A biological warfare between two bacterial viruses in a particular genomic region using a CRISPR-Cas system as a weapon

Alejandro Rubio[1], Andrés Garzón[1], Antonio J. Pérez-Pulido[1,*]

[1]Andalusian Centre for Developmental Biology (CABD, UPO-CSIC-JA). Faculty of Experimental Sciences (Genetics Area), University Pablo de Olavide, 41013, Seville, Spain

Abstract

CRISPR-Cas systems are acquired immunity systems of bacteria and archaea that prevent infection by phages and other mobile genetic elements. It is currently known that this defense system has also been co-opted by viruses. These viruses could use CRISPR-Cas systems to compete against other rival viruses. We have discovered a virus in the bacterium *Acinetobacter baumannii* that incorporates a CRISPR-Cas system into an integration hotspot of the host genome. Once integrated, this CRISPR-Cas system could prevent the infection of the most frequent viruses in this bacterial species, especially one that competes with the CRISPR-Cas system itself for the same integration site. This latter virus is prevalent in strains of the species belonging to the so-called Global Clone 2, which causes the most frequent outbreaks worldwide. According to the World Health Organisation, *A. baumannii* is a top-priority opportunistic pathogen with an increasing number of isolates emerging as multi-drug resistant. The knowledge about this new viral warfare using CRISPR-Cas systems, which are known to limit the entry of antibiotic resistance genes into bacteria, could be useful in the fight against the infections they cause.

Introduction

Acinetobacter baumannii is a bacterium that causes endemic infections and hospital-acquired outbreaks with a high mortality and morbidity rate. For this reason, the World Health Organization recently classified it as “priority 1” on a list of pathogenic bacteria for which the development of new antibiotics is urgent (1). The bacterium can colonize the human body and even resist in the hospital environment, and opportunistically causes pneumonia, urinary tract and wound infections and, occasionally, bacteremia (2, 3). Much of the success of *A. baumannii* can be attributed to its genome plasticity, which mutates rapidly under adverse conditions and stress (4).

Strains of *A. baumannii* can be classified according to clonal groups based on several conserved genes (5), and two adjacent clonal groups reflect genomes arising from a recent common ancestor. In one of the reference schemes (the Institut Pasteur MLST), ST2 corresponds to Global Clone 2, which is by far the dominant group of this species, causing a large proportion of outbreaks worldwide (6).

As found in 40% of bacterial genomes, *A. baumannii* posesses the acquired CRISPR-Cas defense system against bacterial viruses (bacteriophages, known as phages) and other mobile genetic elements (7). However, this system appears in only 13% of its sequenced

genomes (8, 9). These CRISPR-Cas systems are mainly targeted against phages, but also against plasmids to a lesser extent (10). Finally, it has been discussed that they could also be targeted against phage-plasmids (8), which are phages that can act as plasmids under certain conditions (11).

CRISPR-Cas systems are highly diverse prokaryotic acquired immunity systems, currently classified into 2 classes, 6 types and 33 subtypes, which differ in the architecture of their components and their sequence specificity (12). They are composed of genes that are part of the different stages of this immune system and are generically called *cas* (CRISPR-associated genes) (13). When the phage DNA enters the host, its sequence is cleaved into fragments that are added to the CRISPR array by the Cas1 and Cas2 proteins in a process called adaptation. During this step, phage sequences of about 32 nucleotides, known as spacers, are recognized and stored in a CRISPR array, separated by sequence repeats of similar length. The CRISPR array is then transcribed during the expression stage, resulting in transcripts that are cleaved into their corresponding spacers. Finally, during the interference stage, the spacer recognizes the complementary sequence in the target phage, called protospacer, which leads to the elimination of the viral infection by cleaving the phage DNA. This last step requires a nuclease, which in class I CRISPR-Cas systems is a protein complex in which the Cas3 protein stands out as the nuclease responsible for cleaving the phage genome. In addition, for the identification of the protospacer, it is essential to recognize a sequence of a few nucleotides (often only 2 or 3) just at the 5' position of the protospacer, known as protospacer adjacent motif (PAM) sequence (14).

A. baumannii mainly presents class I CRISPR-Cas systems, divided into 2 type I subtypes: I-Fa (also classified as I-F2) and I-Fb (also classified as I-F1) (8, 15–17). Remarkably, Global Clone 2 does not usually present CRISPR-Cas systems (9).

Phage-encoded CRISPR-Cas systems have also been discovered in recent years (18). The most characterized of these is the phage ICP1 of *Vibrio cholerae*, which recognizes another phage with which it competes (known as satellite phage). This satellite phage needs ICP1 to complete its lytic cycle, and ICP1 uses an I-F type CRISPR-Cas system to defend against the satellite phage (19).

Phages that escape the bacterial defense systems and integrate into the bacterial genome, entering the so-called lysogenic cycle, can remain as prophages until they resume lytic cycle, in which they replicate and can leave the bacterium (20). These prophages are usually integrated into tRNA and tmRNA genes, as well as other genomic elements including integrative plasmids, and genomics islands. It is facilitated by the specificity of their integrases (21). Genomic islands are sometimes grouped in so-called archipelagos, which have been little studied so far (22, 23).

We have found an archipelago in an *A. baumannii* tmRNA gene, prominently featuring a prophage that appears to compete for that site with one of the bacterial CRISPR-Cas systems. Remarkably, this same CRISPR-Cas system is sometimes carried by a phage-plasmid. All of this suggests that both phages may be competing for the bacterium, and that the phage-plasmid may leave the CRISPR-Cas system integrated into the bacterium to gain the advantage in that armed race between phages.

Materials and methods

Pangenome construction

The pangenome used was the one constructed with 9,696 genomes in Rubio et al., 2023 (8). All genomes are structurally and functionally annotated, as well as the reference genes of the pangenome itself. The CRISPR-Cas systems of each genome, as well as the spacers and repeats of the CRISPR arrays were also obtained from the aforementioned work. In some cases the CRISPR-Cas systems were checked manually, especially in the I-Fa subtype, finding some additional ones that were mostly incomplete (lacked CRISPR array or some or most of the cas genes). The latter systems were labeled as ambiguous or incomplete.

Phage search

Prophages were searched in all the genomes using Phigaro version 2.3.0 with default parameters and the *abs* mode (56). Only phages ≥9Kb in length were considered. Then, prophages were grouped by similarity using the MeShClust version 3.0 program with the parameters *-v* (total initial sequences) and *-b* (1/4 of total initial sequences). To combine similar phage genomes with the sequence oriented in the two possible reading directions, we performed a similarity search with BLASTN by comparing the reference sequences of each cluster with each other. Then, clusters whose reference sequence shared at least 90% identity and coverage (with respect to the shortest sequence) were combined. Finally, the shortest sequence from each final group was selected as the reference sequence.

The phage-plasmid was searched for in contigs of no more than 130,000 Kb, in which it was found with at least 90% identity and 70% sequence coverage with respect to a reference sequence that had it (GenBank: NZ_UFOI01000014.1), to which the CRISPR-Cas system was masked. CRISPR-Cas I-Fa and restriction-methylation systems were manually checked on these sequences, searching for genes *hsdR*, *hsdS* and *hsdM*.

The full version of the phage P1virus was obtained by taking the nucleotide sequence from the phage integrase (gene homologs of GenBank: CU468230.2) to the direct repeat of the *ssrA* gene (GTTCAACTCCCGCCATCTCCACCA). Incomplete versions of the phage P1virus were taken by searching for the phage integrase and checking that the genome had at least 40% of the complete phage sequence.

The rest of the genomic islands integrated in *ssrA* were obtained by searching for the core gene (gene present in all the genomes of the species) most likely to be the one that should appear next to *ssrA* (gene homologs of GenBank: CAM88546.1), and extracting the different gene clusters that could be found in between them.

Phage annotation

The phage-plasmid was found in independent contigs in the genomes analyzed. After performing a similarity search on the NCBI-Blast web portal, several plasmids were found

with more than 90% identity and coverage, with and without the CRISPR-Cas I-Fa system. PHASTEST was used to annotate the genome of this phage (24).

VPF-Class tool (no version) was used to annotate the genre of the phage integrated in *ssrA* (P1virus) (25). PhANNs was used to annotate the putative viral genes of the different genomic islands integrated in *ssrA* (26).

Molecular phylogenies

Molecular phylogenies were performed using the assembly of complete bacterial or phage genomes, always using an outgroup to root the trees. Phylogenies were performed with the web tool Realphy 1.13 (27) and the trees were visualized with the ggtree library of the R programming language.

Retrieval of CRISPR repeats and spacers

Initially, the CRISPR array repeats were extracted from the output of the CRISPRCasFinder tool (28). Their orientation was taken from the 'potential_direction' output attribute. Since most of the repeats had a value equal to 'unknown', the different variants of the repeats were reviewed manually and those that were similar were grouped together, transforming those that were in reverse-complementary direction. In this way, the direction of all CRISPR arrays could be determined.

Subsequently, the spacers were extracted, already with the direction suggested by the flanking repeats. In this way, sequence logos could be made using the ggseqlogo library of the R language.

Discovery of protospacers

Protospacers were searched by performing a similarity search with BLASTN and the *blastn-short* option turned on, using a threshold of ≥95% sequence identity and 100% spacer coverage.

Results

Acinetobacter baumannii has two different subtypes of CRISPR-Cas I-F systems

We analyzed an *A. baumannii* pangenome composed of 9,696 genomes, in which we searched for CRISPR-Cas systems. Two different configurations of CRISPR-Cas systems were found, both type I-F. Specifically, 9.5% of the genomes had the I-F1 subtype (also referenced as I-Fb in this species), 3.5% presented a system similar to the I-F2 subtype (also referenced as I-Fa), and 1.6% were classified as ambiguous, since they only had isolated CRISPR arrays or some isolated *cas* genes (Fig. 1A). The I-Fa system presented 2 independent CRISPR arrays, as previously known (9). The difference between both subtypes was given precisely by the number of CRISPR arrays, but also by the order of the

cas genes and the genomic integration site. Furthermore, the amino acid sequence of the encoded Cas proteins only share about 30% identity between the homologs of both subtypes. In addition, the Cas1, Cas8f and Cas7f proteins of the I-Fa subtype are notably shorter than the corresponding homologous proteins of the I-Fb system.

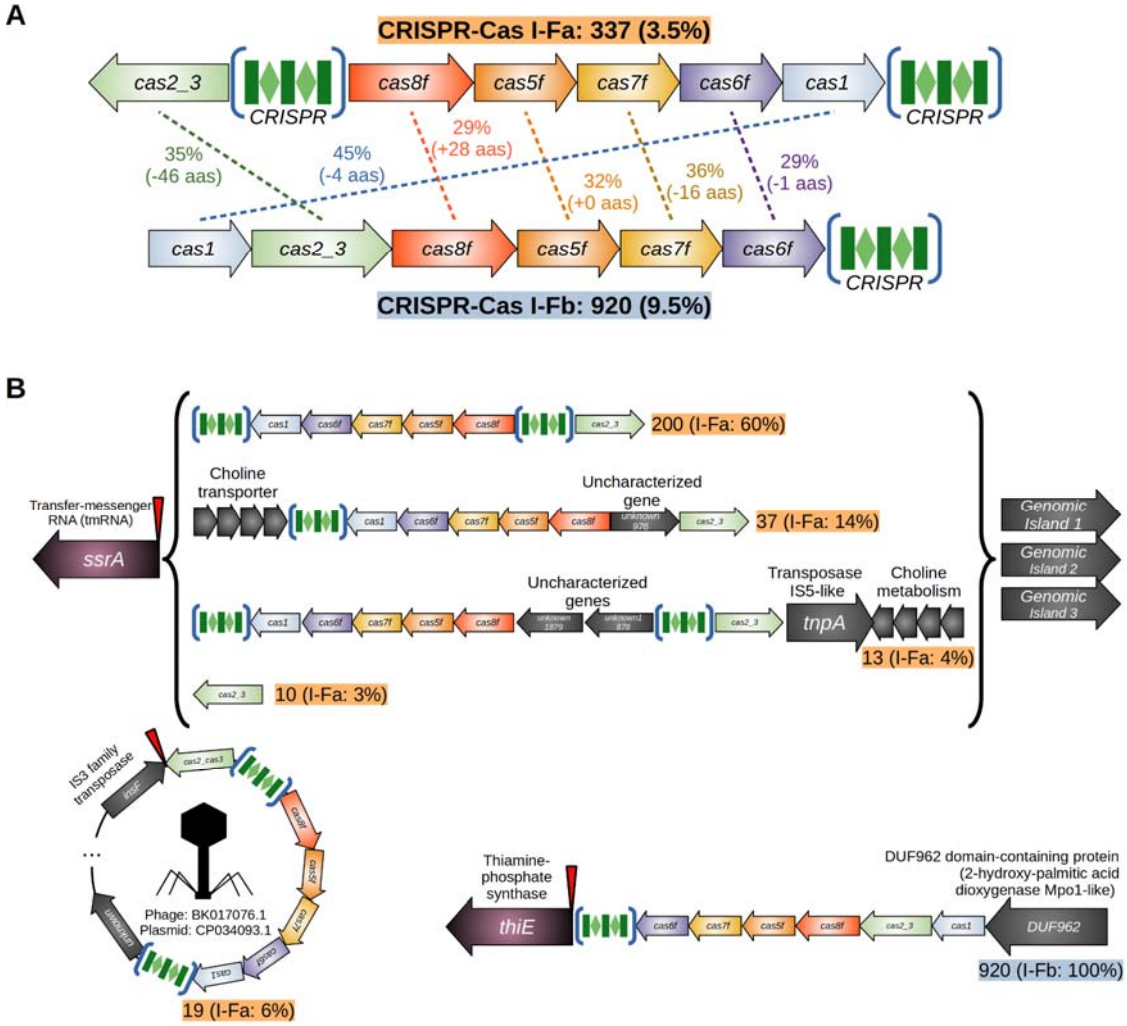


Fig. 1. CRISPR-Cas I-F systems in *Acinetobacter baumannii*. (A) Gene structure of the I-Fa and I-Fb CRISPR-Cas systems. Next to the type of CRISPR-Cas system, the number of genomes that have it and the percentage with respect to the total number of genomes analyzed are shown. In the middle of both CRISPR-Cas systems, the percentage of identity between the homologous proteins of both systems is shown, and in parentheses the number of amino acids that the Cas proteins of subtype I-Fa had more or fewer than I-Fb. (B) Genomic regions in which CRISPR-Cas systems are found. The I-Fa subtype is mostly integrated behind a tmRNA called *ssrA*. But the latter is also found downstream of *ssrA*, under a cluster of genes including a choline transporter, together with a transposase and two uncharacterized genes, with only one of the *cas* genes (*cas2_3*). Finally, it also appears in a circular molecule, together with another transposase. The latter configuration is also found in a plasmid in the NCBI RefSeq database (CP034093.1) and in several phages (BK017076.1; 30% coverage). The I-Fb subtype is always located between the *thiE* gene and an

uncharacterized gene encoding a protein with the DUF962 domain. The red inverted triangles indicate the genomic islands insertion site.

The I-Fb system appeared always integrated at exactly the same site, between the *thiE* gene (encoding a Thiamine-phosphate synthase) and a gene encoding an uncharacterized protein (Fig. 1B). However, the I-Fa system is found to be integrated with different frequencies at different sites. In almost two-thirds of the genomes with this subtype, the CRISPR-Cas system is integrated near the 5' end of the *ssrA* gene, a transfer-messenger RNA (tmRNA). In 14% of the genomes, the CRISPR-Cas I-Fa system is displaced by a cluster of 14 genes including a choline transporter. Less frequently, the system is interrupted by 2 uncharacterized genes and at the opposite end of *ssrA* is an IS5-like transposase, followed by the choline transporter cluster. Moreover, we found 10 genomes in which only the *cas2_3* gene appears next to the *ssrA* gene. In all these cases, other gene clusters are usually found downstream of the CRISPR-Cas I-Fa system, in the form of genomic islands, in addition to the one containing the choline transporter.

Finally, we found that the CRISPR-Cas I-Fa system also appeared in episomes of about 110 Kb, not assembled with the bacterial chromosome sequence, together with an IS3 family transposase and a number of other genes. When we searched for sequences similar to these episomes in the NCBI RefSeq database, we found several plasmids, which did or did not contain the CRISPR-Cas system (Suppl. Fig. S1). Furthermore, when we analyzed its sequence, we found that it could also be a phage (105 of its 112 genes are annotated as putative phage genes). This latter, together with the previous result, suggests that it could actually be a phage-plasmid that occasionally carries a CRISPR-Cas I-Fa system.

The rest of the CRISPR-Cas I-Fa systems appear in other less frequent locations or their position cannot be determined, due to their fragmentation in the genomes analyzed. It is important to note that the *cas2_3* gene is the only one that appears in all the CRISPR-Cas I-Fa system configurations found. The protein it encodes has about 30% identity with that of the *Vibrio cholera* phage ICP1 (RefSeq:AGG09390.1), which was one of the first phages known to carry a CRISPR-Cas I-F system, and the best characterized to date (19, 29).

The *ssrA* gene supports an archipelago for the integration of multiple genomic islands

Since tRNA and tmRNA genes are often hotspots for integration of genomic islands in bacteria, we wanted to know whether other elements, in addition to the CRISPR-Cas I-Fa system, were integrated into *ssrA*. Thus, we found that other gene clusters appear integrated in this gene, with different frequencies (Fig. 2). The most frequent were two, the cluster containing the choline transporter, and another with a putative prophage. This prophage has a length of about 82 Kb, with a CG content of 36.5% (*A. baumannii* presents an average of 38.5%), has 90 genes, and is classified as a P1virus phage. A proof of the virus integration is that a direct repeat is always found at the end of the virus, with the same sequence as the 3' end of the *ssrA* gene, probably representing a cognate att (phage attachment) sequence (see Materials and Methods). This prophage includes a toxin-antitoxin system that could help its persistence in bacteria, with a RelE/StbE family addiction module toxin and the corresponding mRNA interferase RelE/StbE antitoxin. The prophage is found in 4,597

genomes (47% of all genomes analyzed), although only about half of them were complete between the *ssrA* gene and the distal att sequence.

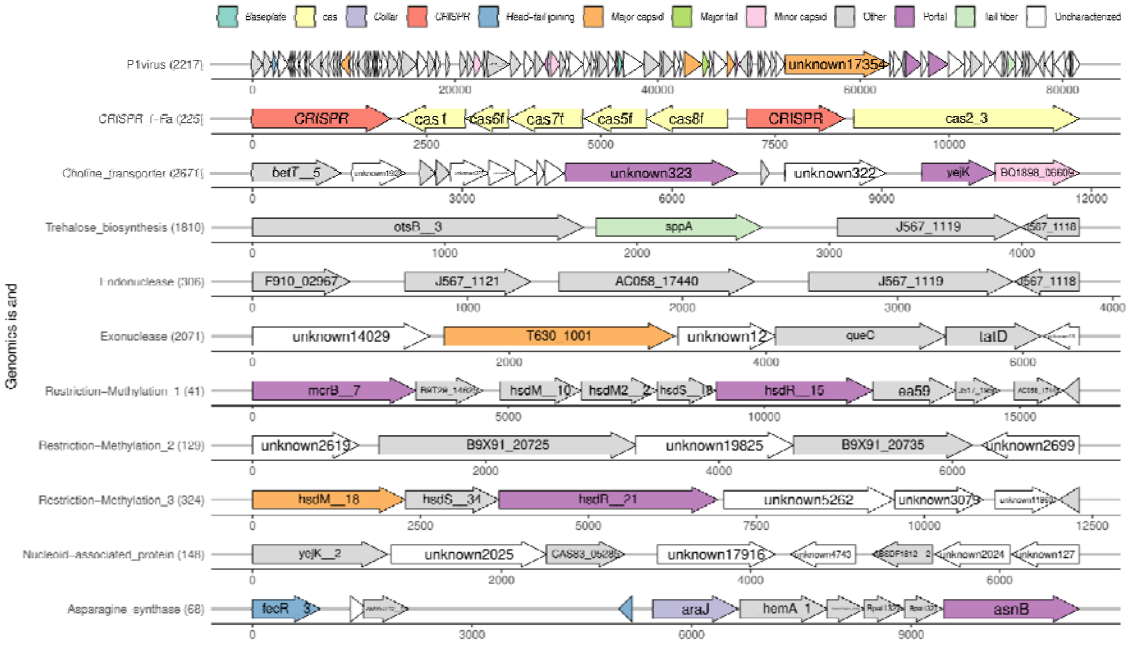


Fig. 2. *ssrA* integrated genomic islands. Colors other than white represent predicted phage genes organized by class, except for red and yellow, which have been used for elements of the CRISPR-Cas I-Fa system. The numbers in parentheses accompanying genomic island names represent the absolute frequency of that genomic island next to the *ssrA* gene.

Other genomic islands integrated at the same site are a cluster of 4 genes with trehalose biosynthesis genes, another with arginine synthesis genes, 2 especially frequent with endonucleases or exonucleases, and 3 different islands that seem to contain restriction-methylation systems, and that could be part of the bacterial defensome, as well as the CRISPR-Cas system. Some of the genes in these genomic islands are similar to different phage genes.

P1virus and the CRISPR-Cas system do not usually coexist in the same genome

These genomic islands sometimes appear displaced by others and may take up different positions between the *ssrA* gene and the next core gene that encodes a hypothetical lipoprotein (Fig. 3A). In fact, this bacterial core gene only appears next to *ssrA* in 3 of the genomes analyzed. However, the P1virus prophage appears to take up only the first position (next to *ssrA*), probably driven by the presence of the att site in the *ssrA* gene. On the contrary, the genomic island with the choline transporter may take up either the first or the second position. In fact, the P1virus appears as the only genomic island in more than 2,000 genomes, or followed by other genomic islands in lower frequency (Fig. 3B). In contrast, the CRISPR-Cas system also appears alone or preceded by the island with the choline transporter. Notably, the P1virus and the CRISPR-Cas I-Fa system do not usually appear

together in the same genome, and when they do, the virus appears mostly in its incomplete version (Fig. 3C). Specifically, they co-occur in 37 genomes (Fisher exact test, *p-value* = 7.39e-60). It is also noteworthy that 31 of the genomes have both CRISPR-Cas I-Fa and I-Fb systems, exactly the number that would be expected by chance (Fisher exact test, *p-value* = 0.92).

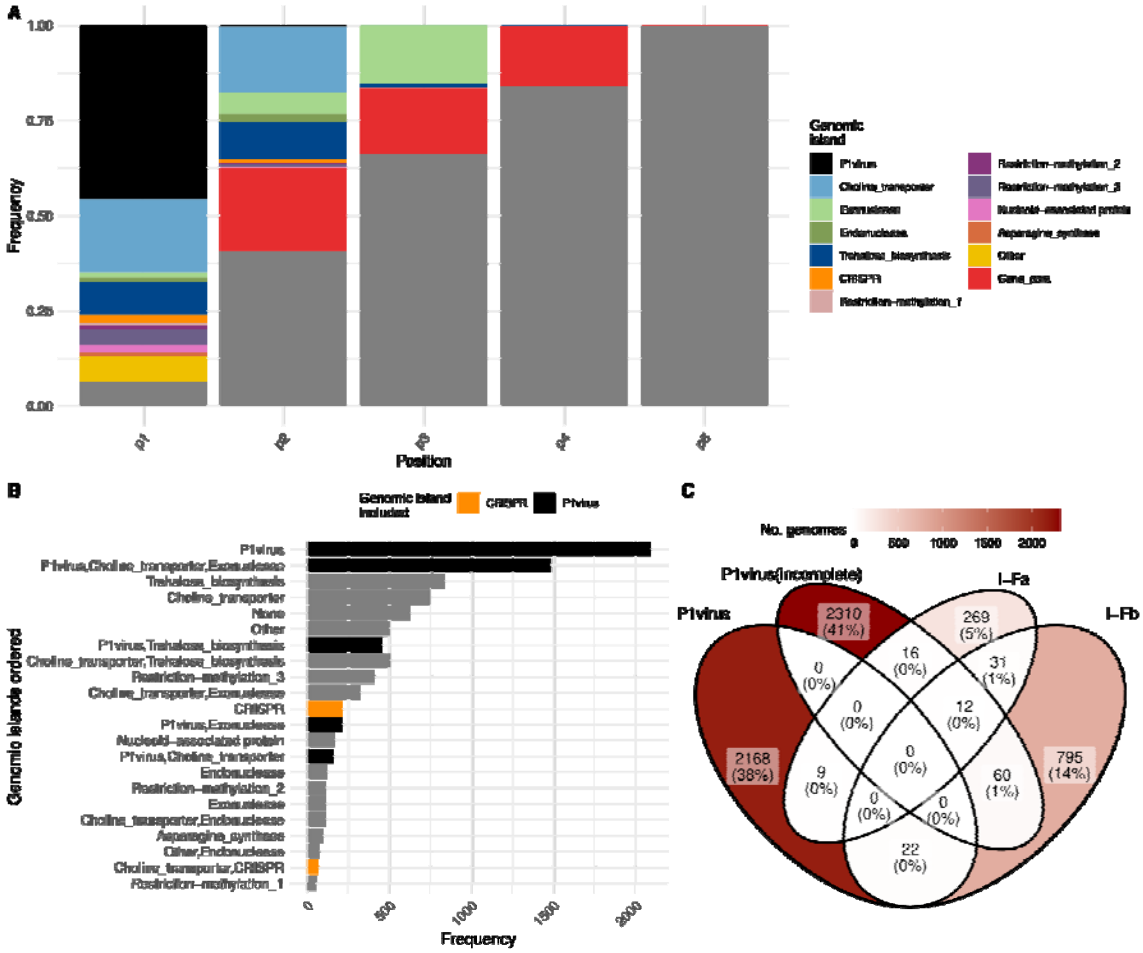


Fig. 3. Position of the genomic islands integrated in the *ssrA* gene. (A) Frequency of each genomic island at each sequential position (p1 corresponds to the closest to *ssrA* and p5 to the farthest). 'Other' indicates that the genomic island is none of those studied here. The core gene is the bacterial gene that should appear next to *ssrA*, if there were no integrated genomic island. When the genomic island could not be determined, it was highlighted in gray. (B) Frequency of the different combinations of genomic islands found. Only the most frequent combinations are shown. (C) Number of genomes having the P1virus, or the CRISPR-Cas I-Fa or I-Fb systems, including group overlapping.

The non-coexistence of these two genomic islands could be due to the evolutionary divergence of the genomes containing them. To discard this hypothesis, we studied the distribution of both islands in the different MLST clonal groups of *A. baumannii*. The ST2 group was the one with the highest number of strains in our pangenome, as was to be expected due to its already known great evolutionary success (Fig. 4A). And particularly in this group, about two thirds of the genomes have the prophage P1virus. On the other hand,

the CRISPR-Cas I-Fa system was more represented in the ST79 group, although this group also presents genomes with P1virus (Fig. 4B).

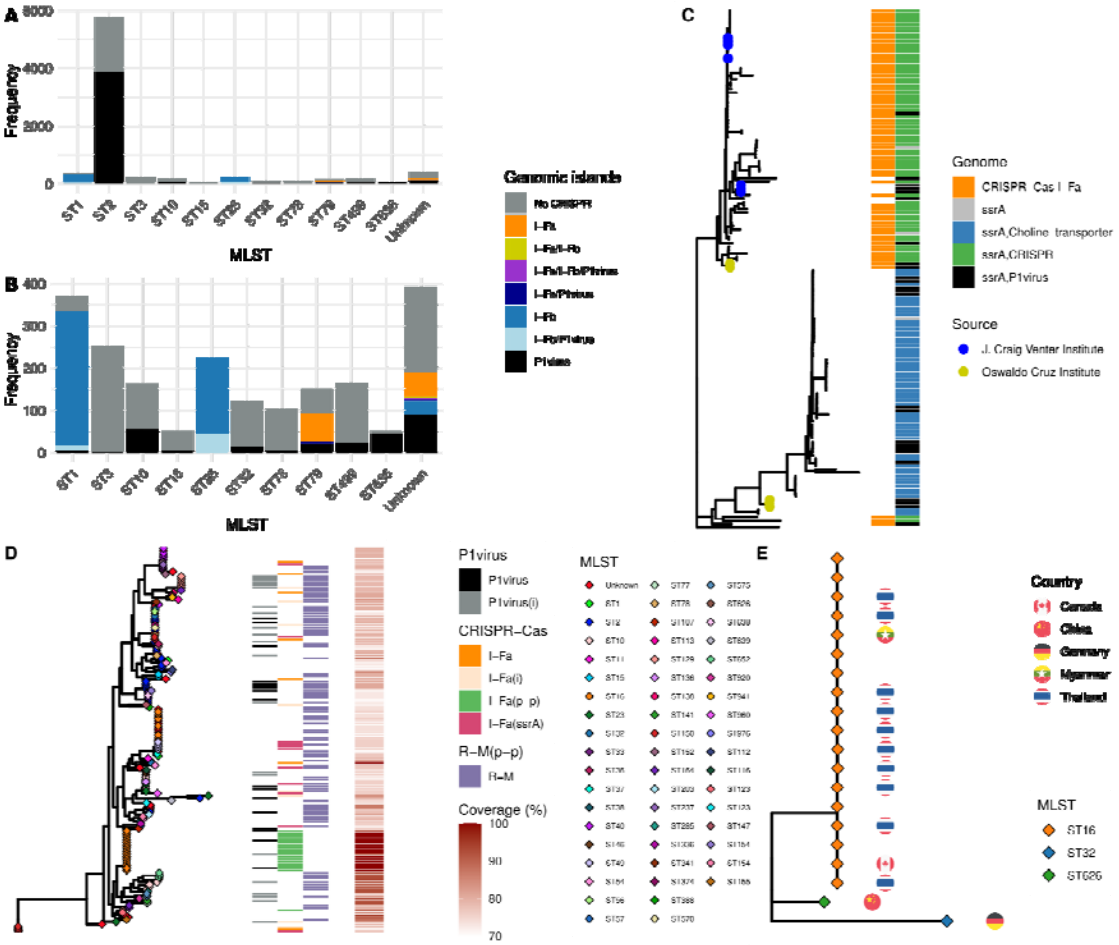


Fig. 4. Distribution of genomes by MLST, and phylogeny of ST79 and the phage-plasmid. (A) Distribution of genomic islands along the most abundant MLST groups of *A. baumannii*. “Unknown” corresponds to genomes to which it has not been possible to assign an MLST group. (B) Same figure as (A) but without showing ST2, to highlight the rest of the groups. (C) Molecular phylogeny of all genomes of the ST79 group, using as outgroup the reference genome of *A. baumannii*, which belongs to the ST46 group (the distance of this outgroup has been divided by 10, for better visualization). (D) Molecular phylogeny of all phage-plasmid sequences found, from 70% sequence coverage with respect to a complete one. It is shown: the MLST groups in which the phage has been found, whether the genome carrying the phage-plasmid also has the P1virus (i=incomplete), or a CRISPR-Cas I-F system (i=incomplete, only a CRISPR array or a *cas* gene has been found; p-p=the CRISPR-Cas system is located in the phage-plasmid; ssra=the CRISPR-Cas system is located next to the gene *ssrA*), or a restriction-methylation system (R-M). It is also shown the coverage of the phage-plasmid sequence that was found. (E) Molecular phylogeny of all phage-plasmids found that carry the CRISPR-Cas I-Fa system in their sequence. The MLST groups in which the phage has been found and the country of isolation are shown.

Thus, we took the ST79 group as the one that most frequently presented genomes with one or the other of the two genomic islands (CRISPR-Cas I-Fa or P1virus). By comparing the

genomes of the group containing P1virus with those containing the CRISPR-Cas I-Fa system we were able to verify that the main difference involved the genes of the opposite genomic island. Mostly, the presence of one of the islands seems to exclude the other, even in evolutionarily very close genomes, or isolated at the same place (Fig. 4C). Specifically, of the 152 genomes in this group, 73 carry the CRISPR-Cas system, 25 carry the phage P1virus and only 5 carry both, noting that in none of these 5 cases does the CRISPR-Cas system appear integrated in the *ssrA* archipelago.

Here, two genomes sequenced at the same site (Oswaldo Cruz Institute) are noteworthy, given that the P1virus appears to have displaced the CRISPR-Cas I-Fa system downstream. However, the spacers that these genomes have do not recognize the P1virus sequence. In addition, two more genomes sequenced at the same institute, also present the phage P1virus, but have already lost the CRISPR-Cas I-Fa system. All this would support the incompatibility of both genomic islands, even in very close genomes.

The phage-plasmid has an alternative defense system and appears in diverse genomes

On the other hand, we also wanted to know what type of genomes the phage-plasmid that sometimes carries the CRISPR-Cas I-Fa system appeared in. For this purpose, we searched for the phage-plasmid sequence, considering a threshold of 70% sequence coverage (at least 77 Kb). The sequence was found in 164 genomes, again in the form of independent episomes. This putative phage-plasmid appears in several different clonal groups and has been isolated in different countries, with or without the defense system (Fig. 4DE). Within these genomes, those carrying the CRISPR-Cas system usually do not have the P1virus. Only six genomes had both elements, but in 4 of them the CRISPR-Cas system was present in the phage-plasmid (not integrated with the *ssrA* gene), and in the others the system was incomplete (some or all of the *cas* genes were missing).

Interestingly, most of the remaining phage-plasmids carry a type I restriction-methylation system that appears to replace the CRISPR-Cas I-Fa system. In addition, two of these phage-plasmids present the same att sequence that appears to use the P1virus. This att site was found next to the restriction-methylation system, suggesting that this new found defense system could also become integrated into the *ssrA* gene.

The CRISPR-Cas I-Fa systems of *A. baumannii* show common features regardless of their site of integration

To analyze the possible differences between the two CRISPR-Cas I-F subtypes, we counted the number of spacers in each of them. In the case of I-Fa, since it had two CRISPR arrays, these were analyzed separately. We considered CRISPR 1 array (C1a) the one closest to the *cas1* gene and CRISPR 2 array (C2a) the one closest to the *cas2_3* gene. Thus, the CRISPR arrays with the highest average number of spacers were C1a and those of the CRISPR-Cas I-Fb system (Fig. 5A). Within the CRISPR-Cas I-Fa system, C1a has almost twice as many spacers on average as C2a, something that holds true for the I-Fa systems associated with the choline transporter genomic island and the transposon. However, the trend is the opposite in CRISPR arrays carrying the phage-plasmid. This could suggest that the CRISPR array grows, acquiring more spacers, when the CRISPR-Cas system is

integrated into the bacterial chromosome, especially in the C1a. However, the low number of spacers appearing in the phage-plasmid would be expected due to the space limitations it would have in its capsid.

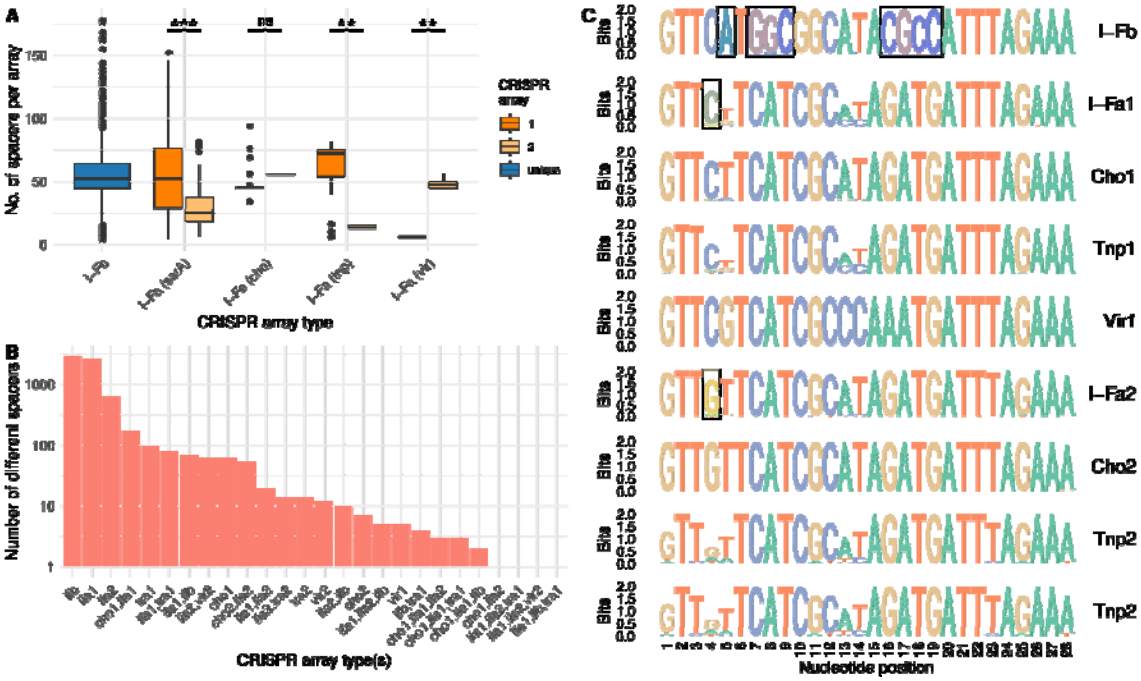


Fig. 5. Number and type of spacers, and consensus sequence of repeats. (A) Number of spacers per CRISPR array separated by CRISPR-Cas system type; ifa=CRISPR-Cas I-Fa, ifb=CRISPR-Cas I-Fb, cho=choline transporter near to the ssrA, tnp=transposon near to the CRISPR-Cas I-Fa, vir=CRISPR-Cas system in the phage-plasmid, 1=C1a near to *cas1*, 2=C2a near to *ca3_2*, unique=array of I-Fb. Asterisks highlight pairs of CRISPR arrays with significant differences in the median according to a Wilcoxon rank-sum test (from *≤0.05 to ***≤1e-16). (B) Number of non-redundant spacers separated according to the type of CRISPR-Cas system to which the CRISPR array belongs. When the same spacer appears in different types, they are separated by comma. The y-axis is in logarithmic scale. (C) Logo of the sequence repeats of CRISPR arrays separated by CRISPR-Cas system type. The differential I-Fb nucleotides have been shaded with blue boxes, and those distinguishing the two I-Fa arrays in yellow.

The number of different spacers is highest in the CRISPR-Cas I-Fb system, followed closely by C1a (Fig. 5B). However, C2a has almost 5 times fewer than the other two. As expected, some spacers are common among the different systems, especially between C1a or C2a of the different I-Fa variants.

Finally, the nucleotide sequence of the repeats of the CRISPR-Cas systems were analyzed, and it was found that the I-Fb system differs from the I-Fa in 8 of its 28 nucleotides (Fig. 5C). However, the repeats of the two I-Fa systems differed by only 1 nucleotide (2 nucleotides in the case of phage-plasmid C1a). This fact supports the idea that the C1a of all the different configurations of the CRISPR-Cas I-Fa system would have a common origin, as would all the C2a, on the other hand. Based on these results, the different CRISPR C1a CRISPRs will

be considered as the same type of CRISPR-Cas I-Fa1 array from now on, and the same with the different C2a CRISPR.

CRISPR-Cas I-Fa system spacers recognize all P1virus variants

The prophage P1virus is integrated into the *ssrA* gene and is present in a complete form in 2,199 genomes, with an average length of 79,989±7,846 bp. All of them can be considered as variants of the phage. Moreover, the genomes analyzed have other prophages, which are also found in the form of different variants. To compare the interaction of the CRISPR-Cas systems with P1virus and with the rest of the prophages found in the bacteria, all the phages found in the *A. baumannii* genomes were obtained and grouped by sequence similarity, finding a total of 330 different phages. Of these, only the 40 that appeared in more than 100 genomes were taken for the following analyses.

As expected, when analyzing the genomes in which such phages appear, including P1viruses, it can be seen that these tend to be those lacking CRISPR-Cas systems (Suppl. Fig. S2). In fact, 18% of the genomes with a CRISPR-Cas I-Fa system, and 11% of those with the I-Fb type did not have any of these prophages. Whereas this was only the case in the 5% of genomes lacking CRISPR-Cas systems.

We then analyzed which phage sequences were recognized by the spacers of the 3 different types of the CRISPR arrays. The first thing that was found is that all P1virus variants were recognized by the I-Fa C1a spacers, and almost all variants (96%) by the C2a spacers (Fig. 6; Suppl. Fig. S3). However, hardly any P1virus variants are targeted by spacers of the CRISPR-Cas I-Fb system (6%). In addition, almost 85% of C1a genomes contain spacers against phage P1virus, as do 67% of C2a genomes. And these spacers account for 4-5% of the different spacers in these CRISPR arrays. The second thing to note is that while the I-Fb spacers seem to recognize many of the different phages (31 out of 41, if we consider 100% of their variants), these spacers do not target hardly any variants of two phages: the referred P1virus and the Phage_1022.

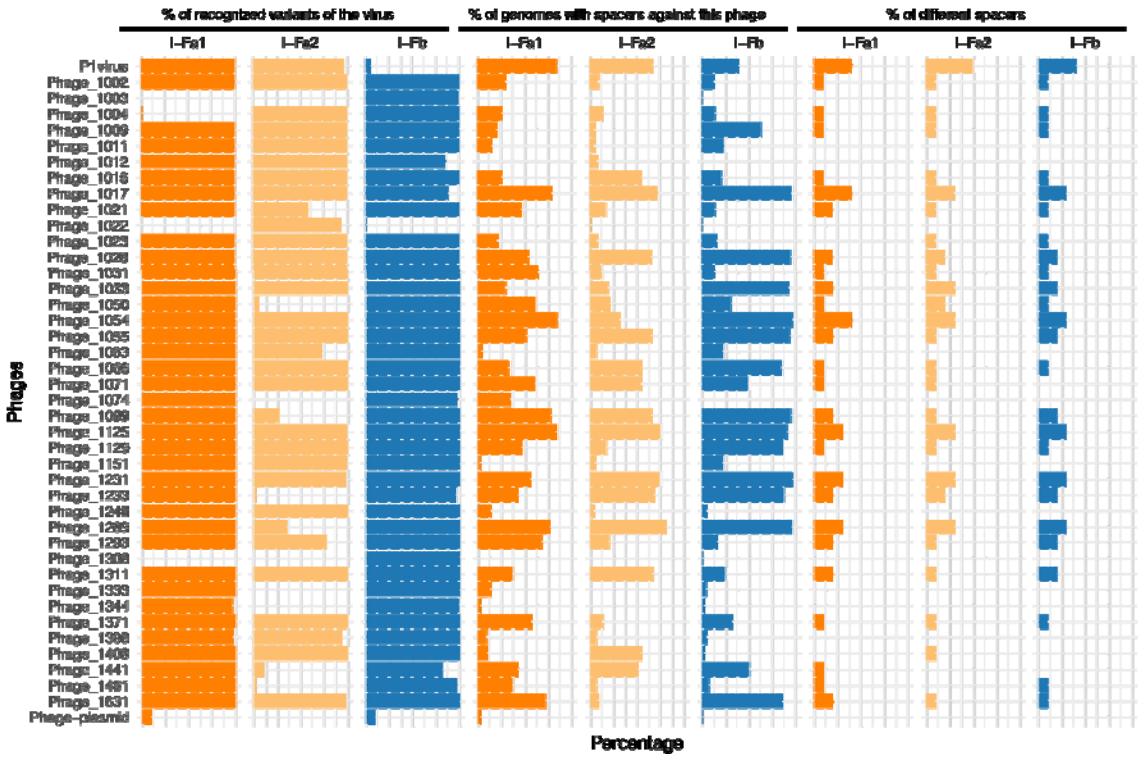


Fig. 6. Spacers versus phages separated by CRISPR array type. The first block of 3 columns shows the proportion of different variants of each phage that are targeted by spacers of each type. Those in the second block show the proportion of bacterial genomes that have spacers against each of the phages. And those in the third block show the percentage of non-redundant spacers of each CRISPR type with respect to the total number of different spacers of that type. In all cases the X-axis shows the ratio between 0-100%, except in the third block, where it is shown between 0-10%, due to its lower values.

Notably, 44% of the CRISPR-Cas I-Fa array spacers recognize all 41 of these phages. However, only 18% of the spacers of CRISPR-Cas I-Fb systems recognize these phages. This suggests some specialization on frequent phages by the I-Fa system. On the other hand, phage-plasmid spacers target 17 phages, among which P1virus is not. But this could be due to the low variability of phage-plasmid CRISPR arrays that we have or have been able to classify efficiently (5 different spacers in C1a and 76 in C2a). Although another explanation could be that the spacers against this virus were only acquired when the CRISPR-Cas system is integrated into the bacterial chromosome.

Finally, phage-plasmid variants were rarely targeted by any of the CRISPR-Cas systems. However, some phage-plasmid variants were recognized by a few C1a spacers. The main spacer that targeted the phage-plasmid came precisely from C1a of the phage-plasmid itself. This spacer matches 100% identity with the coding region of the *cas2_3* gene of the CRISPR-Cas I-Fa system. In fact, it highlights the idea that the homologous region of the gene *cas2_3* from the CRISPR-Cas I-Fb system is also the most conserved region of that gene, and in both *cas2_3* genes, the protospacer has the PAM sequence conserved (CC) almost 100% (Suppl. Fig. S4).

P1virus attempts to escape the CRISPR-Cas I-Fa system

The fact that multiple spacers were found against multiple sequence variants of the virus suggests that P1virus could be escaping the CRISPR-Cas system by fixing mutations in the protospacer or the corresponding PAM sequence. This would in turn force the bacterium to generate new spacers against the same virus.

To test this, we searched for the protospacers targeted by the spacers of the different types of CRISPR arrays. First, we searched for those that recognized sequences of the prophage P1virus. In the case of the CRISPR-Cas I-Fb system very few protospacers were found, as had been seen before, but most of them conserved the expected PAM sequence CC (Fig. 7A). In the case of C2a more protospacers were found, but most recognized the same sequence, again retaining the PAM sequence with the CC consensus. However, in the case of C1a the number of spacers increased almost 7-fold with respect to C2a and although the main PAM sequence was still CC, it was no longer as conserved and both positions could change to A or T, mainly in the position -1 of the PAM sequence (Fig. 7B). This suggests that the acquisition of mutations in the phage PAM sequence is favored to try to evade a cognate CRISPR-Cas system that seems to be specialized against it. This would also be supported by the number of protospacers present in the phage P1virus, which could also suggest that these are mutations to escape the bacterial defense system.

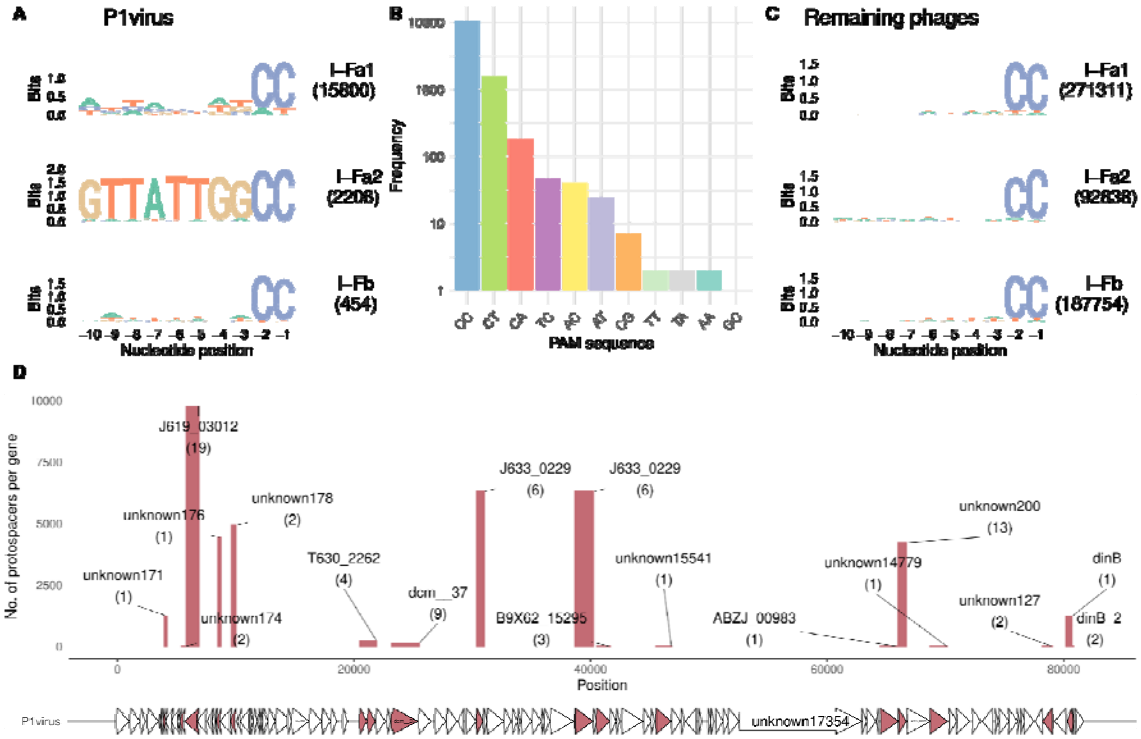


Fig. 7. Spacers that recognize bacterial phages. (A) PAM sequence (10 nucleotides) upstream of the protospacers of phage P1Virus. The number in parentheses shows the number of different protospacers found. (B) The same as in (A) but with the rest of the bacterial phages. (C) Frequencies of consensus PAM sequences in the protospacers of phage P1virus. The y-axis is in logarithmic scale. (D) Number of protospacers found in the

different genes of the phage P1virus. The number in parentheses below the gene name shows the number of protospacers that are different for that gene.

When the protospacers of the other bacterial prophages were analyzed, a high number of them were found in both CRISPR-Cas I-Fb and I-Fa1 systems, and approximately half in C2a (Fig. 7C). In addition, the PAM sequence was maintained as CC in most of the protospacers.

Finally, to map the protospacers of the phage P1virus, all phage genes appearing in the bacterial pangenome were taken and compared with the spacers of the CRISPR-Cas I-Fa system. Thus, protospacers could be found in 17 phage genes, with several protospacers of different sequences appearing in most genes (Fig. 7D).

Discussion

A small proportion of genomes of the pathogenic bacterium *A. baumannii* have long been known to exhibit two subtypes of type I-F CRISPR-Cas systems (8, 9). These were named as subtypes I-Fa and I-Fb. According to the most recent classification of CRISPR-Cas systems, the I-Fa subtype has been classified as I-F2, and the I-Fb subtype as I-F1. However, as we have found here, subtype I-Fa presents the *cas8f* gene, contrary to what has been described for subtype I-F2 (12). However, these differences in the established configurations of these defense systems are commonly found in virus-borne CRISPR-Cas systems (18).

The knowledge of CRISPR-Cas systems in pathogenic bacteria, such as *A. baumannii*, is of interest. On the one hand, the presence of these systems can prevent the acquisition of resistance and virulence genes by the bacterium, having been found that genomes with I-Fa present a lower number of these genes than I-Fb, and even less than genomes without CRISPR-Cas systems (8, 15, 16). However, this relationship has not been confirmed in other studies (9). On the other hand, in the context of the currently flourishing phage therapy (use of phages to treat bacterial infections), it could be of special interest to know whether the strain that is producing the infection has effective defense systems against the phage to be used (30).

CRISPR-Cas systems are part of the accessory genome of prokaryotes, and it is known that they can be acquired or lost depending on the environment (8, 31). The manner in which these are acquired is not yet known. They have been associated with transposons and phages, which could serve as vectors to facilitate the horizontal transfer of these defense systems (32–34).

We have found a CRISPR-Cas I-Fa system associated with a phage-plasmid, and this same defense system also appears to be sometimes associated with a transposon, and much more frequently integrated into a bacterial tmRNA. This integration site is a genomic archipelago in which different genomic islands are integrated. Although some harbor alternative restriction-methylation defense systems, the site is mostly colonized by a prophage. When present, the prophage seems to exclude the CRISPR-Cas I-Fa system, and vice versa. In a previous study, in which this genomic island was denoted as fGR4, the integration site was not found to be a tmRNA gene, but the P1virus toxin-antitoxin system and the presence of methylases were found (35). In another study, in which they named the

island G13, they only mentioned the presence of a bacteriophage and methylation-restriction systems, in addition to other island genes (22).

The only phages characterized so far to contain a CRISPR-Cas system, which they use to defend itself against a second phage, are the *V. chloreae* phages ICP1 and vibriophage 1.161 (19, 36). In the case of ICP1, both phages appear to compete with each other using the bacterium as a battlefield. The CRISPR-Cas system presented by ICP1 is also of type I-F, and presents a similar configuration to I-Fa of *A. baumannii* (except that *cas1* is rearranged), including the two CRISPR arrays. In addition, here we have found sequence similarity between their *cas* genes. However, so far it has not been reported that ICP1 integrated its CRISPR-Cas system into the bacterial genome, something that could be useful in the aforementioned phage warfare. In this case, the satellite virus is characterized as a phage-inducible chromosomal island (PICI). PICI are phage-derived satellite elements that integrate into the host chromosome, but are reactivated and excised upon phage infection and do not encode any virion components, as they hijack those of the other phage (known as helper) (37). PICI are maintained in the bacteria, among other advantages because they prevent infection and lysis of the helper phage. The phage P1virus found here could constitute a new PICI. In fact, some prediction tools only find an incomplete phage in its sequence. However, it should be noted that several genes encoding capsid proteins are found in it.

We have found that the I-Fa system has a large number of spacers against the phage P1virus, especially from array C1a. This array seems to acquire spacers against the different variants of the virus, which probably arise to try to escape from the defense system. The array seems to acquire spacers against P1virus after the CRISPR-Cas I-Fa system is integrated into the bacterium. In addition, the array C1a appears to acquire new spacers against common phages in the bacterium, which could also positively impact the success of the phage-plasmid by eliminating competitors.

Another way for phages to escape CRISPR-Cas systems, in addition to modifying their protospacer sequences, is by mutating the PAM sequence. This is something that has long been reported, and would be an efficient way to prevent recognition of the protospacer (38, 39). Specifically, phage ICP1 is less efficient in lysing *V. cholerae* when its satellite phage has non-canonical PAM sequences (29). Here we have seen that, although it may be a strategy common to all *A. baumannii* viruses recognized by CRISPR-Cas systems, it seems of particular relevance in the case of phage P1virus and the C1a array of I-Fa system. In addition, the presence of multiple spacers spanning different regions of the P1virus sequence may ensure virus neutralization by the CRISPR-Cas system, as has been demonstrated in other cases (29, 39, 40).

However, at least in the hospital environment, it seems that the phage-plasmid carrying the CRISPR-Cas system has not been able to gain the upper hand. ST2 strains of *A. baumannii* are the most commonly isolated in outbreaks of this bacterium, whereas ST79 and several other strains, which more frequently carry the CRISPR-Cas I-Fa system, are isolated more rarely, probably because their infections have a better prognosis (6, 41). This could be in agreement with the already described weaknesses that carrying such a large sequence as a CRISPR-Cas system by a virus may entail, thus requiring alternative systems (42).

However, the CRISPR-Cas I-Fa system, and more specifically thanks to its array C1a, seems to be an efficient defense against the most frequent phages that integrate into bacteria, apart from P1viruses. This, indirectly, would also benefit the phage-plasmid that carries this defense system.

The CRISPR-Cas I-F system of the phage ICP1 suppresses the antiphage function of the satellite but the mechanism used for this is still under study (43). We have found that some phage-plasmids carrying the CRISPR-Cas I-Fa system carry spacers against the most conserved region of the coding sequence of the *cas2_3* gene, both of the I-Fa subtype itself and of the I-Fb subtype. This could be a clue to their function.

This specific spacer appears in the array C1a of the phage-plasmid, which has a smaller number of spacers. In addition, the C1a arrays have repeats that differ from the C2a arrays in the nucleotide in the fourth position. And it has already been shown that variations in the sequence of the repeats can lead the CRISPR system to act as a regulator of endogenous genes, including *cas* genes (44). This is common in so-called mini-arrays, composed of 2 repeats (sometimes with a degenerate sequence) and a single spacer, and usually appear in the vicinity of complete CRISPR-Cas systems. These mini-arrays are often found in phages and are sometimes associated with blocking the superinfection of the phage itself. The fact that they are short CRISPR arrays close to the main CRISPR array could explain why in some phages such as ICP1 and the one described here, two CRISPR arrays with an unequal number of spacers appear.

In addition, the phage-plasmid sometimes appears to carry alternative defense systems to CRISPR-Cas, such as a restriction-methylation system. Alternative systems to fight against its rival virus have also been described in the case of phage ICP1, but with a recombinase system involved (45, 46). This fact suggests that both viruses are adapting in a guns and hire race in which each one can impose variants, which are eventually overcome by the other virus and force both to continue adapting (40, 42).

In conclusion, it seems increasingly clear that CRISPR-Cas systems can be used in phage arms races. Although, at least in the case reported here with *A. baumannii*, it seems not to be enough to succeed in the artificial hospital environment to which we have brought the pathogenic bacteria.

Acknowledgements

We thank C3UPO for the HPC support.

Funding

This work has been supported by MCIN/AEI/ PID2020-114861GB-I00 (Agencia Estatal de Investigación / Ministry of Science and Innovation of the Spanish Government), and by the European Regional Development Fund and the Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía (PY20_00871).

Competing Interests

The authors declare they have no competing interests.

Data and Materials Availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

The code used to analyze the data is available in the repository:
https://github.com/UPOBioinfo/cancun

References

1. E. Tacconelli, *et al.*, Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).
2. A. Y. Peleg, H. Seifert, D. L. Paterson, Acinetobacter baumannii: emergence of a successful pathogen. *Clin. Microbiol. Rev.* **21**, 538–582 (2008).
3. J. M. Cisneros, *et al.*, Bacteremia due to Acinetobacter baumannii: epidemiology, clinical findings, and prognostic features. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **22**, 1026–1032 (1996).
4. C. M. Harding, S. W. Hennon, M. F. Feldman, Uncovering the mechanisms of Acinetobacter baumannii virulence. *Nat. Rev. Microbiol.* **16**, 91–102 (2018).
5. K. A. Jolley, J. E. Bray, M. C. J. Maiden, Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* **3**, 124 (2018).
6. M. Hamidian, S. J. Nigro, Emergence, molecular mechanisms and global spread of carbapenem-resistant Acinetobacter baumannii. *Microb. Genomics* **5**, e000306 (2019).
7. K. S. Makarova, *et al.*, Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
8. A. Rubio, *et al.*, Analysis of bacterial pangenomes reduces CRISPR dark matter and reveals strong association between membranome and CRISPR-Cas systems. *Sci. Adv.* **9**, eadd8911 (2023).
9. G. Yadav, R. Singh, In silico analysis reveals the co-existence of CRISPR-Cas type I-F1 and type I-F2 systems and its association with restricted phage invasion in Acinetobacter baumannii. *Front. Microbiol.* **13** (2022).
10. E. L. Mangas, *et al.*, Pangenome of Acinetobacter baumannii uncovers two groups of genomes, one of them with genes involved in CRISPR/Cas defence systems associated with the absence of plasmids and exclusive genes for biofilm formation. *Microb. Genomics* **5** (2019).
11. E. Pfeifer, J. A. Moura de Sousa, M. Touchon, E. P. C. Rocha, Bacteria have numerous distinctive groups of phage-plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Res.* **49**, 2655–2673 (2021).
12. E. V. Koonin, J. S. Gootenberg, O. O. Abudayyeh, Discovery of Diverse CRISPR-Cas Systems and Expansion of the Genome Engineering Toolbox. *Biochemistry* (2023) https:/doi.org/10.1021/acs.biochem.3c00159 (May 17, 2023).
13. L. A. Marraffini, E. J. Sontheimer, CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* **11**, 181–190 (2010).
14. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiol. Read. Engl.* **155**, 733–740 (2009).

15. M. Tyumentseva, *et al.*, Genomic and Phenotypic Analysis of Multidrug-Resistant Acinetobacter baumannii Clinical Isolates Carrying Different Types of CRISPR/Cas Systems. *Pathogens* **10**, 205 (2021).
16. Y. Wang, *et al.*, CRISPR-Cas in Acinetobacter baumannii Contributes to Antibiotic Susceptibility by Targeting Endogenous AbaI. *Microbiol. Spectr.* **10**, e0082922 (2022).
17. A. M. A. Silva, *et al.*, Analysis of CRISPR/Cas Genetic Structure, Spacer Content and Molecular Epidemiology in Brazilian Acinetobacter baumannii Clinical Isolates. *Pathog. Basel Switz.* **12**, 764 (2023).
18. B. Al-Shayeb, *et al.*, Diverse virus-encoded CRISPR-Cas systems include streamlined genome editors. *Cell* **185**, 4574-4586.e16 (2022).
19. K. D. Seed, D. W. Lazinski, S. B. Calderwood, A. Camilli, A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489–491 (2013).
20. C. Howard-Varona, K. R. Hargreaves, S. T. Abedon, M. B. Sullivan, Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J.* **11**, 1511–1520 (2017).
21. K. P. Williams, Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res.* **30**, 866–875 (2002).
22. P. P. Di Nocera, F. Rocco, M. Giannouli, M. Triassi, R. Zarrilli, Genome organization of epidemic Acinetobacter baumanniistrains. *BMC Microbiol.* **11**, 224 (2011).
23. C. Bouyioukos, S. Reverchon, F. Képès, From multiple pathogenicity islands to a unique organized pathogenicity archipelago. *Sci. Rep.* **6**, 27978 (2016).
24. D. S. Wishart, *et al.*, PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Res.* **51**, W443–W450 (2023).
25. J. C. Pons, *et al.*, VPF-Class: taxonomic assignment and host prediction of uncultivated viruses based on viral protein families. *Bioinforma. Oxf. Engl.* **37**, 1805–1813 (2021).
26. V. A. Cantu, *et al.*, PhANNs, a fast and accurate tool and web server to classify phage structural proteins. *PLOS Comput. Biol.* **16**, e1007845 (2020).
27. F. Bertels, O. K. Silander, M. Pachkov, P. B. Rainey, E. van Nimwegen, Automated Reconstruction of Whole-Genome Phylogenies from Short-Sequence Reads. *Mol. Biol. Evol.* **31**, 1077–1088 (2014).
28. D. Couvin, *et al.*, CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).
29. A. C. McKitterick, K. N. LeGault, A. Angermeyer, M. Alam, K. D. Seed, Competition between mobile genetic elements drives optimization of a phage-encoded CRISPR-Cas system: insights from a natural arms race. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **374**, 20180089 (2019).
30. S. A. Strathdee, G. F. Hatfull, V. K. Mutalik, R. T. Schooley, Phage therapy: From biological mechanisms to future directions. *Cell* **186**, 17–31 (2023).
31. E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. *Nat. Rev. Genet.* **21**, 119–131 (2020).
32. A. Butiuc-Keul, A. Farkas, R. Carpa, D. Iordache, CRISPR-Cas System: The Powerful Modulator of Accessory Genomes in Prokaryotes. *Microb. Physiol.*, 1–16 (2021).
33. J. E. Peters, K. S. Makarova, S. Shmakov, E. V. Koonin, Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl. Acad. Sci.* **114**, E7358–E7366 (2017).
34. G. Faure, *et al.*, CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.* (2019) https:/doi.org/10.1038/s41579-019-0204-7.
35. A. P. Chan, *et al.*, A novel method of consensus pan-chromosome assembly and large-scale comparative analysis reveal the highly flexible pan-genome of Acinetobacter baumannii. *Genome Biol.* **16**, 143 (2015).
36. C. M. Boyd, *et al.*, Bacteriophage ICP1: A Persistent Predator of Vibrio cholerae. *Annu. Rev. Virol.* **8**, 285–304 (2021).
37. A. Fillol-Salom, *et al.*, Phage-inducible chromosomal islands are ubiquitous within the

bacterial universe. *ISME J.* **12**, 2114–2128 (2018).
38. C. L. Sun, *et al.*, Phage mutations in response to CRISPR diversification in a bacterial population. *Environ. Microbiol.* **15**, 463–470 (2013).
39. M. A. Schelling, G. T. Nguyen, D. G. Sashital, CRISPR-Cas effector specificity and cleavage site determine phage escape outcomes. *PLOS Biol.* **21**, e3002065 (2023).
40. B. J. O'Hara, Z. K. Barth, A. C. McKitterick, K. D. Seed, A highly specific phage defense system is a conserved feature of the Vibrio cholerae mobilome. *PLoS Genet.* **13**, e1006838 (2017).
41. J. Garnacho-Montero, *et al.*, Acinetobacter baumannii in critically ill patients: Molecular epidemiology, clinical features and predictors of mortality. *Enferm. Infecc. Microbiol. Clin.* **34**, 551–558 (2016).
42. Z. K. Barth, M. H. Nguyen, K. D. Seed, A chimeric nuclease substitutes a phage CRISPR-Cas system to provide sequence-specific immunity against subviral parasites. *eLife* **10**, e68339 (2021).
43. M. Zhang, *et al.*, Mechanistic insights into DNA binding and cleavage by a compact type I-F CRISPR-Cas system in bacteriophage. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2215098120 (2023).
44. S. A. Shmakov, *et al.*, Widespread CRISPR-derived RNA regulatory elements in CRISPR-Cas systems. *Nucleic Acids Res.*, gkad495 (2023).
45. M. H. T. Nguyen, Z. Netter, A. Angermeyer, K. D. Seed, A phage weaponizes a satellite recombinase to subvert viral restriction. *Nucleic Acids Res.* **50**, 11138–11153 (2022).
46. Z. Netter, C. M. Boyd, T. V. Silvas, K. D. Seed, A phage satellite tunes inducing phage gene expression using a domesticated endonuclease to balance inhibition and virion hijacking. *Nucleic Acids Res.* **49**, 4386–4401 (2021).

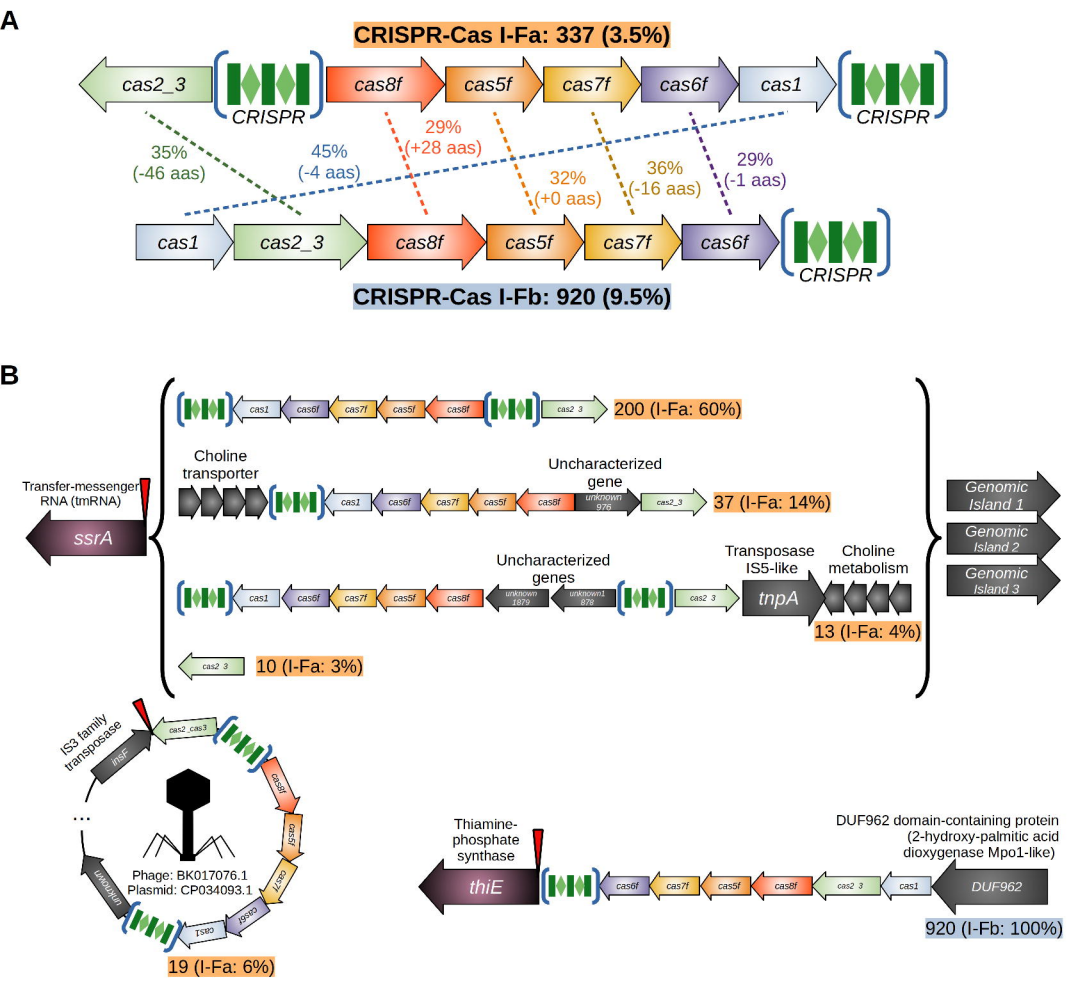
Supplementary Files

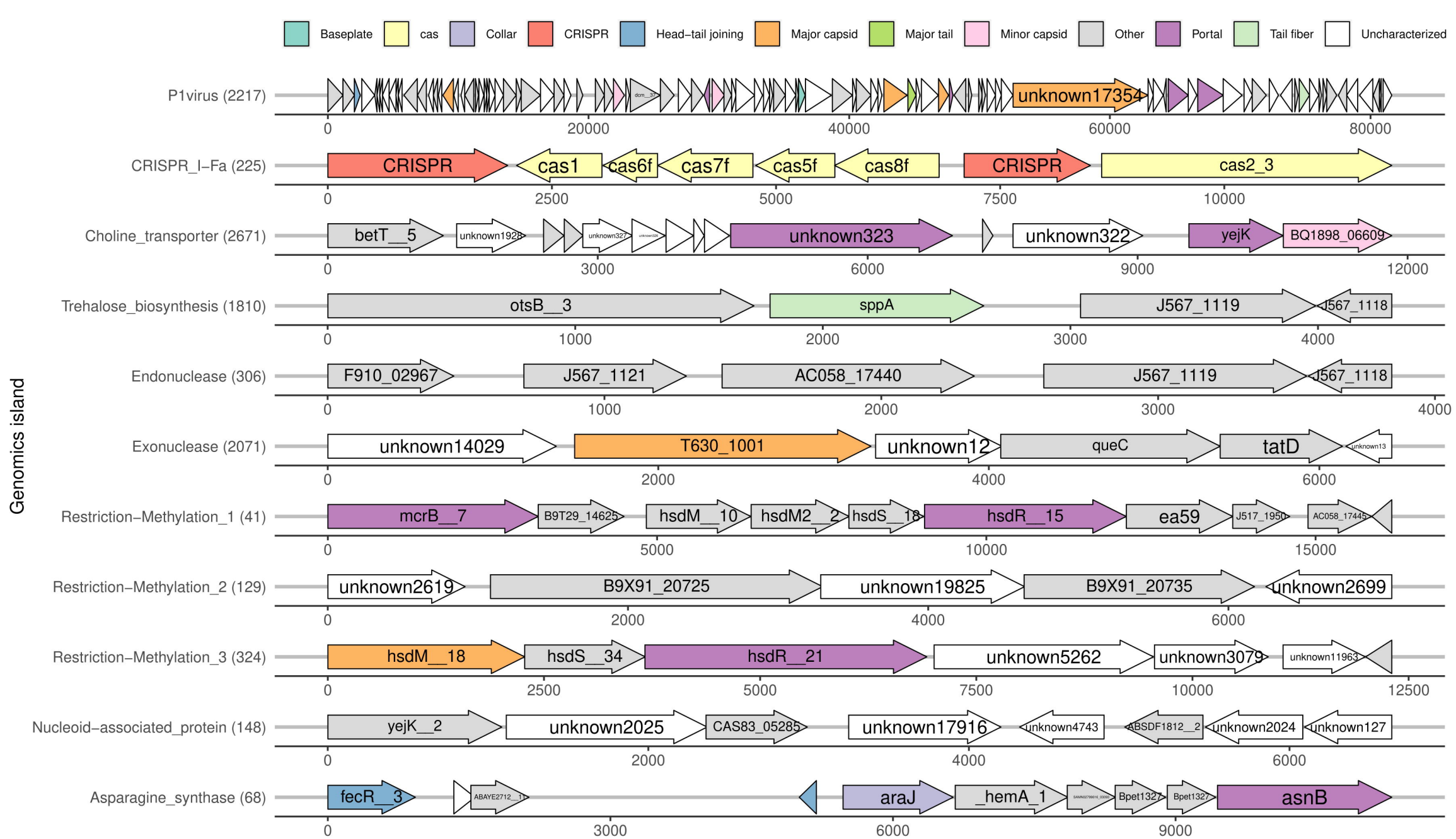
Suppl. Fig. S1. Phage containing the CRISPR I-Fa system. (A) Results when searching the NCBI Blast web portal. All results found are phage containing the CRISPR-Cas I-Fa system or not (gap at the 3' end). (B) Phage predicted by the phage prediction web tool PHASTEST, by analyzing the phage-plasmid sequence.

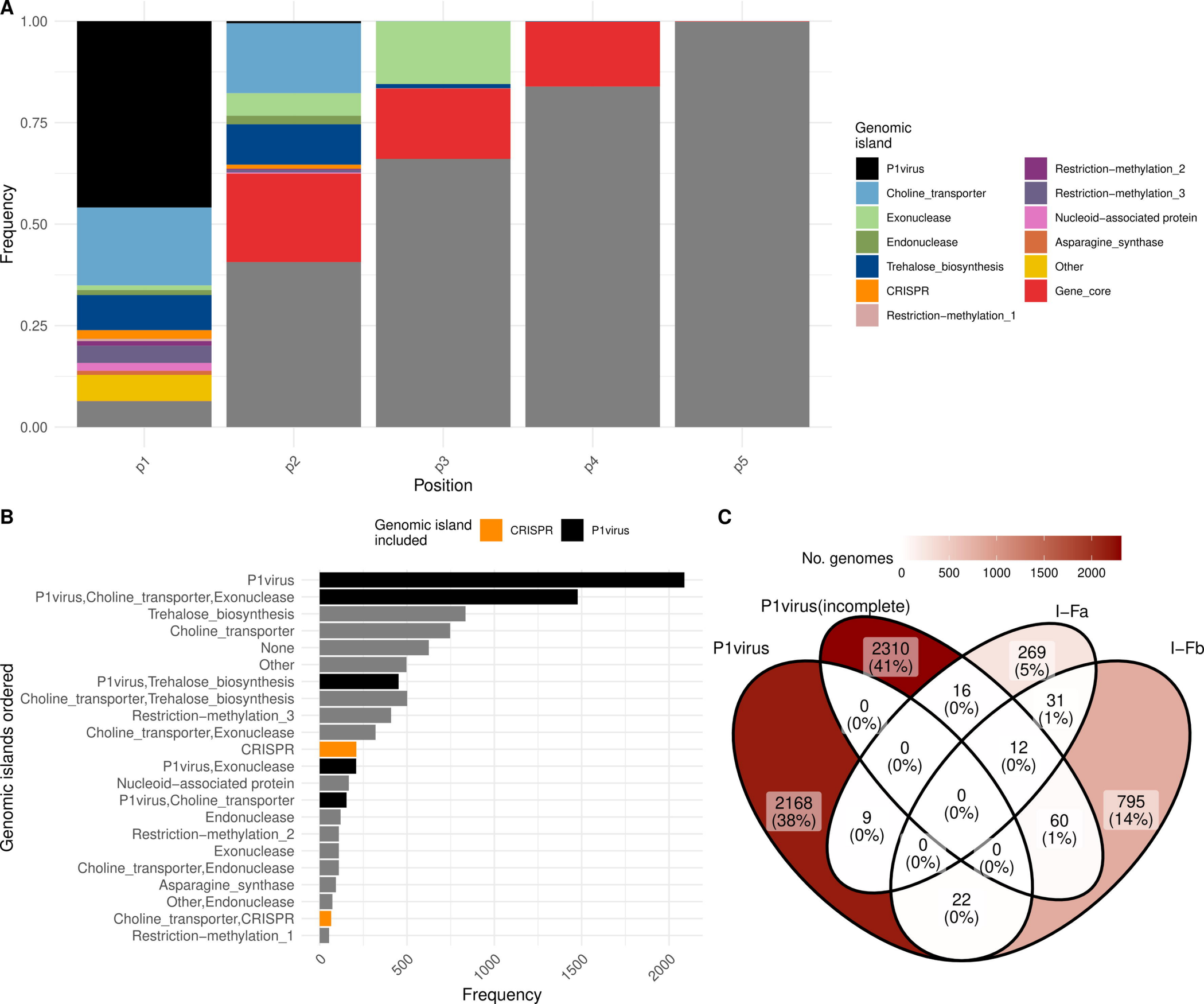
Suppl. Fig. S2. Presence/absence of the most frequent phages in *A. baumannii* genomes.

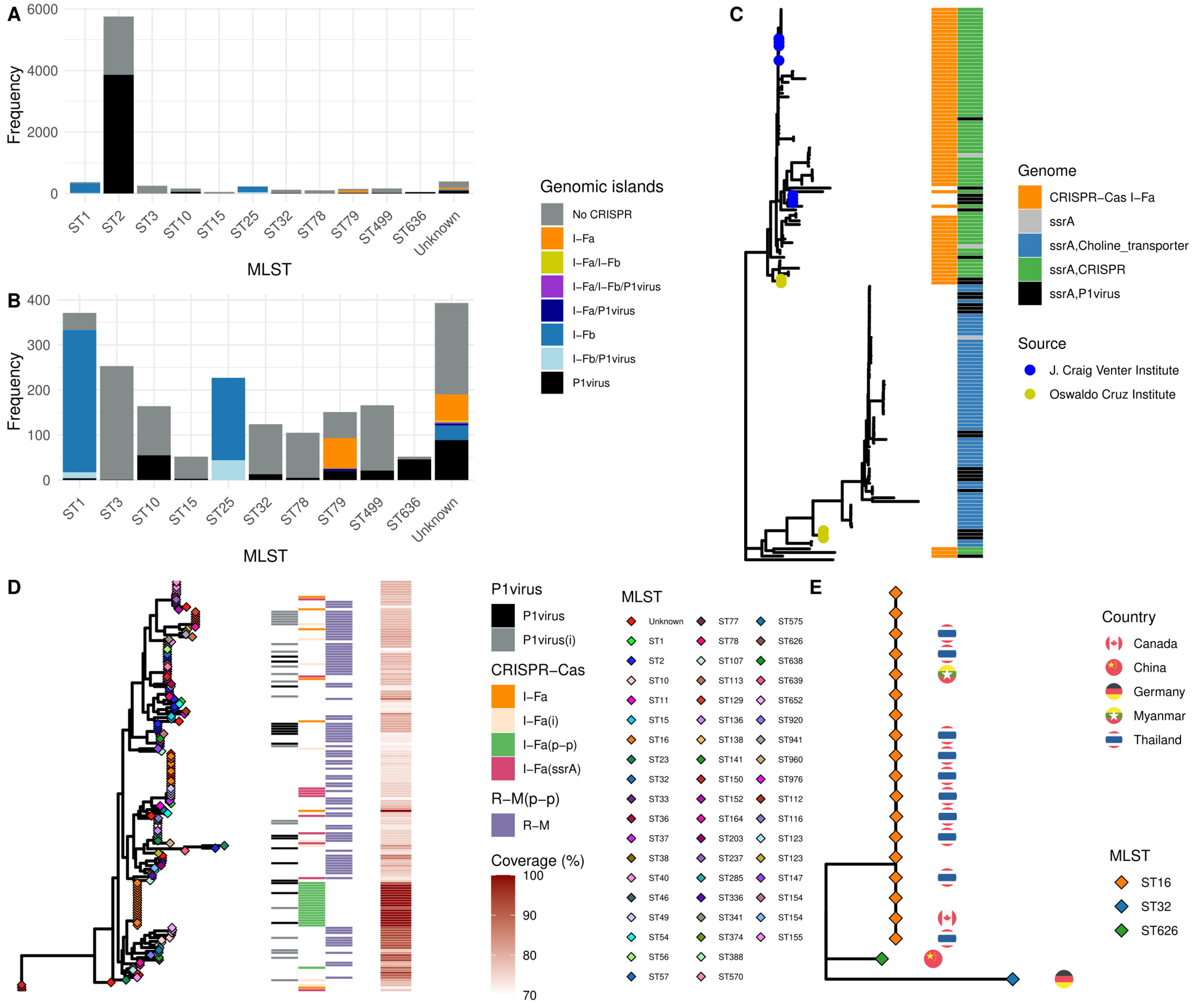
Suppl. Fig. S3. Same fig.6 but using absolute values instead of proportions.

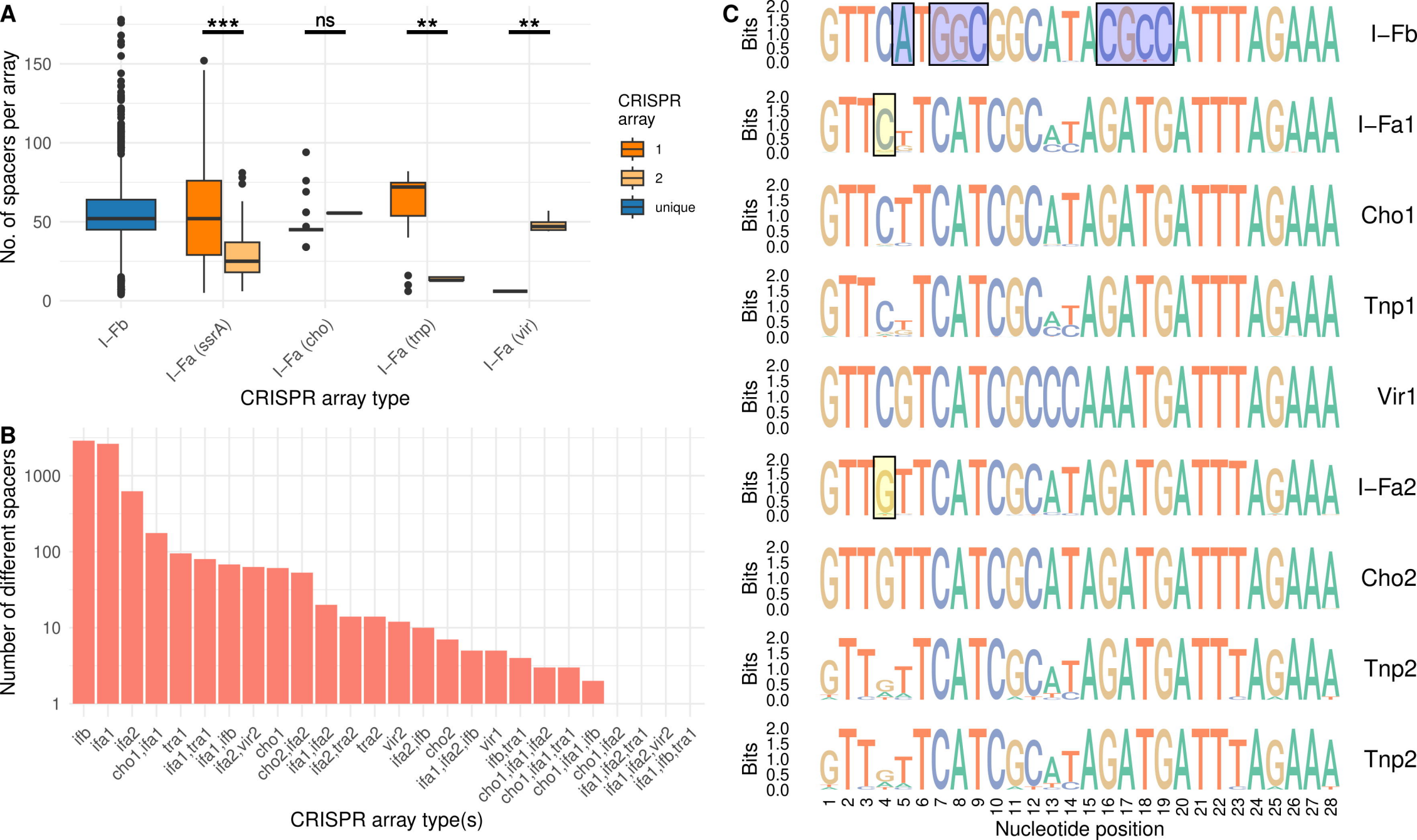
Suppl. Fig. S4. Alignment of the *cas2_3* reference coding sequences of the CRISPR-Cas I-Fa and I-Fb systems. The region of the putative phage-plasmid proto-spacer containing the I-Fa system is highlighted in red, and the corresponding PAM sequence in green.

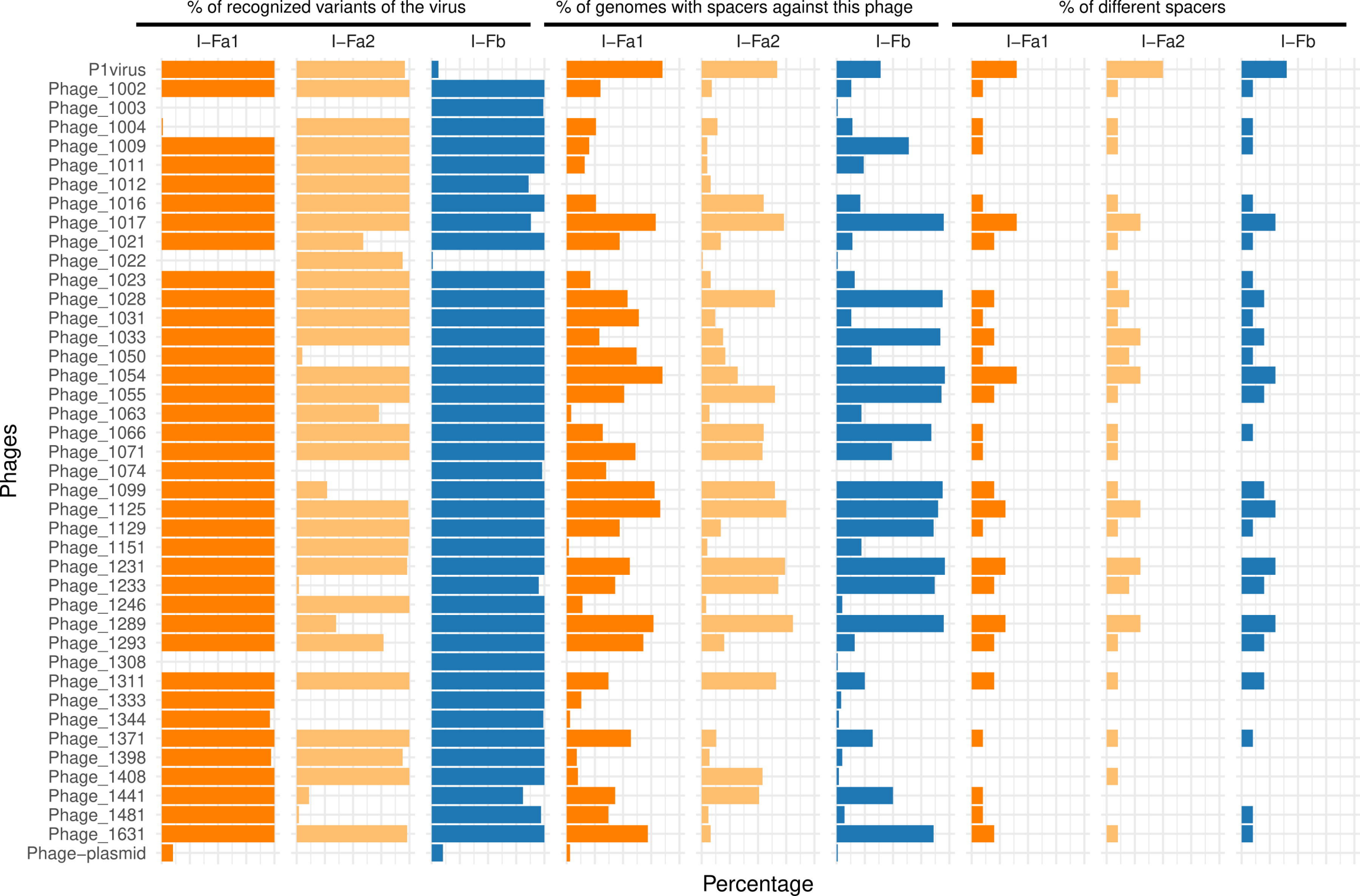

A
CRISPR-Cas I-Fa: 337 (3.5%)
cas2_3
CRISPR
cas8f
cas5f
cas7f
cas6f
cas1
CRISPR
35%
(-46 aas)
45%
(-4 aas)
29%
(+28 aas)
32%
(+0 aas)
36%
(-16 aas)
29%
(-1 aas)
cas1
cas2_3
cas8f
cas5f
cas7f
cas6f
CRISPR
CRISPR-Cas I-Fb: 920 (9.5%)
B
Transfer-messenger RNA (tmRNA)
ssrA
200 (I-Fa: 60%)
Choline transporter
Uncharacterized gene
37 (I-Fa: 14%)
Uncharacterized genes
Transposase IS5-like
Choline metabolism
tnpA
13 (I-Fa: 4%)
10 (I-Fa: 3%)
Genomic Island 1
Genomic Island 2
Genomic Island 3
IS3 family transposase
Phage: BK017076.1
Plasmid: CP034093.1
19 (I-Fa: 6%)
Thiamine-phosphate synthase
thiE
DUF962 domain-containing protein (2-hydroxy-palmitic acid dioxygenase Mpo1-like)
DUF962
920 (I-Fb: 100%)

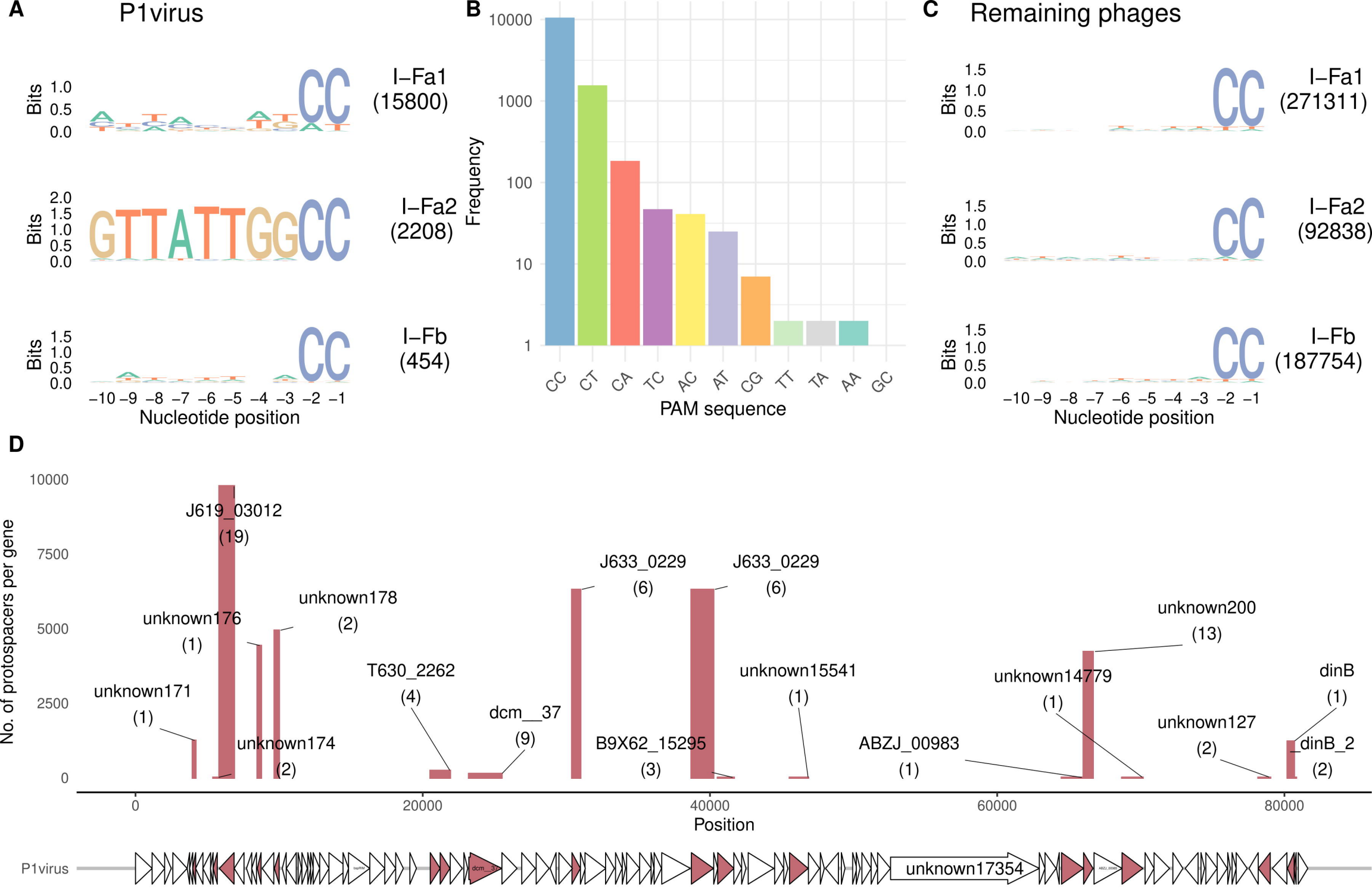

Baseplate
cas
Collar
CRISPR
Head−tail joining
Major capsid
Major tail
Minor capsid
Other
Portal
Tail fiber
Uncharacterized
Genomics island
P1virus (2217)
unknown17354
0
20000
40000
60000
80000
CRISPR_I−Fa (225)
CRISPR
cas1
cas6f
cas7f
cas5f
cas8f
CRISPR
cas2_3
0
2500
5000
7500
10000
Choline_transporter (2671)
betT__5
unknown1928
unknown323
unknown322
yejK
BQ1898_06609
0
3000
6000
9000
12000
Trehalose_biosynthesis (1810)
otsB__3
sppA
J567_1119
J567_1118
0
1000
2000
3000
4000
Endonuclease (306)
F910_02967
J567_1121
AC058_17440
J567_1119
J567_1118
0
1000
2000
3000
4000
Exonuclease (2071)
unknown14029
T630_1001
unknown12
queC
tatD
unknown13
0
2000
4000
6000
Restriction−Methylation_1 (41)
mcrB__7
B9T29_14625
hsdM__10
hsdM2__2
hsdS__18
hsdR__15
ea59
J517_1950
AC058_17445
0
5000
10000
15000
Restriction−Methylation_2 (129)
unknown2619
B9X91_20725
unknown19825
B9X91_20735
unknown2699
0
2000
4000
6000
Restriction−Methylation_3 (324)
hsdM__18
hsdS__34
hsdR__21
unknown5262
unknown3079
unknown11963
0
2500
5000
7500
10000
12500
Nucleoid−associated_protein (148)
yejK__2
unknown2025
CAS83_05285
unknown17916
unknown4743
ABSDF1812__2
unknown2024
unknown127
0
2000
4000
6000
Asparagine_synthase (68)
fecR__3
ABAYE2712__11
araJ
__hemA_1
Bpet1327
Bpet1327
asnB
0
3000
6000
9000


A
1.00
0.75
0.50
0.25
0.00
Frequency
p1
p2
p3
p4
p5
Position
Genomic island
P1virus
Choline_transporter
Exonuclease
Endonuclease
Trehalose_biosynthesis
CRISPR
Restriction−methylation_1
Restriction−methylation_2
Restriction−methylation_3
Nucleoid−associated protein
Asparagine_synthase
Other
Gene_core
B
Genomic island included
CRISPR
P1virus
Genomic islands ordered
P1virus
P1virus,Choline_transporter,Exonuclease
Trehalose_biosynthesis
Choline_transporter
None
Other
P1virus,Trehalose_biosynthesis
Choline_transporter,Trehalose_biosynthesis
Restriction−methylation_3
Choline_transporter,Exonuclease
CRISPR
P1virus,Exonuclease
Nucleoid−associated protein
P1virus,Choline_transporter
Endonuclease
Restriction−methylation_2
Exonuclease
Choline_transporter,Endonuclease
Asparagine_synthase
Other,Endonuclease
Choline_transporter,CRISPR
Restriction−methylation_1
0
500
1000
1500
2000
Frequency
C
No. genomes
0
500
1000
1500
2000
P1virus(incomplete)
I−Fa
P1virus
I−Fb
2310 (41%)
269 (5%)
0 (0%)
16 (0%)
31 (1%)
0 (0%)
12 (0%)
2168 (38%)
9 (0%)
0 (0%)
60 (1%)
795 (14%)
0 (0%)
0 (0%)
22 (0%)


A
Frequency
6000
4000
2000
0
ST1
ST2
ST3
ST10
ST15
ST25
ST32
ST78
ST79
ST499
ST636
Unknown
MLST
B
Frequency
400
300
200
100
0
ST1
ST3
ST10
ST15
ST25
ST32
ST78
ST79
ST499
ST636
Unknown
MLST
Genomic islands
No CRISPR
I–Fa
I–Fa/I–Fb
I–Fa/I–Fb/P1virus
I–Fa/P1virus
I–Fb
I–Fb/P1virus
P1virus
C
Genome
CRISPR–Cas I–Fa
ssrA
ssrA,Choline_transporter
ssrA,CRISPR
ssrA,P1virus
Source
J. Craig Venter Institute
Oswaldo Cruz Institute
D
P1virus
P1virus
P1virus(i)
CRISPR–Cas
I–Fa
I–Fa(i)
I–Fa(p–p)
I–Fa(ssrA)
R–M(p–p)
R–M
Coverage (%)
100
90
80
70
MLST
Unknown
ST1
ST2
ST10
ST11
ST15
ST16
ST23
ST32
ST33
ST36
ST37
ST38
ST40
ST46
ST49
ST54
ST56
ST57
ST77
ST78
ST107
ST113
ST129
ST136
ST138
ST141
ST150
ST152
ST164
ST203
ST237
ST285
ST336
ST341
ST374
ST388
ST570
ST575
ST626
ST638
ST639
ST652
ST920
ST941
ST960
ST976
ST112
ST116
ST123
ST123
ST147
ST154
ST154
ST155
E
Country
Canada
China
Germany
Myanmar
Thailand
MLST
ST16
ST32
ST626


A
No. of spacers per array
150
100
50
0
***
ns
**
**
I–Fb
I–Fa (ssrA)
I–Fa (cho)
I–Fa (tnp)
I–Fa (vir)
CRISPR array type
CRISPR array
1
2
unique
B
Number of different spacers
1000
100
10
1
ifb
ifa1
ifa2
cho1,ifa1
tra1
ifa1,tra1
ifa1,ifb
ifa2,vir2
cho1
cho2,ifa2
ifa1,ifa2
ifa2,tra2
tra2
vir2
ifa2,ifb
cho2
ifa1,ifa2,ifb
vir1
ifb,tra1
cho1,ifa1,ifa2
cho1,ifa1,tra1
cho1,ifa1,ifb
cho1,ifa2
ifa1,ifa2,tra1
ifa1,ifa2,vir2
ifa1,ifb,tra1
CRISPR array type(s)
C
Bits
I–Fb
I–Fa1
Cho1
Tnp1
Vir1
I–Fa2
Cho2
Tnp2
Tnp2
Nucleotide position


% of recognized variants of the virus
% of genomes with spacers against this phage
% of different spacers
I−Fa1
I−Fa2
I−Fb
I−Fa1
I−Fa2
I−Fb
I−Fa1
I−Fa2
I−Fb
Phages
P1virus
Phage_1002
Phage_1003
Phage_1004
Phage_1009
Phage_1011
Phage_1012
Phage_1016
Phage_1017
Phage_1021
Phage_1022
Phage_1023
Phage_1028
Phage_1031
Phage_1033
Phage_1050
Phage_1054
Phage_1055
Phage_1063
Phage_1066
Phage_1071
Phage_1074
Phage_1099
Phage_1125
Phage_1129
Phage_1151
Phage_1231
Phage_1233
Phage_1246
Phage_1289
Phage_1293
Phage_1308
Phage_1311
Phage_1333
Phage_1344
Phage_1371
Phage_1398
Phage_1408
Phage_1441
Phage_1481
Phage_1631
Phage−plasmid
Percentage


A
P1virus
Bits
I–Fa1
(15800)
I–Fa2
(2208)
I–Fb
(454)
Nucleotide position
B
Frequency
PAM sequence
CC
CT
CA
TC
AC
AT
CG
TT
TA
AA
GC
C
Remaining phages
I–Fa1
(271311)
I–Fa2
(92838)
I–Fb
(187754)
D
No. of protospacers per gene
J619_03012
(19)
unknown176
(1)
unknown171
(1)
unknown174
(2)
unknown178
(2)
T630_2262
(4)
dcm__37
(9)
J633_0229
(6)
B9X62_15295
(3)
J633_0229
(6)
unknown15541
(1)
ABZJ_00983
(1)
unknown14779
(1)
unknown200
(13)
unknown127
(2)
dinB
(1)
dinB_2
(2)
Position
P1virus
unknown17354