ISME

ARTICLE

The CRISPR-Cas systems were selectively inactivated during evolution of *Bacillus cereus* group for adaptation to diverse environments

Ziqiang Zheng[1] · Yulan Zhang[1] · Zhiyu Liu[1] · Zhaoxia Dong[1] · Chuanshuai Xie[1] · Alejandra Bravo[2] · Mario Soberón[2] · Jacques Mahillon[3] · Ming Sun[1] · Donghai Peng[1]

Abstract

CRISPR-Cas systems are considered as barriers to horizontal gene transfer (HGT). However, the influence of such systems on HGT within species is unclear. Also, little is known about the impact of CRISPR-Cas systems on bacterial evolution at the population level. Here, using *Bacillus cereus* sensu lato as model, we investigate the interplay between CRISPR-Cas systems and HGT at the population scale. We found that only a small fraction of the strains have CRISPR-Cas systems (13.9% of 1871), and most of such systems are defective based on their gene content analysis. Comparative genomic analysis revealed that the CRISPR-Cas systems are barriers to HGT within this group, since strains harboring active systems contain less mobile genetic elements (MGEs), have lower fraction of unique genes and also display limited environmental distributions than strains without active CRISPR-Cas systems. The introduction of a functional CRISPR-Cas system into a strain lacking the system resulted in reduced adaptability to various stresses and decreased pathogenicity of the transformant strain, indicating that *B. cereus* group strains could benefit from inactivating such systems. Our work provides a large-scale case to support that the CRISPR-Cas systems are barriers to HGT within species, and that in the *B. cereus* group the inactivation of CRISPR-Cas systems correlated with acquisition of MGEs that could result in better adaptation to diverse environments.

Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and its associated proteins (Cas) systems function

✉ Ming Sun
m98sun@mail.hzau.edu.cn

✉ Donghai Peng
donghaipeng@mail.hzau.edu.cn

[1] State Key Laboratory of Agricultural Microbiology, College of Life Science and Technology, Huazhong Agricultural University, Wuhan, 430070 Hubei, China

[2] Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apdo. postal 510-3, Cuernavaca, 62250 Morelos, Mexico

[3] Laboratory of Food and Environmental Microbiology, Earth and Life Institute, UCLouvain, Croix du Sud, 2 - L7.05.12, B-1348 Louvain-la-Neuve, Belgium

as an antiviral defense pathway in prokaryotes [1, 2]. A typical CRISPR-Cas system consists of three functional components: an operon containing a set of *cas* genes, a leader sequence, and a CRISPR DNA array. The *cas* genes are usually required for new spacer acquisition, target recognition, and degradation of foreign nucleic acids [3, 4]. The classification of CRISPR-Cas systems is based on their signature of *cas* genes, the organization of *cas* operons, and the direct repeats present in the CRISPR arrays [5]. Six types of CRISPR-Cas systems have been characterized: types I, III, and IV of class 1 [6], and types II, V, and VI of class 2 [7].

The discovery and characterization of CRISPR-Cas, especially in members of Class 2 systems, has led to a revolution in genome editing and engineering technologies [7–9]. Beyond its immune function, a number of recent observations revealed unexpected roles for CRISPR-Cas systems in gene expression, autoimmunity, chromosomal segregation or rearrangement, and DNA repairing, although most of the mechanisms involved in these functions still remain unclear [10]. Of particular interest was the discovery of the involvement of CRISPR-Cas systems on the regulation of bacterial virulence [11–13].

The roles for CRISPR-Cas systems on horizontal gene transfer (HGT) have been widely investigated. It was primary proposed that CRISPR-Cas system may prevent the acquisition of advantageous mobile genetic elements (MGEs) [4], because of its immune function against acquisition of foreign DNA. A series of studies support this hypothesis and demonstrate that CRISPR-Cas limits HGT through conjugation and transformation. For example, the CRISPR loci in *Staphylococci* counteracts multiple routes of HGT to limit the spread of antibiotic resistance genes [4]. Similarly, in *Streptococcus pneumoniae* the CRISPR-Cas prevents the acquisition of virulence related genes [11], and in *Enterococcus faecalis* it was reported that the potential for DNA acquisition could be maximized by attenuating its CRISPR-Cas system [14, 15]. Previously, a bioinformatic study showed that there is no detectable influence of CRISPR-Cas systems on the HGT rate over evolutionary timescales among species [16]. However, new evidence provides an opposite view, suggesting that CRISPR-Cas facilitates the HGT by transduction [17–19]. For instance, the CRISPR-Cas system in *Pectobacterium atrosepticum* facilitates acquisition of non-phage genetic material by phage-mediated transduction [18]. Also, it was shown that phages can recombine with spacers from CRISPR-Cas loci leading to transfer CRISPR-adjacent genes promoting HGT in *Staphylococci* [19].

Despite the relevance of CRISPR-Cas systems as antiviral defense strategy in bacteria, the influence of CRISPR-Cas on HGT has remained a matter of debate. The existing studies mostly were conducted with individual strains, but these analyses cannot truly reflect the relationship between CRISPR-Cas systems and HGT within species at the population scale. In addition, the CRISPR-Cas systems have been proposed as important factors contributing for the evolution of virulence in several pathogens [11, 14–16, 18], but the impact of such systems on bacterial evolution at population level is also poorly understood. The increasing availability of sequenced genomes will facilitate the study and understanding of the role interplayed among CRISPR-Cas systems and HGT, and its impact on bacterial evolution within species.

Members of the *Bacillus cereus* group are ubiquitous and highly versatile bacteria. It has been proposed that they have evolved in ecotypes since they differ mainly in their plasmid DNA content [20]. The plasmids present in this group of bacteria harbor different toxin genes, conferring diverse host specificities such as *Bacillus thuringiensis*, that is an important entomopathogen used worldwide as an effective bio-insecticide to control various pests [21]; *Bacillus anthracis*, which cause anthrax and *B. cereus* that may cause food-borne gastroenteritis [22, 23]. The large number of available genomes in this group provides a good model to study the relationship among CRISPR-Cas systems and HGT at the population level.

In this study, we determined the prevalence, diversity, and phylogenetic distribution of CRISPR-Cas systems present within 1871 genomes from *B. cereus* group. Our results show that most of the strains analyzed did not contain any CRISPR-Cas system or harbor defective systems. Combined with large-scale comparative genomic analysis, we concluded that the CRISPR-Cas systems are barriers to HGT within this group of bacteria. Even more, we evaluated the impact of CRISPR-Cas systems at the population level, showing that in *B. cereus* group the inactivation of such systems correlates with MGEs acquisition that could provide genetic traits for better adaptation to diverse environments.

Materials and methods

Identification of CRISPR-Cas systems among *B. cereus* group

The genome sequences of the 1871 strains from *B. cereus* group including their chromosome and plasmids sequences were retrieved from GenBank (ftp://ftp.ncbi.nlm.nih.gov/genomes). Cas proteins prediction was performed through BLASTP using genome-wide amino acid sequences of the 1871 strains against known Cas proteins, with an identity threshold of 30% and dual coverage over 60% [24]. The predicted Cas proteins and spacers were verified with CRISPR-finder [25]. Classification of CRISPR-Cas systems was based on the *cas* gene organization as previously reported [6, 26]. The information of the 260 *B. cereus* group strains found to contain CRISPR-Cas systems are listed in Table S1.

Phylogenetic analysis

The sequence of Cas7 proteins were clustered by using the MCL algorithm with an inflation value of 1.5, after an all-against-all BLASTP search with an E value of $<10^5$ [27]. The single-copy core protein sequences of each cluster were aligned with MUSCLE [28]. All alignments were concatenated using an in-house Perl script. ML phylogenetic tree was constructed using FastTree software [29] with bootstrap support values calculated from 1000 replicates.

Analysis of environmental niches distribution

Information of the original ecological niches of 423 selected strains from the original 1871 strains was collected from NCBI (https://www.ncbi.nlm.nih.gov/) or from published papers (Table S2). We included a total of 423 strains which information on niche isolation was available. The niches

include targets, hosts and/or sources, which were classified into six major groups, including: the animal-associated group, that is composed by fish, insects, mouse, nematode and earthworm subgroups; the human-associated group, that is composed by human and human-excrement subgroups; the plant-associated group, that is composed by plant, plant core, plant leaf, plant rhizosphere and plant root subgroups; the fungi-associated group; the soil-associated group; and others group, which includes food, industrial product and waste water subgroups.

Prediction of transposases, plasmid essential replication proteins, and prophage proteins

The protein sequences of the 423 selected stains described above were constructed as target sequence database. A complete profile of transposases downloaded from TnpPred [30] was searched against this database by using HMMER [31] with sequence and domain threshold E values of $<10^5$. The predicted transposases are listed in Table S3. Replication essential proteins [32] were used to predict plasmids (TXT S1). TubZ model (TXT S2) is obtained from the Pfam database [33], and the TubZ protein sequences were obtained with the same HMMER search command as transposase prediction. The other sequences of essential plasmid replication proteins were obtained using BLASTP with an identity threshold above 30% and a dual coverage over 60%. All the essential plasmid replication proteins predicted are listed in Table S4. Prophage elements were predicted by PHASTER [34] and are listed in Table S5.

Unique gene diversity calculations

The presence and absence of genes in the 423 selected genomes described above was determined by using the Roary software [27]. The genes shared by any two strains refer to the core genes, the sum of genes contained by both strains refers the total genes, and the difference of core genes and total genes refers to the unique genes. The ratio of number of unique genes against total genes of each genome pair refers the unique gene diversity value. The unique gene diversity values were calculated in groups of strains with complete, incomplete and none CRISPR-Cas systems that are listed in Table S6.

Growth conditions for bacterial strains and nematode *Caenorhabditis elegans*

Bacterial strains and plasmids used in this work are listed in Table S7. All *B. cereus* group strains were grown at 28 °C in Luria-Bertani (LB) medium and the competent cells were transformed by electroporation [35]. *Escherichia coli* strains

were grown at 37 °C in LB medium. The nematode *C. elegans* Bristol N2 strain was grown at 20 °C as previously described [36].

Biofilm formation assays

Biofilm formation was evaluated in 96-well plates as described before [37].

Determination of sporulation rate

Overnight bacterial cultures (100 μl, $OD_{600} = 0.3$) were spread on solid LB-agar plates and allowed them to sporulate for 5 days at 28 °C. The spores were harvested and suspended in 1.5 mL ice-cold 0.85% NaCl, and two aliquots of 500 μL of spore suspensions were prepared. The aliquot "*a*" was heated at 80 °C for 10 min to kill the vegetative cells and then kept on ice for 10 min; while the aliquot "*b*" was just kept on ice. Serial tenfold dilutions were prepared and 100 μl of each dilution was plated onto LB-agar plates. Growing colonies were counted after 24 h of incubation, and the sporulation rates were calculated as the percentage of colony forming units (CFU) of aliquot "*a*" (CFU-*a*) divide by the CFU of aliquot "*b*" (CFU-*b*) for each strain.

Salt tolerance assay

Overnight bacterial cultures in LB medium (100 μl, $OD_{600} = 0.3$) were spread on solid LB-agar medium supplemented with 1, 2, 4, or 6% NaCl. The bacterial colonies were counted after 12 h, and the bacterial surviving rates on these media were calculated using the 1% NaCl medium as reference.

pH tolerance assay

One milliliter sample of the bacterial cultures in LB medium ($OD_{600} = 0.6$) were collected. The pellet was washed with sterilized ddH$_2$O three times and suspended in 1 mL sterilized ddH$_2$O adjusted at pH 4.0–10.0, followed by incubation for 30 min with agitation at 220 rpm. The buffers used to establish the optimum pH and to maintain pH stability were as follows: sodium acetate buffer (0.2 mol/L NaAc and 0.2 mol/L HAc, pH 4.0–5.0), phosphate buffer (0.06 mol/L Na$_2$HPO$_4$ and 0.06 mol/L KH$_2$PO$_4$, pH 6.0–8.0), and Gly-NaOH buffer (0.2 mol/L glycine and 0.2 mol/L NaOH, pH 9.0–10.0) [38]. The pellets were washed with sterilized ddH$_2$O three times, suspended in 0.1 mL sterilized ddH$_2$O again, and spread on the solid LB-agar medium. The bacterial colonies were counted after 12 h incubation. The surviving rates of bacteria were calculated, using suspension incubated at pH 7.0 as reference.

In vivo colonization assays

The in vivo colonization assays for *B. cereus* group strains were conducted as described before [39]. A dose of 10^3 spores from each strain was used to feed *C. elegans* L4 worms in 96-well plates.

Life-span assays of nematode *C. elegans*

A total of 50–100 L4 stage N2 worms were transferred to fresh NG plates spread with 30 µl ($OD_{600} = 0.3$) spore/crystals mixtures of *B. thuringiensis* strains, with 25 µM FUDR to prevent eggs from hatching. NG plates spread with *E. coli* strain OP50 were used as controls. The worms were scored for alive/dead every 12 h at 20 °C until 96 h. The survival fractions were calculated as the number of living nematodes to that of all nematodes added in each plate, and survival curves were compared using *t*-test with statistical significance set at $p < 0.05$.

Results

CRISPR locus analysis and its distribution among *B. cereus* group strains

To determine the distribution and diversity of CRISPR-Cas systems in *B. cereus* group, we retrieved from GenBank all the 1871 whole-genome sequences available from this group. The search of CRISPR-Cas elements in the available genomes revealed that only 260 strains (*ca*. 13.9%) contained type I CRISPR-Cas system related elements, while only one strain of *B. cytotoxicus*, AFSSA_08CEB44bac, contained type II as revealed by the presence of a Cas9-like protein (Table S1). Interestingly, no other type systems were found among these strains and most of the strains possessed only one CRISPR-Cas structure, with the exception of *B. cytotoxicus* CH_23 (Table S1). The occurrence of CRISPR-Cas system varied greatly among the ecotypes of this group (Table 1). The

highest proportion was found in *B. cytotoxicus* where 86% of the analyzed genomes (14 genomes available) contain CRISPER-Cas elements, while only one *B. anthracis* strain from a total of 188 *B. anthracis* genomes analyzed was found to contain CRISPR-Cas elements. In the case of the *B. cereus* strains, 9% of 862 genomes analyzed showed CRISPER-Cas elements, while regarding to the *B. thuringiensis* strains, we found that 26% of strains have CRISPR-Cas elements from a total of 445 genomes analyzed (Table 1). Interestingly, only 30 of the 252 CRISPR-Cas system loci (11%) (8 of them could not be classified) were located on the chromosome, while the majority (88%) and were plasmid-borne (Table S1), suggesting that these CRISPR-Cas systems were mainly acquired by HGT.

Based on the signature genes and gene arrangements, two different intact subtypes I were identified (subtypes I-B and I-C), and one variant of subtype I-B system was also found (named subtype I-B-like) (Table S1 and Fig. 1a). In subtype I-B like, the *cas8* is replaced by *casX* encoding a hypothetical protein (Fig. 1a). Among these complete systems, subtype I-C was the most abundant, which accounted for 18.85% of all identified systems (Fig. 1b), and it was found in 49 strains across five ecotypes of bacteria (Table S1, Fig. 1b). The subtype I-B system was only present in 27 strains, while the I-B-like system was found in 20 strains. The incomplete systems accounted for 63.08% of all identified systems (Fig. 1b and detailed information in Table S1).

The number of spacers varied from 10 to 44 for subtype I-C, 0 to 32 for subtype I-B, and 10 to 34 for I-B like system (Table S1). A total of 4788 spacers were identified covering 1916 unique types (Dataset 1). However, only 701 out of the 1916 unique types of spacer have targets in the GenBank nucleotide sequence database (Dataset 2). Among them, 436 matched with sequences of plasmids, 57 with bacteriophage genes, and 288 with chromosome sequences (Dataset 2 and Fig. S1). Most of the spacers have sequence similarity with plasmids and phages, consistent with the CRISPR-Cas function as a protective immune system against acquiring foreign nucleic acids. The high number of

Table 1 Distribution of the CRISPR-Cas system types among *B. cereus* group members.

Taxonomic group	Genomes analyzed	Type I system	Type II system	Type III system	Type IV system	Type V system	Type VI system	Occurrence of type I system
B. anthracis	188	1 (0+1)	0	0	0	0	0	0.5%
B. cereus	862	75 (22+53)	0	0	0	0	0	8.7%
B. cytotoxicus	14	12 (8+4)	1 (0+1)	0	0	0	0	85.7%
B. mycoides	57	5 (1+4)	0	0	0	0	0	8.8%
B. pseudomycoides	104	29 (29+0)	0	0	0	0	0	27.9%
B. thuringiensis	445	116 (14+102)	0	0	0	0	0	26.1%
B. toyonensis	201	22 (2+20)	0	0	0	0	0	10.9%
Total	**1,871**	**260 (76+184)**	**1 (0+1)**	**0**	**0**	**0**	**0**	**13.9%**

The labels in red and blue in column of "Type I system" represent number of complete and incomplete CRISPR-Cas systems, respectively.
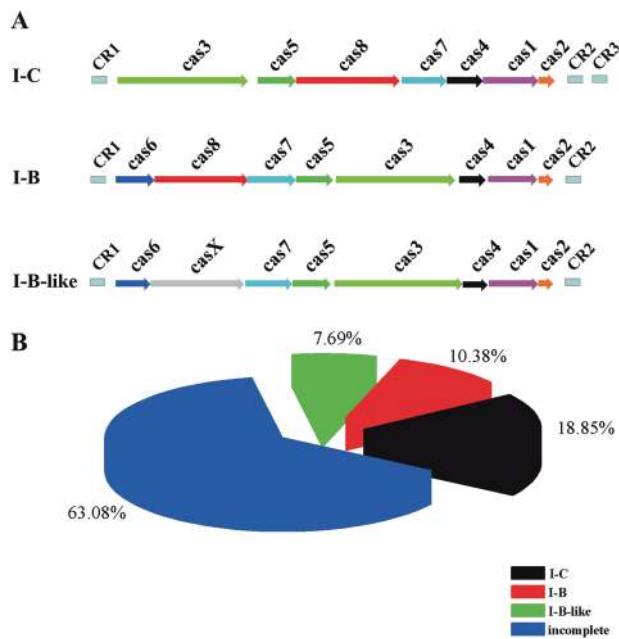
Fig. 1 Analysis of CRISPR-Cas systems in the *B. cereus* group. a Genetic organization of type I CRISPR-Cas systems found in *B. cereus* group. Only the representative complete subtypes are shown. **b** Proportion of *B. cereus* group strains with complete or incomplete CRISPR-Cas systems. Subtypes I-B, I-B-like and I-C are complete systems, and all systems, which lack any *cas* genes compared to the complete systems are defined as incomplete.

spacers that matched with chromosome sequences is very interesting. Some of the targets in the chromosome seemed to be conserved, and the three most common targets are collagen-like protein (30 hits), MoxR family ATPase (18 hits), and cell division protein FtsK (15 hits) (Dataset 2 and Table S8). These 288 unique spacers that matched chromosome sequences, include 34 self-targeting spacers (Fig. S2 and Dataset 1), which may raise an issue of autoimmunity in the host. The incorporation of self-chromosomal DNA into the CRISPR-Cas systems has been documented in bacteria and it was proposed that this may induce autoimmune responses resulting in high fitness cost [40]. Thus, it was proposed that the autoimmune responses could explain the high frequency and abundance of incomplete CRISPR-Cas systems across some prokaryotes [40]. Moreover, there are 239 intra-species targeting and 85 inter-species targeting spacers (Fig. S2 and Dataset 1). It has been reported that the inter-species targeting spacers have a negative effect on mating success in *Archaea* and may limit inter-species HGT [41].

Most CRISPR-Cas systems display a genetic inactivation phenomenon

Sixty-three percent (164 of 260) of the identified CRISPR-Cas systems were considered as not functional based on the

different gene deletions observed in the CRISPR-Cas systems of *B. cereus* group (Table S1). These defective systems derived from the intact subtypes I-C or I-B, displaying deletion of certain *cas* genes (Fig. 2a). The exception was the defective type II system that was identified in *B. cytotoxicus* strain AFSSA_08CEB44bac (Table S1). Based on the Cas protein composition and organization, these 164 defective systems were divided into five variants: I-C d1, I-C d2, I-C d3, I-C d4 derived from intact subtype I-C, and I-B d1 derived from intact subtype I-B or I-B like (Fig. 2a and Table S1). Specifically, in variant I-C d1 (identified in 47 strains), the *cas4* and *cas1* genes of intact subtype I-C were replaced by two *csd2* genes and one regulator *albA* gene, that encode proteins, which have different functions than Cas4 and Cas1. We performed BLASTP analysis of these *csd2* and *albA* genes finding that *csd2* is related to a CRISPR-associated protein, with high similarity to Cas7 a backbone protein that forms nucleic acid cleavage complexes. While the *albA* gene encodes a transcriptional regulator, containing a putative DNA-binding domain (AlbA_2), whose function is unknown. Variant I-C d2 (identified in 87 strains) lacks *cas1*, *cas2*, and *cas4* genes. Variant I-C d3 (identified in only one strain) showed a truncated CRISPR-Cas system lacking *cas3*, *cas5*, and the N-terminal end of *cas8* compared with the intact subtype I-C. Variant I-C d4 (identified in 24 strains) lost most of the elements and only has *cas5*, *cas8*, and *cas7* genes. Similarly, variant I-B d1 (identified in five strains) showed only *cas3*, *cas5*, and *cas7* genes with deletions of CRISPR-Cas system elements at both upstream and downstream regions of the intact subtype I-B locus (Fig. 2a).

To understand the pathways that may have lead to these defective systems, a phylogenetic tree was constructed based on Cas7 protein sequence, since this protein is present in all defective variants and intact systems. As shown in Fig. 2b, the systems are grouped in two major clades, one containing all subtype I-C and I-C variants while the other contains all subtype I-B, I-B like, and I-B variants (detailed shown in Fig. S3). Moreover, inside each major clade, with few exceptions, most systems in the same subtype or same variants formed separate branches. These results strongly suggest that the defective CRISPR-Cas systems evolved from intact systems, with a lineage relationship respective to the degeneration pathways of I-C and I-B subtypes (Fig. 2a).

Together, the Cas protein organization profiles and phylogenetic relationship analysis, suggest that the degeneration of CRISPR-Cas systems in the *B. cereus* group may have followed two main inactivation pathways (Fig. 2c): in the inactivation pathway for subtype I-C variants (Fig. 2c), the *cas1* and *cas4* genes, which are responsible for acquiring new spacers, were replaced by *csd2* and *albA* genes (variant I-C d1). In a subsequent event, variant I-C d2 was formed, where the *csd2* and *albA* genes together with *cas2* were lost. Finally,
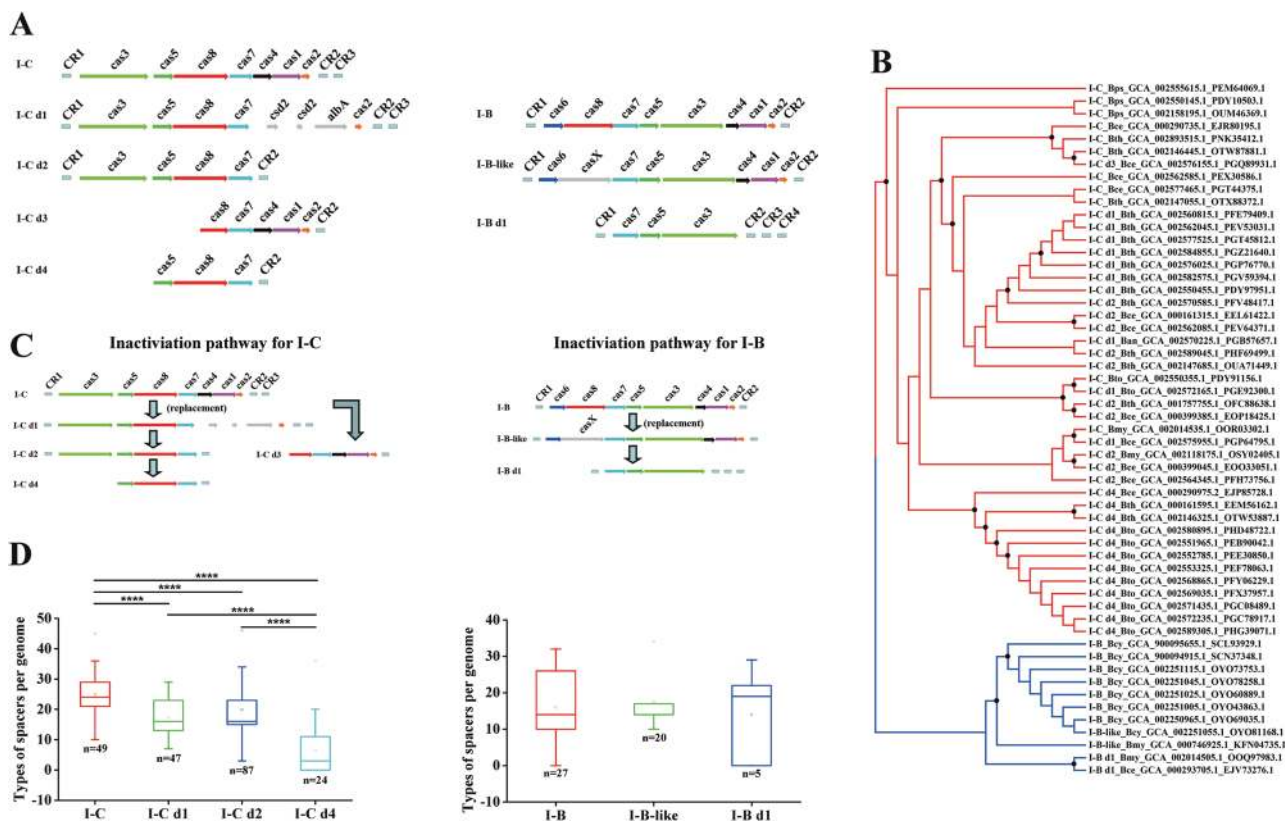
Fig. 2 Genetic inactivation of CRISPR-Cas system commonly occurs in B. cereus group. a Comparison of the genetic organization of variants of type I CRISPR-Cas systems. All the sequences were retrieved from NCBI and their accession numbers are listed in Table S1. Only representative strains are shown. **b** The maximum-likelihood (ML) phylogenetic tree based on the sequence of Cas7 protein was constructed by FastTree. The strains with intact or variants of subtypes I-C systems are shown in red branches, while the intact or variants of subtypes I-B systems are shown in blue branches. Nodes supported with a bootstrap values ≥50% are indicated with a black dot. Ecotypes names of *B. cereus* group are abbreviated as follows: Ban for *B. anthracis*, Bce for *B. cereus*, Bcy for *B. cytotoxicus*, Bmy for *B. mycoides*, Bps for *B. pseudomycoides*, Bth for *B. thuringiensis*, and Bto for *B. toyonensis*. The numbers, such as GCA_002560815.1 and PFE79409.1 are the GenBank assembly accessions of genomes and GenBank accessions of Cas7 proteins, respectively. **c** Proposed inactivation pathways of type I CRISPR-Cas system in *B. cereus* group. The predicted inactivation pathways for subtypes I-C or subtypes I-B variants are shown. **d** Differences of spacer types among strains with variants of CRISPR-Cas system in the *B. cereus* group. The spacers of variant I-C d3 were excluded because only one strain contains this variant. All the spacers are listed in Dataset 1.

the *cas3* gene was deleted leading to variant I-C d4 (Fig. 2c). In the case of variant I-C d3, *cas3*, *cas5*, and *cas8* genes were deleted (Fig. 2c). This variant may be an accidental event, since it was observed in only one strain. In the inactivation pathway for subtype I-B variants (Fig. 2c), replacement of *cas8* gene by *casX* resulted in subtype I-B-like, then a directly derived CRISPR-Cas system (variant I-B d1) evolved, where *cas6*, *cas8*, *cas4*, *cas1*, and *cas2* were deleted without further intermediate steps (Fig. 2c). Interestingly, in variant I-B d1, the *cas* genes responsible for acquiring new spacers were also lost. Overall, inactivation pathway for subtype I-C variants may represent the major degeneration route of CRISPR-Cas systems in *B. cereus* group, since this was observed in 208 of the 260 strains that contain CRISPR-Cas systems (Table S1).

Spacer acquisition is one of the most important functions of CRISPR-Cas systems and the spacer diversity in CRISPR array reflects the interaction frequency and course between bacterial host and invasive nucleic acids. In *B. cereus* group,

there are more spacers and higher diversity of spacer compositions in strains with intact subtype I-C system compared with those containing the I-C defective systems (Figs. 2d, S4 and Table S1). In particular, the types of spacers diminished gradually from subtype I-C to variants I-C d1–d2 and then to variant I-C d4, in accordance with the proposed deletion pathway of *cas* genes in this subtype group (Fig. 2d).

Strains with intact CRISPR-Cas system contain restricted niche distribution, less MGEs, and less variable genes

The CRISPR-Cas system has been proposed as barriers to HGT preventing acquisition of MGEs [4]. Therefore, one possibility is that intact CRISPR-Cas system may prevent acquisition of advantageous exogenous genes to adapt to complex environments. To confirm this hypothesis, 96 genomes that contain intact CRISPR-Cas systems, 164

genomes that contain defective variant systems, and 163 genomes without any CRISPR-Cas elements (randomly selected from the 1871 genomes reported in GenBank) were analyzed regarding their content of DNA obtained by HGT.

We first analyzed the distribution of strains containing complete or incomplete systems at the population level. A maximum-likelihood (ML) phylogenetic tree based on the core genome of the above strains was constructed by FastTree. Phylogenomic analysis divided these strains into six clades (Fig. 3), which are in accordance with the population structure previously described for the *B. cereus* group [42]. In addition, *B. cytotoxicus* and *B. pseudomycoides* were included into the new clades 5 and 6, respectively. Interestingly, the intact CRISPR-Cas systems can be found in all six clades, while the incomplete systems are only distributed in clades 1, 2, 3, and 4. However, in each of these clades, the distribution of strains with incomplete systems is more widely represented than that of strains containing intact systems (Fig. 3). These data indicate that the defective CRISPR-Cas systems have more widely distribution than that of the intact systems at the population level in the *B. cereus* group. Even more, strains that contain same CRISPR-Cas variant do not form separate branches in the population structure, but are scattered in each clade, indicating that inactivation of CRISPR-Cas systems are related to HGT, rather than inherited from parental lineages.

Strains of *B. cereus* group can be isolated from diverse environments as pathogenic or nonpathogenic bacteria [43, 44], making this group a powerful system for investigating the role of CRISPR-Cas systems in niche adaptability. The environmental distribution of strains was obtained for those strains where information was available regarding their isolation source (see "Materials and methods"). At least 21 kinds of different environments, such as human, animal, plant, fungi, soil, and others were identified (Fig. 4a). Interestingly, strains with incomplete and without CRISPR-Cas systems occupied much more diverse environments than strains with intact CRISPR-Cas systems (Fig. 4b), indicating that the CRISPR-Cas systems may limit their environmental distribution.

HGT plays important roles in bacterial adaptation to different environments [45]. Inactivation of CRISPR-Cas systems could explain HGT differences of the *B. cereus* group. To determine the capacity of HGT, we analyzed the content of MGEs, including putative transposases, replication proteins of plasmids, and prophage elements in the different strains. The results showed that the content of putative transposases (Fig. 5a and Table S3), plasmid replication proteins (Fig. 5b and Table S4), and prophage elements (Fig. 5c and Table S5) in strains with intact CRISPR-Cas systems were significantly lower than those of strains with incomplete or lacking CRISPR-Cas systems.

Some genes, such as virulence factors are carried by mobile elements. Furthermore, genes may be able to persist in MGEs through hitch-hiking with beneficial genes or alleles ensuring high frequencies via selection of MGEs [46] and transfer between organisms [47]. To evaluate the direct effects of CRISPR-Cas system on HGT, we analyzed the variation of unique genes of strains with complete, incomplete, and none CRISPR-Cas systems in the *B. cereus* group. These analyses showed that variation of unique genes is significantly lower in the strains with complete CRISPR-Cas systems ($n = 93$) than in strains with incomplete ($n = 164$, $p < 0.0001$) or none CRISPR-Cas system groups ($n = 135$, $p < 0.05$) (Fig. 5d). Together, our results suggest that the CRISPR-Cas system reduces HGT frequency within *B. cereus* group.

Active CRISPR-Cas system displays negative impact on the adaptation of *B. thuringiensis* under extreme conditions

To further test the hypothesis that bacteria with active CRISPR-Cas system may reduce their environmental adaptation, a complete subtype I-C system form *B. cereus* strain VD115 (Figs. 6a, S5) was cloned (Supplementary Text, Fig. S6). We selected to clone subtype I-C from VD115 strain to examine its function since this is a functional system (Supplementary Text, Fig. S7) and also because inactivation pathway for subtype I-C variants represents the major inactivation pathway of CRISPR-Cas systems in the *B. cereus* group (Fig. 2). We confirmed that the PAM site of VD115 subtype I-C system is TTN (Fig. 6b). We constructed derivatives vectors of pHT-304 containing synthetic VD115 spacer 1 (Table S9) with upstream 5′-TTN or four random 5′-NNN PAMs, and then transformed these vectors into strain VD115. The transformation efficiency of plasmids containing spacer 1 accompanied by 5′-TTN were significantly lower than those accompanied by non-5′-TTN (Fig. 6b), confirming that the PAM site of VD115 subtype I-C system is TTN.

Then the effect of the active CRISPR-Cas system from *B. cereus* strain VD115 on the adaptation to different environments was evaluated when transformed into a wild-type *B. thuringiensis* strain YBT-1518 lacking CRISPR-Cas resulting in BMB1685 strain. We selected to work with *B. thuringiensis* YBT-1518 strain because it lacks CRISPR-Cas systems and also because it is toxic to the nematode *C. elegans* [39] allowing the analysis of CRISPR-Cas on pathogenicity. We first analyzed the functionality of CRISPR-Cas system when expressed in *B. thuringiensis* BMB1685 strain. To detect if this CRISPR locus protects from exogenous DNA, we introduced synthetic spacers 1, 8, 10, 13, and 14 with the 5′-TTC PAM sites into the vector pHT-304 (Table S10). Then, these constructed plasmids and
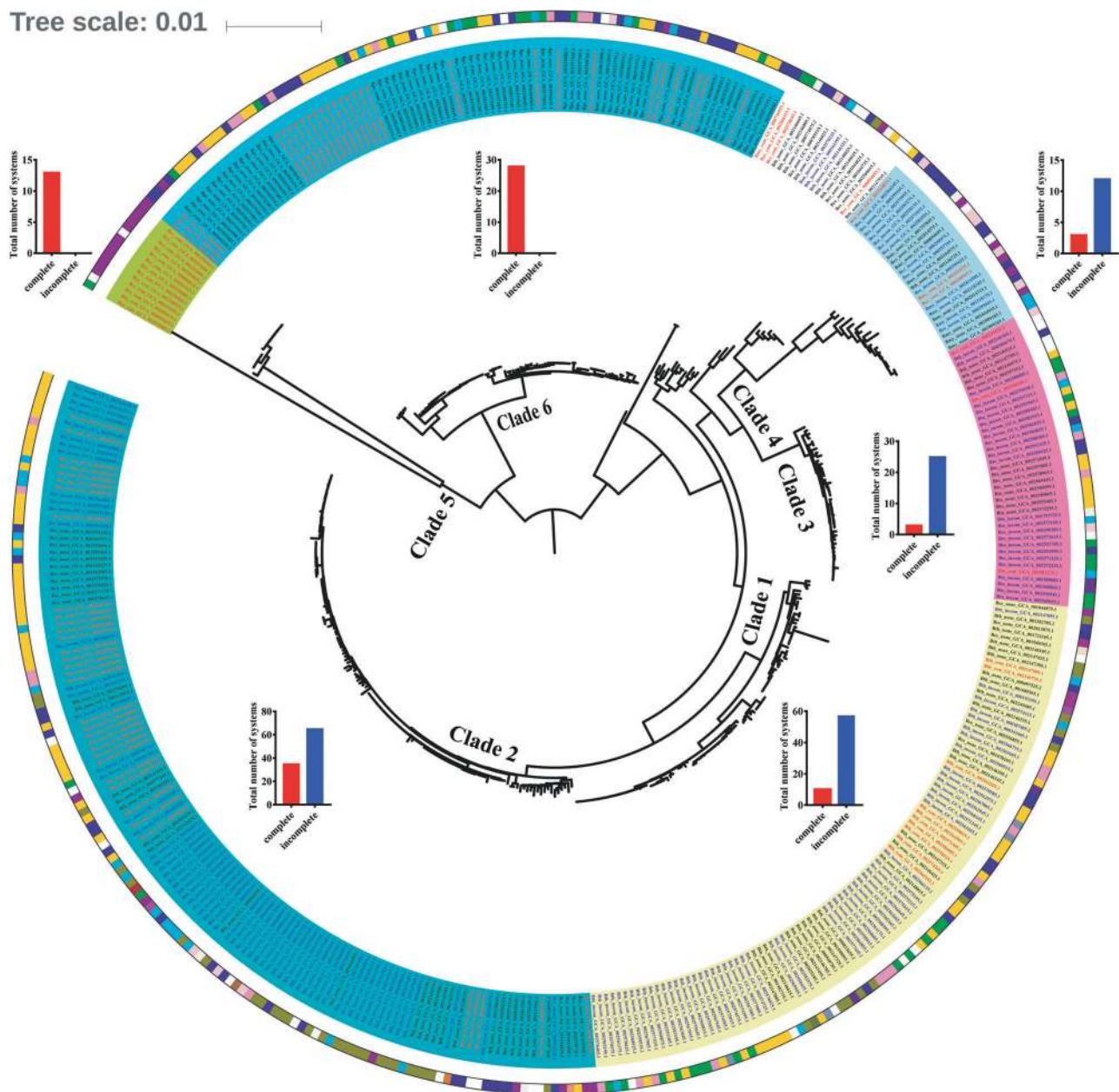
Fig. 3 The distribution of *B. cereus* group strains with active- or inactive-CRISPR-Cas systems on population structure. The ML tree was constructed by FastTree based on the protein sequences of single-copy core genes. Bootstrap support values were calculated from 1000 replicates. Strain names were abbreviated as described in Fig. 2. The background colors of the labels represent the clades: yellow, green, fuchsia, light green, olive, and dark green for clades 1, 2, 3, 4, 5, and 6, respectively. The colors of labels represent states of CRISPR-Cas system: red, blue, and black for complete, incomplete and none, respectively. Colors of the circle next to the labels represent the bacterial niches or hosts as follow: olive drab for earthworm Haplotaxida, peach puff for fish, indigo for Coleoptera, chocolate for Diptera, sienna for Homoptera, olive for Lepidoptera, coral for mouse, dark violet for nematode, dark orchid for nematode and insect, pink for human, lavender blush for human excrement, cyan for plant, brown yellow for plant core, green for plant leaf, light yellow for plant rhizosphere, violet for plant root, blue for soil, purple for food, fire brick for industrial product, navy for waste water, and white for unknown information. Inserts show column charts showing total numbers of the complete (red bars) and incomplete systems (blue bars) in each Clade.

the pHT-304 vector, were transformed into strain VD115 and recombinant BMB1685. The results showed that the CRISPR locus in both strains reduced the transformation efficiency of plasmids containing spacers flanked by 5′-TTC (Fig. 6c, d).

The genome editing function of the cloned subtype I-C system in BMB1685 was also analyzed. The nematicidal *cry6Aa* gene [48], which codes for a 54-kDa protein active against nematodes, was selected as target. For this purpose, a plasmid containing a small guide RNA

Fig. 4 The environmental niches distribution of *B. cereus* group strains with active- or inactive-CRISPR-Cas systems. a Total number of strains adapted to the different main types of ecological niches. The strains are grouped by complete, incomplete, and no type I CRISPR-Cas systems. **b** Comparison of the ratio of occupied ecological niches for strains with complete, incomplete, or no CRISPR-Cas systems. The data are shown as proportion of the total niches colonized by all strains in each group.
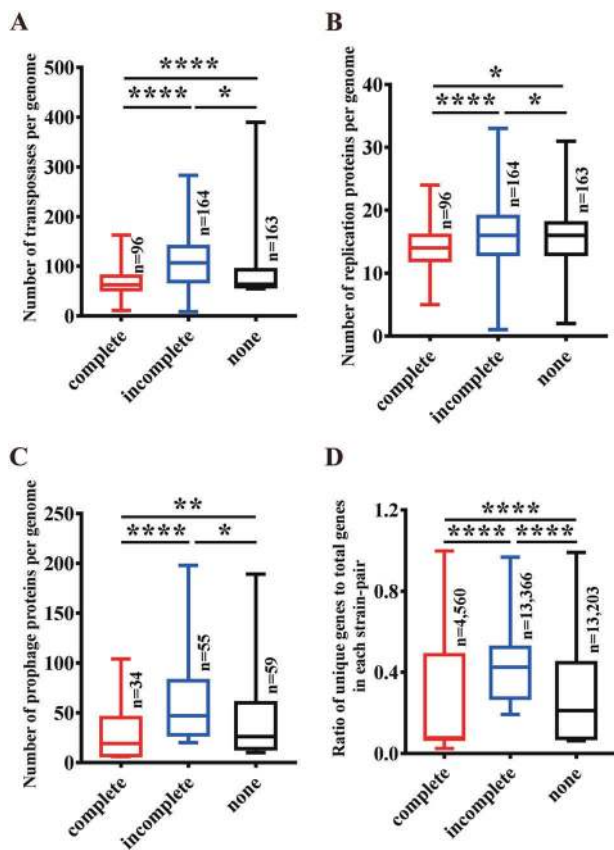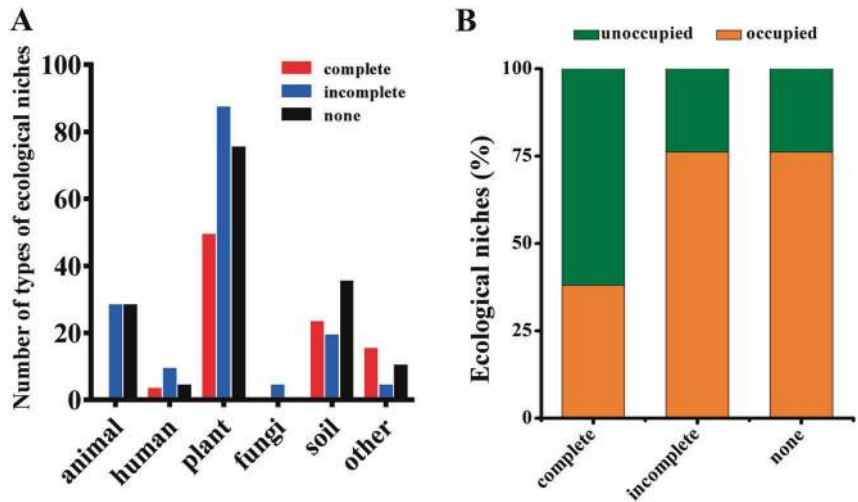
Fig. 5 Distribution of mobile genetic elements and unique genes among selected *B. cereus* group strains. The MGEs including transposases (**a**), plasmid replication essential proteins (**b**), and prophage proteins (**c**) contents among selected strains with complete, incomplete or no CRISPR-Cas systems are shown. The contents were compared by calculating the numbers of each kind of MGEs per genome of each group. **d** The comparison of unique gene diversity value among *B. cereus* group strains. Unique gene diversity was compared by calculating the rate of unique genes per genome for all the selected genomes. The significance of the differences among samples was evaluated using two sample Wilcoxon–Mann–Whitney test at, $p < 0.0001$ (****) $p < 0.001$ (***), $p < 0.01$ (**) and $p < 0.05$ (*).

(sgRNA) for *cry6Aa* gene was constructed as described in Supplementary text (Fig. S8A) and introduced into BMB1685 strain. Sixty-nine colonies were tested by PCR, and 98.5% of the colonies displayed a smaller band (568 bp) in agarose gel electrophoresis that represents the deletion in *cry6Aa* gene (Fig. S8B). The sequence of the PCR products confirmed the 198 bp deletion on *cry6Aa* gene as designed (Fig. 6e). Finally, analysis of the expression of Cry6Aa protein showed that this 54 kDa protein band was not detected in SDS-PAGE in the strains with the 198 bp deletion (Fig. 6f). Taking together, above results showed that cloned VD115 subtype I-C CRISPR-Cas system is functional in the *B. thuringiensis* transformant BMB1685 strain.

In order to analyze the effect of CRISPR-Cas system expression in the environmental adaptation of *B. thuringiensis* BMB1685 strain, we first analyzed bacterial growth in LB medium. As shown in Fig. 7a, the growth curves no differences on growth during their logarithmic growth phase were observed among YBT-1518, control BMB1684, and BMB1685 containing the complete subtype I-C system. However, after 20 h of growth, BMB1685 showed significant lower biomass till the end of the experiment. These results suggest that subtype I-C system decreases the growth of bacteria during stationary phase. In addition, the sporulation rate of BMB1685 showed a significant decrease compared with that of the *B. thuringiensis* YBT-1518 and control BMB1684 strains (Fig. 7b).

Biofilm is a community of surface-associated bacteria, which is generally composed of exopolysaccharides, proteins, and some nucleic acids to protect bacteria from severe environmental conditions [49]. To detect the effect of CRISPR-Cas system on biofilm formation, the biofilm formation ability of the *B. thuringiensis* BMB1685, YBT-1518, and BMB1684 strains was compared. On the basis of the OD_{595} of the soluble dye, a significant lower
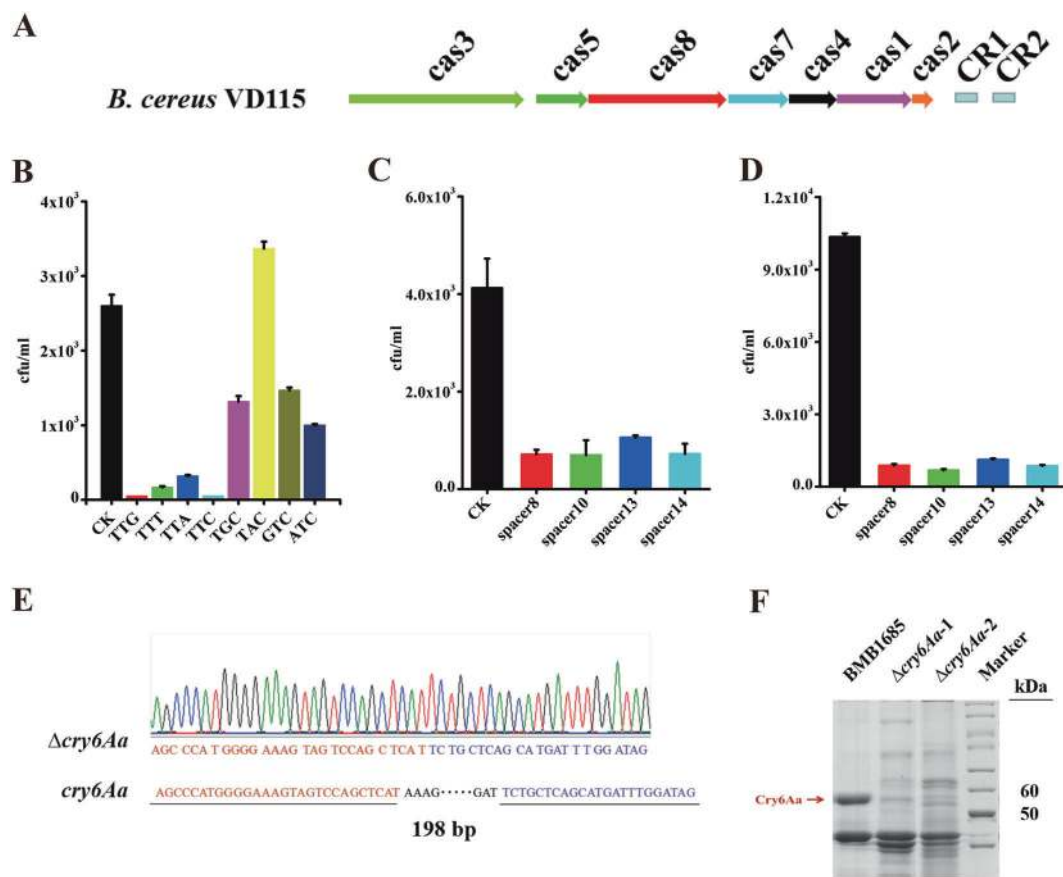
Fig. 6 Functional analysis of the subtype I-C CRISPR-Cas system from *B. cereus* strain VD115. a The genetic organization of *cas* operon in subtype I-C CRISPR-Cas system loci from *B. cereus* VD115. **b** Plasmids derivatives of pHT-304 containing synthetic VD115 spacer 1 accompanied by 5′-TTN or 5′-NNN PAMs (TGC, TAC, GTC and ATC) were transformed into *B. cereus* VD115 strain and the transformation efficiency of plasmids was analyzed and showed as the number of CFU/mL. The immunity assays of subtype I-C CRISPR-Cas locus in *B. cereus* VD115 (**c**) and in *B. thuringiensis* BMB1685 (**d**) strains. The immunity was assessed via transformation of plasmids containing spacers flanked by a 5′-TTC PAM site in both strains. The transformation was conducted via electroporation, and the results were expressed as the number of CFU/mL. The pHT-304 vector was used as negative control in all assays. **e** T-A cloning and sequencing demonstrated the deletion region of *cry6Aa* gene. A loss of 198 bps in the edited *cry6A* (Δ*cry6Aa*) is shown, compared with the wild type *cry6Aa* gene in BMB1685 strain. **f** The Cry6Aa protein production in BMB1685 strain and in selected *cry6Aa* deletion mutants was analyzed by SDS-PAGE. The red arrow points to the band of the complete Cry6Aa protein.

staining was found in the BMB1685 strain transformed with the complete subtype I-C system, indicating that CRISPR-Cas system negatively affects the biofilm formation ability (Figs. 7c and S9).

To determine the effect of CRISPR-Cas system on the tolerance to salt or pH, we tested the survival rate of cells grown on LB plates supplemented with different salt concentrations or at various pH, from pH 4 to pH 10. The results showed that bacteria without of CRISPR-Cas system grew better at 2% salt condition compared to the BMB1685 strain transformed with CRISPER-Cas system (Fig. 7d). Similarly, the growth of BMB1685 was not affected after incubation at pH 7.0 and pH 8.0 while cell survival of this strain decreased significantly after incubation at extreme acidic (pH 5.0 and pH 4.0) and extreme alkaline (pH 9.0 and pH 10.0) conditions, compared to YBT-1518 and BMB1684 strains (Fig. 7e). These data suggest that the CRISPR-Cas system negatively impacts bacteria tolerance to extreme stresses.

To confirm the above observations, we introduced the cloned VD115 CRISPR/Cas system into another model *B. thuringiensis* strain BMB171 that lacks CRISPR/Cas system. Similar to BMB1685 strain that was transformed with the complete subtype I-C system, (Fig. 7), the introduction of a functional subtype I-C system in BMB171 strain resulted in reduced bacteria growth, sporulation rate, biofilm formation ability, and adaptability to various stress, including acid, alkaline, and high salts tolerance (Fig. S10). We also confirmed by qPCR analysis that the introduction of CRISPR-Cas system leading to the observed phenotypes do not depend on the cost of wide defects on gene expression by analyzing the expression of two widely used
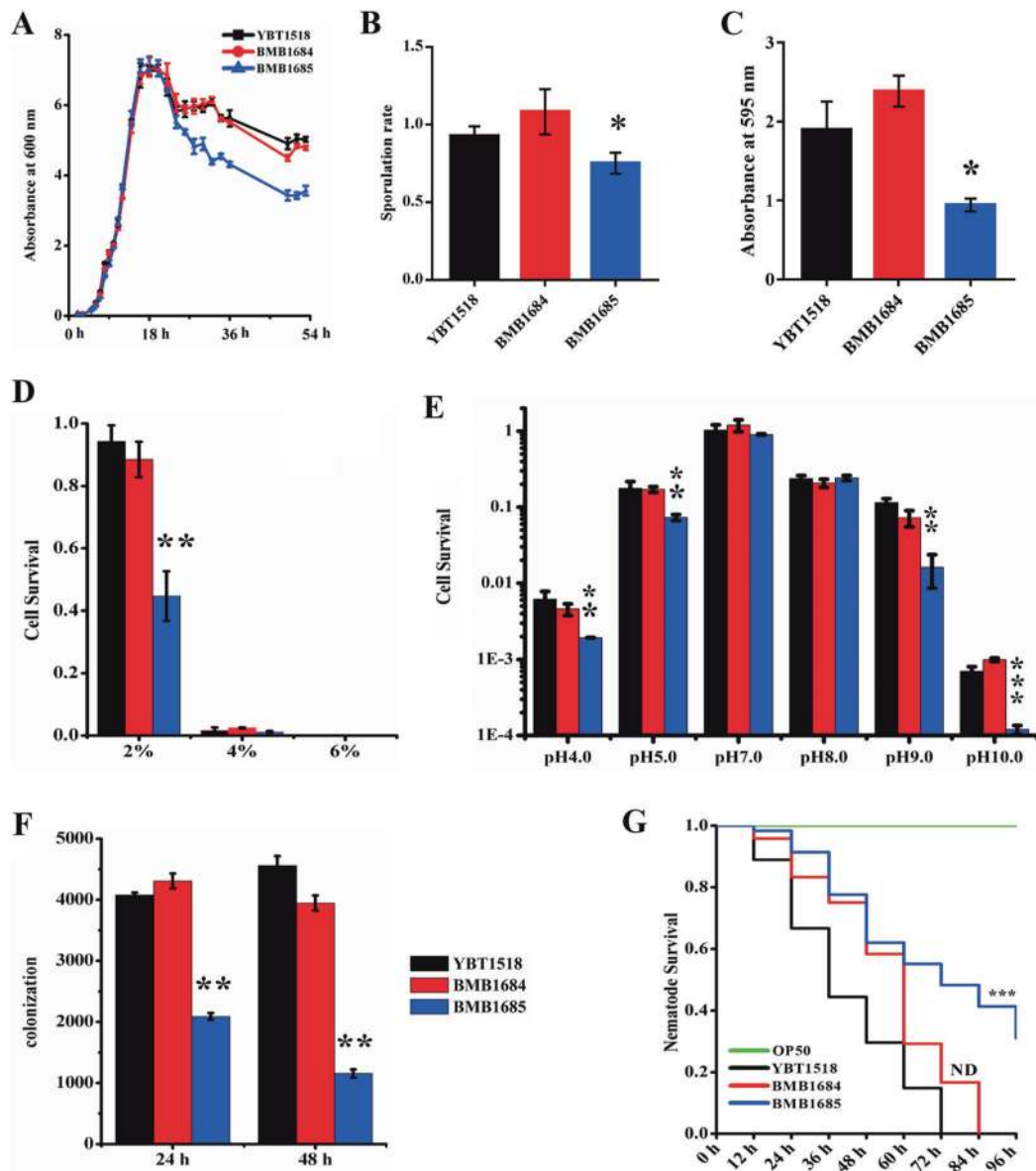
Fig. 7 Impact of active CRISPR-Cas system on the adaptation to different environmental stresses and into its nematode host. **a** Growth curve of *B. thuringiensis* transformed with subtype I-C CRISPR-Cas system (BMB1685) or same *B. thuringiensis* strain without transformation (YBT-1518) or transformed with empty vector (BMB1684). The optical absorbance value at 600 nm of these *B. thuringiensis* strains grown in LB medium was analyzed successively from 0 to 54 h. **b** Sporulation rate of bacteria. Sporulation rates were calculated by the CFU of spore-containing aliquots heated at 80 °C for 10 min divide by CFU of the aliquots that were kept on ice before growing in LB-agar plates for 24 h at 28 °C. **c** Biofilm formation ability of bacteria. The data are expressed as the absorbance at 595 nm of wells treated with different *B. thuringiensis* strains after staining. **d** Salt tolerance of bacteria. The data are expressed as survival rate of the different *B. thuringiensis* strains exposed to high-salt environments containing 2%, 4 and 6% NaCl. **e** pH tolerance of bacteria. The data are expressed as survival rate of *B. thuringiensis* strains exposed to different pH environments. In all assays, BMB1684 strain containing empty pBAC44 vector was used as control. **f** Colonization of the different *B. thuringiensis* strains in *C. elegans*. A total of 10^3 spores of each *B. thuringiensis* strain were added to 30–50 L4 stage N2 nematodes as the only food source. The colonies of *B. thuringiensis* were counted after analyzing the samples isolated from infected worms after 24 and 48 h. The BMB1684 strain containing empty pBAC44 vector was used as control. **g** Life-span assay of *C. elegans* feeding with *B. thuringiensis* with or without of subtype I-C CRISPR-Cas system. The survival rates of L4 stage N2 worms (50–100 for each well) fed with spore/crystal mixture of the different *B. thuringiensis* strains were calculated after different incubation times. The BMB1684 strain and standard food *E. coli* OP50 were used as controls. The error bars represent the standard deviations from the mean values of three independent experiments. The differences of recombinant strains against YBT-1518 wild type was evaluated using two-sample *t*-test. A double asterisk indicates $p < 0.01$, single asterisk indicates $p < 0.05$, lacking of any symbol and ND indicates no significant difference.

reference genes in *B. cereus* group strains, *gatB* and *rpoA* [50], showing that these two housekeeping genes have similar expression level at the same growth stage in strains with or without active CRISPR-Cas system (detailed information was shown in the Supplementary Text, Fig. S11). Thus, we concluded that the active CRISPR-Cas system increased the fitness cost of host, and displays negative impact on host adaptation.

Active CRISPR-Cas system interferes with *B. thuringiensis* pathogenicity

B. thuringiensis YBT-1518 is a nematicidal strain, which can complete its life cycle in *C. elegans* after colonization of the worm intestine [39]. To determine the effect of CRISPR-Cas system on the adaptation of YBT-1518 in its host, we analyzed the colonization capacity of BMB1685 strain expressing CRISPR Cas I-C system in *C. elegans* after 24 and 48 h post infections. The results showed that the number of *B. thuringiensis* colonies obtained from worms infected with of BMB1685 was significantly reduced in comparison with worms infected with either YBT-1518 or BMB1684 strains (Fig. 7f). Colonization of *B. thuringiensis* in the nematode intestine leads to nematode mortality. To analyze the effect of CRISPR-Cas system on the pathogenicity of *B. thuringiensis* BMB1685, we determined the survival of *C. elegans* fed with the different *B. thuringiensis* strains. The results show that nematodes fed with BMB1685 have higher survival rates than that those fed with YBT-1518 and BMB1684 strains in the mortality assays of *C. elegans* (Fig. 7g).

Discussion

The main conclusion of our work is that the CRISPR-Cas systems are barriers to HGT within *B. cereus* group of bacteria correlating with wider environmental distribution. In accordance, strains harboring active CRISPR-Cas systems have limited environmental distributions (Figs. 3, 4), less MGEs and lower number of unique genes in their genomes (Fig. 5), as compared with strains without active CRISPR-Cas systems.

We found that *B. cereus* group harbor only CRISPR-Cas systems subtypes I-C and I-B, which agrees with the trend observed in all bacteria genome sequences characterized showing that type I systems are the most frequently found in bacteria (60%) [6]. It is interesting to note that it was proposed that subtype I-C seems to be a derivative of subtype I-B [26]. Our data show that the distribution rate of CRISPR-Cas systems is largely different among ecotypes of *B. cereus* group. Also, that the distribution seems to depend on the environmental niches of certain strains (Fig. 3). For

example, *Bacillus* strains with complete CRISPR-Cas systems showed reduced distribution in different ecological niches (Fig. 4a), while many of the *B. thuringiensis*, *B. cereus* and *B. cytotoxicus* strains, which can be isolated from more diverse ecological niches, have incomplete CRISPER-Cas system or lack these systems (Fig. 4). CRISPR-Cas systems are also widely found in 87% Archaea and 45% bacterial genomes, most of which are bacterial pathogens (http://crispr.i2bc.paris-saclay.fr/) and are found in specific environmental niches. Thus, we proposed that Archaea and bacterial pathogens from single environmental niches may have larger distribution of complete functional CRISPR-Cas systems, while Archaea or bacteria from open or diverse environments may have less active systems. A reasonable explanation is that inactivation of the CRISPR-Cas systems allowed higher frequency of HGT to obtain selective genetic traits for better adaptability to diverse environments. This proposition is in agreement with the proposal that pangenomes resulted from an adaptive evolution, where HGT positively increases the possibility to acquire advantageous genes allowing prokaryotic species to migrate to new niches [51]. However, we cannot ignore the defense role of CRISPR-Cas systems in protecting the genome from phages and transposons invasion explaining that most spacer targets actually come from integrated elements such as phages and transposons as found in this work (Dataset 2 and Fig. S1) and previously reported [52]. The fact that the CRISPR–Cas systems in most of *B. cereus* group strains are often lost or inactivated (Figs. 1, 2), together with the data that strains harboring active systems have less MGEs (Fig. 5), strongly suggest that inactivation of CRISPR-Cas systems should be an evolution trade-off between individual self-protection and better adaptability of the *B. cereus* group population.

Another important observation was that the inactivation pathways of CRISPR-Cas systems in *B. cereus* group mainly involves the replacement or deletion of *cas* genes that are responsible for acquiring new spacers (Fig. 2). The spacer acquisition is important for the immune function of CRISPR-Cas systems, and the spacer diversity in CRISPR array reflects the interaction frequency between the bacteria and the invasive nucleic acids [1]. In accordance, more spacers and higher diversity of spacer compositions were found in strains with intact CRISPER-Cas systems compared with those lacking CRISPER-Cas systems or containing the incomplete ones (Fig. 2d). Losing the spacer acquisition ability but keeping other components may have some special biological significance for certain strains (164 of 260), since CRISPR-Cas systems have been shown to possess several other unexpected roles beyond its immune function [10]. However, the factors that determined such inactivation pathways of CRISPR-Cas systems in *B. cereus* group, and

the detailed mechanisms and significance involved remain unclear and need further studies. Other inactivation models of CRISPR-Cas systems include loss of nuclease components or repression of the expression of nucleases [14, 15] and attenuating the function by diverse anti-CRISPR proteins [53]. In addition, loss of CRISPR-Cas elements through recombination has been reported [54].

It has been shown that CRISPR-Cas systems limited antibiotic resistance in the pathogens *Klebsiella pneumoniae* [55] and *E. coli* [56]. Perhaps the benefit of acquiring new MGEs for better adaptation to new environments outweighed the costs of phage infection, and additional anti-phage systems may be present in these bacteria. BLASTP analysis and domain searching revealed that other anti-phages systems including restriction (R)-modification (M) systems, Druantia, and Gabija systems, etc [57–60], are widely distributed within *B. cereus* group strains (Table S11). These results suggest that inactivation of CRISPR-Cas systems resulted in enhanced HGT that could have detrimental effects due to enhanced infection with phages and transposons, which may be overcome by the acquisition of additional and different antiphage protection system (such as DISARM, Druantia, Gabija, Lamassu, Septu, Thoeris, or Zorya-I systems). In accordance with this hypothesis, the number of such antiphages systems is much higher in the strains that have incomplete CRISPR-Cas systems than those with a complete system (Fig. S12). These data may explain why the number of putative transposases, prophage elements, plasmids, and the number of unique genes per genome are significantly higher in the in strains with incomplete CRISPR-Cas than those strains lacking CRISPR-Cas (Fig. 5).

Thus, deletion of CRISPR-Cas systems may have co-evolved with the acquisition of alternative immune systems. We indeed showed that the functional CRISPR-Cas system in BMB1685 strain reduces the adaptation of this transformant bacterium under both host infection and environmental conditions (Fig. 7). We have shown that spacers identified could target multiple chromosomal genes as self-targeting, intra-species targeting, or inter-species targeting (Fig. S2 and Dataset 1). The self-targeting spacers may raise the issue of auto-immunity in the host. Thus, the observed effects in BMB1685 strain, may indicate activation of autoimmunity responses affecting the performance of BMB1685 on the different conditions analyzed. We do not know which genes were targets in these assays and further studies will help to better understand the disadvantages of expressing such CRISPR-Cas system in BMB1685. While the inter-species targeting spacers has proposed may limit inter-species HGT [41]. Therefore, the large numbers of intra-species targeting may limit the HGT among strains in the *B. cereus* group. Thus, the autoimmunity responses and intra-species targeting may explain why most *B. cereus*

group strains evolve to inactivate CRISPR-Cas systems. Our work puts forward a novel adaptive strategy employed by bacteria via inactivation of CRISPR-Cas systems by deletion of certain key components, reducing autoimmunity responses, and facilitating the acquisition of potentially beneficial MGEs.

As pathogens for insects and nematodes, *B. thuringiensis* produces diverse virulence factors [39, 61–65]. Production of multiple virulence factors benefits the populations by facilitating host infection, that represent more resources. Comparative genome analysis performed with a highly insecticidal *B. thuringiensis* HD-1 strain [66] as reference, revealed that *B. thuringiensis* strains with active CRISPR-Cas systems contain less virulence factors than strains without active CRISPR-Cas systems (Fig. S13). This information indicates that the CRISPR-Cas systems prevent the transfer of virulence genes within pathogens. The CRISPR-Cas systems in *S. pneumoniae* and *Streptococcus pyogenes* were also shown to prevent acquisition of virulence factors [11, 67]. Thus, the negative effect of CRISPR-Cas systems for pathogenicity and adaptability may be a common phenomenon, suggesting that the strategy of CRISPR-Cas systems inactivation could improve the adaptability of various bacterial pathogens to their hosts.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol. 2014;12:479–92.

2. Barrangou R, Marraffini LA. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol Cell. 2014;54:234–44.

3. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315:1709–12.

4. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science. 2008;322:1843–5.

5. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR–Cas systems. Nat Rev Microbiol. 2011;9:467–77.

6. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol. 2015;13:722–36.

7. Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, et al. Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol. 2017;15:169–82.

8. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 Is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 2015;163:759–71.

9. Barrangou R, Doudna JA. Applications of CRISPR technologies in research and beyond. Nat Biotechnol. 2016;34:933–41.

10. Faure G, Makarova KS, Koonin EV. CRISPR-Cas: complex functional networks and multiple roles beyond adaptive immunity. J Mol Biol. 2019;431:3–20.

11. Bikard D, Hatoum-Aslan A, Mucida D, Marraffini LA. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. Cell Host Microbe. 2012;12:177–86.

12. Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature. 2013;497:254–7.

13. Heidrich N, Hagmann A, Bauriedl S, Vogel J, Schoen C. The CRISPR/Cas system in Neisseria meningitidis affects bacterial adhesion to human nasopharyngeal epithelial cells. RNA Bio. 2019;16:390–6.

14. Palmer KL, Gilmore MS. Multidrug-resistant enterococci lack CRISPR-Cas. mBio. 2010;1:e00227-10.

15. Hullahalli K, Rodrigues M, Nguyen UT, Palmer K. An attenuated CRISPR-Cas system in Enterococcus faecalis permits DNA acquisition. mBio. 2018;9:e00414-18.

16. Gophna U, Kristensen DM, Wolf YI, Popa O, Drevet C, Koonin EV. No evidence of inhibition of horizontal gene transfer by CRISPR-Cas on evolutionary timescales. ISME J. 2015;9:2021–7.

17. Du Toit A. Viral infection: CRISPR-Cas enhances HGT by transduction. Nat Rev Microbiol. 2018;16:186.

18. Watson BNJ, Staals RHJ, Fineran PC. CRISPR-Cas-mediated phage resistance enhances horizontal gene transfer by transduction. mBio. 2018;9:e02406-17.

19. Varble A, Meaden S, Barrangou R, Westra ER, Marraffini LA. Recombination between phages and CRISPR-cas loci facilitates horizontal gene transfer in Staphylococci. Nat Microbiol. 2019;4:956–63.

20. Cohan FM. Bacterial species and speciation. Syst Biol. 2001;50:513–24.

21. Schnepf E, Crickmore N, Van Rie J, Lereclus D, Baum J, Feitelson J, et al. Bacillus thuringiensis and its pesticidal crystal proteins. Microbiol Mol Biol Rev. 1998;62:775–806.

22. Kolsto AB, Tourasse NJ, Okstad OA. What sets Bacillus anthracis apart from other Bacillus species? Annu Rev Microbiol. 2009;63:451–76.

23. Kovac J, Miller RA, Carroll LM, Kent DJ, Jian J, Beno SM, et al. Production of hemolysin BL by Bacillus cereus group isolates of dairy origin is associated with whole-genome phylogenetic clade. BMC Genomics. 2016;17:581.

24. Yang J, Wang L, Ji X, Feng Y, Li X, Zou C, et al. Genomic and proteomic analyses of the fungus Arthrobotrys oligospora provide insights into nematode-trap formation. PLoS Pathog. 2011;7:e1002179.

25. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35:W52–7.

26. Makarova KS, Wolf YI, Koonin EV. The basic building blocks and evolution of CRISPR-CAS systems. Biochem Soc Trans. 2013;41:1392–400.

27. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3.

28. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

29. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26:1641–50.

30. Riadi G, Medina-Moenne C, Holmes DS. TnpPred: a web service for the robust prediction of prokaryotic transposases. Comp Funct Genomics. 2012;2012:678761.

31. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res. 2018;46:W200–4.

32. Zheng J, Peng D, Ruan L, Sun M. Evolution and dynamics of megaplasmids with genome sizes larger than 100 kb in the Bacillus cereus group. BMC Evol Biol. 2013;13:262.

33. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. Nucleic Acids Res. 2012;40:D290–301.

34. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44:W16–21.

35. Peng D, Luo Y, Guo S, Zeng H, Ju S, Yu Z, et al. Elaboration of an electroporation protocol for large plasmids and wild-type strains of Bacillus thuringiensis. J Appl Microbiol. 2009;106:1849–58.

36. Stiernagle T. Maintenance of C. elegans. WormBook. 2006:1–11.

37. O'Toole GA. Microtiter dish biofilm formation assay. J Vis Exp. 2011:e2437.

38. Wang M, Du J, Zhang D, Li X, Zhao J. Modification of different pulps by homologous overexpression alkali-tolerant endoglucanase in Bacillus subtilis Y106. Sci Rep. 2017;7:3321.

39. Peng D, Lin J, Huang Q, Zheng W, Liu G, Zheng J, et al. A novel metalloproteinase virulence factor is involved in Bacillus thuringiensis pathogenesis in nematodes and insects. Environ Microbiol. 2016;18:846–62.

40. Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. Self-targeting by CRISPR: gene regulation or autoimmunity? Trends Genet. 2010;26:335–40.

41. Turgeman-Grott I, Joseph S, Marton S, Eizenshtein K, Naor A, Soucy SM, et al. Pervasive acquisition of CRISPR memory driven by inter-species mating of archaea can limit gene transfer and influence speciation. Nat Microbiol. 2019;4:177–86.

42. Zheng J, Gao Q, Liu L, Liu H, Wang Y, Peng D, et al. Comparative genomics of Bacillus thuringiensis reveals a path to specialized exploitation of multiple invertebrate hosts. mBio. 2017;8:e00822-17.

43. Ceuppens S, Boon N, Uyttendaele M. Diversity of *Bacillus cereus* group strains is reflected in their broad range of pathogenicity and diverse ecological lifestyles. FEMS Microbiol Ecol. 2013;84:433–50.

44. Raymond B, Bonsall MB. Cooperation and the evolutionary ecology of bacterial virulence: the *Bacillus cereus* group as a novel study system. Bioessays. 2013;35:706–16.

45. McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH. High frequency of horizontal gene transfer in the oceans. Science. 2010;330:50.

46. Bergstrom CT, Lipsitch M, Levin BR. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. Genetics. 2000;155:1505–19.

47. Rancurel C, Legrand L, Danchin EGJ. Alienness: rapid detection of candidate horizontal gene transfers across the tree of life. *Genes*. 2017;8:E248.

48. Zhang F, Peng D, Cheng C, Zhou W, Ju S, Wan D, et al. *Bacillus thuringiensis* crystal protein Cry6Aa triggers *Caenorhabditis elegans* necrosis pathway mediated by aspartic protease (ASP-1). PLoS Pathog. 2016;12:e1005389.

49. Burmolle M, Ren D, Bjarnsholt T, Sorensen SJ. Interactions in multispecies biofilms: do they actually matter? Trends Microbiol. 2014;22:84–91.

50. Reiter L, Kolsto AB, Piehler AP. Reference genes for quantitative, reverse-transcription PCR in *Bacillus cereus* group strains throughout the bacterial life cycle. J Microbiol Methods. 2011;86:210–7.

51. McInerney JO, McNally A, O'Connell MJ. Why prokaryotes have pangenomes. *Nat Microbiol*. 2017;2:17040.

52. Shmakov SA, Sitnik V, Makarova KS, Wolf YI, Severinov KV, Koonin EV. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. mBio. 2017;8: e01397-17.

53. Pawluk A, Davidson AR, Maxwell KL. Anti-CRISPR: discovery, mechanism and function. Nat Rev Microbiol. 2018;16:12–7.

54. Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, et al. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J Bacteriol. 2008;190:1401–12.

55. Li HY, Kao CY, Lin WH, Zheng PX, Yan JJ, Wang MC, et al. Characterization of CRISPR-Cas systems in clinical *Klebsiella pneumoniae* isolates uncovers its potential association with antibiotic susceptibility. Front Microbiol. 2018;9:1595.

56. Long J, Xu Y, Ou L, Yang H, Xi Y, Chen S, et al. Polymorphism of Type I-F CRISPR/Cas system in *Escherichia coli* of phylogenetic group B2 and its application in genotyping. Infect Genet Evol. 2019;74:103916.

57. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, et al. BREX is a novel phage resistance system widespread in microbial genomes. EMBO J. 2015;34:169–83.

58. Ofir G, Melamed S, Sberro H, Mukamel Z, Silverman S, Yaakov G, et al. DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat Microbiol. 2018;3:90–8.

59. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of antiphage defense systems in the microbial pangenome. Science. 2018;359:eaar4120.

60. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. 2003;31:1805–12.

61. Luo X, Chen L, Huang Q, Zheng J, Zhou W, Peng D, et al. *Bacillus thuringiensis* metalloproteinase Bmp1 functions as a nematicidal virulence factor. Appl Environ Microbiol. 2013;79:460–8.

62. Ruan L, Wang H, Cai G, Peng D, Zhou H, Zheng J, et al. A two-domain protein triggers heat shock pathway and necrosis pathway both in model plant and nematode. Environ Microbiol. 2015;17:4547–65.

63. Liu XY, Ruan LF, Hu ZF, Peng DH, Cao SY, Yu ZN, et al. Genome-wide screening reveals the genetic determinants of an antibiotic insecticide in *Bacillus thuringiensis*. J Biol Chem. 2010;285:39191–200.

64. Du C, Cao S, Shi X, Nie X, Zheng J, Deng Y, et al. Genetic and biochemical characterization of a gene operon for trans-aconitic acid, a novel nematicide from *Bacillus thuringiensis*. J Biol Chem. 2017;292:3517–30.

65. Zhao C, Luo Y, Song C, Liu Z, Chen S, Yu Z, et al. Identification of three Zwittermicin A biosynthesis-related genes from *Bacillus thuringiensis* subsp. *kurstaki* strain YBT-1520. Arch Microbiol. 2007;187:313–9.

66. Zhu L, Peng D, Wang Y, Ye W, Zheng J, Zhao C, et al. Genomic and transcriptomic insights into the efficient entomopathogenicity of *Bacillus thuringiensis*. Sci Rep. 2015;5:14129.

67. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011;471:602–7.