3 (nrdE-IVS3) and IB4 (nrdE-IVS5) (Figure 2). This indicates that the introns in the nrdEF operon have multiple origins and illustrate the dynamic behavior of this genomic locus. The recA intron exhibits characteristics of both the IA1 and IA3 subclasses. Interestingly, BLASTN sequence searches revealed that the *B. cereus* group introns share significant homologies in their catalytic RNA parts mainly with introns from various bacteriophages (Table 3). Some of these phage introns are also inserted in recA and nrdEF homologues. These similarities in sequence and set of host genes between *B. cereus* group and phage group I introns strongly suggest that the chromosomal *B. cereus* group type I introns have been acquired through bacteriophages.

At any site within nrdE, nrdF and recA, the introns shared by subsets of *B. cereus* group strains are virtually identical, as are their flanking coconverted exons. Furthermore, introns are shared by isolates of diverse origins and geographical locations (Tables 1 and 2). These observations imply that either *B. cereus* group organisms obtained the introns from related sources or that introns were acquired in the common ancestor of a particular subset of isolates and were subsequently spread by vertical inheritance. Vertical transmission of introns is very likely for *B. anthracis* due to the clonal nature of this species (53). In order to gain further insights into the evolutionary history of the introns in the *B. cereus* group, their distribution was compared with the phylogenomic relationships of the strains (Figure 3). The phylogeny of the *B. cereus* group strains examined here was inferred based on the conserved chromosomal 'core' sequence (a total of 2.1 Mb). The recA intron is lacking the homing endonuclease gene (HEG) in 20 of the 21 intron-positive strains, with the exception of *B. cereus* AH1134. Interestingly, the 20 isolates share a common ancestor in the tree, which strongly suggests that the recA intron has been inherited by vertical descent since its acquisition in an ancestral strain, which subsequently lost the HEG. An alternative explanation would be to assume independent acquisitions by reverse splicing. In contrast, *B. cereus* AH1134 belongs to a different phylogenetic cluster with isolates that do not carry the intron. The presence of a putative HEG in the *B. cereus* AH1134 intron implies an independent homing event in that isolate (Figure 3). It should be noted that the *B. cereus* AH1134 HEG is peculiar, as it does not match the features of any of the known HEG families and may thus represent a new or unusual type of HEG.

Similar to the recA situation, strains harboring the nrdE-IVS4 element, which also lacks a full HEG, reside in a common part of the *B. cereus* group tree (Figure 3), suggesting vertical inheritance of nrdE-IVS4 among these strains. On the other hand, the nrdE introns located at sites IVS3 and IVS5 usually encode the HEG and are thus potentially capable of homing mobility. Nevertheless, the isolates carrying these introns share common ancestry: nrdE-IVS5 is found in the closely related *B. thuringiensis* Al Hakam and *B. cereus* 03BB108, while nrdE-IVS3 is found in the *B. cereus* and *B. thuringiensis* neighbors to *B. anthracis*. Spread of these introns among the isolates analyzed here would seem more likely to have occurred by vertical transmission than by independent integration events. However, in a recent study by Nord and Sjöberg (48) including an additional set of *B. thuringiensis* strains several mobility events of the nrdEF introns have been inferred from the discrepancies between gene phylogenies, distribution of introns, and nrdE–nrdF intergenic sequences. The distribution of the nrdF (IVS6) intron is more discontinuous across the *B. cereus* group tree, and since this element generally encodes a HEG, this would be more readily explained by invoking several homing transfers. Furthermore, the phylogenetic and intron distribution data reported here and in ref. (48) show that there have been cases of HEG or complete intron losses. HEG loss from a strain is suggested when a HEG-lacking strain is present in a phylogenetic cluster comprising mostly HEG-containing isolates, and similarly the presence of scattered intron-negative strains within clusters containing mainly intron-positive isolates implies intron loss from the former isolates. For example, among the *B. anthracis* neighbors, *B. thuringiensis konkukian* 97-27 has lost the HEG from nrdE-IVS3. Evidence that *B. thuringiensis* Al Hakam has lost the recA intron is given by the fact that it is the only strain lacking the element within the recA intron-containing cluster (Figure 3). These instances of HEG and intron loss in the *B. cereus* group population are in agreement with the proposed evolution cycle of group I introns, which predicts successive intron invasion, deletion and re-invasion (57,58).

Introns inserted in prophages

Another set of group I introns carried by *B. cereus* group bacteria are located within prophages, i.e. bacteriophages

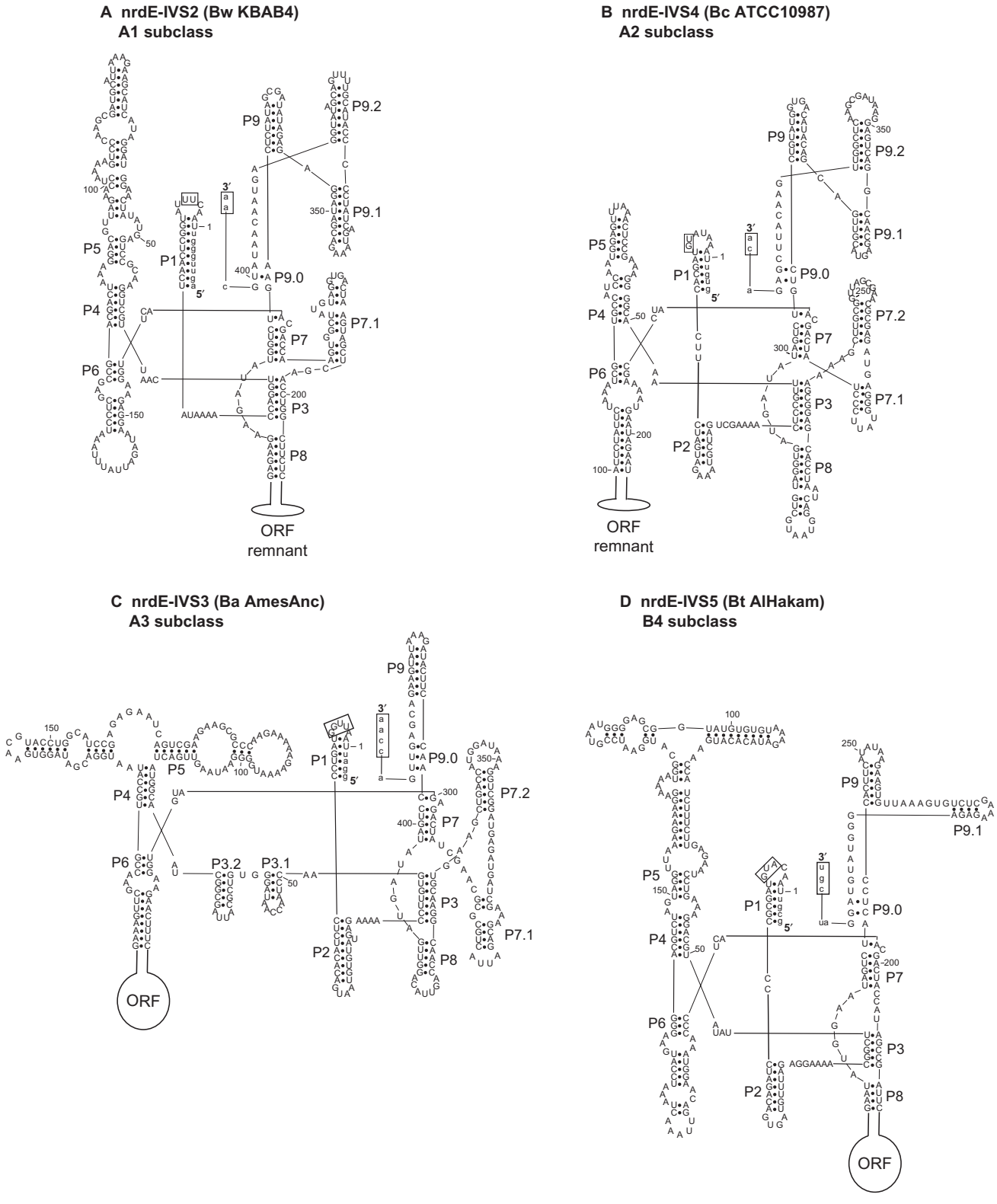

Figure 2. Predicted secondary structures of the group I introns inserted within the nrdE gene of *B. cereus* group strains. The figure shows that the introns belong to different structural classes and are thus from multiple origins. Structure models were predicted using MFOLD (86) and redrawn according to the format defined in (87) using RnaViz (88). Intron sequences are in uppercase letters and exon nucleotides are in lowercase letters. Base pairs are linked by dots. Labels P1 to P9 indicate the group I intron domains. The P1 stem represents the internal guide sequence (IGS) used for recognition of the 5′ splice site. Nucleotides involved in formation of the P10 pairing for 3′ splice-site selection are boxed. Intron-encoded HEG ORFs are not included. Base numbering does not include full HEG sequences, however it includes ORF remnants. *Bacillus cereus* group host strains are given in parentheses after the intron names (other strains sharing the introns are listed in Table 2). Note that the nrdE-IVS3 intron previously reported as belonging to the A2 subclass (49,54) has been reclassified here in the A3 subclass mainly due to the presence of a single P9 stem followed by a short 3′-end.

Table 3. Sequence similarities between *B. cereus* group and bacteriophage group I introns

B. cereus group intron	Partially matching bacteriophage Intron[a,b]
recA	*B. thuringiensis* phage 0305phi8-36 (recA)
nrdE-IVS3	*Bacillus sp.* BSG40 prophage (RIR, nrdE-1)
	B. thuringiensis phage 0305phi8-36 (recA)
nrdE-IVS4	Enterobacterial phage T4 (RIR, nrdB)
	Pseudomonas phage phiKZ
	Bacillus phage SPβ/SPβc2 (RIR, nrdF)
nrdF-IVS6	*Bacillus* phage SPβ/SPβc2 (RIR, nrdF)
	Bacillus sp. BSG40 prophage (RIR, nrdF)
	Bacillus sp. M1918 prophage (RIR, nrdF)
	Staphylococcus phage K (lysin)
	Staphylococcus phage G1 (amidase)
	Staphylococcus bacteriophage 812 (holin)
TMP-IVS1	*Bacillus* phage SPβ/SPβc2 (RIR, nrdE)
	Bacillus sp. M1321 prophage (RIR, nrdE-1 and nrdE-2)
	Bacillus sp. M1918 prophage (RIR, nrdE-1)
	Bacillus sp. BSG40 prophage (RIR, nrdE-2)
	Bacillus sp. M135 prophage (RIR, nrdE)
	Staphylococcus phage PH15 (lysin)
	Enterobacterial phage K1E (large terminase)
	Staphylococcus phage Twort
	Staphylococcus phage U16
	Staphylococcus phage 812
TMP-IVS2	*Bacillus* phage SPβ/SPβc2 (RIR, nrdE)
	Bacillus sp. M1321 prophage (RIR, nrdE-1)
	Bacillus sp. M1918 prophage (RIR, nrdE-1)
	Staphylococcus phage PH15 (lysin)
	Enterobacterial phage K1E (large terminase)
	Staphylococcus phage Twort
	Staphylococcus phage U16
	Staphylococcus phage 812
	Enterobacterial phage RB3 (RIR, nrdB)
	Enterobacterial phage U5 (RIR, nrdB)
	Lactococcus bacteriophage Tuc2009 (major head protein)
TMP-IVS3a	*Staphylococcus hæmolyticus* JCSC1435 phage ϕSH1
Term-IVS1	*Staphylococcus hæmolyticus* JCSC1435 phage ϕSH1
	Lactobacillus delbrueckii bacteriophage JCL1032 (TMP)
Term-IVS2	*Staphylococcus* bacteriophage ROSA
Tail tube	*Staphylococcus* bacteriophage 812 (recA)
	Clostridium botulinum phage c-st

[a]The hits reported here are continuous matches of 80 nt or more to intron catalytic RNAs (i.e. HEG not included) obtained by a BLASTN search with increased match reward (−r 2) and no filtering for low complexity regions (-F F) using the *B. cereus* group introns as queries.
[b]The intron host gene function is given in parentheses, when known. RIR, ribonucleotide reductase.

that are integrated in the bacterial genome. In particular, several introns are inserted in the gene coding for the conserved tape measure protein (TMP) of prophages from five *B. cereus*, one *B. weihenstephanensis* and all the 11 *B. anthracis* isolates (Figure 4). TMP is a large (>1200 amino acids) protein involved in proper assembly and length determination of the phage tail. Introns were found in three different positions within TMP, denoted IVS1, IVS2 and IVS3. It has been previously reported that *B. cereus* E33L contains a prophage homologous to the lambda01 element of *B. anthracis* in its chromosome (45). In this study, we found that a corresponding prophage is also present in the *B. weihenstephanensis* KBAB4 chromosome (approximate genomic coordinates 3 466 500–3 518 000). In the TMP of the lambda01-like phages, *B. weihenstephanensis* KBAB4 shares IVS2 with *B. anthracis*, while it shares an intron at IVS3 with *B. cereus* E33L (TMP-IVS3a). The latter strain also harbors IVS1, which is present at the corresponding site in the TMP homologue of different prophages from *B. cereus* AH820, AH187 and G9842 (Figure 4A and Table 2). In turn, the *B. cereus* AH820 and G9842 phages carry another intron located at the IVS3 site (TMP-IVS3b), which is shared by a different, non-lambda01-like, *B. weihenstephanensis* KBAB4 prophage integrated in the pBWB404 plasmid. In addition to TMP, group I introns were found in a couple of other phage genes. The lambda01-like prophage of *B. cereus* E33L, carrying TMP-IVS1 and TMP-IVS3a, also contains a group I intron (term-IVS1) inserted within the terminase (large subunit) gene, involved in phage genome packaging. A different intron is present within the large terminase gene of an unrelated *B. cereus* 03BB108 prophage (term-IVS2). Finally, the *B. weihenstephanensis* KBAB4 phage region BwK1 (59), part of the pBWB401 plasmid, and the corresponding prophage from *B. cereus* B4264 share a group I intron in the gene encoding the phage tail tube protein, a major structural component of the phage tail. Phages are themselves mobile elements and, as expected, there is no correlation between the phylogenetic relationships of the host strains and the distribution of the prophages and their associated introns. *Bacillus weihenstephanensis* KBAB4 and *B. cereus* G9842 are relatively distantly related from *B. cereus* AH820 and *B. anthracis*, and yet these strains share common phages and/or introns (Figures 3 and 4 and Table 2). The complex distribution pattern of group I introns within prophages emphasizes the fact that bacteriophages are an important vector for lateral transfer of introns in the *B. cereus* group.

To our knowledge, group I introns have never been found on bacterial plasmids. Here, we report the discovery of the TMP-IVS3b and tube elements that are respectively located within the 52.8-kb pBWB404 plasmid and the 417-kb pBWB401 plasmid of *B. weihenstephanensis* KBAB4, although the introns are inserted within prophages. Both introns lack a complete HEG and are thus not mobile, and their presence in multiple strains, including plasmids, is most likely due to passive transmission via integration of the phages into the bacterial genomes. TMP-IVS3b is also found in a chromosomal phage in *B. cereus* AH820 (Figure 5D) and is an example of a group I intron that has been transmitted both to a bacterial chromosome and a bacterial plasmid. The prophages carrying TMP-IVS3b in *B. weihenstephanensis* KBAB4, *B. cereus* AH820 and G9842 are different, and therefore how an immobile intron has been acquired by distinct phages still remains a puzzling question. The host gene (BcerK BAB4_5750 + BcerKBAB4_5751) of the tube intron is part of the BwK1 phage region of *B. weihenstephanensis* KBAB4 and this phage is highly conserved in *B. thuringiensis israelensis* ATCC 35646 [BtI1 phage; (59)] as well as in *B. cereus* B4264; however, it is not known whether the region is part of a plasmid in the latter two isolates. BwK1 also shows weak homology to the left arm of the large atypical *B. thuringiensis* 0305ϕ8-36 phage (59). While *B. cereus* B4264 does harbor the tube intron like BwK1, the element is absent from the corresponding locus in the BtI1 and 0305ϕ8-36 phages, implying that

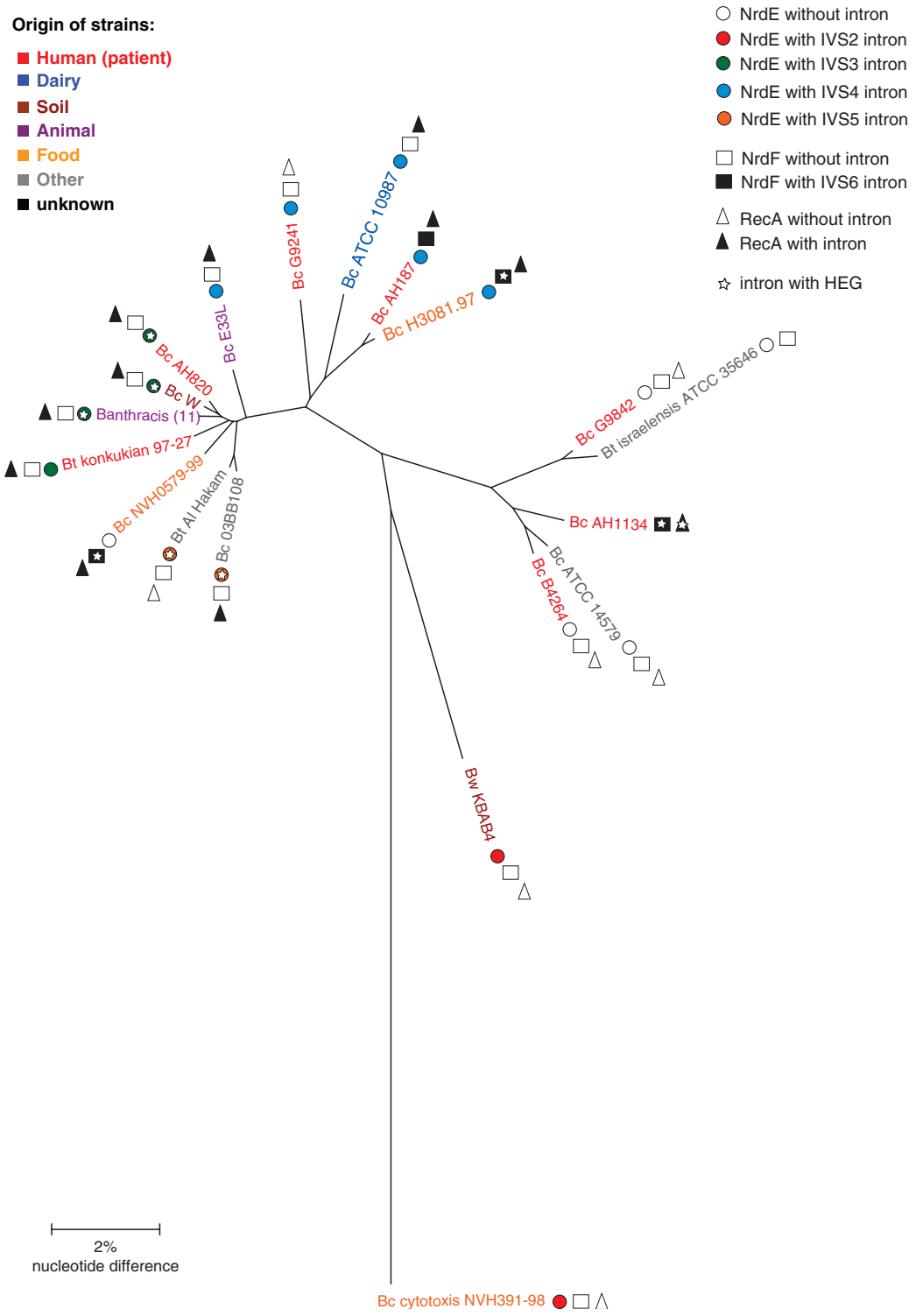

Figure 3. Unrooted phylogenetic tree of the sequenced *B. cereus* group strains. Strains are colored by source of isolation. For each isolate, the presence or absence of group I introns within the recA, nrdE and nrdF genes are indicated. The 11 *B. anthracis* isolates form a clonal complex and are represented here as a single lineage. The tree is based on the concatenation of the sequences conserved among the chromosomes of all strains (a total of 2 128 496 bp, gaps removed) and was built using the Neighbor-Joining method applied to a distance matrix of pairwise percentages of nucleotide differences between the sequences. All nodes in the tree have a bootstrap support of 100%, based on 1000 replicates.

independent events of intron gain or loss have occurred. Furthermore, the host gene of the tube intron in *B. weihenstephanensis* KBAB4, *B. cereus* B4264 and *B. thuringiensis israelensis* ATCC 35646 is only 35% identical overall to its counterpart in 0305ϕ8-36; however, 23 nt surrounding the intron insertion site are identical, indicating that the intron targets the most highly conserved region within the host gene. Such a behavior has been observed for other

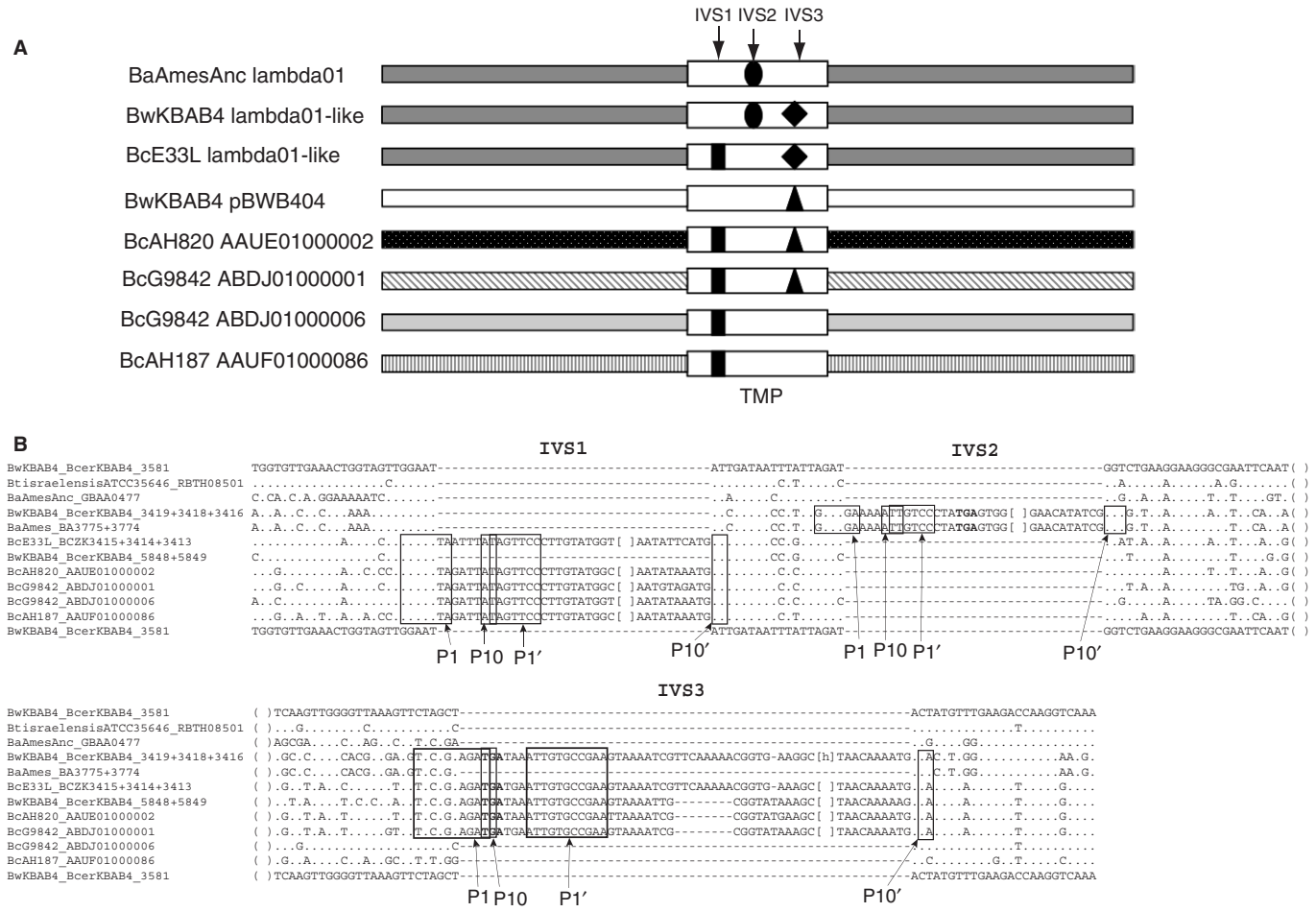

Figure 4. Group I introns within the TMP gene of *B. cereus* group prophages. (**A**) Schematic representation of the distribution of group I introns within the TMP gene of *B. cereus* group prophages. Introns are present at three insertion sites, IVS1, IVS2 and IVS3, indicated by arrows. Identical introns are represented by the same symbol. Homologous prophages are drawn in the same color or hatching pattern. Contig names are given after the strain identifiers. (**B**) Multiple sequence alignment of the intron insertion sites within the TMP gene. The sequences of intron-less TMPs were included for comparison. Gene or contig names are given after the strain identifiers. Sequences are represented as described in the legend to Figure 1. The central part of the TMP gene is not shown and is replaced with empty parentheses.

group I and group II introns and has been interpreted as a strategy used by introns to maximize their spread (1,32,60).

The catalytic RNAs of the *B. cereus* group phage introns belong to either the IA1 or IA2 structural subclasses and exhibit nucleotide sequence and structure homologies with group I introns from bacteriophages of various Gram-positive and Gram-negative bacteria (Figure 5 and Table 3), and they are not related to the *B. cereus* group recA and nrdEF introns. Furthermore, a detailed comparative sequence and structure analysis of the *B. cereus* group elements revealed some of the fundamental aspects of group I intron evolution. At each of the IVS1 and IVS2 sites within TMP the introns are 92–99% identical among the *B. cereus* group strains. Furthermore, the IVS1 and IVS2 elements are also closely related to each other. They are 80% identical overall and their secondary structures are typical of the IA2 subclass (Figure 5A and B). Interestingly, their central core region including domains P3 to P7 is more conserved (90–92% identity) and most of the divergence covers the P1, P2, P8 and P9 domains. The IVS1 and IVS2 insertion site sequences are slightly different and one of the obvious differences between the TMP-IVS1 and TMP-IVS2 introns lies within the P1 domain, which has evolved to form an IGS with either IVS1 or IVS2 (Figures 4B, 5A and B). This illustrates how homologous group I introns can adapt to different homing sites via their peripheral domains. An opposite situation is seen at the TMP-IVS3 site. Introns located at this site are clearly different, although related, and can be divided in two groups which are 67–75% identical: the introns shared by *B. cereus* E33L and *B. weihenstephanensis* KBAB4 chromosomal lambda01-like phages on one hand (TMP-IVS3a), and the introns from *B. cereus* AH820 and G9842 and the phage on the pBWB404 plasmid of *B. weihenstephanensis* KBAB4 on the other hand (TMP-IVS3b; Figure 4A). All these TMP-IVS3 elements belong to the IA1 class of group I introns (Figure 5C and D). However, in this case, the peripheral domains are more conserved than the central core (which shows only 54–62% sequence identity overall) and, in particular, their IGS and P1 domains are virtually identical, allowing these introns to recognize the same target site (Figure 4B). Here, the data

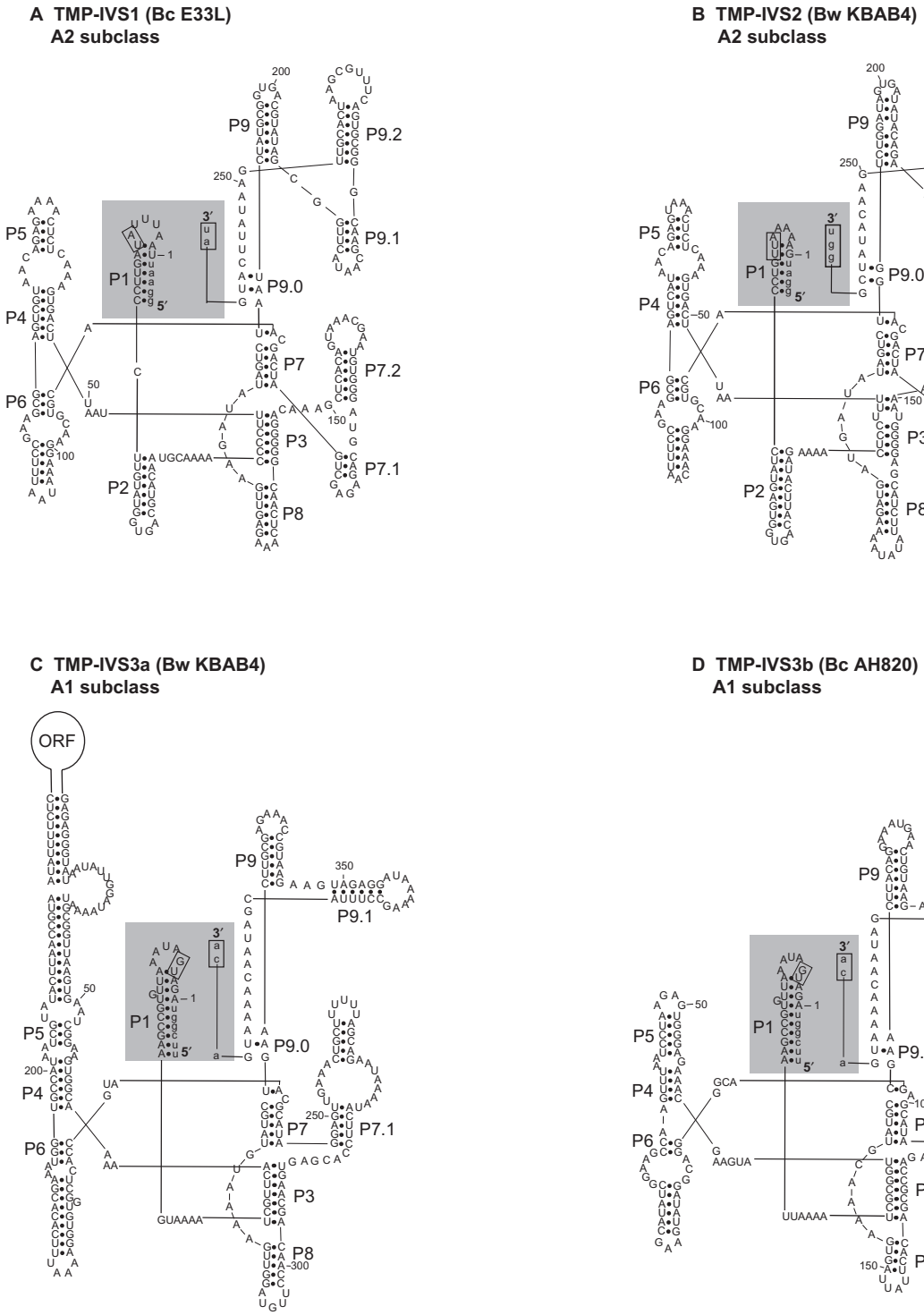

Figure 5. Predicted secondary structures of the group I introns inserted within the TMP gene of *B. cereus* group prophages. The figure illustrates intron adaptation to specific target sites. (**A** and **B**) Similar introns inserted in different sites (IVS1 and IVS2) within TMP; (**C** and **D**) different introns inserted within the same site (IVS3). Introns are represented as described in the legend to Figure 2 and, in addition, the P1 domain and bases forming the P10 pairing (boxed) involved in splice-site recognition are shaded in gray. The P1 and P10 pairings are different between the IVS1 and IVS2 introns, while they are identical between IVS3a and IVS3b. *Bacillus cereus* group host strains are given in parentheses after the intron names (other strains sharing the introns are listed in Table 2).

show how different introns can adapt to the same homing site. Finally, it should be noted that the *B. cereus* E33L intron located within the phage terminase gene (term-IVS1), although unique to that isolate, is actually homologous to the TMP-IVS3 introns (65–77% identity overall), except that its P1 domain is completely different, in agreement with the fact that the homing sites of the terminase and TMP-IVS3 introns are distinct.

This gives another example of intron adaptation to different homing sites. Furthermore, the *B. cereus* E33L term-IVS1 intron, along with the TMP-IVS3a intron located in the chromosomal lambda01-like phage of *B. weihenstephanensis* KBAB4, are the only *B. cereus* group phage introns that encode a HEG. Remarkably, although these introns are homologous, their HEGs are unrelated and belong to two different families of endonucleases, the GIY-YIG and the H-N-H family, respectively, based on conserved protein motifs (37). This indicates that similar introns have been invaded by different HEGs conferring them mobility. Indeed, HEGs are themselves selfish mobile entities that have invaded various introns and inteins (protein introns) and have therefore taken part in the evolution and spread of these elements (61,62).

Additional features common to all *B. cereus* group type I introns

A couple of features are common to the chromosomal and prophage group I introns. Firstly, all introns located within the recA, nrdE, nrdF, TMP, terminase and tail tube genes are inserted in-frame with the host gene, except nrdE-IVS5. Many of the introns, in particular all five nrdEF elements, start with a stop codon TAA, TAG or TGA, while the others contain stop codons within the first 27 nt, which would terminate translation in the P1 or P2 stems upstream of the central catalytic core (Figures 1 and 4B), with the exception of term-IVS2 for which translation would terminate 108 bases inside the intron. These properties are general among group I introns from bacteriophages and prokaryotic organisms in which transcription is coupled to translation, and reflect the selection pressure to prevent the translating ribosomes from running into the intron core, which would impair intron folding and splicing (20,63). Second, in accordance with the cyclic model of group I intron evolution, many introns have lost their HEG, which is evidenced by the fact that introns lacking the HEG have remains of the ORF. The exact position of the HEG is variable, but it is usually located within the loop of a particular intron domain. Therefore, internal deletion of the HEG does not disrupt the overall secondary structure of the intron, as in *B. thuringiensis konkukian* 97-27 nrdE-IVS3 and *B. cereus* AH187 nrdE-IVS4. Interestingly, in the case of the recA and TMP-IVS3 introns, the deleted region extends beyond the HEG and includes a piece of the intron RNA, but the deletion still maintains the overall structure and, presumably, the ability to splice (Figure 5). Indeed, the recA intron in *B. anthracis* does splice efficiently *in vitro* and *in vivo* (47). Therefore, even though the intron is degenerating, there is a strong selection pressure to maintain intron activity.

It should be mentioned that an additional group I intron in the *B. cereus* group is specifically associated with an IS element, forming a composite ribozyme called an IStron, Bc*ISt1* (46). This element will not be discussed here and a comparative IStron analysis and survey will be reported elsewhere.

DISTRIBUTION AND EVOLUTION OF GROUP II INTRONS IN THE *B. CEREUS* GROUP

Considering all *B. anthracis* strains as one and the same, *B. cereus* group genomes contain varying numbers of full-length group II introns, ranging from none in five strains to 12 or 13 elements in the two emetic isolates *B. cereus* AH187 and H3081.97, respectively [Table 1; (46,52)]. Among the 24 intron-carrying isolates examined, nine introns are shared between different subsets of strains isolated from various sources and locations (Table 2). In agreement with the general distribution of bacterial group II introns, *B. cereus* group elements are inserted in a broad set of genes from the chromosome and plasmids, and in several instances introns are 'intergenic' [i.e. located outside predicted protein-coding ORFs; however, 'intergenic' introns may be within the noncoding part of the host gene's mRNA from which they splice, (52)]. The predicted functions of the host genes are very diverse including, e.g. DNA metabolism, cell surface protein, nucleotidyltransferase, asparagine synthase and a number of hypothetical genes of unknown function. The presence of group II introns in four conserved genes involved in DNA metabolism, namely the DNA polymerase (polC), DNA repair genes of the radC and mutS families and a DNA helicase of the uvrD/rep family, is intriguing, as this kind of genes is usually the refuge for group I introns [see Group I section above; (1,32,49)]. In contrast to group I elements, no *B. cereus* group type II introns were found within identifiable prophages and many elements are located in extrachromosomal plasmids. Although it is common for bacterial group II introns to be associated with other types of mobile or transferable elements which may act as vectors for intron dissemination (16,41), none of the introns identified in the *B. cereus* group genomes analyzed here is inserted within IS elements or conjugation genes, although two instances have been reported in other isolates (64,65). These observations imply that the original donor elements may have been lost from the strains examined here and that plasmids rather than transposable elements may be a major source for intron spreading within the *B. cereus* group. This is supported by the fact that a number of introns (*B.a*.I1, *B.a*.I2, *B.c*.I1, *B.c*.I2 and *B.c*.I11) are shared between plasmids harbored by several strains belonging to separate phylogenetic lineages (Table 2 and Figure 3). For example, *B.a*.I1 and *B.a*.I2 are located within the same host genes in the pXO1 plasmid of *B. anthracis* and the virtually identical pBCXO1 plasmid of *B. cereus* G9241 (66), while *B.c*.I11 is present in the pPER272 plasmid of *B. cereus* AH820 and a related pCER270-like plasmid of *B. cereus* H3081.97 (unpublished data). Indeed, plasmid transfer within the *B. cereus* group has been demonstrated in different environments (67,68). After entering a new strain via a plasmid, the intron can subsequently integrate into the bacterial chromosome by retrohoming or retrotransposition. This is exemplified by *B.c*.I2. This intron is found in two identical copies, one in the chromosome and one in the large plasmid in *B. cereus* AH820, *B. cereus* ATCC 10987 and the emetic *B. cereus* AH187 and H3081.97 strains [Table 2 and (52)], and is inserted in

the same chromosomal and in the same plasmidic loci in all four strains. *Bacillus cereus* AH820 is phylogenetically separated from the other three isolates (Figure 3), which implies lateral acquisition of the intron, most likely via the plasmid, and subsequent invasion of the chromosome. Conversely, *B. anthracis* and *B. cereus* G9241 lack *B.c.*I2 in both the chromosomal and plasmidic host genes. In addition to horizontal transfer, vertical transmission of group II introns also occurs (8). Vertical inheritance is suggested when an intron is found in the corresponding host genes of isolates sharing a common evolutionary ancestor. This could be the case for *B.c.*I3 copies shared between the very closely related *B. cereus* 03BB108 and *B. thuringiensis* Al Hakam, as well as for *B.c.*I5, which is present only in the phylogenetic cluster comprising *B. cereus* ATCC 10987 and the emetic *B. cereus* AH187 and H3081.97 strains (Figure 3). Even though *B.c.*I5 is located on plasmids, the plasmids carried by *B. cereus* ATCC 10987, AH187 and H3081.97 have a common backbone and likely derive from a common ancestor [Figure 7C and (69)]. *B.c.*I2 is present in these strains and thus may also have been vertically transmitted among them, and this added to its horizontal transmission to or from *B. cereus* AH820, as discussed above, makes *B.c.*I2 an example of an intron which was both vertically and horizontally acquired. Other examples are provided by the *B.a.*I1 and *B.a.*I2 elements, which are most likely transmitted by vertical descent among the clonal *B. anthracis* isolates, and were transferred horizontally to or from *B. cereus* G9241.

In terms of intron structure, the secondary structure is used to divide the group II introns into subclasses, which correlate with the phylogenetic origin of the introns (70). In the *B. cereus* group genomes, the vast majority of the group II introns belong to the B class (68 out of 77), in particular the B2-like subclass, which is widespread among the strains, while eight others belong to the A1 class, most of those being found in *B. cereus* G9842 and 03BB108 (Figure 6). Even though an increasing number of introns from the C class, which specifically target transcriptional terminators (71), are being identified in diverse bacterial species, the phylogenetically divergent *B. cereus cytotoxis* NVH391-98 strain is thus far the only known *B. cereus* group isolate harboring a group II intron of class C. As opposed to most group I introns, group II elements do not generate a translation stop codon at their very 5′ end and all start with the consensus motif GUGYG (Y is C or U). Stop codons are predicted to occur deeper into the 5′ intron end, sometimes after more than 100 bases. This reflects the fact that group I and group II introns follow distinct folding pathways, as their structures and splicing mechanisms are fundamentally different (10,72). How group II introns avoid interference from ribosomes has not been particularly studied, as opposed to their group I counterparts.

ORF-less group II introns lacking the internal intron-encoded RT are infrequent in prokaryotes and only four ORF-less elements are present amongst the *B. cereus* group genomes examined in this study, one in *B. cereus* G9842 (*B.c.*I13) and three in the pE33L466 plasmid of *B. cereus* E33L [*B.c.*I7; an additional one has been

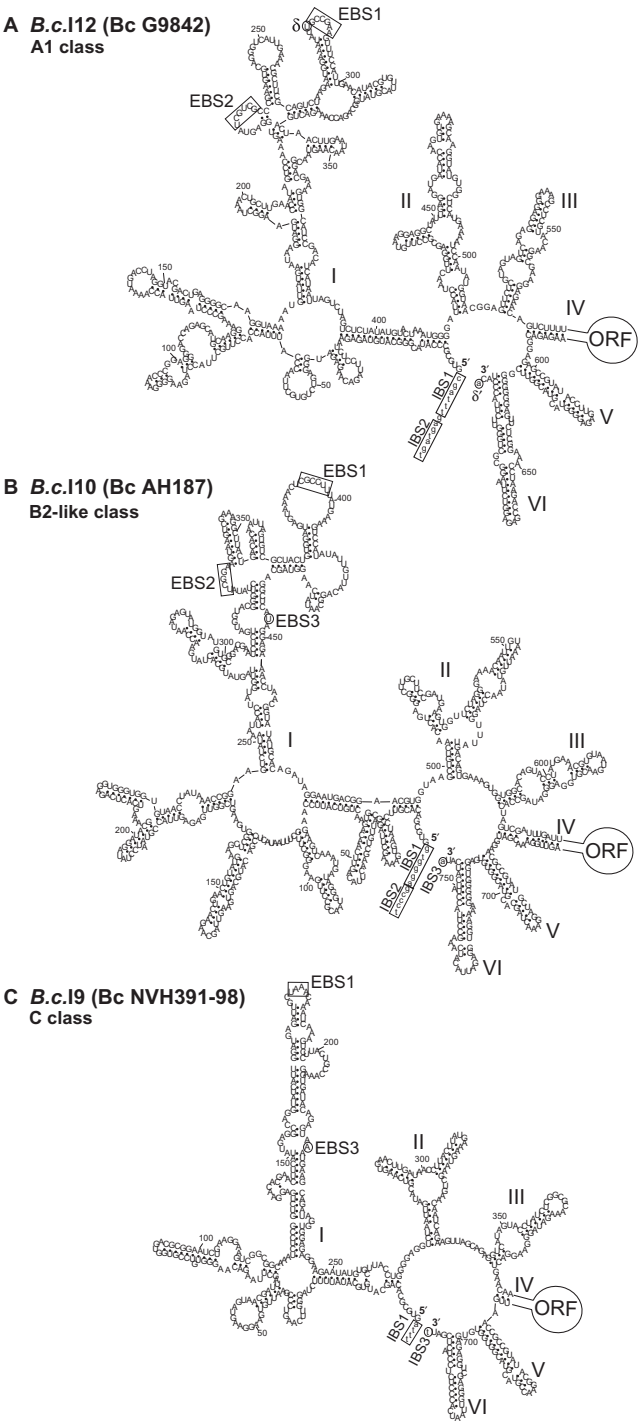

Figure 6. Predicted secondary structures of selected group II introns from *B. cereus* group bacteria, illustrating various intron classes found in these organisms. *Bacillus cereus* group host strains are given in parentheses after the intron names (other strains sharing *B.c.*I10 are listed in Table 2). Roman letters I–VI indicate the six functional RNA domains. The intron-encoded multifunctional RT ORF, located within domain IV, is not included, thus base numbering does not include the ORF sequence. Intron sequences are in uppercase letters and exon nucleotides are in lowercase letters. Base pairs are linked by dots. Potential exon-binding sites (EBS1, EBS2 and EBS3 or δ′) and their corresponding intron-binding sites (IBS1, IBS2 and IBS3 or δ) involved in base-pairings used for splice-site recognition are boxed. Note that the δ–δ′ pairing in class A introns is analogous to the EBS3–IBS3 pairing in other classes and that there is no EBS2–IBS2 pairing in class C elements.

reported in *B. thuringiensis kurstaki* HD73 (65)]. Interestingly, the three *B.c.*I7 copies are nearly identical and the presence of similar ORF-less introns in multiple sites implies that these may be actually mobile (4,7,16). Since the RT ORF is required for both splicing and mobility *in vivo*, the presence of a complete *B.c.*I7 intron in the *B. cereus* E33L genome could give the potential for activation of the ORF-less introns by the protein supplied in-*trans* by the complete element, as demonstrated in the cyanobacterium *Trichodesmium erythræum* (73). A few intron elements are particularly interesting with respect to either their distribution or their molecular features, and will be described subsequently: *B.c.*I1, *B.c.*I10 and *B.th.*I3.

*B.c.*I1, a widely distributed intergenic intron

A copy of the *B.c.*I1 group II intron is found in nine loci distributed among six different *B. cereus* group isolates from three continents (Tables 1 and 2). All *B.c.*I1 copies are >99% identical to each other. The intron is inserted either in the bacterial chromosome or a plasmid, or both, and most of the insertion sites are unique to a given strain (Figure 7A). Based on the available genomic data, *B.c.*I1 thus represents one of the most widespread group II introns in the *B. cereus* group with respect to diversity of strains and genomic locations (Table 2). Strikingly, *B.c.*I1 is always located outside coding regions and is part of the 3′ untranslated region (3′ UTR) of the mRNA transcript, as confirmed experimentally in *B. cereus* ATCC 10987 (52), since a transcriptional terminator structure is predicted downstream of the intron but not between the gene and the intron (the intron is within 120 bp of the stop codon). Group II intron homing sites generally span ~30 bp (from −20 to +10 relative to the insertion point), which are recognized and bound by the intron-encoded multifunctional RT ORF (4,9). A subset of these exon nucleotides, the intron-binding motifs IBS1, IBS2 and IBS3, are implicated in base-pairings with the complementary exon-binding sites EBS1, EBS2 and EBS3 in the intron RNA, which are necessary for splicing and reverse splicing. As can be seen in Figure 7A, all *B.c.*I1 genomic loci in the various strains are well conserved from positions −22 to +11, indicating that the intron has disseminated via site-specific retrohoming, and that its homing site covers 33 bp, as previously suggested (52). Interestingly, the RT of *B.c.*I1 lacks the endonuclease domain which is required for homing by the typical target-primed reverse transcription (TPRT) pathway. This implies that *B.c.*I1 mobility must have occurred following an alternative pathway. One possibility may be the use of DNA replication forks and nascent strands to prime reverse transcription, as demonstrated for the *Sinorhizobium meliloti* RmInt1 (*Sr.me.*I1) intron, which also lacks endonuclease activity (43). The widespread distribution of *B.c.*I1 demonstrates that introns lacking endonuclease function are still able to propagate efficiently within a bacterial population. A search for potential *B.c.*I1 homing sites in the various *B. cereus* group genomes (using the sequences shown in Figure 7A as queries) indicated that 0–4 possible targets are present in a given strain and that target sites are predominantly present in the 3′ UTR of a small number of genes. The genes are unrelated and the nature and function of the *B.c.*I1 site are unknown and deserve further investigation. By targeting specifically this noncoding motif, *B.c.*I1 uses a strategy that would minimize the impact of intron presence on host protein expression, but the intron may have an effect on the mRNA structure or stability. Two other group II introns inserted outside coding regions were identified in the *B. cereus* group genomes: *B.c.*I7 present in the pE33L466 plasmid of *B. cereus* E33L and in *B. cereus* AH1134, and *B.c.*I9 which is unique to *B. cereus cytotoxis* NVH391-98. In contrast to *B.c.*I1, these introns are usually located at long distances (>300 bp) from predicted coding sequences, and *B.c.*I7 is inserted in variable ectopic sites, indicating that the intron does not target a highly conserved motif (data not shown). Therefore, these intergenic elements exhibit features distinct from *B.c.*I1.

Numerous group II intron copies in emetic *B. cereus* isolates

The presence of 12 and 13 group II introns in the two closely related emetic *B. cereus* AH187 and H3081.97 strains, respectively, constitute the highest number of such elements for the *B. cereus* group so far. Even though strains from other bacterial species, in particular cyanobacteria, can exhibit similar or even larger group II intron content, this number is above the average, as bacterial organisms usually carry between 0 and 5 group II introns (8,16). Remarkably, many of the introns in *B. cereus* AH187 and H3081.97 are virtually identical in sequence and represent copies of four different introns, among them *B.c.*I1 and notably *B.c.*I10, which has multiplied six and seven times within the genome of *B. cereus* AH187 and H3081.97, respectively. Four of the *B.c.*I10 insertion loci are common to both emetic strains, while the others are strain-specific implying either independent mobility events or differential intron loss in the two isolates. In contrast to *B.c.*I1 discussed earlier, the conservation of the insertion sites of *B.c.*I10 is limited to the intron-binding motifs IBS1, IBS2 and IBS3 (Figure 7B). Absence of conservation beyond these motifs can be taken as evidence for retrotransposition of *B.c.*I10 into ectopic sites. Furthermore, the genomic locations of *B.c.*I10 in *B. cereus* AH187 and H3081.97 show a DNA strand bias correlating with the orientation of chromosome replication (Figure 7C). This strongly suggests that *B.c.*I10 mobility is also correlated with DNA replication. Although the RT of *B.c.*I10 does contain the endonuclease domain, which should potentially enable the intron of retrohoming by the usual target-primed reverse transcription (TPRT) pathway, use of a replication fork-based mechanism is a means to invade divergent ectopic sites that would not be recognized and cleaved by the intron-encoded RT protein, as shown for the Ll.LtrB (*L.l.*I1) intron of *Lactococcus lactis* (44). It should noted that, in fact, the correlation between intron distribution and the chromosomal DNA strand replication bias is not limited to *B.c.*I10 and extends to all group II intron copies in *B. cereus* AH187 and H3081.97 (Figure 7C). This could indicate that different introns have used a similar mobility pathway coupled to DNA

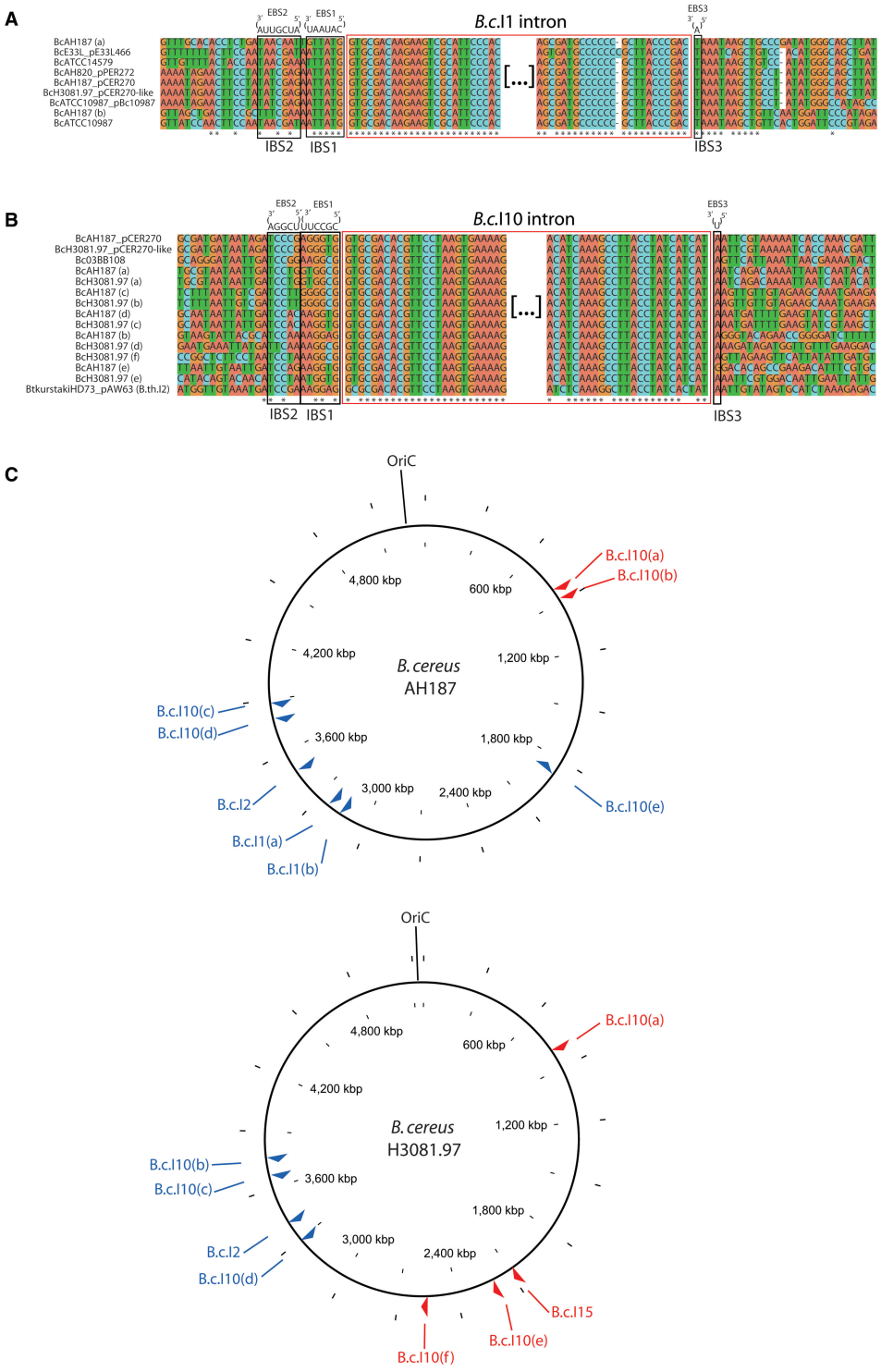

Figure 7. Insertion sites and genomic distribution of the *B.c.*I1 and *B.c.*I10 group II introns in *B. cereus* group genomes. (**A** and **B**) Multiple sequence alignments of the insertion sites of *B.c.*I1 and *B.c.*I10, respectively. The intron-binding sites (IBS1, IBS2 and IBS3) in the exons involved in base-pairings with the complementary exon-binding sites (EBS1, EBS2 and EBS3) in the intron RNA upon splicing and reverse splicing are indicated (IBS1, IBS2 and IBS3 are boxed in black). EBS1, EBS2 and EBS3 are identical in all intron copies. For the introns, delimited by a red box, only the first few nucleotides at the 5′- and 3′-ends are shown and are separated by '[. . .]'. Positions identical in all sequences are marked with asterisks below the alignments. Multiple chromosomal intron copies are distinguished by a letter in parentheses after the strain identifiers and for plasmidic copies plasmid names are given after an underscore. For *B.c.*I10, the related ORF-less intron *B.th.*I2 from the pAW63 plasmid of *B. thuringiensis kurstaki* HD73 (65) has been included for comparison. (**C**) Circular representations of the chromosomes of the emetic *B. cereus* AH187 and H3081.97 strains showing the locations of the group II introns present in these strains. Since the chromosomal sequences of these strains are unfinished, pseudochromosomes were assembled using the *B. anthracis* Ames Ancestor chromosome as a reference (see the 'Strain phylogeny reconstruction' section in Supplementary Material for details). Introns inserted in the forward and reverse DNA strands are represented by red and blue arrowheads, respectively. Multiple copies of the same intron are distinguished by a letter in parentheses. *OriC* indicates the putative origin of replication. The circular representations were generated using CGView (89).

replication to settle within the two emetic isolates. Finally, although *B.c*.I10 has invaded the two emetic genomes, only a single copy is present in another, phylogenetically unrelated strain, *B. cereus* 03BB108, isolated from dust (74). This difference is intriguing since seven of a total of nine host genes and target sites occupied by *B.c*.I10 in both emetic strains combined are also present (in corresponding locations), but unoccupied, in *B. cereus* 03BB108, indicating that the absence of intron is not due to a lack of potential insertion sites. However, in three cases, the corresponding target site in *B. cereus* 03BB108 exhibits a substitution at one of the positions that are absolutely conserved (the first base of IBS2 or the last base of IBS1, see Figure 7B); a mutation that might prevent intron insertion. In the four other loci, the sequences in *B. cereus* 03BB108 are either identical to those in *B. cereus* AH187 and H3081.97 or vary at nonconserved positions. The lack of spreading of *B.c*.I10 within *B. cereus* 03BB108 might indicate that the intron has been recently acquired and has not had time to multiply or might also be related to the biology or physiology of the strain, which provides less favorable conditions for the intron, or in which the intron activity is reduced.

An intron inserted within a newly described retroelement in *B. thuringiensis*

The *B.th*.I3 group II intron in *B. thuringiensis israelensis* ATCC 35646 is inserted within a gene encoding a RT (RBTH_06733 + RBTH_06731), and remarkably the 5′ side of the target site, including the IBS1 and IBS2 motifs, is the RT active site (AGGTATGCTGAT GACT, which corresponds to the RYADD motif at the amino acid level; Figure 8). This target specificity is shared with the *M.a*.I5 and *UA*.I3 introns from the archaeon *Methanosarcina acetivorans* and the uncultured archaeal species GZfos32G12, respectively, and the *P.l*.I2 intron from the enterobacterium *Photorhabdus luminescens*, which specifically insert into the catalytic motif of RT ORFs from other related group II introns, creating twintrons [nested introns; (60,75)]. Indeed, the *B.th*.I3, archaeal and *Photorhabdus* introns share a significant sequence homology (58–67% identity) throughout their entire lengths (1.8 kb) and all belong to the B-like structural class, to which none of the other *B. cereus* group elements belong. These similarities strongly suggest a common origin of the *B.th*.I3, *M.a*.I5, *UA*.I3 and *P.l*.I2 introns. The RT ORF of these introns lack the endonuclease domain and conservation of the insertion sites is limited to the IBS1 and IBS2 RNA-binding motifs in the 5′ exon, which makes likely that these introns use a pathway based on reverse splicing at DNA replication forks for mobility (43). Unlike for the archaeal and *Photorhabdus* introns, the host gene of *B.th*.I3 is not the RT ORF of another group II intron but belongs to a different kind of retroelement. Phylogenetic analysis of RT proteins indicated that RBTH_06733 + RBTH_06731 is distantly related to RTs of retrons, branching at the base of the retron cluster (data not shown). Retrons are genetic retroelements that produce 'multicopy single-stranded DNA' (msDNA), which is synthesized by the RT and is made up of a DNA strand covalently linked to an RNA strand. In addition to the RT ORF, retrons consist of two divergently oriented loci which encode the DNA and RNA strand of the msDNA, respectively (76,77). These loci could not be identified in the sequence upstream of the RBTH_06733 + RBTH_06731 RT and the sequence could not be folded as a retron secondary structure. In addition, the region surrounding RBTH_06733 + RBTH_06731 is flanked by nearly identical 25-bp direct repeats, a characteristic not found in retrons. The repeats are features reminiscent of transposable elements and the region in-between the repeats would define an intergenic segment of 1482 bp (Figure 8). No similar sequence was found in the other *B. cereus* group genomes or in the public databases. The target region of *B.th*.I3 may therefore represent a novel type of bacterial retroelement, possibly transposable, which needs to be further characterized experimentally.

Genomic rearrangements involving introns

Group II introns are generally not involved in major genomic rearrangements because they are usually present in low copy numbers and do not mediate the mobilization of flanking sequences during the mobility event. However, even though *B. cereus* group chromosomes are highly syntenic, in a couple of instances group II introns appear to have been the target of genomic modifications (see Supplementary Material).

GENERAL DISCUSSION

In this article, we reported on the distribution and molecular evolutionary features of group I and group II introns

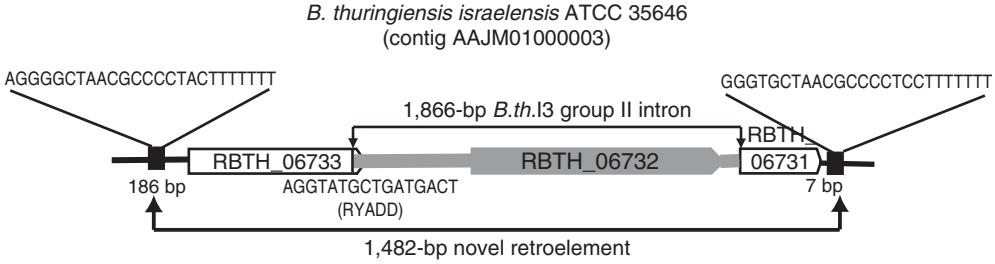

Figure 8. Schematic representation illustrating the *B.th*.I3 group II intron in *B. thuringiensis israelensis* ATCC 35646, which is inserted within a newly described retroelement. Predicted RT ORFs are drawn as boxes containing the gene names. The retroelement and the intron are delimited by arrows. The intron is in gray and is inserted at the catalytic site of the retroelement's RT (RBTH_06733 + RBTH_06731), corresponding to the RYADD motif at the amino-acid level. The 25-bp direct repeats flanking the retroelement are indicated.

in the 29 sequenced strains of the closely related *Bacillus* species forming the *B. cereus* group. The results extend those from previous intron analyses in these bacteria (46,48,49) and give a general overview of the intron evolution and spread within a worldwide-distributed Gram-positive bacterial population, while other population surveys have focused on group I introns in proteobacteria (28), cyanobacteria (29,30) and phages (17,18,31,32,78), and group II introns in *Escherichia coli* (79) and the order Rhizobiales [including *Rhizobium* and *Sinorhizobium* species; (80)]. The main findings regarding introns in the *B. cereus* group are: (i) the presence of nearly identical introns in strains of different origins or geographical locations (Tables 1 and 2); (ii) the presence of group I and group II introns from multiple structural classes and evolutionary origins (Figures 2, 5 and 6); (iii) for group I introns, bacteriophages are an important source for lateral dissemination among isolates (Table 3). Once integrated in a bacterial or phage genome, group I introns often lose their HEG mobility gene and then become vertically inherited or passively enter the bacterial genome via prophage integration (Figures 3 and 4) and (iv) for group II introns, plasmids appear to be the major vectors of horizontal spread, and introns can then subsequently invade the host genome via independent events of retrohoming and retrotransposition, and then may be vertically transmitted (Table 2 and Figure 7). An efficient strategy for propagation within genomes involves the coupling of group II intron mobility to host DNA replication.

The general theme of all intron surveys, including ours, is the large variability in intron content among bacterial species, subspecies and strains (28–30,78–80). Species from the same genus and strains from the same species can show very different intron compositions, unless the species is highly clonal like *B. anthracis*. Furthermore, there is no correlation between intron content and strain origin (Figure 3 and Table 1). These sporadic and variable distributions have been attributed to the way group I and group II introns spread among bacteria. The chance a given strain has to receive an intron depends on the environment and microbial community it resides in, and a fortuitous opportunity to obtain specific introns from specific plasmid, transposon or phage donors. In addition, the molecular features of the host strain are also important for intron acquisition, in particular the availability of suitable target sites. It is also likely that the physiological conditions in the host are determinant for intron maintenance and activity, as are population genetics factors which can drive the cycle of intron invasion and loss, as demonstrated for group I introns (57,58). The molecular and biological properties of the host strain are certainly an underlying factor responsible for the differences in copy numbers of the same intron in various strains. On the other hand, some of these differences could also reflect the fact that isolates have obtained the intron at different times during their evolution. Due to the general sporadic nature of intron distribution, it should be emphasized that the intron content of a species cannot be predicted. For example, for the Bacilli species *B. subtilis* and *B. halodurans*, a single strain has been sequenced, namely *B. subtilis* 168, which carries no group II intron at all, while *B. halodurans* C-125 contains a number of group II introns belonging to the unusual C class, which specifically insert after transcription terminators. No conclusion regarding the intron distributions in these species can be drawn until multiple isolates have been examined.

The finding of virtually identical introns shared between strains from different sources, locations, dates of isolation and/or evolutionary lineages is puzzling (Tables 1 and 2, Figure 3). High similarities between introns of strains from various origins and/or separate phylogenetic branches have been observed in other organisms (30,79,80) and are usually interpreted as recent horizontal transfer events. Nevertheless, how these events occurred are not trivial to explain, especially when the strains implicated have been isolated from different continents and decades apart, such as e.g. *B. cereus* ATCC 10987 (Canada, 1930) and *B. cereus* E33L (Africa, 1996; Table 1). The various isolates may have acquired the introns from widespread donors, and thus they do not have to be in direct contact with each other to receive the same mobile element. Another possibility may be that the donors and/or the recipient organisms have been disseminated worldwide by human and/or animal movements and transport. Alternatively, the high conservation of introns in different strains may also suggest a strong selection pressure to maintain the sequence and structure for efficient splicing and mobility. This could be the case for group I introns inserted in essential host genes, which presumably need to be highly efficient. As hypothesized for the tRNALeu(UAA) group I intron in cyanobacteria (30), the fact that some introns are highly conserved and maintained in multiple strains could also suggest that they might confer some kind of selective advantage to the host bacterium under specific conditions, for example, as a regulatory mechanism for protein expression or avoidance of R-loop formation in RNAs (19,81). Some of the arguments discussed earlier, such as recent transfers or strong selective pressure, could also be put forward to explain the presence of multiple identical group II intron copies within a given genome.

The distribution patterns of group II introns in the *B. cereus* group are in agreement with the general distribution in bacterial genomes and the fact that they behave as selfish mobile retroelements rather than introns (in the eukaryotic sense, i.e. only functioning as splicing elements) (4,7,8,16). Based on the available data, group II introns in the *B. cereus* group do not seem to be particularly associated with IS elements, which is in contrast with *E. coli* and the Rhizobiales (79,80). Rather, *B. cereus* group type II introns preferentially use plasmids as dissemination vectors. Indeed, group II introns are found in different types or families of large plasmids carried by *B. cereus* group bacteria, including the plasmids sharing a common backbone with the pXO1 plasmid of *B. anthracis* (the pXO1-like plasmids pBc10987, pPER272 and PCER270), the pXO2-like plasmid pAW63, the pBC210-pE33L466 group and pBMB67 (45,64,65,69). Interestingly, no introns have been identified in small plasmids (<15 kb in size) replicating by the rolling circle (RC) mechanism (82), which are common in the *B. cereus* group (83). In fact, all plasmidic group II introns available at the group II

intron database (55) are exclusively present in large plasmids (26–1300 kb) that are likely to replicate by non-RC-like mechanisms. Whether there is an intrinsic molecular or biochemical reason for this due to the plasmid size or replication mode has not been investigated. In this study, we have reported group I introns that are inserted within prophages integrated in plasmids, showing that several genetic elements can combine to spread these introns. It should be noted that, unlike group II introns, no group I introns in any bacterium have been identified in standalone plasmid genes that are not part of prophages. Here again, whether this is simply because group I introns preferentially target highly conserved housekeeping genes that are normally not encoded on plasmids, or whether there is an underlying functional or mechanistic reason, e.g. involving the DNA break repair process, is not clear. It is also not known why group II introns are not found in bacteriophages unlike group I elements. Overall, the fact that group I introns are mainly spread by phages, while group II introns are mostly spread by plasmids is probably due to the specific molecular mechanisms and requirements for splicing and mobility that these different elements have.

Another slight discrepancy between the distribution of *B. cereus* group type II introns and the general distribution of introns in bacteria was the finding of group II introns in conserved DNA metabolism and repair genes, a type of genes which is usually the target of group I intron elements. Interestingly, in four strains of thermophilic *Geobacillus kaustophilus*, *G. stearothermophilus* and *Bacillus caldolyticus* the essential housekeeping recA gene is interrupted by a group II intron (84,85). It has been suggested that the element serves as an intron in this locus rather than a selfish mobile retroelement (although it encodes an RT ORF) and that its activity might be related to some regulation of the recA gene expression under high temperature. In comparison to bacterial group II introns, bacterial group I elements behave more like 'splice-only' introns, since they have a propensity of losing their internal mobility HEG gene, retaining only the splicing activity, and target a restricted set of highly conserved genes, a situation which is reminiscent of organellar group II introns (16). Why group I introns are preferentially inserted into DNA metabolism or other essential phage genes or chromosomal tRNA/rRNA genes has been extensively debated (1).

The compilation of group II introns in bacterial genomes by Dai and Zimmerly (16) revealed that bacterial group II introns have a general tendency to become fragmented. Fragmentation can occur at the 5′- or 3′-ends and may result for example from incomplete reverse transcription or genomic rearrangements, such as insertions of IS elements. In agreement with this trend, partial introns were identified in about half of the sequenced *B. cereus* group isolates (usually one or two fragments per strain, considering all *B. anthracis* strains as one, and some fragments are shared among strains; data not shown); however, group II intron fragmentation in the *B. cereus* group does not seem to be as prevalent as in *E. coli* or the Rhizobiales (79,80). To our knowledge, fragmentation of bacterial group I introns has not been specifically reported, in contrast to bacterial group II elements, and this is another similarity with organellar group II introns and their behavior as 'splice-only' introns (16). This absence of fragmentation reflects the fact that group I and organellar group II introns are inserted in essential genes whose function and integrity must be maintained.

Our survey of group I and group II introns in the *B. cereus* group has confirmed the general patterns of intron behavior and evolution observed in other bacteria, while pointing out a few differences, providing examples based on a relatively large amount of genomic data and the virtually full complement of introns in a given strain. We have identified a number of group I and group II elements that are not annotated in the genomic database records of the *B. cereus* group, in particular the ORF-less introns. This, added to the finding of a novel retroelement in *B. thuringiensis*, indicates that there may be a number of unannotated genetic elements yet to be discovered in genome sequences. The survey described in this article revealed that the diversity and spread of mobile introns in the *B. cereus* group is more common than previously recognized. Our study gives a snapshot of the genomic picture of introns in the *B. cereus* group population as we can infer it from the data currently available. Some questions remain open, such as the origin and maintenance of the introns in these bacteria and experimental data are needed to confirm the activity of the elements. The information reported here should be relevant to the study of other bacterial organisms.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Fredrik B. Stabell for helpful comments on the article. The work was supported by the Norwegian Functional Genomics (FUGE II) and Consortium for Advanced Microbial Sciences and Technologies (CAMST) platform of the Research Council of Norway. Funding to pay the Open Access publication charge was provided by the Norwegian Functional Genomics (FUGE II) platform of the Research Council of Norway.

Conflict of interest statement. None declared.

REFERENCES

1. Edgell,D.R., Belfort,M. and Shub,D.A. (2000) Barriers to intron promiscuity in bacteria. *J. Bacteriol.*, **182**, 5281–5289.
2. Haugen,P., Simon,D.M. and Bhattacharya,D. (2005) The natural history of group I introns. *Trends Genet.*, **21**, 111–119.
3. Hougland,J.L., Piccirilli,J.A., Forconi,M., Lee,J. and Herschlag,D., (2006) In Gesteland,R. F., Cech,T.R. and Atkins,J.F. (eds), *The RNA World, 3rd edn, Vol. Cold Spring Harbor monograph series; Vol. 43*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 133–205.
4. Lambowitz,A.M. and Zimmerly,S. (2004) Mobile group II introns. *Annu. Rev. Genet.*, **38**, 1–35.
5. Lehmann,K. and Schmidt,U. (2003) Group II introns: structure and catalytic versatility of large natural ribozymes. *Crit. Rev. Biochem. Mol. Biol.*, **38**, 249–303.

6. Pyle,A.M. and Lambowitz,A.M. (2006) In Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds), *The RNA World, 3rd edn, Vol. Cold Spring Harbor monograph series; Vol. 43*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 469–505.
7. Robart,A.R. and Zimmerly,S. (2005) Group II intron retroelements: function and diversity. *Cytogenet. Genome Res.*, **110**, 589–597.
8. Toro,N. (2003) Bacteria and Archaea group II introns: additional mobile genetic elements in the environment. *Environ. Microbiol.*, **5**, 143–151.
9. Toro,N., Jimenez-Zurdo,J.I. and Garcia-Rodriguez,F.M. (2007) Bacterial group II introns: not just splicing. *FEMS Microbiol. Rev.*, **31**, 342–358.
10. Woodson,S.A. (2005) Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.*, **15**, 324–330.
11. Stahley,M.R. and Strobel,S.A. (2006) RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr. Opin. Struct. Biol.*, **16**, 319–326.
12. Vicens,Q. and Cech,T.R. (2006) Atomic level architecture of group I introns revealed. *Trends Biochem. Sci.*, **31**, 41–51.
13. Mueller,J.E., Clyman,J., Huang,Y.J., Parker,M.M. and Belfort,M. (1996) Intron mobility in phage T4 occurs in the context of recombination-dependent DNA replication by way of multiple pathways. *Genes Dev.*, **10**, 351–364.
14. Smith,D., Zhong,J., Matsuura,M., Lambowitz,A.M. and Belfort,M. (2005) Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes Dev.*, **19**, 2477–2487.
15. Cho,Y., Qiu,Y.L., Kuhlman,P. and Palmer,J.D. (1998) Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl Acad. Sci. USA*, **95**, 14244–14249.
16. Dai,L. and Zimmerly,S. (2002a) Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic Acids Res.*, **30**, 1091–1102.
17. Foley,S., Bruttin,A. and Brüssow,H. (2000) Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus* bacteriophages. *J. Virol.*, **74**, 611–618.
18. Sandegren,L. and Sjöberg,B.M. (2004) Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages: evidence for recent transfer of old introns. *J. Biol. Chem.*, **279**, 22218–22227.
19. Chen,Y., Klein,J.R., McKay,L.L. and Dunny,G.M. (2005) Quantitative analysis of group II intron expression and splicing in *Lactococcus lactis*. *Appl. Environ. Microbiol.*, **71**, 2576–2586.
20. Sandegren,L. and Sjöberg,B.M. (2007) Self-splicing of the bacteriophage T4 group I introns requires efficient translation of the pre-mRNA *in vivo* and correlates with the growth state of the infected bacterium. *J. Bacteriol.*, **189**, 980–990.
21. Everett,K.D., Kahane,S., Bush,R.M. and Friedman,M.G. (1999) An unspliced group I intron in 23S rRNA links *Chlamydiales*, chloroplasts, and mitochondria. *J. Bacteriol.*, **181**, 4734–4740.
22. Cui,X. and Davis,G. (2007) Mobile group II intron targeting: applications in prokaryotes and perspectives in eukaryotes. *Front Biosci.*, **12**, 4972–4985.
23. Jung,H.S. and Lee,S.W. (2006) Ribozyme-mediated selective killing of cancer cells expressing carcinoembryonic antigen RNA by targeted trans-splicing. *Biochem. Biophys. Res. Commun.*, **349**, 556–563.
24. Karberg,M., Guo,H., Zhong,J., Coon,R., Perutka,J. and Lambowitz,A.M. (2001) Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat. Biotechnol.*, **19**, 1162–1167.
25. Wei,M.Q., Mengesha,A., Good,D. and Anne,J. (2008) Bacterial targeted tumour therapy-dawn of a new era. *Cancer Lett.*, **259**, 16–27.
26. Disney,M.D., Childs,J.L. and Turner,D.H. (2004) New approaches to targeting RNA with oligonucleotides: inhibition of group I intron self-splicing. *Biopolymers*, **73**, 151–161.
27. Nesbø,C.L. and Doolittle,W.F. (2003) Active self-splicing group I introns in 23S rRNA genes of hyperthermophilic bacteria, derived from introns in eukaryotic organelles. *Proc. Natl Acad. Sci. USA*, **100**, 10806–10811.
28. Paquin,B., Heinfling,A. and Shub,D.A. (1999) Sporadic distribution of tRNA(Arg)CCU introns among alpha-purple bacteria: evidence for horizontal transmission and transposition of a group I intron. *J. Bacteriol.*, **181**, 1049–1053.
29. Rudi,K. and Jakobsen,K.S. (1997) Cyanobacterial tRNALeu(UAA) group I introns have polyphyletic origin. *FEMS Microbiol. Lett.*, **156**, 293–298.
30. Paquin,B., Kathe,S.D., Nierzwicki-Bauer,S.A. and Shub,D.A. (1997) Origin and evolution of group I introns in cyanobacterial tRNA genes. *J. Bacteriol.*, **179**, 6798–6806.
31. Kwan,T., Liu,J., DuBow,M., Gros,P. and Pelletier,J. (2005) The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc. Natl Acad. Sci. USA*, **102**, 5174–5179.
32. Lazarevic,V. (2001) Ribonucleotide reductase genes of *Bacillus* prophages: a refuge to introns and intein coding sequences. *Nucleic Acids Res.*, **29**, 3212–3218.
33. Landthaler,M. and Shub,D.A. (2003) The nicking homing endonuclease I-BasI is encoded by a group I intron in the DNA polymerase gene of the *Bacillus thuringiensis* phage Bastille. *Nucleic Acids Res.*, **31**, 3071–3077.
34. Jaeger,L., Michel,F. and Westhof,E. (1996) In Eckstein,F. and Lilley,D.M.J. (eds), *Catalytic RNA*, *Vol. 10*. Springer, Berlin, pp. 33–51.
35. Michel,F. and Westhof,E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.*, **216**, 585–610.
36. Mueller,J.E., Smith,D. and Belfort,M. (1996) Exon coconversion biases accompanying intron homing: battle of the nucleases. *Genes Dev.*, **10**, 2158–2166.
37. Stoddard,B.L. (2005) Homing endonuclease structure and function. *Q. Rev. Biophys.*, **38**, 49–95.
38. Birgisdottir,A.B. and Johansen,S. (2005) Site-specific reverse splicing of a HEG-containing group I intron in ribosomal RNA. *Nucleic Acids Res.*, **33**, 2042–2051.
39. Roman,J. and Woodson,S.A. (1998) Integration of the *Tetrahymena* group I intron into bacterial rRNA by reverse splicing in vivo. *Proc. Natl Acad. Sci. USA*, **95**, 2134–2139.
40. Belhocine,K., Yam,K.K. and Cousineau,B. (2005) Conjugative transfer of the *Lactococcus lactis* chromosomal sex factor promotes dissemination of the Ll.LtrB group II intron. *J. Bacteriol.*, **187**, 930–939.
41. Klein,J.R. and Dunny,G.M. (2002) Bacterial group II introns and their association with mobile genetic elements. *Front Biosci.*, **7**, d1843–1856.
42. Seetharaman,M., Eldho,N.V., Padgett,R.A. and Dayie,K.T. (2006) Structure of a self-splicing group II intron catalytic effector domain 5: parallels with spliceosomal U6 RNA. *RNA*, **12**, 235–247.
43. Martinez-Abarca,F., Barrientos-Duran,A., Fernandez-Lopez,M. and Toro,N. (2004) The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. *Nucleic Acids Res.*, **32**, 2880–2888.
44. Zhong,J. and Lambowitz,A.M. (2003) Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J.*, **22**, 4555–4565.
45. Rasko,D.A., Altherr,M.R., Han,C.S. and Ravel,J. (2005) Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiol. Rev.*, **29**, 303–329.
46. Tourasse,N.J., Helgason,E., Økstad,O.A., Hegna,I.K. and Kolstø,A.B. (2006) The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. *J. Appl. Microbiol.*, **101**, 579–593.
47. Ko,M., Choi,H. and Park,C. (2002) Group I self-splicing intron in the recA gene of *Bacillus anthracis*. *J. Bacteriol.*, **184**, 3917–3922.
48. Nord,D. and Sjöberg,B.M. (2008) Unconventional GIY-YIG homing endonuclease encoded in group I introns in closely related strains of the *Bacillus cereus* group. *Nucleic Acids Res.*, **36**, 300–310.
49. Nord,D., Torrents,E. and Sjöberg,B.M. (2007) A functional homing endonuclease in the *Bacillus anthracis nrdE* group I intron. *J. Bacteriol.*, **189**, 5293–5301.
50. Robart,A.R., Montgomery,N.K., Smith,K.L. and Zimmerly,S. (2004) Principles of 3′ splice site selection and alternative splicing for an unusual group II intron from *Bacillus anthracis*. *RNA*, **10**, 854–862.
51. Stabell,F.B., Tourasse,N.J., Ravnum,S. and Kolstø,A.B. (2007) Group II intron in *Bacillus cereus* has an unusual 3′ extension and

splices 56 nucleotides downstream of the predicted site. *Nucleic Acids Res.*, **35**, 1612–1623.
52. Tourasse,N.J., Stabell,F.B., Reiter,L. and Kolstø,A.B. (2005) Unusual group II introns in bacteria of the *Bacillus cereus* group. *J. Bacteriol.*, **187**, 5437–5451.
53. Van Ert,M.N., Easterday,W.R., Huynh,L.Y., Okinaka,R.T., Hugh-Jones,M.E., Ravel,J., Zanecki,S.R., Pearson,T., Simonson,T.S., U'Ren,J.M. *et al.* (2007) Global genetic population structure of *Bacillus anthracis*. *PLoS ONE*, **2**, e461.
54. Zhou,Y., Lu,C., Wu,Q.J., Wang,Y., Sun,Z.T., Deng,J.C. and Zhang,Y. (2008) GISSD: Group I intron sequence and structure database. *Nucleic Acids Res.*, **36(Database issue)**, D31–D37.
55. Dai,L., Toor,N., Olson,R., Keeping,A. and Zimmerly,S. (2003) Database for mobile group II introns. *Nucleic Acids Res.*, **31**, 424–426.
56. Chen,I., Christie,P.J. and Dubnau,D. (2005) The ins and outs of DNA transfer in bacteria. *Science*, **310**, 1456–1460.
57. Goddard,M.R. and Burt,A. (1999) Recurrent invasion and extinction of a selfish gene. *Proc. Natl Acad. Sci. USA*, **96**, 13880–13885.
58. Burt,A. and Koufopanou,V. (2004) Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr. Opin. Genet. Dev.*, **14**, 609–615.
59. Hardies,S.C., Thomas,J.A. and Serwer,P. (2007) Comparative genomics of *Bacillus thuringiensis* phage 0305phi8-36: defining patterns of descent in a novel ancient phage lineage. *Virol. J.*, **4**, 97.
60. Dai,L. and Zimmerly,S. (2003) ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. *RNA*, **9**, 14–19.
61. Chevalier,B.S. and Stoddard,B.L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.*, **29**, 3757–3774.
62. Gimble,F.S. (2000) Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol. Lett.*, **185**, 99–107.
63. Öhman-Hedén,M., Åhgren-Stålhandske,A., Hahne,S. and Sjöberg,B.M. (1993) Translation across the 5′-splice site interferes with autocatalytic splicing. *Mol. Microbiol.*, **7**, 975–982.
64. Chao,L., Qiyu,B., Fuping,S., Ming,S., Dafang,H., Guiming,L. and Ziniu,Y. (2007) Complete nucleotide sequence of pBMB67, a 67-kb plasmid from *Bacillus thuringiensis* strain YBT-1520. *Plasmid*, **57**, 44–54.
65. Van der Auwera,G.A., Andrup,L. and Mahillon,J. (2005) Conjugative plasmid pAW63 brings new insights into the genesis of the *Bacillus anthracis* virulence plasmid pXO2 and of the *Bacillus thuringiensis* plasmid pBT9727. *BMC Genomics*, **6**, 103.
66. Hoffmaster,A.R., Ravel,J., Rasko,D.A., Chapman,G.D., Chute,M.D., Marston,C.K., De,B.K., Sacchi,C.T., Fitzgerald,C., Mayer,L.W. *et al.* (2004) Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *Proc. Natl Acad. Sci. USA*, **101**, 8449–8454.
67. Van der Auwera,G.A., Timmery,S., Hoton,F. and Mahillon,J. (2007) Plasmid exchanges among members of the *Bacillus cereus* group in foodstuffs. *Int. J. Food Microbiol.*, **113**, 164–172.
68. Yuan,Y.M., Hu,X.M., Liu,H.Z., Hansen,B.M., Yan,J.P. and Yuan,Z.M. (2007) Kinetics of plasmid transfer among *Bacillus cereus* group strains within lepidopteran larvae. *Arch. Microbiol.*, **187**, 425–431.
69. Rasko,D.A., Rosovitz,M.J., Økstad,O.A., Fouts,D.E., Jiang,L., Cer,R.Z., Kolstø,A.B., Gill,S.R. and Ravel,J. (2007) Complete sequence analysis of novel plasmids from emetic and periodontal *Bacillus cereus* isolates reveals a common evolutionary history among the *B. cereus*-group plasmids, including *Bacillus anthracis* pXO1. *J. Bacteriol.*, **189**, 52–64.
70. Toor,N., Hausner,G. and Zimmerly,S. (2001) Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA*, **7**, 1142–1152.
71. Robart,A.R., Seo,W. and Zimmerly,S. (2007) Insertion of group II intron retroelements after intrinsic transcriptional terminators. *Proc. Natl Acad. Sci. USA*, **104**, 6620–6625.
72. Waldsich,C. and Pyle,A.M. (2007) A folding control element for tertiary collapse of a group II intron ribozyme. *Nat. Struct. Mol. Biol.*, **14**, 37–44.
73. Meng,Q., Wang,Y. and Liu,X.Q. (2005) An intron-encoded protein assists RNA splicing of multiple similar introns of different bacterial genes. *J. Biol. Chem.*, **280**, 35085–35088.
74. Hoffmaster,A.R., Hill,K.K., Gee,J.E., Marston,C.K., De,B.K., Popovic,T., Sue,D., Wilkins,P.P., Avashia,S.B., Drumgoole,R. *et al.* (2006) Characterization of *Bacillus cereus* isolates associated with fatal pneumonias: strains are closely related to *Bacillus anthracis* and harbor *B. anthracis* virulence genes. *J. Clin. Microbiol.*, **44**, 3352–3360.
75. Rest,J.S. and Mindell,D.P. (2003) Retroids in archaea: phylogeny and lateral origins. *Mol. Biol. Evol.*, **20**, 1134–1142.
76. Lampson,B., Inouye,M. and Inouye,S. (2001) The msDNAs of bacteria. *Prog. Nucleic Acid Res. Mol. Biol.*, **67**, 65–91.
77. Lampson,B.C., Inouye,M. and Inouye,S. (2005) Retrons, msDNA, and the bacterial genome. *Cytogenet. Genome Res.*, **110**, 491–499.
78. Stankovic,S., Soldo,B., Beric-Bjedov,T., Knezevic-Vukcevic,J., Simic,D. and Lazarevic,V. (2007) Subspecies-specific distribution of intervening sequences in the *Bacillus subtilis* prophage ribonucleotide reductase genes. *Syst. Appl. Microbiol.*, **30**, 8–15.
79. Dai,L. and Zimmerly,S. (2002b) The dispersal of five group II introns among natural populations of *Escherichia coli*. *RNA*, **8**, 1294–1307.
80. Fernandez-Lopez,M., Muñoz-Adelantado,E., Gillis,M., Willems,A. and Toro,N. (2005) Dispersal and evolution of the *Sinorhizobium meliloti* group II RmInt1 intron in bacteria that interact with plants. *Mol. Biol. Evol.*, **22**, 1518–1528.
81. Niu,D.K. (2007) Protecting exons from deleterious R-loops: a potential advantage of having introns. *Biol. Direct*, **2**, 11.
82. Khan,S.A. (2005) Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid*, **53**, 126–136.
83. Andrup,L., Jensen,G.B., Wilcks,A., Smidt,L., Hoflack,L. and Mahillon,J. (2003) The patchwork nature of rolling-circle plasmids: comparison of six plasmids from two distinct *Bacillus thuringiensis* serotypes. *Plasmid*, **49**, 205–232.
84. Chee,G.J. and Takami,H. (2005) Housekeeping *recA* gene interrupted by group II intron in the thermophilic *Geobacillus kaustophilus*. *Gene*, **363**, 211–220.
85. Ng,B., Nayak,S., Gibbs,M.D., Lee,J. and Bergquist,P.L. (2007) Reverse transcriptases: intron-encoded proteins found in thermophilic bacteria. *Gene*, **393**, 137–144.
86. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
87. Cech,T.R., Damberger,S.H. and Gutell,R.R. (1994) Representation of the secondary and tertiary structure of group I introns. *Nat. Struct. Biol.*, **1**, 273–280.
88. De Rijk,P., Wuyts,J. and De Wachter,R. (2003) RnaViz 2: an improved representation of RNA secondary structure. *Bioinformatics*, **19**, 299–300.
89. Stothard,P. and Wishart,D.S. (2005) Circular genome visualization and exploration using CGView. *Bioinformatics*, **21**, 537–539.