



CRISPR-Cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa*

Rachel M. Wheatley ¹ · R. Craig MacLean ¹

Received: 18 June 2020 / Revised: 6 November 2020 / Accepted: 26 November 2020 / Published online: 21 December 2020
© The Author(s) 2020. This article is published with open access

Abstract

CRISPR-Cas systems provide bacteria and archaea with an adaptive immune system that targets foreign DNA. However, the xenogenic nature of immunity provided by CRISPR-Cas raises the possibility that these systems may constrain horizontal gene transfer. Here we test this hypothesis in the opportunistic pathogen *Pseudomonas aeruginosa*, which has emerged as an important model system for understanding CRISPR-Cas function. Across the diversity of *P. aeruginosa*, active CRISPR-Cas systems are associated with smaller genomes and higher GC content, suggesting that CRISPR-Cas inhibits the acquisition of foreign DNA. Although phage is the major target of CRISPR-Cas spacers, more than 80% of isolates with an active CRISPR-Cas system have spacers that target integrative conjugative elements (ICE) or the conserved conjugative transfer machinery used by plasmids and ICE. Consistent with these results, genomes containing active CRISPR-Cas systems harbour a lower abundance of both prophage and ICE. Crucially, spacers in genomes with active CRISPR-Cas systems map to ICE and phage that are integrated into the chromosomes of closely related genomes lacking CRISPR-Cas immunity. We propose that CRISPR-Cas acts as an important constraint to horizontal gene transfer, and the evolutionary mechanisms that ensure its maintenance or drive its loss are key to the ability of this pathogen to adapt to new niches and stressors.

Introduction

CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated protein) systems are adaptive immune systems that provide heritable immunity against foreign DNA, and are widespread in bacterial and archaeal genomes [1–3]. CRISPR-Cas systems are able to incorporate segments of invading DNA, such as fragments of bacteriophage or mobile genetic elements, as spacers in CRISPR loci [4]. Active systems must contain a set of Cas genes that enable the CRISPR arrays to be transcribed and processed into short CRISPR RNAs (crRNAs) [5]. These crRNAs contain a single spacer and must be bound to the Cas endonuclease. This complex uses crRNA base

complementarity to recognise and degrade DNA from elements containing the spacer sequence upon subsequent re-infection of the cell [6]. In effect, CRISPR-Cas systems provide a molecular memory of past infections and provide bacteria and archaea with adaptive immunity against foreign DNA [1, 4, 6].

Horizontal gene transfer (HGT) plays an important role in bacterial evolution [7] and is a major source of genome expansion [8]. CRISPR-Cas systems were first recognised for their role as phage defence mechanisms [3, 9–11] and can provide protection by preventing lysogenic conversion [12, 13], which is an important mechanism of HGT [14–16]. Although CRISPR-Cas systems can target parasitic genetic elements, such as lytic phage, the xenogenic immunity provided by CRISPR-Cas may constrain HGT more broadly [17–20]. There is growing recognition that CRISPR-Cas systems target other mobile genetic elements [17, 21, 22]. It is suggested that CRISPR-Cas may play a very general role in preventing HGT by targeting integrative conjugative elements (ICE) and plasmids, or DNA that is acquired by transformation. Experimental studies have shown in a number of systems that CRISPR-Cas can prevent HGT over short time scales [3, 11, 21, 23]. For example, in *Staphylococcus epidermidis*, CRISPR-Cas

Supplementary information The online version of this article (<https://doi.org/10.1038/s41396-020-00860-3>) contains supplementary material, which is available to authorised users.

✉ Rachel M. Wheatley
rachel.wheatley@zoo.ox.ac.uk

¹ Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

systems possessing a spacer that targets a highly conserved nickase present on staphylococcal conjugative plasmids have been shown to be successful in preventing plasmid transformation [21]. Bioinformatic studies, on the other hand, have produced conflicting results on the importance of CRISPR-Cas in HGT over longer time scales [18, 21, 23–27]. For example, a recent bioinformatics study found no evidence of a correlation between CRISPR-Cas activity and the frequency of HGT [24]. Furthermore, a genome-wide correlation analysis reported that the presence of CRISPR-Cas systems constrains the acquisition of antibiotic resistance genes in only a sub-set of bacterial pathogens [18]. In summary, the role of CRISPR-Cas is well established from an experimental point of view, but the long-term consequences of this interference are not as clearly understood. In this paper, we address this problem by investigating the relationship between CRISPR-Cas systems and HGT in the opportunistic pathogen *Pseudomonas aeruginosa*.

P. aeruginosa genomes are large (typically 6–7 Mbp), and ~50% of sequenced *P. aeruginosa* genomes have been predicted to possess an active CRISPR-Cas system [18, 28]. Three major CRISPR-Cas system types (I-F, I-E and I-C) have been identified in *P. aeruginosa* [28], and *P. aeruginosa* genomes contain a large repertoire of mobile genetic elements, including phages, transposons, ICE and plasmids [29, 30]. ICE are modular mobile genetic elements that can integrate into a host genome and be vertically propagated through cell replication or transfer horizontally following excision from the chromosome [31, 32]. ICE and plasmids both use the same type IV secretion system for conjugative transfer [31, 33–35], and the difference between ICE and plasmids comes from their ability to integrate into the chromosome. Like plasmids, ICE contain cargo genes [29, 36–38], and *P. aeruginosa* ICE have been implicated in a range of traits including xenobiotic compound degeneration [39], antibiotic resistance [32, 40–42], and virulence formation [43]. Although plasmids and ICE share many similarities, ICE are abundant in *P. aeruginosa*, whereas plasmids are thought to be comparatively rare.

While it is straightforward to understand the benefits of CRISPR based immunity to obligate genetic parasites, such as lytic phage, many mobile genetic elements can be either parasitic or beneficial, depending on conditions. For example, the acquisition of prophage can improve *Pseudomonas* metabolism and increase competitive ability [16, 44], but prophage entry into the lytic cycle leads to cell lysis and death. Similarly, ICE and plasmids carry genes that can allow *Pseudomonas* to exploit new niches, such as novel metabolites or eukaryotic hosts, or resist stresses, such as heavy metals and antibiotics, but the acquisition of these elements also tends to be associated with costs that can generate selection against carriage [32, 45–47]. Given

these costs and benefits, it is difficult to predict whether CRISPR-Cas systems should target these elements. The diversity and plasticity of *P. aeruginosa* genomes combined with the high variability of CRISPR-Cas presence makes *P. aeruginosa* a very useful species to study for evidence of CRISPR-Cas mediated HGT inhibition. Previous work has shown that *P. aeruginosa* CRISPR-Cas systems are associated with small genome size [28], and reduced mobile sulphonamide resistance genes [18], and it has been suggested that *P. aeruginosa* is an example of a bacterial pathogen where CRISPR-Cas does play a recognisable role in HGT. However, the broader impacts of CRISPR-Cas on genome divergence have not been investigated in this species in detail. Here we analyse 300 high-quality assembled genomes (including 201 complete genomes) of *P. aeruginosa* to test the hypothesis that CRISPR-Cas constrains HGT, and to identify mobile genetic elements that are targeted by CRISPR-Cas. There is growing evidence that anti-CRISPR (Acr) genes play an important role in antagonising CRISPR-Cas [48–51], and our analysis also tests the hypothesis that Acr genes negate the impact of CRISPR-Cas on HGT.

Materials and methods

Genomic data

P. aeruginosa genome sequences were downloaded from NCBI RefSeq (https://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/Pseudomonas_aeruginosa/) (Supplementary Table S1). These genomes are considered complete or assembled to a high level; complete genome [201], chromosome [39] or scaffolds of 7 or fewer [52]. Genome metadata was downloaded in parallel (genome size, guanine-cytosine (GC) content, number of coding sequences (CDS)), along with isolation niche and geographical location of sample collection (Supplementary Table S1). Multi-locus sequence typing (MLST) was carried out using MLST software that scans against PubMLST typing schemes (<https://pubmlst.org/>) [53, 54]. Genome annotation was carried out using prokka [55].

CRISPR-Cas and anti-CRISPR annotation

CRISPRCasFinder was used to predict the presence of CRISPR arrays and cognate Cas proteins [56]. CRISPR-CasFinder assigns evidence levels to putative CRISPR loci on a 1–4 level scale [56], using an algorithm to measure CRISPR repeat conservation based on Shannon's entropy and produce an EBcons (entropy-based conservation) index. Evidence level 4 was used as the cut-off for annotating CRISPR loci (Supplementary Table S1), and the details of this algorithm and evidence level system are described in

[56]. CasFinder version 2.0 of CRISPRCasFinder was used to identify and type Cas systems in genomes with predicted CRISPR loci (Supplementary Table S1) [56, 57]. Acr genes were identified by screening genomes against type I-F and type I-E Acr sequences in the Acr database [51]. At the date of analysis there were no type I-C specific Acr sequences in the Acr database [51]. Type I-C genomes were screened for Acr against recently published type I-C Acr sequences by the Bondy-Denomy group; Leon et al. [58] and Marino et al. [59]. CRISPR-Cas systems were predicted to be functionally active if they were annotated to possess a CRISPR array, cognate Cas genes and the absence of Acr. All Cas systems were identified to be type I-F, I-E or I-C with the exception of one type U annotation [56] (Supplementary Table S1). This type U annotation genome was excluded from downstream analysis.

Spacer target identification

A unique spacer set ($n = 2123$) was generated by clustering spacer sequences identified in CRISPRCasFinder [56] for all CRISPR(+) genomes with CD-HIT [52, 60] using a 95% sequence identity threshold as used in previous studies [18]. Blastn was used to predict spacer targets by screening unique spacers against four databases: (1) Phage genomes (2) ICE, plasmid and conjugative transfer gene sequences (3) Resistance genes and (4) Virulence genes. Blastn hits with at least 95% sequence identity to a spacer and at least 95% sequence coverage were accepted as predicted spacer targets. This threshold was previously defined in a spacer analysis study by Shmakov et al., based on control analysis against false positive predictions comparing prokaryotic to eukaryotic virus targeting [17]. The phage genomes used in this study were downloaded from NCBI (<ftp://ftp.ncbi.nih.gov/refseq/release/viral/>), and the resulting database contained 12,182 genome sequences. Phage genomes clustered into lytic, temperate and non-lytic groups as described in [61] were used to characterise the types of phages being targeted. The ICE, plasmid and conjugative transfer genes sequences were compiled from three locations. ICE sequences were downloaded from the ICEberg 2.0 database of bacterial integrative and conjugative elements containing 552 sequences [62]. Plasmid sequences were downloaded from a curated database of plasmid sequences containing 10,892 complete plasmid sequences [63]. Details of how this plasmid database has been curated are given in Brooks et al. [63]. Conjugative transfer gene sequences (*tra* genes, *trb* genes and type IV secretion system genes) were downloaded from annotated *P. aeruginosa* genes in NCBI gene [64]. Acquired resistance gene sequences were downloaded from the ResFinder database of acquired antimicrobial resistance genes [65], and virulence genes were downloaded from the Virulence Factor Database [66, 67].

The spacers per genome were analysed for GC content and the prediction of phage or ICE and conjugative transfer system targeting. The focus of this downstream analysis was on ICE and phage, rather than plasmids. The abundance of ICE and prophage in *P. aeruginosa* genomes make them good targets to study with regards to CRISPR-Cas system correlations. Plasmids are thought to be comparatively rare, and the sample size of genomes in this study combined with the varied levels of assembly (~2/3 complete genomes) provides a dataset that we believe is not well suited to assess correlations between plasmid presence and CRISPR-Cas. As such, plasmids have not been included in the spacers per genome or intra-ST variability analysis. An excellent recent bioinformatic study by O'Meara et al. details a broad-scale analysis of plasmid carriage and CRISPR across bacterial species [19]. The GC of each spacer was calculated using a Perl script available on GitHub [68]. The average spacer GC per genome was then calculated using awk in command line. The spacer sequences per genome were searched against the phage, ICE and conjugative transfer gene datasets as previously outlined using blastn. Blastn hits with at least 95% sequence identity to a spacer and at least 95% sequence coverage were accepted as predicted spacer targets [18]. All spacers were blasted against their source genome to identify self-targeting spacers, as described in Nobrega et al. [69], and using the previously defined blastn spacer search parameters.

Analysis of intra-ST CRISPR variability

Five *P. aeruginosa* STs were identified with variable presence or absence of CRISPR-Cas systems: ST111, ST262, ST274, ST277 and ST2619. Complete CRISPR(+) and CRISPR(-) genome representatives of ST111, ST262, ST277 and ST2619 were aligned in Mauve (progressive Mauve alignment with default settings) [70] using the GenBank (.gb) files downloaded from NCBI. The complete genome sequences used for alignment are indicated in Supplementary Table S2. Unique regions annotated to contain phage or ICE in the GenBank annotation file were highlighted on the Mauve alignments. To quantify the influence of phage and ICE, we systematically searched for phage and ICE in our genomes with intra-ST CRISPR variability. The identification of prophage regions was carried out using PHASTER [71], which was used to predict a total number of prophage regions within each genome (Supplementary Table S2). The estimation of ICE abundance was carried out using a blastn search of the coding sequence annotation files (.ffn) obtained from prokka [55] against the database of ICE sequences [62] and conjugative transfer system genes in *P. aeruginosa*. From this, the number of CDS predicted to represent these conjugative

elements within each genome was standardised per genome Mb. A genome size standardised measure of ICE abundance was used so that conjugative element integration could be compared regardless of the already apparent genome size bias between CRISPR(−) and CRISPR(+) isolates.

Unique CRISPR(−) regions identified from Mauve alignment were extracted in nucleotide (.fasta) format. These were compiled to produce a single (.fasta) file containing all unique regions present in each CRISPR(−) genome compared to their CRISPR(+) counterpart within an ST (Supplementary Table S2). Blastn was used to predict whether the CRISPR(+) spacer sequences had targeting identity to these unique CRISPR(−) regions. Blastn hits with a minimum *e* value of 0.01 were accepted as predicted spacer targets [18]. Putative identity of the predicted spacer targets was taken from the annotated .gb file for each genome. The identity of hypothetical protein-encoding gene targets was further characterised using NCBI blast, searching for coding regions with high homology to the target (>90% identity across whole length). Targets that have been further characterised in this way have been indicated (^a) in Table 2.

Statistical analysis

Statistical testing was done using built-in methods in R (t.test, cor.test) [72]. An unpaired two-tailed *t*-test was used to test the association between CRISPR-Cas system presence and genome size, and CRISPR-Cas system presence and GC content. A paired sample two-tailed *t*-test was used to test the association between genome GC and spacer GC. Pearson's correlation coefficient test was used to test the correlation between phage targeting spacers and total spacers within a genome. For intra-ST analyses, differences between CRISPR(−) and CRISPR(+) genomes were analysed using a paired sample one-tailed *t*-test.

Results and discussion

Phylogenetic distribution of CRISPR-Cas in collection of *P. aeruginosa* genomes

A collection of 300 *P. aeruginosa* genomes were downloaded from NCBI RefSeq [73]. These genomes spanned a large number of Sequence Type (ST)s: 113 defined STs (271 genomes) and 29 genomes had an undefined ST (Supplementary Table S1). The isolation niches of these genomes were clinical (232/300), environmental (49/300) and data not available on entry (19/300) (Supplementary Table S1). A total of 149 of the 300 genomes were predicted to possess CRISPR arrays and accompanying type I-F, I-E or I-C Cas genes [56] (Supplementary Table S1). Degenerate systems were identified in 26 genomes that

possessed a CRISPR array but lacked cognate Cas genes (Supplementary Table S1). The proportion of genomes with predicted CRISPR-Cas systems compared to those without is in line with previous studies that have analysed a larger dataset (>600) of *P. aeruginosa* genomes [18, 28].

CRISPR-Cas systems may lose their effectiveness by the acquisition of Acr proteins [74]. Acr proteins originate from phage genomes [49], and inhibit targeting by CRISPR-Cas systems through a variety of distinct mechanisms [12, 49, 50]. Screening these genomes against an Acr database identified Acr genes in 20/149 genomes with a CRISPR-Cas system, leaving 129 genomes that were predicted to encode functional CRISPR-Cas systems (CRISPR(+) genomes). CRISPR(+) genomes spanned 56 defined STs, with 12 genomes of undefined ST. We also identified Acr genes in genomes lacking CRISPR loci and/or Cas genes (Supplementary Table S1). Interestingly, the presence of active CRISPR-Cas systems was variable in some STs, whereas other STs consisted entirely of either CRISPR(+) or CRISPR(−) genomes (Supplementary Table S1). For a more detailed analysis of the phylogenetic distribution of CRISPR-Cas systems in *P. aeruginosa*, an excellent study was carried out in 2015 by van Belkum et al. [28].

Relationship between CRISPR-Cas systems and genome size

HGT is the key source of genome expansion in bacteria [8]; for example, ~99% of the genes in γ -proteobacteria (including *P. aeruginosa*) are predicted to be acquired by HGT [75]. Given this tight link between gene acquisition and HGT, active CRISPR-Cas systems should be associated with smaller genomes if CRISPR-Cas constrains HGT. In agreement with previous work [28], we found that CRISPR-Cas systems were associated with smaller *P. aeruginosa* genome size (Fig. 1A) [28]. To further test the hypothesis that CRISPR-Cas constrains gene acquisition, we compared how genome size varied between *P. aeruginosa* genomes with or without Acr genes. Genomes with both CRISPR-Cas systems and Acr genes (CRISPR(+)/Acr) were significantly larger than CRISPR(+) genomes with no Acr genes (Fig. 1B). As Acr proteins inhibit CRISPR systems, this difference in genome size may support the hypothesis that active CRISPR-Cas systems limit HGT. Genomes lacking CRISPR loci and/or Cas genes were also found to contain Acr genes (CRISPR(−)/Acr), and were slightly larger than genomes that were both CRISPR(−) and Acr negative (Fig. 1C). We speculate whether this might be representative of strains that have acquired a large number of lysogenic phage, leading to both Acr acquisition and genome expansion. Alternatively, we add that correlations between Acr presence and genome size could also be due to strain differences in promiscuity towards mobile genetic

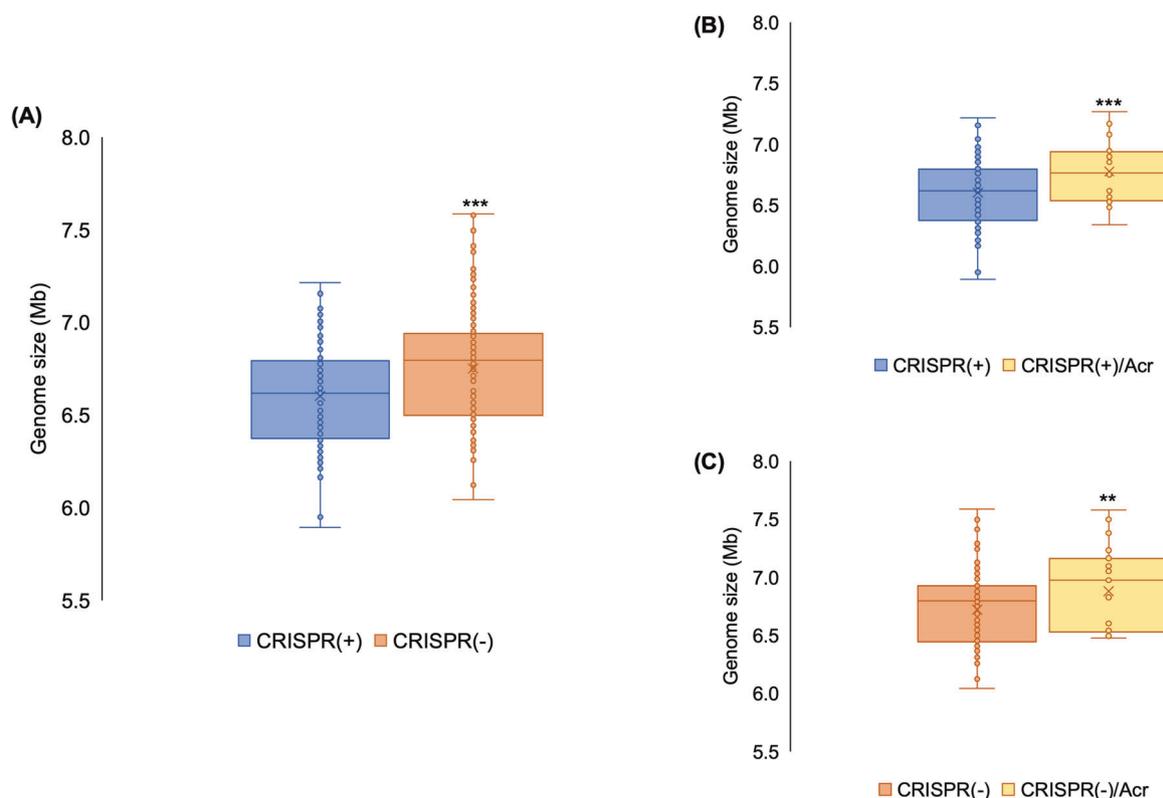


Fig. 1 CRISPR and genome size. **A** The size of CRISPR(+) and CRISPR(-) genomes; CRISPR(+) defined as genomes which contain a functionally predicted active CRISPR-Cas system and CRISPR(-) defined as genomes lacking CRISPR-Cas and/or carrying Acr genes. The means of the two groups were significantly different at $p < 0.01$ (two-tailed t -test). **B** The size of genomes containing an active

CRISPR-Cas system or both CRISPR-Cas and Acr genes. The means of the two groups were significantly different at $p < 0.01$ (two-tailed t -test). **C** The size of genomes that lack a CRISPR-Cas system that either carry or lack Acr genes. The means of the two groups were significantly different at $p < 0.05$ (two-tailed t -test).

elements. Due to the vectoring of Acr genes by mobile genetic elements, a strain with greater promiscuity towards mobile genetic elements may be more likely to contain an Acr. The observation that genomes with Acr presence are larger than those without, regardless of the presence or absence of CRISPR-Cas (Fig. 1B, C), also supports this interpretation of results.

Does CRISPR-Cas block the acquisition of potentially costly lower GC content elements?

Mobile genetic elements usually have a lower GC content than their bacterial hosts [76, 77], suggesting that HGT should be associated with reduced GC content. For example, the overall GC content of *P. aeruginosa* genomes is typically between 65 and 67% [29], which is very high in relation to most bacteria [78]. Regions of the genome with low GC are indicative of the presence of recently acquired mobile elements [79, 80]. We found that the presence of CRISPR-Cas was associated with higher genomic GC content, which is consistent with the hypothesis that CRISPR-Cas restricts the acquisition of foreign DNA (Fig. 2A). If CRISPR-Cas

restricts the acquisition of mobile elements with low GC content, then we would also expect the GC content of CRISPR loci spacers to be low relative to the rest of the genome. Our findings support this; whilst genome GC stratifies by size, spacer GC was always lower than the genome-wide GC content, and the average difference in GC composition was 5% (Fig. 2B). Across bacterial species there is a correlation between genome size and GC content [81], suggesting that an association between functional CRISPR-Cas systems and GC content may be a spurious correlation driven by small size of CRISPR(+) genomes. However, we found that the association between CRISPR-Cas presence and GC bias still held true after correcting for variation in genome size (Fig. 2C). Although a complex number of factors can influence genome GC bias, our results suggest that CRISPR-Cas systems may influence GC by preventing the acquisition of low GC elements.

What are CRISPR-Cas loci spacers targeting?

We identified a total of 2123 unique spacers across the CRISPR(+) *P. aeruginosa* genomes, and set out to

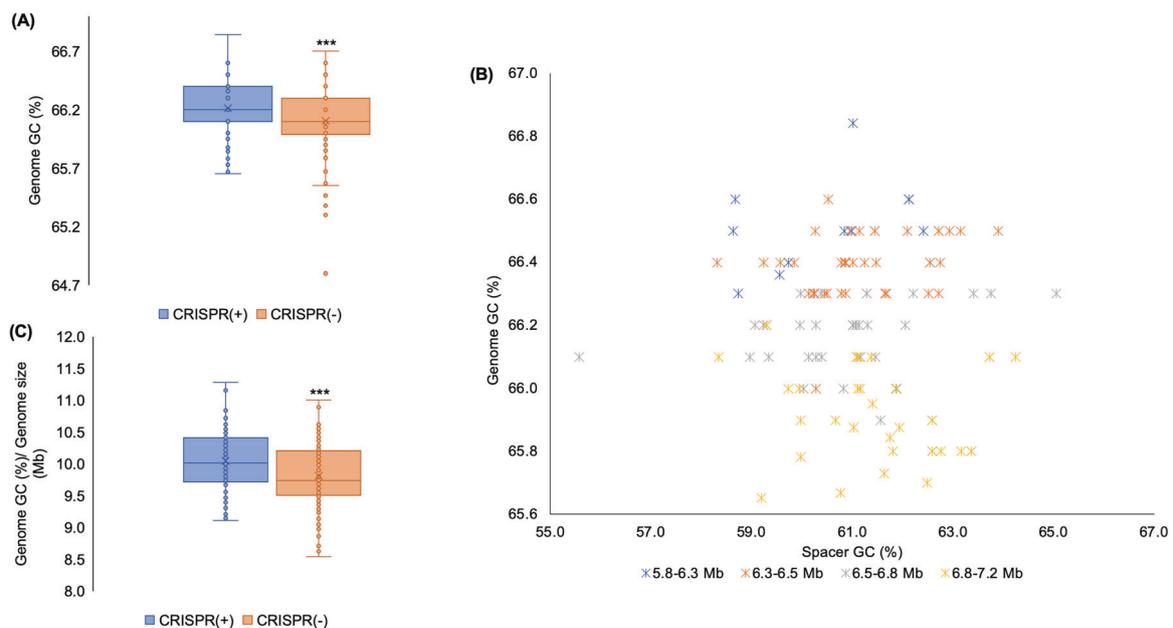


Fig. 2 CRISPR-Cas systems and GC content. **A** Shows GC content in CRISPR(–) vs. CRISPR(+) genomes. Active CRISPR-Cas systems were associated with high GC content (two-tailed t -test $p < 0.01$). **B** Shows a comparison of the GC content of spacers with genomic GC composition in genomes predicted to contain a functional CRISPR-

Cas system. Spacer GC content was always lower than genome-wide GC composition (paired sample t -test $p < 0.01$). **C** Shows the GC content of *Pseudomonas* genomes, standardised according to genome size. CRISPR-Cas systems were associated with higher genome GC after correcting for genome size (two-tailed t -test $p < 0.01$).

characterise the proportion of this spacerome targeting phage, ICE, plasmids, and conjugative transfer genes, virulence factors and resistance genes (Fig. 3A). These screening categories were chosen to address key suggested targets of CRISPR-Cas systems (i.e. phage and additional mobile genetic elements) and other potentially important genes of clinical relevance (i.e. resistance genes and virulence genes) that are also known to be carried on mobile genetic elements. Phage encounter is considered a strong evolutionary pressure for retaining CRISPR-Cas systems [12, 82] and, as expected, a large proportion of spacers (30.52%) were predicted to target phage DNA (Fig. 3A). The types of phage being targeted were further classified based on temperate, lytic and non-lytic phage genome groups (Fig. 3A) [61]. Temperate phage genomes were most commonly predicted to be targeted by the unique *P. aeruginosa* spacers (Fig. 3A). Interestingly, the number of spacers targeting phage was positively well correlated to the total spacer number within each genome (Fig. 3B, C). All CRISPR(+) genomes contained spacers predicted to target phage and the mean proportion of spacers targeting phage per genome was 31.34% (Fig. 3C).

A smaller proportion of spacers (5.61%) were predicted to target ICE, plasmids and conjugative transfer genes (Fig. 3A). One unique spacer (0.05%) was predicted to target the *crpP* ciprofloxacin resistance gene [83], and this spacer was only found in one genome. Similarly, one unique spacer (0.05%) was predicted to target a virulence

gene, which was a pyochelin dihydroaeruginosic acid synthetase gene (*pchE*), and this spacer was only found in one genome. The remaining spacers (63.78% of collection) had no identifiable target in our database searches. The targets of spacers outside of the characterised sequences for phage and other mobile genetic elements has recently been explored by Shmakov et al. [17], who suggest a large proportion of CRISPR ‘dark matter’ may represent uncharacterised mobile genetic material. The known databases of ICE, phage and plasmids clearly under-represent the diversity of mobile elements that can be transferred to *Pseudomonas*. It is highly likely that expansions in genome sequencing, genome annotations and characterisations of gene functions (e.g. hypothetical proteins) will continue to expand these databases and our knowledge of spacer targeting alongside them.

A self-targeting spacer analysis was carried out by blasting spacers against their source genome, and we identified self-targeting spacers in 70/300 genomes. Self-targeting spacers were common in genomes predicted to contain inactivated CRISPR-Cas systems due to the presence of Acr genes (17/20 genomes) or the absence of cognate Cas genes (10/26 genomes). More surprisingly, we found self-targeting spacers in 43/129 genomes predicted to contain a functional CRISPR-Cas system. Self-targeting spacers were strongly associated with the presence of type I-F CRISPR-Cas systems (37/43 genomes). Interestingly, recent work shows that this strong association between I-F

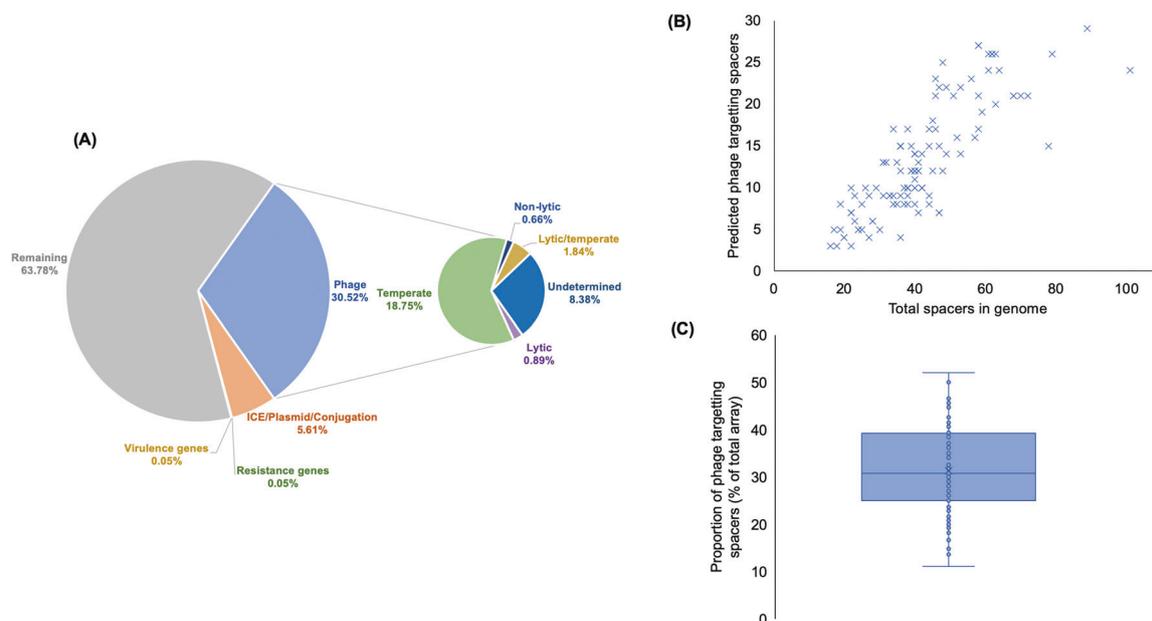


Fig. 3 The CRISPR spacerome across *P. aeruginosa*. **A** Shows the predicted targets of the unique spacers in CRISPR(+) genomes across searches in databases for phage, ICE, plasmid, and conjugative genes, resistance genes and virulence genes. The types of phage being targeted were further classified into temperate, lytic and non-lytic genome groups. **B** Shows for each CRISPR(+) genome the total number of

identified spacers and the number of these predicted to target phage. Predicted phage targeting spacers were strongly correlated to total genome spacer size ($r = 0.81$, $p < 0.00001$, Pearson's correlation coefficient test). **C** Shows this data as the proportion of phage targeting spacers out of the total array for each CRISPR(+) genome.

CRISPR-Cas and self-targeting spacers is found across bacterial genomes [69]. The preservation of CRISPR-Cas systems in these genomes suggests cryptic mechanisms may exist that can protect bacteria from the detrimental effects of self-targeting auto-immunity.

Conjugative elements are a common target of *P. aeruginosa* CRISPR systems

We screened the spacers in each CRISPR(+) genome against a database of ICE and conjugative transfer genes to assess the prevalence of spacers targeting conjugative elements. Spacers targeting ICE or conjugative transfer system genes were widespread, occurring in 111/129 CRISPR(+) genomes (Supplementary Table S1), spanning 45/56 (~80%) of the defined CRISPR(+) STs in our collection (Fig. 4A). Crucially, spacers targeting conserved components of the conjugative machinery (*tra* genes, *trb* genes, type IV secretion system genes [84–86]) were found in 26/56 (46%) of STs that were associated with predicted functional CRISPR-Cas systems (Fig. 4B). Although these spacers made up a small fraction of the overall spacerome, the distribution of these spacers and the high conservation of the conjugative machinery implies that CRISPR-Cas systems likely play an important role in preventing the acquisition of conjugative elements. In addition to these spacers, we identified spacers that target

genes associated with ICE [62] in 43/56 (77%) of the defined CRISPR(+) STs in our collection (Fig. 4B). Eleven STs with CRISPR-Cas systems contained no spacers predicted to target ICE or conjugative transfer system genes (Fig. 4A).

CRISPR-Cas systems prevent the acquisition of prophage and ICE

Our preceding analyses have focused on broad-scale correlations between genome content and CRISPR-Cas activity across the diversity of *P. aeruginosa*. As a complementary approach, we focused in greater detail on five STs in our genome collection where the presence of predicted functional CRISPR-Cas systems was variable (Table 1). Since different isolates within the same ST are very closely related to each other, these STs provide the opportunity to investigate the evolutionary consequences of CRISPR-Cas activity in much greater detail, with less potential for confounding variables to obscure the effects of CRISPR-Cas. Consistent with our previous analyses, we found CRISPR(–) genomes were larger than their CRISPR(+) counterparts within an ST (Table 1).

We next investigated whether CRISPR-Cas systems could be linked to reduced presence of mobile genetic elements in these STs. We aligned complete genome representatives of CRISPR(+) and CRISPR(–) from each

Fig. 4 Spacers targeting ICE or conjugative transfer system genes. **A** Shows the combined number of ICE or conjugative transfer system targeting spacers per genome per ST across the 56 defined CRISPR(+) STs. **B** Shows the distribution of ICE and conjugative transfer system specific targeting spacers for the 45/56 defined STs they are present in, given as the mean per genome per ST.

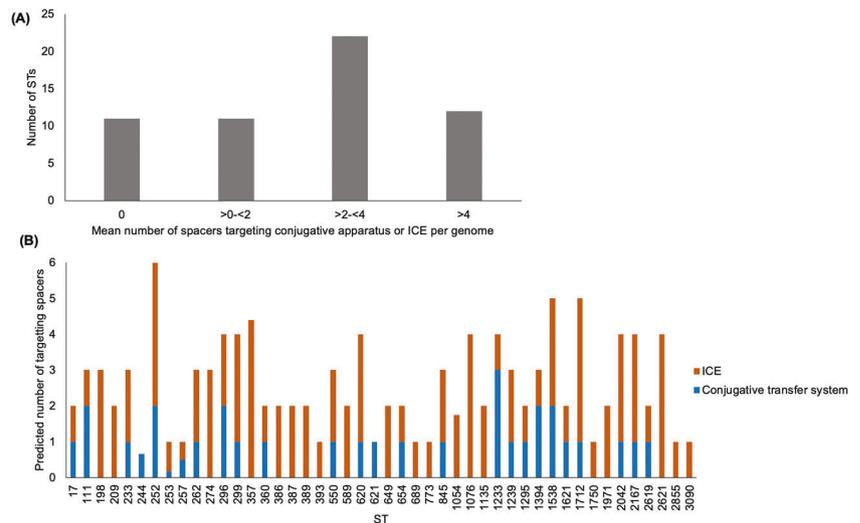


Table 1 *P. aeruginosa* STs with variable presence of CRISPR-Cas systems: ST111, ST262, ST274, ST277 and ST2619.

ST	Number of CRISPR(+) genomes	Mean CRISPR(+) genome size/ Mb	±S.E.M	Number of CRISPR(-) genomes	Mean CRISPR(-) genome size/ Mb	±S.E.M
111	1	6.65	–	8	7.19	0.082
262	6	6.44	0.004	2	6.94	0.005
274	1	6.45	–	2	6.65	0.034
277	7	6.88	0.059	3	6.98	0.049
2619	1	6.24	–	1	6.61	–

Table shows the number of CRISPR(+) and CRISPR(-) genomes in each ST, and the size (Mb) of these genomes (mean if >1 genome). CRISPR(-) genomes were larger than their CRISPR(+) counterparts (paired sample *t*-test $p < 0.01$). It should be noted ST262 genomes are from the same bioproject (Supplementary Information Table 2).

ST and searched for unique regions containing phage or ICE, both of which make an important contribution to HGT in *Pseudomonas* (Fig. 5) [7, 37, 87]. This alignment clearly suggested that the presence of CRISPR-Cas systems was associated with reduced occurrence of phage and ICE between pairs of closely related isolates (Fig. 5). To quantify this, we systematically searched for ICE and phage in all genomes in the five STs with variable presence or absence of CRISPR systems in our dataset (Supplementary Table S2).

We found that predicted functional CRISPR-Cas systems were associated with a lower relative abundance of predicted ICE genes in four of five STs (Fig. 6A), and the only exception (ST277) was one of the eleven STs that lacked any spacers predicted to target ICE or conjugative genes (Supplementary Table S2). Similarly, we found that predicted functional CRISPR-Cas systems were associated with a lower number of prophage regions (Fig. 6B) (Supplementary Table S2). Once again, ST277 was the exception to this general trend, and we speculate that CRISPR-Cas may have been recently gained or lost in this ST, given the small difference in CRISPR(+) and CRISPR(-) genome size.

Spacers in CRISPR(+) genomes map to mobile elements present in closely related CRISPR(-) genomes

Finally, we investigated whether the spacers found in isolates with a predicted functional CRISPR-Cas system match to mobile elements that are present in the genomes of closely related isolates that lack CRISPR-Cas. We aligned CRISPR(+) and CRISPR(-) complete genome representatives in ST111, ST262, ST277 and ST2619 (Fig. 5), and extracted regions unique to the CRISPR(-) genomes. We found that 10–20% of spacers from the CRISPR(+) genomes matched targets present in the unique CRISPR(-) genome regions (Table 2). Spacers that mapped to unique regions present in the CRISPR(-) genomes had a diversity of targets, including both phage and conjugative genes (Table 2). For example, ST111 CRISPR(-) genomes are rich in ICE (Fig. 6A), and ST111 CRISPR(+) genomes carry spacers that target conjugative transfer machinery genes (Table 2). The identification of CRISPR(+) spacers that map to mobile elements in their closely related CRISPR(-) counterparts provides compelling evidence that

Fig. 5 Mauve alignment of the complete CRISPR(+) and CRISPR(-) genome representatives of ST111, ST262, ST277 and ST2619 [70]. Mauve alignment colours indicate colinear gene blocks. All eight complete genome representatives were composed of a chromosome and had no plasmids identified. Unique regions are shown by blank blocks of genes. Counterintuitively, a blank block of genes in a CRISPR(+) genome indicates genes that are absent from the corresponding CRISPR(-) genome, and vice versa. Unique regions were annotated according to the presence of phage (P), ICEs (I) or Cas genes (Cas). The genomes aligned are highlighted in Supplementary Table S2 and are complete genome sequences containing one chromosome. ST274 has been excluded from this alignment due to lack of a complete genome CRISPR(+) representative.

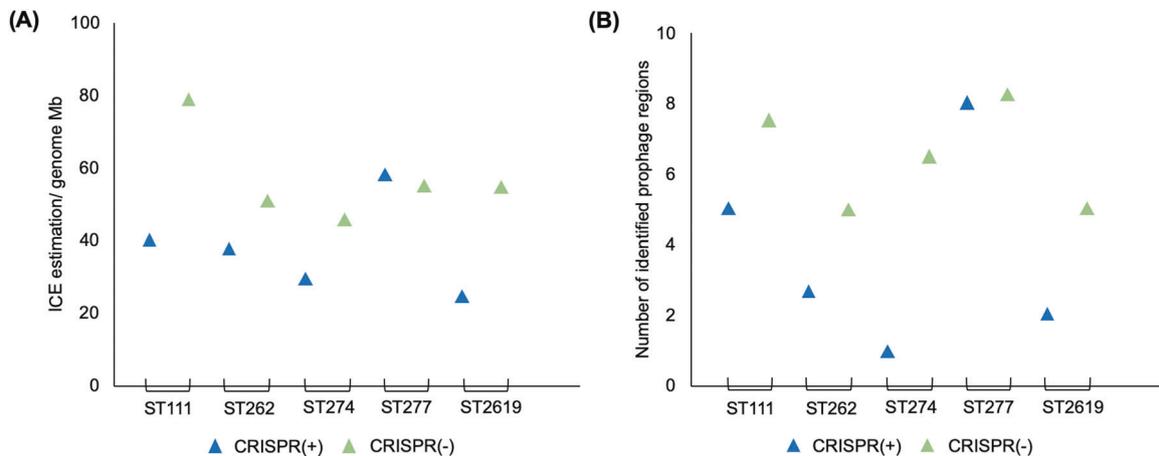
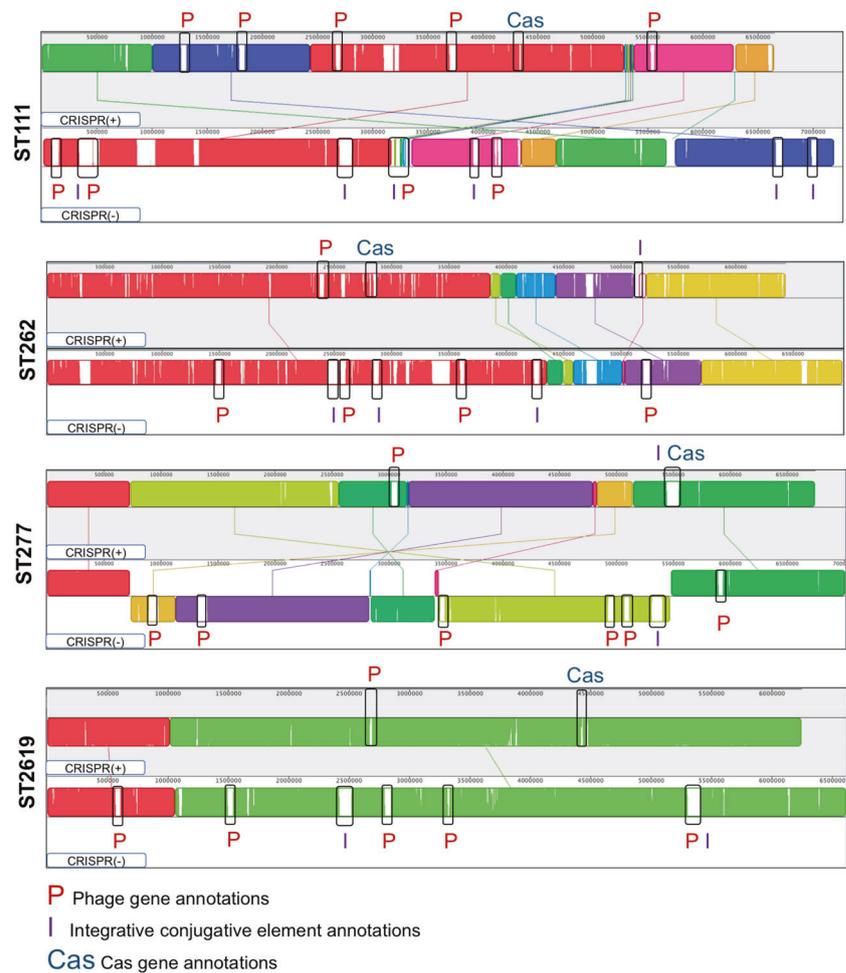


Fig. 6 Comparative abundance of ICE and prophage regions in closely related CRISPR(+) and CRISPR(-). **A** Shows the abundance of ICE genes (/genome Mb). CRISPR(-) genomes contained a higher abundance of ICE genes compared to their CRISPR(+) counterparts (paired sample t -test $p < 0.05$). **B** Shows the identification of prophage regions. CRISPR(-) genomes contained a higher number of identified prophage regions compared to their CRISPR(+) counterparts (paired sample t -test $p < 0.05$).

an absence of CRISPR inhibition of HGT has contributed to genome divergence and expansion in these lineages. To our knowledge, this is the first analysis of CRISPR-Cas in relation to intra-ST *P. aeruginosa* genome evolution in such a way.

It is important to note that these results can be interpreted by two alternative models. Firstly, negative correlations between signatures of HGT and CRISPR-Cas abundance could result from a direct function of CRISPR-Cas systems

Table 2 The number of CRISPR(+) spacers predicted to target unique regions present in the CRISPR (–) genomes for the complete genome representatives of ST111, ST262, ST277 and ST2619 that were aligned (Fig. 5) (Supplementary Table S2).

Sequence Type	Total number of spacers	Number of CRISPR(+) spacers found in blastn of CRISPR(–) unique genomic regions	Predicted targeting identity of spacers (given as putative gene encoded product)
ST111	27	4	(1) Intergenic region (2) Restriction endonuclease-like protein ^a (3) Conjugative transfer system protein (4) Conjugative transfer system protein
ST262	42	7	(1) Phage protein (possible phage transcriptional regulator) ^a (2) Hypothetical protein (3) Integrating conjugative element protein (4) Hypothetical protein (5) Two binding sites of intergenic regions (6) Phage protein (hypothetical protein) ^a (7) Regulatory/ putative secretion system component
ST277	41	8	(1) Hypothetical protein (2) Intergenic region (3) Hypothetical protein (4) Phage protein (virion structural protein) ^a (5) Phage protein (structural protein) ^a (6) Phage protein (hypothetical protein) ^a (7) Hypothetical protein (8) DUF-domain containing protein
ST2619	38	4	(1) DUF-domain containing protein (2) S49 family peptidase (3) Conjugative plasmid adhesion PilV (4) DUF-domain containing protein

The predicted targets of the spacers is taken from the annotation of each genome. Two spacers mapped to regions annotated as intergenic, however it is possible that these regions have been missed by genome annotation or had degenerate coding sequences [17].

^aTargets annotated as hypothetical proteins whose identity was further investigated using NCBI blast to search for characterised homologous coding regions (>90% identity across whole length).

in limiting HGT, such that mobile genetic elements targeted by spacers are blocked from transfer into CRISPR(+) genomes, but not CRISPR(–) genomes. Alternatively, beneficial mobile genetic elements have been demonstrated to drive the loss of CRISPR-Cas systems from bacterial genomes through selection against the fitness costs of retention [22, 88]. As such, it is important to consider that negative correlations between CRISPR-Cas abundance and HGT can result from both CRISPR-Cas limiting HGT, or from HGT driving the loss of CRISPR-Cas. However, in both interpretations, the function of CRISPR-Cas systems in targeting mobile genetic material is key in driving this genome divergence apparent between CRISPR(+) and CRISPR(–) strains.

Conclusions

Broad scale comparisons across the diversity of *P. aeruginosa* reveal that CRISPR-Cas systems are associated with smaller genome size (Fig. 1) and a higher GC content (Fig. 2), which is indicative of reduced acquisition of low GC mobile elements. To gain insights into how CRISPR-

Cas systems may have contributed to genome divergence in *P. aeruginosa*, we focused on comparing the genomes of closely related strains that are CRISPR(+) or CRISPR(–). Phage are a main target of *P. aeruginosa* spacers (Fig. 3), and CRISPR-Cas systems were associated with a reduced abundance of prophage (Fig. 6B). Most isolates with a functional CRISPR-Cas system carry spacers that target either ICE or the conserved conjugative transfer apparatus used by ICE and conjugative plasmids (Fig. 4), and CRISPR-Cas systems were also associated with a reduced abundance of ICE (Fig. 6A). These comparisons between closely related isolates demonstrate clear-cut differences between CRISPR(+) and CRISPR(–) genomes, and they provide an important complement to the broad-scale analyses that focus on the association between CRISPR-Cas and genome composition at larger phylogenetic scales [18, 28] and experimental studies that investigate the influence of CRISPR-Cas on the transfer of individual elements [18, 19, 89]. Collectively our results provide further evidence to support the hypothesis that CRISPR-Cas can act as an important constraint on HGT in bacteria [10, 18, 28] and *P. aeruginosa* is an example of an important bacterial pathogen where this seems to be the case [18, 28].

Although CRISPR-Cas systems were initially characterised as phage defence mechanisms [9, 10], it is becoming increasingly clear that CRISPR-Cas systems can target additional mobile genetic elements [19–21, 90]. Why does CRISPR-Cas target these elements? On the one hand, the tight correlation between the number of spacers targeting phage and total spacer count suggests that spacers targeting mobile genetic elements may simply be acquired as a non-selected by-product of CRISPR-Cas systems that rapidly acquire spacers that target invading phage. Alternatively, it is possible that selection favours the acquisition of spacers that target mobile elements due to the costs associated with these elements. It is compelling that we identified a large number of spacers that target the conjugative transfer apparatus, suggesting that selection has favoured the acquisition of these spacers, perhaps as a result of fitness costs [91] or increased susceptibility to phage [87, 92] and toxins [93] associated with the expression of conjugative machinery. In line with our findings, experimental work has shown CRISPR-Cas systems in *S. epidermidis* can be effective at preventing plasmid transmission [21], and Acr genes have recently been identified on plasmids that can help overcome this immunity [94].

Our search for spacers that target mobile elements was based on searching databases that represent known ICE, phage and plasmids. These databases clearly under-represent the diversity of mobile elements that can be transferred to *Pseudomonas* [17]. Given this, alongside the stringent search parameters used to avoid false positive discovery, our study provides a conservative estimate of the mobile genetic element targeting of CRISPR-Cas system spacers in *P. aeruginosa*.

Plasmids are widespread in bacteria, and they play a key role in HGT [95, 96]. At a broad scale, plasmid carriage is higher in CRISPR(–) genomes compared to CRISPR(+) counterparts [19], suggesting that CRISPR systems play an important role in constraining the transfer of plasmids between bacteria. Plasmids have played an important role in acquisition of resistance genes in *P. aeruginosa*, including carbapenemases [97–101], highlighting the importance of understanding constraints to plasmid transfer in this pathogen. Systematic surveys of the abundance of plasmids in *P. aeruginosa* are lacking, in part due to the challenges of identifying plasmids from short read assemblies, but it is clear that plasmids are much less abundant than ICE, as *P. aeruginosa* genomes typically contain multiple ICE (Fig. 5). Interestingly, many *P. aeruginosa* plasmids lack conjugative genes [46], suggesting that conjugative plasmids may be restricted to a sub-set of the diversity of *P. aeruginosa* [101] or carry genes that overcome CRISPR-Cas immunity.

CRISPR-Cas systems are present in an estimated ~85–90% of archaea, but are significantly less prevalent in bacteria [1]. The molecular mechanisms of CRISPR-Cas are

clear [4, 5], but studies have produced conflicting results on the importance of CRISPR-Cas to genome evolution [18, 21, 23–27]. This study contributes to a growing body of evidence supporting the hypothesis that CRISPR-Cas systems limit HGT [18, 21, 27, 28, 94, 102]. The simplest explanation for this association is that CRISPR-Cas systems restrict the acquisition of mobile genetic elements. However, recent experimental work has shown that CRISPR-Cas systems can be lost as a result of selection to eliminate immunity to newly acquired mobile elements that confer a fitness benefit [22, 88]. Our work supports that *Pseudomonas* is faced with a clear trade-off between CRISPR-Cas systems, which protect against genetic parasites, and HGT, which facilitates adaptation to new ecological niches and stressors. An important challenge for future work will be to understand the extent to which mobile elements impose selection for the loss of CRISPR-Cas in natural settings.

Acknowledgements This project was supported by Wellcome Trust Grant 106918/Z/15/Z held by RCM. We would like to thank Mike Brockhurst, Alvaró San Millán, and reviewers for comments on the manuscript.

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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science*. 2010;327:167–70.
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. Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie*. 2015;117:119–28.
5. Hille F, Charpentier E. CRISPR-Cas: biology, mechanisms and relevance. *Philos Trans R Soc B: Biol Sci*. 2016;371:20150496.

6. Westra ER, Buckling A, Fineran PC. CRISPR–Cas systems: beyond adaptive immunity. *Nat Rev Microbiol.* 2014;12:317–26.
7. McNally A, Oren Y, Kelly D, Pascoe B, Dunn S, Sreecharan T, et al. Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. *PLoS Genet.* 2016;12:e1006280.
8. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature.* 2000;405:299.
9. Stern A, Mick E, Tirosh I, Sagy O, Sorek R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res.* 2012;22:1985–94.
10. Cady KC, Bondy-Denomy J, Heussler GE, Davidson AR, O’Toole GA. The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *J Bacteriol.* 2012;194:5728–38.
11. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuys RJ, Snijders AP, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science.* 2008;321:960–4.
12. Trasanidou D, Gerós AS, Mohanraju P, Nieuwenweg AC, Nobrega FL, Staals RH. Keeping crisper in check: diverse mechanisms of phage-encoded anti-crisprs. *FEMS Microbiol Lett.* 2019;366:fnz098.
13. Szczepankowska A. Role of CRISPR/cas system in the development of bacteriophage resistance. *Adv Virus Res.* 2012;82:289–338.
14. Touchon M, Bernheim A, Rocha EP. Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J.* 2016;10:2744–54.
15. Bobay L-M, Touchon M, Rocha EP. Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci.* 2014;111:12127–32.
16. Davies EV, James CE, Kukavica-Ibrulj I, Levesque RC, Brockhurst MA, Winstanley C. Temperate phages enhance pathogen fitness in chronic lung infection. *ISME J.* 2016;10:2553–5.
17. 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.
18. Shehreen S, Chyou T-y, Fineran PC, Brown CM. Genome-wide correlation analysis suggests different roles of CRISPR-Cas systems in the acquisition of antibiotic resistance genes in diverse species. *Philos Trans R Soc B.* 2019;374:20180384.
19. O’Meara D, Nunney L. A phylogenetic test of the role of CRISPR-Cas in limiting plasmid acquisition and prophage integration in bacteria. *Plasmid.* 2019;104:102418.
20. Westra ER, Van Houte S, Gandon S, Whitaker R. The ecology and evolution of microbial CRISPR–Cas adaptive immune systems. *Phil Trans R Soc B.* 2019;374:101.
21. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. *Science.* 2008;322:1843–45.
22. Jiang W, Maniv I, Arain F, Wang Y, Levin BR, Marraffini LA. Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS Genet.* 2013;9:e1003844.
23. 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.
24. 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;9:2021–27.
25. Kos VN, Déraspe M, McLaughlin RE, Whiteaker JD, Roy PH, Alm RA, et al. The resistome of *Pseudomonas aeruginosa* in relationship to phenotypic susceptibility. *Antimicrobial Agents Chemother.* 2015;59:427–36.
26. Turton JF, Wright L, Underwood A, Witney AA, Chan Y-T, Al-Shahib A, et al. High-resolution analysis by whole-genome sequencing of an international lineage (sequence type 111) of *Pseudomonas aeruginosa* associated with metallo-carbapenemases in the United Kingdom. *J Clin Microbiol.* 2015;53:2622–31.
27. Mackow NA, Shen J, Adnan M, Khan AS, Fries BC, Diago-Navarro E. CRISPR-Cas influences the acquisition of antibiotic resistance in *Klebsiella pneumoniae*. *PLoS ONE.* 2019;14:e0225131.
28. van Belkum A, Soriaga LB, LaFave MC, Akella S, Veyrieras J-B, Barbu EM, et al. Phylogenetic distribution of CRISPR–Cas systems in antibiotic-resistant *Pseudomonas aeruginosa*. *mBio.* 2015;6:e01796–15.
29. Klockgether J, Cramer N, Wiehlmann L, Davenport CF, Tümmler B. *Pseudomonas aeruginosa* genomic structure and diversity. *Front Microbiol.* 2011;2:150.
30. Botelho J, Grosso F, Peixe L. Antibiotic resistance in *Pseudomonas aeruginosa*—Mechanisms, epidemiology and evolution. *Drug Resistance Updates.* 2019;44:100640.
31. Johnson CM, Grossman AD. Integrative and conjugative elements (ICEs): what they do and how they work. *Annu Rev Genet.* 2015;49:577–601.
32. Botelho J, Schulenburg H. The role of integrative and conjugative elements in antibiotic resistance evolution. *Trends Microbiol.* 2020. <https://doi.org/10.1016/j.tim.2020.05.011>.
33. Wozniak RA, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, Garriss G, et al. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet.* 2009;5:e1000786.
34. Guglielmini J, Quintais L, Garcillán-Barcia MP, de La Cruz F, Rocha EP. The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet.* 2011;7:e1002222.
35. Guglielmini J, de La Cruz F, Rocha EP. Evolution of conjugation and type IV secretion systems. *Mol Biol Evol.* 2013;30:315–31.
36. Weiss E, Spicher C, Haas R, Fischer W. Excision and transfer of an integrating and conjugative element in a bacterial species with high recombination efficiency. *Sci Rep.* 2019;9:1–14.
37. Burrus V, Waldor MK. Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol.* 2004;155:376–86.
38. Rodríguez-Blanco A, Lemos ML, Osorio CR. Integrating conjugative elements as vectors of antibiotic, mercury, and quaternary ammonium compound resistance in marine aquaculture environments. *Antimicrobial Agents Chemother.* 2012;56:2619–26.
39. Obi CC, Yayla S, De Gannes V, Berres ME, Walker J, Pavelec D, et al. The integrative conjugative element *clc* (ICE_{clc}) of *Pseudomonas aeruginosa* JB2. *Front Microbiol.* 2018;9:1532.
40. Botelho J, Grosso F, Peixe L. Unravelling the genome of a *Pseudomonas aeruginosa* isolate belonging to the high-risk clone ST235 reveals an integrative conjugative element housing a *bla* GES-6 carbapenemase. *J Antimicrobial Chemother.* 2018;73:77–83.
41. Botelho J, Roberts AP, León-Sampedro R, Grosso F, Peixe L. Carbapenemases on the move: it’s good to be on ICEs. *Mobile DNA.* 2018;9:37.
42. Ding Y, Teo JW, Drautz-Moses DI, Schuster SC, Givskov M, Yang L. Acquisition of resistance to carbapenem and macrolide-mediated quorum sensing inhibition by *Pseudomonas aeruginosa* via ICE Tn4371 6385. *Commun Biol.* 2018;1:1–10.
43. Subedi D, Kohli GS, Vijay AK, Willcox M, Rice SA. Accessory genome of the multi-drug resistant ocular isolate of *Pseudomonas aeruginosa* PA34. *PLoS ONE.* 2019;14:e0215038.
44. Tariq MA, Everest FL, Cowley LA, Wright R, Holt GS, Ingram H, et al. Temperate bacteriophages from chronic *Pseudomonas aeruginosa* lung infections show disease-specific changes in host range and modulate antimicrobial susceptibility. *MSystems.* 2019;4:e00191–18.

45. San Millan A, MacLean RC. Fitness costs of plasmids: a limit to plasmid transmission. *Microbial Transmission*. 2019;5:65–79.
46. San Millan A, Toll-Riera M, Qi Q, Betts A, Hopkinson RJ, McCullagh J, et al. Integrative analysis of fitness and metabolic effects of plasmids in *Pseudomonas aeruginosa* PAO1. *ISME J*. 2018;12:3014–24.
47. San Millan A, Peña-Miller R, Toll-Riera M, Halbert Z, McLean A, Cooper B, et al. Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nat Commun*. 2014;5:1–11.
48. Maxwell KL. Phages fight back: inactivation of the CRISPR-Cas bacterial immune system by anti-CRISPR proteins. *PLoS Pathogens*. 2016;12:e1005282.
49. Bondy-Denomy J, Garcia B, Strum S, Du M, Rollins MF, Hidalgo-Reyes Y, et al. Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature*. 2015;526:136–9.
50. Hwang S, Maxwell KL. Meet the Anti-CRISPRs: widespread protein inhibitors of CRISPR–Cas systems. *CRISPR J*. 2019;2:23–30.
51. Dong C, Hao G-F, Hua H-L, Liu S, Labena AA, Chai G, et al. Anti-CRISPRdb: a comprehensive online resource for anti-CRISPR proteins. *Nucleic Acids Res*. 2018;46:D393–8.
52. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150–2.
53. Seemann T. MLST. 2014. <https://github.com/tseemann/mlst>.
54. Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinforma*. 2010;11:1–11.
55. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.
56. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res*. 2018;46:W246–51.
57. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJ, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*. 2019;18:1–17.
58. Leon LM, Park AE, Borges AL, Zhang J, Bondy-Denomy J. Mobile element warfare via crspr and anti-crspr in *Pseudomonas aeruginosa*. 2020. <https://www.biorxiv.org/content/10.1101/2020.06.15.151498v1>.
59. Marino ND, Zhang JY, Borges AL, Sousa AA, Leon LM, Rauch BJ, et al. Discovery of widespread type I and type V CRISPR–Cas inhibitors. *Science*. 2018;362:240–2.
60. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22:1658–9.
61. England WE, Kim T, Whitaker RJ. Metapopulation structure of CRISPR–Cas immunity in *Pseudomonas aeruginosa* and its viruses. *MSystems*. 2018;3:e00075–18.
62. Liu M, Li X, Xie Y, Bi D, Sun J, Li J, et al. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res*. 2019;47:D660–5.
63. Brooks L, Kaze M, Sistro M. A curated, comprehensive database of plasmid sequences. *Microbiol Resour Announcements*. 2019;8:e01325–18.
64. NCBI. Gene. Bethesda (MD): National Library of Medicine (US) NCFBI. 2004. <https://www.ncbi.nlm.nih.gov/gene/>.
65. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrobial Chemother*. 2012;67:2640–4.
66. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res*. 2019;47:D687–92.
67. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res*. 2005;33:D325–D328.
68. Meneghin J. get_GC_content.pl. 2009. https://git-r3lab.uni.lu/IMP/IMP/blob/7e33be824a54c224c8052fc5faefb464f09549e4/src/get_GC_content.pl.
69. Nobrega FL, Walinga H, Dutilh BE, Brouns SJ. Prophages are associated with extensive, tolerated CRISPR–Cas auto-immunity. 2020. <https://www.biorxiv.org/content/10.1101/2020.03.02.973784v1>.
70. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004;14:1394–403.
71. 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.
72. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. <https://www.R-project.org/>.
73. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 2007;35:D61–5.
74. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*. 2013;493:429–32.
75. Lerat E, Daubin V, Ochman H, Moran NA. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol*. 2005;3:e130.
76. Rocha EP, Danchin A. Base composition bias might result from competition for metabolic resources. *Trends Genet*. 2002;18:291–4.
77. Nishida H. Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids. *Int J Evol Biol*. 2012;2012:342482.
78. Bohlin J, Eldholm V, Pettersson JH, Brynildsrud O, Snipen L. The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes. *BMC Genom*. 2017;18:151.
79. San Millan A, Toll-Riera M, Qi Q, MacLean RC. Interactions between horizontally acquired genes create a fitness cost in *Pseudomonas aeruginosa*. *Nat Commun*. 2015;6:1–8.
80. Stover CK, Pham XQ, Erwin A, Mizoguchi S, Warrener P, Hickey M, et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*. 2000;406:959.
81. Almpanis A, Swain M, Gatherer D, McEwan N. Correlation between bacterial G+ C content, genome size and the G+ C content of associated plasmids and bacteriophages. *Microbiol Genom*. 2018;4:e000168.
82. Westra ER, van Houte S, Oyesiku-Blakemore S, Makin B, Broniewski JM, Best A, et al. Parasite exposure drives selective evolution of constitutive versus inducible defense. *Curr Biol*. 2015;25:1043–9.
83. Chávez-Jacobo VM, Hernández-Ramírez KC, Romo-Rodríguez P, Pérez-Gallardo RV, Campos-García J, Gutiérrez-Corona JF, et al. CrpP is a novel ciprofloxacin-modifying enzyme encoded by the *Pseudomonas aeruginosa* pUM505 plasmid. *Antimicrobial Agents Chemother*. 2018;62:e02629–17.
84. Lawley T, Klimke W, Gubbins M, Frost L. F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett*. 2003;224:1–15.
85. Flores-Ríos R, Moya-Beltrán A, Pareja-Barrueto C, Arenas-Salinas M, Valenzuela S, Orellana O, et al. The type IV secretion system of ICEAfe1: formation of a conjugative pilus in *Acidithiobacillus ferrooxidans*. *Front Microbiol*. 2019;10:30.
86. Christie PJ, Whitaker N, González-Rivera C. Mechanism and structure of the bacterial type IV secretion systems. *Biochimica et Biophysica Acta (BBA)-Mol Cell Res*. 2014;1843:1578–91.

87. Brockhurst MA, Harrison E, Hall JP, Richards T, McNally A, MacLean RC. The ecology and evolution of pangenomes. *Curr Biol*. 2019;29:R1094–103.
88. Rollie C, Chevallereau A, Watson BN, Chyou T-y, Fradet O, McLeod I, et al. Targeting of temperate phages drives loss of type I CRISPR–Cas systems. *Nature*. 2020;578:149–53.
89. Common J, Morley D, Westra ER, van Houte S. CRISPR–Cas immunity leads to a coevolutionary arms race between *Streptococcus thermophilus* and lytic phage. *Philos Trans R Soc B*. 2019;374:20180098.
90. Faure G, Makarova KS, Koonin EV. CRISPR–Cas: complex functional networks and multiple roles beyond adaptive immunity. *J Mol Biol*. 2019;431:3–20.
91. Porse A, Schønning K, Munck C, Sommer MO. Survival and evolution of a large multidrug resistance plasmid in new clinical bacterial hosts. *Mol Biol Evolution*. 2016;33:2860–73.
92. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett*. 2016;363:fnw002.
93. Ho BT, Basler M, Mekalanos JJ. Type 6 secretion system–mediated immunity to Type 4 secretion system–mediated gene transfer. *Science*. 2013;342:250–3.
94. Mahendra C, Christie KA, Osuna BA, Pinilla-Redondo R, Kleinstiver BP, Bondy-Denomy J. Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer. *Nature. Microbiology*. 2020;5:620–9.
95. Acman M, van Dorp L, Santini JM, Balloux F. Large-scale network analysis captures biological features of bacterial plasmids. *Nat Commun*. 2020;11:1–11.
96. Halary S, Leigh JW, Cheaib B, Lopez P, Bapteste E. Network analyses structure genetic diversity in independent genetic worlds. *Proc Natl Acad Sci*. 2010;107:127–32.
97. Bonnin RA, Poirel L, Nordmann P, Eikmeyer FG, Wibberg D, Pühler A, et al. Complete sequence of broad-host-range plasmid pNOR-2000 harbouring the metallo- β -lactamase gene bla VIM-2 from *Pseudomonas aeruginosa*. *J Antimicrobial Chemother*. 2013;68:1060–5.
98. Botelho J, Grosso F, Peixe L. Characterization of the pJB12 plasmid from *Pseudomonas aeruginosa* reveals Tn6352, a novel putative transposon associated with mobilization of the blaVIM-2-harboring In58 integron. *Antimicrobial Agents Chemother*. 2017;61:e02532–16.
99. Haines AS, Jones K, Cheung M, Thomas CM. The IncP-6 plasmid Rms149 consists of a small mobilizable backbone with multiple large insertions. *J Bacteriol*. 2005;187:4728–38.
100. Xiong J, Alexander DC, Ma JH, Déraspe M, Low DE, Jamieson FB, et al. Complete sequence of pOZ176, a 500-kilobase IncP-2 plasmid encoding IMP-9-mediated carbapenem resistance, from outbreak isolate *Pseudomonas aeruginosa* 96. *Antimicrobial Agents Chemother*. 2013;57:3775–82.
101. Cazares A, Moore MP, Hall JP, Wright LL, Grimes M, Emond-Rhéault J-G, et al. A megaplasmid family driving dissemination of multidrug resistance in *Pseudomonas*. *Nat Commun*. 2020;11:1–13.
102. Seecharran T, Kalin-Manttari L, Koskela K, Nikkari S, Dickins B, Corander J, et al. Phylogeographic separation and formation of sexually discrete lineages in a global population of *Yersinia pseudotuberculosis*. *Microbial Genom*. 2017;3:e000133.