ACCESS MICROBIOLOGY
an open research platform

RESEARCH ARTICLE
Hegde *et al.*, *Access Microbiology* 2023;5:000599.v4
DOI 10.1099/acmi.0.000599.v4

MICROBIOLOGY SOCIETY
OPEN DATA OPEN ACCESS

Identification and characterization of two CRISPR/Cas systems associated with the mosquito microbiome

Shivanand Hegde[1,*,†,‡], Hallie E. Rauch[2,†], Grant L. Hughes[1] and Nikki Shariat[3]

Abstract

The microbiome profoundly influences many traits in medically relevant vectors such as mosquitoes, and a greater functional understanding of host–microbe interactions may be exploited for novel microbial-based approaches to control mosquito-borne disease. Here, we characterized two novel clustered regularly interspaced short palindromic repeats (CRISPR)/Cas systems in *Serratia* sp. Ag1, which was isolated from the gut of an *Anopheles gambiae* mosquito. Two distinct CRISPR/Cas systems were identified in *Serratia* Ag1, CRISPR1 and CRISPR2. Based on *cas* gene composition, CRISPR1 is classified as a type I-E CRISPR/Cas system and has a single array, CRISPR1. CRISPR2 is a type I-F system with two arrays, CRISPR2.1 and CRISPR2.2. RT-PCR analyses show that all *cas* genes from both systems are expressed during logarithmic growth in culture media. The direct repeat sequences of CRISPRs 2.1 and 2.2 are identical and found in the arrays of other *Serratia* spp., including *S. marcescens* and *S. fonticola*, whereas CRISPR1 is not. We searched for potential spacer targets and revealed an interesting difference between the two systems: only 9% of CRISPR1 (type I-E) targets are in phage sequences and 91% are in plasmid sequences. Conversely, ~66% of CRISPR2 (type I-F) targets are found within phage genomes. Our results highlight the presence of CRISPR loci in gut-associated bacteria of mosquitoes and indicate interplay between symbionts and invasive mobile genetic elements over evolutionary time.

DATA SUMMARY

All the methods and data required for the reproduction of this work have been provided here. The accession numbers for the genomes of the bacterial isolates used in this study are JQEI00000000 (Serratia sp. Ag1) and JQEJ00000000 (Serratia sp. Ag2). No supporting external data were generated for this work.

INTRODUCTION

Host-associated microbes play a crucial role in the physiology, diseases and immunity of their host. In mosquitoes, gut-associated microbes profoundly affect their host and these altered phenotypes influence vectoral capacity and vector competence [1–6]. Bacteria are abundant constituents of the gut microbiome of mosquitoes [7–10], but metagenomic studies have also found bacteriophage associated with these vectors [8–12], and it would be reasonable to expect interplay between these microbes, given their co-occurrence. While microbe–microbe interactions within the gut alter bacterial community structure and colonization [13, 14], less is known regarding the interactions between bacterial communities and bacteriophage, although signatures of these encounters can be inferred from bacterial genomes.

Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas systems are present in approximately 45% of sequenced bacterial genomes, and 90% of archaeal genomes [15]. In their canonical function, they act as a small RNA-driven adaptive immune system that provides defence against exogenous nucleic acids, namely bacteriophage and plasmids [16, 17]. CRISPR/

Received 28 March 2023; Accepted 31 July 2023; Published 11 August 2023
Author affiliations: [1]Department of Vector Biology and Tropical Disease Biology, Liverpool School of Tropical Medicine, Centre for Neglected Tropical Disease, Liverpool, UK; [2]Department of Biology, Gettysburg College, Gettysburg, PA, USA; [3]Department of Population Health, University of Georgia, Athens, GA, USA.
***Correspondence:** Shivanand Hegde, s.m.hegde@keele.ac.uk
Keywords: *Anopheles*; CRISPR loci; microbiome; type I-F CRISPR; bacteriophage.
Abbreviations: BLAST, Basic Local Alignment Search Tool; crRNAs, CRISPR RNAs; MEGA, Molecular Evolutionary Genetics Analysis.
‡Present address: School of Life Sciences, University of Keele, Newcastle, UK.
†These authors contributed equally to this work

Cas systems have two components, a suite of *cas* genes and a CRISPR array [18]. The latter comprise direct repeat sequences ranging from 21 to 48 nucleotides in length that separate highly variable spacer sequences of similar lengths [19]. Spacers are commonly derived from foreign nucleic acids and are added in a polar manner to the CRISPR array, with the newest spacers being found closest to the leader sequence, which is directly upstream of the first repeat containing regulatory elements necessary for adaptation.

CRISPR immunity takes place in three distinct steps. First, new spacers are acquired and added to the array as the prokaryote adapts to a new invader [16, 20, 21]. Second, the array is transcribed, and the resulting transcript processed to produce mature CRISPR RNAs (crRNAs) [21, 22]. Third, the crRNA guides an endonuclease to its complementary target nucleic acid, thereby resulting in degradation, or interference, of the target [17, 21]. Various *cas* gene products are required for each of these steps.

CRISPR/Cas systems can be separated into two distinct classes and into further subtypes, depending on the complement and organization of *cas* genes [23]. Class 1, type I systems are defined by the inclusion of Cas3 as the effector endonuclease responsible for cleaving target DNAs. Within type I, there are seven subtypes, I-A–I-G [24–26]. Subtypes I-E (e.g. found in *Escherichia coli* and *Salmonella enterica*) and I-F (e.g. found in *Yersinia pseudotuberculosis* and *Pectobacterium atrosepticum*) differ slightly from each other. Type I-E has a distinct Cas2 protein, whereas in type I-F, Cas2 and Cas3 form a chimeric protein. Further, type I-F systems also lack Cas11, which forms part of the type I-E effector complex [22, 27]. In the family *Enterobacteriaceae*, CRISPR/Cas systems belong almost exclusively to either type I-E or type I-F [22, 27].

In addition to their well-characterized role in prokaryote adaptive immunity, alternative functions have also been attributed to some CRISPR/Cas systems [20]. These include roles in biofilm formation, host avoidance and symbiosis, and highlight the important biological roles of these systems in pathogenic bacteria, as well as other bacterial species [20, 28]. Given this, and the recent explosion in genome editing capabilities of *cas* genes, there is a drive to discover new CRISPR/Cas systems in a wide array of prokaryote genomes. CRISPR/Cas systems in host-associated microbiomes have mainly been examined in the context of human and plant microbiomes [29–32], while investigations in invertebrates are lacking. Studies focused on bacteria that play integral roles in the human microbiome have revealed important roles for CRISPR/Cas in viral resistance and mitigation of foreign genetic material [32–35]. Although CRISPR/Cas technology has been applied for genome editing of mosquito vector hosts and their microbiomes [36–38], characterizing native CRISPR loci in the gut bacteria of mosquitoes has not been attempted so far.

To determine interactions between the gut-associated bacteria of mosquitoes and bacteriophage over evolutionary time, we examined the genomic signature of CRISPR/Cas systems in Ag1, a *Serratia* strain previously isolated from *Anopheles gambiae* mosquitoes [39]. We found that Ag1 harbours two type I CRISPR systems and further classification revealed that they belong to subtypes I-E and I-F. We also examined the origins of the spacer region, thereby identifying past infections of the bacterial host, and characterized the expression of the *cas* genes. Our results indicate the presence of CRISPR/Cas systems in symbiotic bacteria associated within invertebrates and highlight the complexity of microbial interactions within the mosquito gut.

METHODS

Culturing and nucleic acid isolation

The origins of the bacterial isolates *Serratia* sp. Ag1 and *Serratia* sp. Ag2 [JQEI00000000 (Serratia sp. Ag1) and JQEJ00000000 (Serratia sp. Ag2)] used in this study were described previously [39]. Total genomic DNA was isolated from overnight cultures of Ag1 and Ag2 using the Genome Wizard kit (Promega, WI, USA) and following the manufacturer's protocol. DNA pellets were resuspended in 200 µl of molecular grade water and stored at −20 °C. Bacterial strains were cultured in LB broth to log phase and to stationary phase and total RNA was isolated using TRIzol (Life Technologies, CA, USA) and resuspended in 20 µl molecular-grade water. RNA was treated with 1-unit DNase (Life Technologies, CA, USA) and reisolated with TRIzol. Pellets were resuspended in 20 µl molecular grade water and stored at −20 °C.

RT-PCR expression analyses

A total of 100 ng total RNA was used to generate cDNA in a 20 µl reaction using a qScript mastermix (QuantaBio, MA, USA) that contained random hexamers. Reverse transcription was performed in a PCR machine with the following parameters: 22 °C for 5 min, 42 °C for 30 min, 85 °C for 5 min and 4 °C hold. For a non-RT control, reactions were set up in duplicate but without RT enzyme. The cDNAs were diluted 1:10 and 2 µl of each was used for subsequent PCR reactions with one unit of *Taq* polymerase (New England Biolabs, MA USA), 200 uM dNTPs (New England Biolabs, MA, USA) and 1× standard *Taq* polymerase buffer in a 25 µl reaction. The primers used for RT-PCR analysis of *cas* genes are listed in Table 1. Following initial denaturation for 3 min at 95 °C, the PCR conditions were as follows: 20 cycles (16S control PCR) or 25 cycles (*cas* genes) of 95 °C for 30 s, annealing at 57 °C for 30 s and an extension at 72 °C for 30 s. A total of 5 µl of the PCR reaction was imaged by gel electrophoresis. The RT-PCR experiments were run independently twice (different bacterial cultures), and on one of these, the PCR step was performed twice on the same cDNA. The gel (Fig. 4) is from one of the independent experiments. Densitometry was not performed as the difference between the log and stationary phases was clear.

Table 1. Primers used in this study

	Marker	Orientation	Primer sequence (5′–3′)	Annealing temp. (°C)
Type I-E	*cas3*	Forward	GCTAATCTCACGATGCAACTGC	58
		Reverse	CATATAAGGCCGCCTCGGT	58
	cse1	Forward	TGGTAATGTATCCAACGCTGGG	58
		Reverse	ATGCCGTTATCCGCCAACAG	58
	cse2	Forward	CAAGTTCTCTAGAGCCGAACGA	58
		Reverse	CCATTGTGGGGTTGTCTGCT	58
	cas6e	Forward	AATTTCAAGACAAGATTGGCCAACA	58
		Reverse	GCCCTTGCCAATACCATGTTTAAAG	58
	cas7	Forward	GCCGCCATGTTAACCAATGAG	58
		Reverse	CCATCGCCTCACCACATTGAG	58
	cas5	Forward	ATGGCTGGCGCAAATGAATG	58
		Reverse	CCACCATCTGAAAGTCACGCA	58
	cas1	Forward	GGAATGGAAGGTAATCGTGTTCGT	58
		Reverse	TTGGTCAGATCACTCAGCTGAAAT	58
	cas2	Forward	AAATGACTTACCACCTGCTGTTC	58
		Reverse	CTCTGTCGGAGAATATTGCATCAAG	58
	CRISPR1_sp1	Forward	TTTCTGCCTCCGCGCCAT	60
	CRISPR1_sp3	Forward	TTCTGTGGTCGTCGTCAGTACO	60
	CRISPR1_sp7	Forward	TTCTCTTAGGGTGCCTGCGC	60
	CRISPR1_sp1_rev	Forward	AAGACTCTGCCGGTAGCGG	60
	CRISPR1_sp3_rev	Forward	GGAAGACGTTTCAGAATATGCGGTA	60
Type I-F	*cas1*	Forward	ATTGCCGCATTCTGGTTAACG	58
		Reverse	CAGCATCACTGCCGTGGTATT	58
	cas3	Forward	GCTCTACAACGGTGCAGGAT	58
		Reverse	TCTTGCCACTTTTCCGTCGC	58
	csy1	Forward	CAGATCAGCCTGGTGACTCAC	58
		Reverse	TTCAACGCCAATGTGGAGAGATAG	58
	csy2	Forward	ATTTCTGGCGGTGAAGCAGG	58
		Reverse	CCTGTAGCCCGTTAATCGTCC	58
	csy3	Forward	CGACGCCGTCTACCTGTAAT	58
		Reverse	GCAATATTGGTGGCATAACGCC	58
	cas6f	Forward	CGTTTGAACAAATACCGGATACCCA	58
		Reverse	AATTCACCATGCTGAATATAAATTCGCATO	58
	CRISPR2_sp1	Forward	AAAGCAGCTGAAGCGTTGAAGC	60
	CRISPR2_sp4	Forward	ATGCGTCGGGTGAGCAACC	60
	CRISPR2_sp8	Forward	AAGCCATGGAACGTGCGGG	60
	CRISPR2_sp1_rev	Forward	AACGCTGGCCATCAGCTTCA	60
	CRISPR2_sp4_rev	Forward	ACAAACGCAGCAAAGAGGTTGC	60

Identification of CRISPR loci, phylogenetic analyses and spacer identification

The assembled Ag1 genome was analysed using CRISPR-Finder [40] to identify both the CRISPR arrays and the *cas* genes. We used the default setting to analyse the Ag1 genome to identify the CRISPR array and *cas* genes. Spacers were extracted from the arrays and analysed using an Excel-based macro [41]. CRISPR Target [42] was used to identify putative spacer matches. Here, we used default parameters for the initial BLAST screen and target database. For initial output display parameters, we used a default score cut-off of 20, 26/32 base pairs. We considered matches to be 24/32 or 24/33 nucleotides for the type I-E and I-F spacers, respectively. For phylogenetic analyses, the coding sequences of both *cas3* genes were translated and BLAST was used to find the top 20 similar sequences from different species. These amino acid sequences were used in MEGA 7 to build phylogenetic trees with a bootstrap value of 1000 [43].

RESULTS

We identified two type I CRISPR/Cas systems in Ag1 and termed them CRISPR1 and CRISPR2. The former has a single CRISPR array and is of the type I-E subtype of CRISPR/Cas systems (Fig. 1a), with direct repeats and spacers that are 28 and 33 nucleotides long, respectively. The CRISPR2 has a *cas* operon associated with the type I-F subtype, and there were two CRISPR arrays associated with this system, which we termed CRISPR2.1 and CRISPR2.2. The direct repeats and spacers in both arrays are 28 and 32 nucleotides in length, respectively. The type I-E repeat sequences fall under cluster 2 and the type I-F direct repeat sequences fall under cluster 1 [44]. These cluster designations follow those described in [44]. The spacer composition of the three CRISPR arrays in Ag1 was analysed and the spacer content of each array was distinct (Fig. 1b). CRISPR2.1 was the longest array and contained 26 different spacers.

Using the cas3 protein sequence from each CRISPR/Cas system, we identified similar protein sequences from other bacterial species and examined their phylogeny. We found a single match to another *Serratia* sp. Ag2, which is closely related to Ag1 [39] (Fig. 2). Otherwise, we did not find any other *Serratia* spp. whose cas3 matched closely to the cas3 of CRISPR1, suggesting that the type I-E system is not broadly present in other *Serratia* spp. The type I-E cas3 was closely related to *Dickeya* spp. and *Klebsiella* spp., and overall there was little divergence among the type I-E cas3 proteins compared to those from the type I-F subtype (Fig. 2). Conversely, we found several *Serratia* spp. that contained cas3 protein sequences of the type I-F subtype, although the sequence from Ag1 was more closely related to some *Yersina* spp. than those *Serratia* spp.

We analysed the CRISPR spacers to determine whether they matched to any exogenous nucleic acids and found a greater number of matches to plasmid and bacteriophage (including prophage sequences) sequences in CRISPR2.1 (50%, 13/26 spacers had matches) and CRISPR2.2 (53%, 8/15) than in CRISPR1 (32%, 6/19) (Fig. 3, Table 2). In both CRISPR2.1 and 2.2, phage targets accounted for the most hits, constituting two-thirds of the identified targets (Fig. 3). Conversely, CRISPR1 had fewer spacer targets that we could identify, with most identified targets of plasmid origin (Fig. 3). When we increased the stringency of the matches

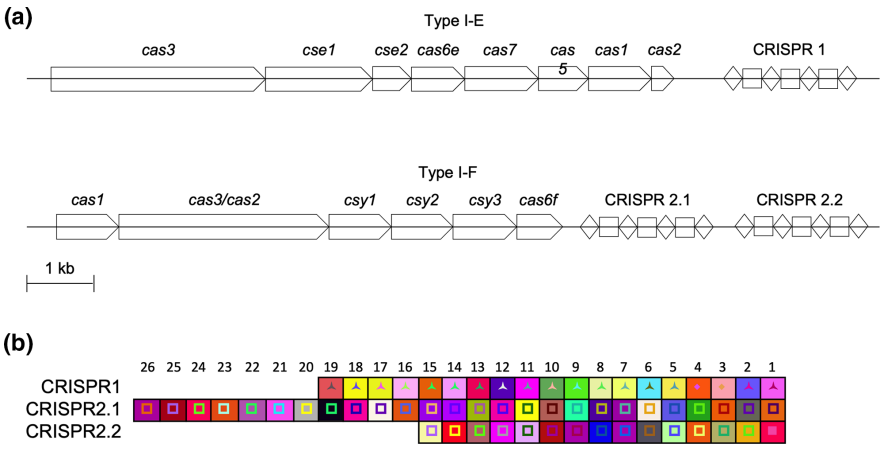

Fig. 1. Organization and expression of the type I-E and type I-F CRISPR/Cas systems of *Serratia* sp. Ag1. (a) All *cas* genes are shown in the forward orientation. Direct repeats in the CRISPR array are shown as black diamonds, while the spacer sequences are represented by white squares. The *cas* genes are scaled to the 1 kb bar shown in the bottom left. (b) Spacer composition of the three CRISPR arrays in Ag1. The unique combination of the background colour and the shape and colour in the foreground represents a single spacer sequence. The three-point star represents a spacer that is 33 nt in length. The inner square represents a 32 nt spacer. The oldest spacer (spacer number 1) is shown to the far right, while the most recently acquired spacer is shown on the far left. The invariant direct repeats have been removed for clarity.

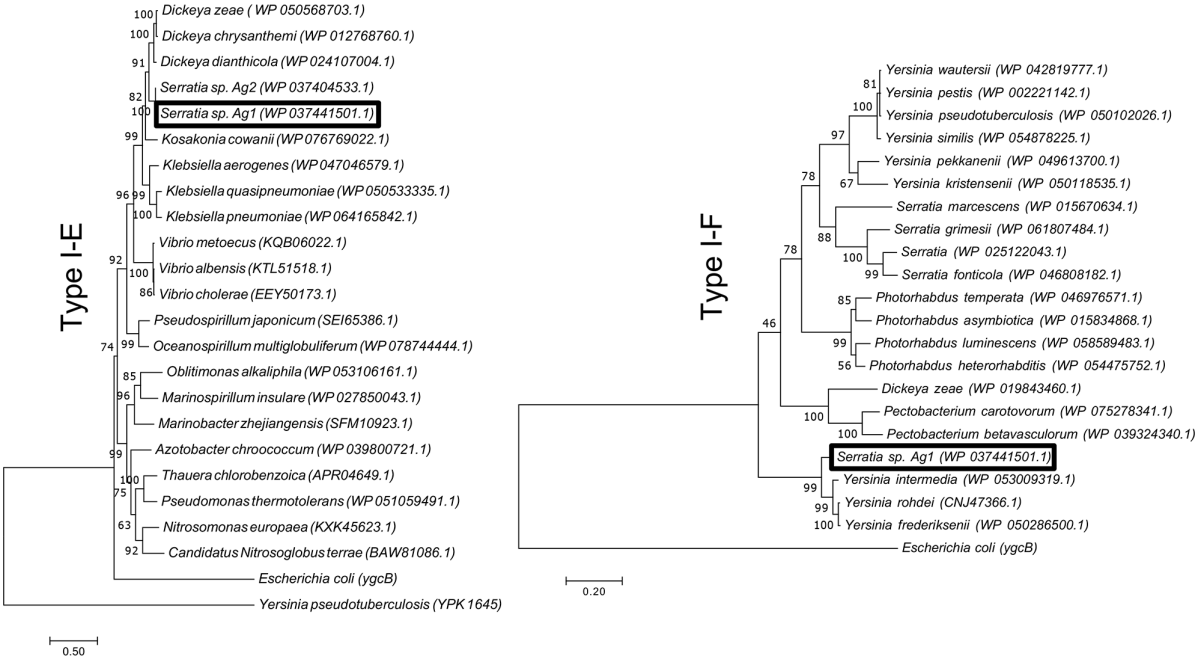

Fig. 2. Phylogenetic analyses of type I-E and type I-F Cas3 from *Serratia* sp. Ag1. Phylogenetic trees show analyses of Cas3 with the top 20 closest BLASTP hits for both trees. Maximum-likelihood trees based on the relevant Cas3 protein are shown with a bootstrap value of 1000. *E. coli* is included as a representative of the type I-E subtype, and *Y. pseudotuberculosis* is included as a representative of type I-F.

to 85% (28/33 nucleotides for the CRISPR1 array, 27/32 for CRISPR2 arrays), the number of hits decreased significantly. Of the remaining 23 spacer targets, only 1 matched to a spacer in CRISPR1 and 22 spacers matched to a phage target. Expression of the *cas* genes from both subtypes was analysed by RT-PCR and for both subtypes the expression of all *cas* genes was greater during log growth than in stationary phase (Fig. 4).

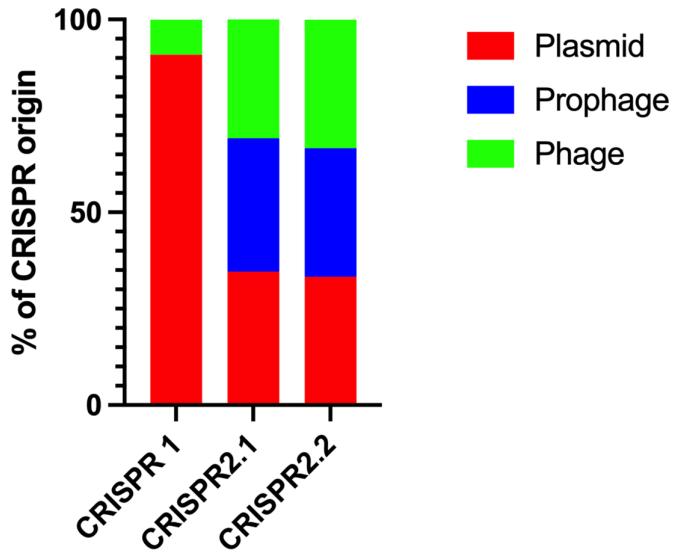

Fig. 3. Origin of exogenous nucleic acid elements in the CRISPR loci: percentage of plasmids, phage and prophage DNA found in the spacer sequences for each CRISPR array in *Serratia* sp. Ag1.

Table 2. Protospacer identity: origin of protospacer sequences matched against plasmids, phage and prophage

	Spacer	Target species	Organism	Protein target	Protein ID	Nucleotide ID	Nucleotide ID (percentage)
CRISPR1	19	Plasmid	*Rhizobium leguminosarum* bv. *viciae* plasmid pRL8	Aminotransferase class V-fold PLP-dependent	WP_011654699.1	25/33	75.76
		Plasmid	*Azospirillum brasilense* strain Az39 plasmid	Helix–turn–helix domain-containing protein	WP_040134386.1	26/33	78.79
		Plasmid	*Arthrobacter* sp. ERGS1:01 plasmid unnamed2	DUF3416 domain-containing protein	WP_082368864.1	24/33	72.73
		Plasmid	*Neorhizobium galegae* chromid pHAMBI540a	DUF1211 domain-containing protein	WP_035996019.1	25/33	75.76
	16	Phage	AJ564013_AJ564013 bacteriophage PY54	Immunity repressor Icd	CAD91791.1	28/33	84.85
	10	Plasmid	*Escherichia coli* strain 08–00022 plasmid pCFSAN004179G	Conjugative transfer relaxase/helicase trai	WP_047088735.1	27/33	81.82
	7	Plasmid	*Enterococcus faecium* strain E1 plasmid	Glycoside hydrolase family 1 protein	WP_002289584.1	26/33	78.79
		Plasmid	*Burkholderia caribensis* MBA4 plasmid	DUF1211 domain-containing protein	WP_035996019.1	25/33	75.76
		Plasmid	*Enterococcus faecium* DO plasmid 3	Beta-glucosidase	YP_006377528.1	26/33	78.79
	5	Plasmid	*Cupriavidus metallidurans* CH34 megaplasmid	Anion permease	WP_011518338.1	26/33	78.79
	1	Plasmid	*Sinorhizobium fredii* NGR234 plasmid	Hypothetical protein	NP_443999.1	24/33	72.73
CRISPR2.1	26	Phage*	*Klebsiella pneumoniae* strain Kp_Goe_149832 plasmid	Phage tail tape measure protein	WP_048292313.1	24/32	75.00
		Plasmid	*Shigella boydii* CDC 3083–94 plasmid	Type II toxin–antitoxin system	WP_000604847.1	25/32	78.13
		Plasmid	*Methylobacterium radiotolerans* JCM 2831 plasmid	MFS transporter	WP_012329600.1	24/32	75.00
		Phage	AF226852_AF226852 *Pseudomonas* phage phi8	P10	AAF63303.1	24/32	75.00
	26	Prophage	*Serratia plymuthica* 4r×13	Phage tail tape measure protein	WP_041417117.1	28/32	87.50
	25	Phage*	*Anoxybacillus amylolyticus* strain DSM 15939 plasmid	DNA primase	WP_084256491.1	25/32	78.13
	24	Prophage	*Yersinia mollaretii* strain IP25089	Baseplate protein	WP_049611672.1	30/32	93.75
	23	Prophage	*Serratia marcescens* strain 907_SMAR 425_37092_653157	Recombinase	WP_049202623.1	29/32	90.63
	21	Phage	KT898134_KT898134 *Aeromonas* phage phiARM81mr	Terminase large subunit	ALN97629.1	31/32	96.88
		Plasmid	*Synechococcus* sp. PCC 7117 plasmid unnamed5	Type I-D CRISPR-associated endonuclease Cas1	WP_065712013.1	26/32	81.25

Continued

Table 2. Continued

	Spacer	Target species	Organism	Protein target	Protein ID	Nucleotide ID	Nucleotide ID (percentage)
		Prophage	*Aeromonas hydrophila* strain Ah-HSP	Terminase	WP_077096195.1	31/32	96.88
		Plasmid	*Rhodobacter sphaeroides* ATCC 17025 plasmid	Gluconate : proton symporter	WP 011910413.1	25/32	78.13
	18	Phage	EU307292_EU307292 *Burkholderia* phage Bups phi1	Putative PAPS reductase/ sulfotransferase	ABY40518.1	27/32	84.38
		Plasmid	*Klebsiella oxytoca* strain 2880STDY5682598	Hypothetical protein	WP_064405428.1	29/32	90.63
	17	Plasmid	*Shigella flexneri* 1 a strain 0228 plasmid	Hypothetical protein	NZ_CP012733.1	26/32	81.25
		Prophage	*Cronobacter sakazakii* strain 699	Phage protein	WP_007882954.1	30/32	93.75
	16	Phage*	*Streptomyces* sp. F2 plasmid	Hypothetical protein	YP_008996313.1	26/32	81.25
	12	Phage	KY709687_KY709687 *Salmonella* phage 29 485	Hypothetical protein	ARB10913.1	27/32	84.38
	9	Plasmid	*Yersinia pseudotuberculosis* IP 32953 plasmid	Type IA DNA topoisomerase	WP_011191426.1	26/32	81.25
	7	Prophage	*Serratia marcescens* 2880STDY5682985	Hypothetical protein	WP_060433331.1	29/32	90.63
		Prophage	*Klebsiella oxytoca* strain 2880STDY5682691	Transcriptional regulator	NZ_FKZE01000009	29/32	90.63
		Prophage	*Sodalis glossinidius* str. morsitans	Hypothetical protein	WP_041867073.1	29/32	90.63
		Plasmid	*Sphingomonas sanxanigenens* DSM 19645	Hypothetical protein	WP_075153410.1	24/32	75.00
	5	Prophage	*Serratia* sp. S4	DUF968 domain-containing protein	WP_017892522.1	31/32	96.88
		Phage	JQ182729 JQ182729 *Enterobacterial* phage mEp390	Hypothetical protein	AFM76141.1	27/32	84.38
		Plasmid	*Dinoroseobacter shibae* DFL 12 plasmid	FAD-dependent oxidoreductase	WP_012187434.1	24/32	75.00
CRISPR2.2	15	Plasmid	*Alteromonas mediterranea* strain CP48 plasmid	Single-stranded DNA-binding protein	WP_071960856.1	27/32	84.38
		Plasmid	*Kangiella geojedonensis* strain YCS-5	Single-stranded DNA-binding protein	WP_046560643.1	29/32	90.63
	13	Phage*	*Methylomonas* sp. DH-1 plasmid	Atp/GTP-binding protein	NP_639742.1	26/32	81.25
		Prophage	*Yersinia aldovae* strain IP23238	Phage capsid protein	WP_049687658.1	29/32	90.63
	8	Phage*	*Klebsiella pneumoniae* strain Kp_Goe_827024 plasmid	Histidine kinase	WP_064023699.1	26/32	81.25
		Prophage	*Klebsiella pneumoniae* strain CHS159	Tail assembly protein	WP_016530340.1	30/32	93.75
	7	Phage*	*Pantoea* sp. At-9b plasmid	Hypothetical protein	WP_013511365.1	29/32	90.63
		Prophage	*Pantoea* sp. At-9b plasmid pPAT9B01	Hypothetical protein	WP_013511365.1	29/32	90.63

Continued

Table 2. Continued

	Spacer	Target species	Organism	Protein target	Protein ID	Nucleotide ID	Nucleotide ID (percentage)
	6	Phage*	*Streptomyces coelicolor* A3(2) plasmid	Atp/GTP-binding protein	NP_639742.1	26/32	81.25
	5	Prophage	*Serratia marcescens* strain 2880STDY5682949	Phage head protein	WP_060427589.1	29/32	90.63
	4	Plasmid	*Microscilla sp. pre1 plasmid*	Ms149	NP_116837.1	26/32	81.25
	1	Plasmid	*Rahnella* sp. J11-6	Noncoding		31/32	96.88

*Denotes that the organism is listed as a plasmid in the NCBI database but is suspected to be a phage due to protein content.

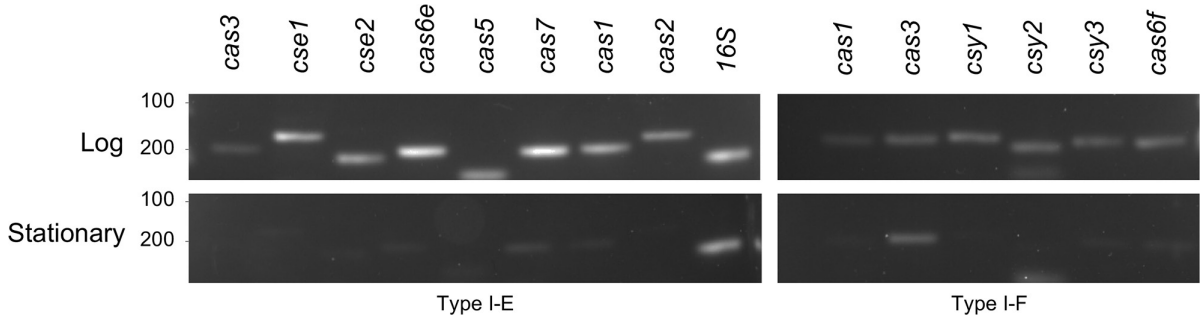

Fig. 4. Expression of the type I-E and type I-F *cas* genes. The RT-PCR analysis of expression of *cas* and *csy* genes in logarithmic and stationary growth phase of *Serratia* sp. Ag1. A 100 base pair ladder was used, and sizes are indicated to the left of the gel images.

DISCUSSION

Bacteria living in complex ecological settings are continuously challenged by predatory viruses. The CRISPR/Cas adaptative immune systems of bacteria protect bacteria from some of these challenges by targeting foreign genetic material such as plasmids and bacteriophage [16, 45]. Here, we provide evidence of CRISPR/Cas systems in the mosquito-associated *Serratia* sp. Ag1, which was isolated from *Anopheles gambiae* [39]. We have identified two type I CRISPR/Cas systems, which are typically found in the family *Enterobacteriaceae* [22, 27]. CRISPR/Cas systems in *Serratia marcescens* have been described previously, and most strains harbour a type I-F or both a type I-E and a type I-F system [27, 34, 46, 47]. One *Serratia* sp. (ATCC39006) contains both these type I systems and also a type III-A system [48].

Analysis of CRISPR spacer sequences in Ag1 confirmed the origin of many spacer sequences. Our results revealed hat 47% (23/49) of the spacer targets that we could identify originated from plasmids, while bacteriophage (phage and prophage) accounted for two-thirds of the matched spacers. Overall, 53% (26/49) of spacers matched to phage or plasmid sequences. This is higher than for other *Enterobacteriaceae*, such as *Salmonella* (12%), *E. coli* (19%) and Shiga toxin-producing *E. coli* (8%) [49] . Extensive spacer sequence analysis has been performed in the genomes of *Enterobacteriaceae*, which are human pathogens and commensals [27, 50–52]. A discrete number of spacers from the *Enterobacteriaceae* members appear to be acquired from extrachromosomal genetic elements, such as plasmids and bacteriophage, while other spacers match to the bacterial host genome in non-prophage regions, although many of the spacers are still of unknown origin [27].

Cas proteins play crucial roles in all three steps of CRISPR/Cas immunity [21, 53]. Our results showed active expression of all *cas* genes during the actively dividing logarithmic growth phage of bacteria and attenuation of all but the type I-F *cas3* gene during stationary phase. This is concordant with previous studies in *E. coli* showing repression of the type I-E *cas3* gene expression during stationary phase compared to log phase [54–56]. Our results for expression analyses of type I-F *cas* gene show continuous expression even in the stationary phase. This result is similar to what was reported in the phytopathogen *Pectobacterium atrosepticum* showing expression of Cas protein in both exponential and stationary growth stages [57, 58]

CRISPR spacer sequences can be used for bacterial subtyping [59]. The presence of the type I-F cas3 in multiple *Serratia* spp. suggests that these genomes likely also contain CRISPR arrays. This would depend on CRISPR arrays being present in all strains of the species and exhibiting strain-to-strain variability that could be exploited for subtyping. Whole-genome sequencing of four *Serratia marcescens* genomes showed that CRISPR/Cas systems were absent in half of these [34]. The prevalence of CRISPR/Cas systems and the diversity of spacer content in other *Serratia* spp. is yet to be determined and would need to be performed to determine the utility of CRISPR typing in this bacterium.

While Cas1 and Cas2 are mainly involved in acquiring the spacers from newly invading phage and foreign genetic material, the cas2/3 ; csy complexes are involved in the priming method for spacer acquisition [60, 61]. Our results show the presence of newly acquired spacer sequences, suggesting that adaptation is occurring actively in these bacteria. Hence, there is the possibility of recurrent encounters between phage and symbiotic bacteria in the mosquito gut. A recent study demonstrated that phage infection can alter bacterial levels in mosquitoes and alter their development in aquatic stages [62]. These phage may also be part of the mosquito gut microbiome, where they interact with gut bacteria and compete for nutritional resources.

The CRISPR/Cas system in bacteria has been explored extensively in terms of its application in different fields, such as human and agriculture diseases [16, 21, 63, 64]. However, analysis of CRISPR loci in the host-associated symbiotic bacteria is limited, especially the role of CRISPR systems in the host–microbe interactions. Apart from anti-viral defence, CRISPR has been shown to be involved in DNA repair, colonization and host immune evasion [65–67]. Hence, by modifying the CRISPR loci the colonization of bacteria in the host environment could be investigated. Such studies are important in deciphering the host–microbe interactions

in complex ecological settings such as the mosquito microbiome. In this regard, further studies are needed to analyse CRISPR loci in the mosquito symbionts and understand the mechanistic basis for CRISPR loci-mediated host–microbe interactions.

Funding information
S.H. was supported by an LSTM Director's Catalyst Fund and Royal Society Research Grant (RGS\R1\231156). G.L.H. was supported by the BBSRC (BB/T001240/1, BB/V011278/1, BB/W018446/1), the UKRI (20197 and 85336), the EPSRC (V043811/1), a Royal Society Wolfson Fellowship (RSWF\R1\180013), the NIHR (NIHR2000907) and the Bill and Melinda Gates Foundation (INV-048598).

Acknowledgements
We thank Naufa Amirani for help with the phylogenetic analyses of cas3 and Phillipe Horvath for providing the macro to visualize the spacer sequences.

Author contributions
H.E.R, S.H and N. S: conceptualization, methodology, investigation, data curation, writing – original draft preparation and visualization. H.E.R and N.S: investigation and data curation. H.E.R, S.H, G.L.H and N. S.: conceptualization methodology, data curation, resources, writing – review and editing and funding. G.L.H and N.S.: conceptualization, resources, writing – review and editing, supervision and funding.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Hegde S, Rasgon JL, Hughes GL. The microbiome modulates arbovirus transmission in mosquitoes. *Curr Opin Virol* 2015;15:97–102.
2. Cansado-Utrilla C, Zhao SY, McCall PJ, Coon KL, Hughes GL. The microbiome and mosquito vectorial capacity: rich potential for discovery and translation. *Microbiome* 2021;9:111.
3. Gabrieli P, Caccia S, Varotto-Boccazzi I, Arnoldi I, Barbieri G, *et al*. Mosquito Trilogy: microbiota, immunity and pathogens, and their implications for the control of disease transmission. *Front Microbiol* 2021;12:630438.
4. Huang W, Wang S, Jacobs-Lorena M. Use of microbiota to fight mosquito-borne disease. *Front Genet* 2020;11.
5. Caragata EP, Tikhe CV, Dimopoulos G. Curious entanglements: interactions between mosquitoes, their microbiota, and arboviruses. *Curr Opin Virol* 2019;37:26–36.
6. Caragata EP, Short SM. Vector microbiota and immunity: modulating arthropod susceptibility to vertebrate pathogens. *Curr Opin Insect Sci* 2022;50:100875.
7. Chandler JA, Liu RM, Bennett SN. RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. *Front Microbiol* 2015;6:185.
8. He X, Yin Q, Zhou L, Meng L, Hu W, *et al*. Metagenomic sequencing reveals viral abundance and diversity in mosquitoes from the Shaanxi-Gansu-Ningxia region, China. *PLoS Negl Trop Dis* 2021;15:e0009381.
9. Shi C, Beller L, Deboutte W, Yinda KC, Delang L, *et al*. Stable distinct core eukaryotic viromes in different mosquito species from Guadeloupe, using single mosquito viral metagenomics. *Microbiome* 2019;7:121.
10. Atoni E, Wang Y, Karungu S, Waruhiu C, Zohaib A, *et al*. Metagenomic virome analysis of culex mosquitoes from Kenya and China. *Viruses* 2018;10:30.
11. Shi C, Liu Y, Hu X, Xiong J, Zhang B, *et al*. A metagenomic survey of viral abundance and diversity in mosquitoes from Hubei province. *PLoS One* 2015;10:e0129845.
12. Batson J, Dudas G, Haas-Stapleton E, Kistler AL, Li LM, *et al*. Single mosquito metatranscriptomics identifies vectors, emerging pathogens and reservoirs in one assay. *Elife* 2021;10:e68353.
13. Kozlova EV, Hegde S, Roundy CM, Golovko G, Saldaña MA, *et al*. Microbial interactions in the mosquito gut determine Serratia colonization and blood-feeding propensity. *ISME J* 2021;15:93–108.
14. Hegde S, Khanipov K, Albayrak L, Golovko G, Pimenova M, *et al*. Microbiome interaction networks and community structure from laboratory-reared and field-collected *Aedes aegypti*, *Aedes albopictus*, and *Culex quinquefasciatus*, mosquito vectors. *Front Microbiol* 2018;9:2160.
15. Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 2007;8:172.
16. Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie* 2015;117:119–128.
17. Barrangou R. CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip Rev RNA* 2013;4:267–278.
18. Makarova KS, Koonin EV. Annotation and classification of CRISPR-Cas systems. *Methods Mol Biol* 2015;1311:47–75.
19. Deveau H, Garneau JE, Moineau S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol* 2010;64:475–493.
20. Mohanraju P, Saha C, van Baarlen P, Louwen R, Staals RHJ, *et al*. Alternative functions of CRISPR-Cas systems in the evolutionary arms race. *Nat Rev Microbiol* 2022;20:351–364.
21. Hille F, Charpentier E. CRISPR-Cas: biology, mechanisms and relevance. *Philos Trans R Soc Lond B Biol Sci* 2016;371:1707.
22. Xue C, Sashital DG. Mechanisms of type I-E and I-F CRISPR-Cas systems in *Enterobacteriaceae EcoSal Plus* 2019;8.
23. Makarova KS, Zhang F, Koonin EV. SnapShot: Class 1 CRISPR-Cas systems. *Cell* 2017;168:946– .
24. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, *et al*. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* 2015;13:722–736.
25. Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, *et al*. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 2011;9:467–477.
26. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, *et al*. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* 2020;18:67–83.
27. Medina-Aparicio L, Dávila S, Rebollar-Flores JE, Calva E, Hernández-Lucas I. The CRISPR-Cas system in *Enterobacteriaceae*. *Pathog Dis* 2018;76.
28. Cui L, Wang X, Huang D, Zhao Y, Feng J, *et al*. CRISPR-*cas3* of *Salmonella* upregulates bacterial biofilm formation and virulence to host cells by targeting quorum-sensing systems. *Pathogens* 2020;9:53.
29. Münch PC, Franzosa EA, Stecher B, McHardy AC, Huttenhower C. Identification of natural CRISPR systems and targets in the human microbiome. *Cell Host Microbe* 2021;29:94–106..
30. Hidalgo-Cantabrana C, Crawley AB, Sanchez B, Barrangou R. Characterization and exploitation of CRISPR loci in *Bifidobacterium longum Front Microbiol* 2017;8:1851.
31. Soto-Perez P, Bisanz JE, Berry JD, Lam KN, Bondy-Denomy J, *et al*. CRISPR-cas system of a prevalent human gut bacterium reveals hyper-targeting against phages in a human virome catalog. *Cell Host Microbe* 2019;26:325–335..

32. Crawley AB, Henriksen ED, Stout E, Brandt K, Barrangou R. Characterizing the activity of abundant, diverse and active CRISPR-Cas systems in lactobacilli. *Sci Rep* 2018;8:11544.
33. Yang L, Li W, Ujiroghene OJ, Yang Y, Lu J, *et al*. Occurrence and diversity of CRISPR loci in *Lactobacillus casei* group. *Front Microbiol* 2020;11:624.
34. Scrascia M, D'Addabbo P, Roberto R, Porcelli F, Oliva M, *et al*. Characterization of CRISPR-Cas systems in *Serratia marcescens* isolated from *Rhynchophorus ferrugineus* (Olivier, 1790) (Coleoptera: Curculionidae). *Microorganisms* 2019;7:368.
35. Toyomane K, Yokota R, Watanabe K, Akutsu T, Asahi A, *et al*. Evaluation of CRISPR diversity in the human skin microbiome for personal identification. *mSystems* 2021;6:e01255-20.
36. Hegde S, Nilyanimit P, Kozlova E, Anderson ER, Narra HP, *et al*. CRISPR/Cas9-mediated gene deletion of the ompA gene in symbiotic Cedecea neteri impairs biofilm formation and reduces gut colonization of *Aedes aegypti* mosquitoes. *PLoS Negl Trop Dis* 2019;13:e0007883.
37. Kistler KE, Vosshall LB, Matthews BJ. Genome engineering with CRISPR-Cas9 in the mosquito *Aedes aegypti*. *Cell Reports* 2015;11:51–60.
38. Chaverra-Rodriguez D, Macias VM, Hughes GL, Pujhari S, Suzuki Y, *et al*. Targeted delivery of CRISPR-Cas9 ribonucleoprotein into arthropod ovaries for heritable germline gene editing. *Nat Commun* 2018;9.
39. Pei D, Hill-Clemons C, Carissimo G, Yu W, Vernick KD, *et al*. Draft genome sequences of two strains of *Serratia* spp. from the midgut of the malaria mosquito Anopheles gambiae. *Genome Announc* 2015;3:e00090-15.
40. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 2007;35:W52–7.
41. Horvath P, Romero DA, Coûté-Monvoisin A-C, Richards M, Deveau H, *et al*. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *J Bacteriol* 2008;190:1401–1412.
42. Biswas A, Gagnon JN, Brouns SJJ, Fineran PC, Brown CM. CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA Biol* 2013;10:817–827.
43. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.
44. Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* 2017;37:67–78.
45. Newsom S, Parameshwaran HP, Martin L, Rajan R. The CRISPR-Cas mechanism for adaptive immunity and alternate bacterial functions fuels diverse biotechnologies. *Front Cell Infect Microbiol* 2020;10:619763.
46. Srinivasan VB, Rajamohan G. Genome analysis of urease positive *Serratia marcescens*, co-producing SRT-2 and AAC(6')-Ic with multidrug efflux pumps for antimicrobial resistance. *Genomics* 2019;111:653–660.
47. Vicente CSL, Nascimento FX, Barbosa P, Ke H-M, Tsai IJ, *et al*. Evidence for an opportunistic and endophytic lifestyle of the *Bursaphelenchus xylophilus*-associated bacteria *Serratia marcescens* PWN146 isolated from wilting Pinus pinaster. *Microb Ecol* 2016;72:669–681.
48. Malone LM, Hampton HG, Morgan XC, Fineran PC. Type I CRISPR-Cas provides robust immunity but incomplete attenuation of phage-induced cellular stress. *Nucleic Acid Res* 2022;50:160–174.
49. Dang TND, Zhang L, Zöllner S, Srinivasan U, Abbas K, *et al*. Uropathogenic *Escherichia coli* are less likely than paired fecal *E. coli* to have CRISPR loci. *Infect Genet Evol* 2013;19:212–218.
50. Shariat N, Timme RE, Pettengill JB, Barrangou R, Dudley EG. Characterization and evolution of *Salmonella* CRISPR-Cas systems. *Microbiology* 2015;161:374–386.
51. Yin S, Jensen MA, Bai J, Debroy C, Barrangou R, *et al*. The evolutionary divergence of Shiga toxin-producing *Escherichia coli* is reflected in clustered regularly interspaced short palindromic repeat (CRISPR) spacer composition. *Appl Environ Microbiol* 2013;79:5710–5720.
52. Hidalgo-Cantabrana C, Sanozky-Dawes R, Barrangou R. Insights into the human virome using CRISPR spacers from microbiomes. *Viruses* 2018;10:479.
53. Sorek R, Lawrence CM, Wiedenheft B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem* 2013;82:237–266.
54. Majsec K, Bolt EL, Ivančić-Baće I. Cas3 is a limiting factor for CRISPR-Cas immunity in *Escherichia coli* cells lacking H-NS. *BMC Microbiol* 2016;16:28.
55. Mitić D, Radovčić M, Markulin D, Ivančić-Baće I. StpA represses CRISPR-Cas immunity in H-NS deficient *Escherichia coli*. *Biochimie* 2020;174:136–143.
56. Westra ER, Pul U, Heidrich N, Jore MM, Lundgren M, *et al*. H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol Microbiol* 2010;77:1380–1393.
57. Vercoe RB, Chang JT, Dy RL, Taylor C, Gristwood T, *et al*. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity Islands. *PLoS Genet* 2013;9:e1003454.
58. Przybilski R, Richter C, Gristwood T, Clulow JS, Vercoe RB, *et al*. Csy4 is responsible for CRISPR RNA processing in *Pectobacterium atrosepticum*. *RNA Biol* 2011;8:517–528.
59. Shariat N, Dudley EG. CRISPRs: molecular signatures used for pathogen subtyping. *Appl Environ Microbiol* 2014;80:430–439.
60. Yosef I, Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* 2012;40:5569–5576.
61. Swarts DC, Mosterd C, van Passel MWJ, Brouns SJJ. CRISPR interference directs strand specific spacer acquisition. *PLoS One* 2012;7:e35888.
62. Tikhe CV, Dimopoulos G. Phage therapy for mosquito Larval control: a proof-of-principle study. *mBio* 2022;13:e0301722.
63. Zhu H, Li C, Gao C. Applications of CRISPR–Cas in agriculture and plant biotechnology. *Nat Rev Mol Cell Biol* 2020;21:661–677.
64. Liu Z, Dong H, Cui Y, Cong L, Zhang D. Application of different types of CRISPR/Cas-based systems in bacteria. *Microb Cell Fact* 2020;19.
65. Varble A, Campisi E, Euler CW, Maguin P, Kozlova A, *et al*. Prophage integration into CRISPR loci enables evasion of antiviral immunity in Streptococcus pyogenes. *Nat Microbiol* 2021;6:1516–1525.
66. Babu M, Beloglazova N, Flick R, Graham C, Skarina T, *et al*. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol Microbiol* 2011;79:484–502.
67. Veesenmeyer JL, Andersen AW, Lu X, Hussa EA, Murfin KE, *et al*. NilD CRISPR RNA contributes to *Xenorhabdus nematophila* colonization of symbiotic host nematodes. *Mol Microbiol* 2014;93:1026–1042.

Peer review history

VERSION 3

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000599.v3.3

Giuseppina Mariano; University of Surrey, UNITED KINGDOM

Date report received: 31 July 2023
Recommendation: Accept

Comments: The work presented is clear and the arguments well formed.

SciScore report

https://doi.org/10.1099/acmi.0.000599.v3.1

iThenticate report

https://doi.org/10.1099/acmi.0.000599.v3.2

Author response to reviewers to Version 2

Dear Editor,

Thanks for editing this manuscript and providing valuable feedback. As suggested, we have now included the parameters used in the bioinformatics analysis of CRISPR array and other analysis. We have added that in the main manuscript in the method sections. We have added the details in the section "*Identification of CRISPR loci, phylogenetic analyses, and spacer identification*".The added information is highlighted below in red.

The assembled Ag1 genome was analysed using CRISPR-Finder (40) to identify both the CRISPR arrays and the *cas*genes. We have used default setting to analyse the Ag1 genome to identify the CRISPR array and *cas*genes. Spacers were extracted from the arrays and analyzed using an Excel-based macro (41). CRISPR Target (42) was used to identify putative spacer matches. Here, we used default parameters for the initial BLAST screen and target database. For initial output display parameters, we used default score cut-off of 20, 26/32 base pairs.

VERSION 2

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000599.v2.3

Giuseppina Mariano; University of Surrey, UNITED KINGDOM

Date report received: 25 July 2023
Recommendation: Minor Amendment

Comments: Dear Authors, thank you for addressing the comments from the referees. Before I can accept this work, I kindly ask you in your methods to clearly indicate for each bioinformatic tool you used, that you have done so by using default parameters (or any other parameters you have used). You have addressed this in your response to referees but I'd like you to include this information in the methods for the sake of helping other researchers that may want to use your papers and methods for analysis in other organisms. Whilst it seems intuitive, a less expert reader may not find the lack of specification to be a clear sign that default parameters were used.

SciScore report

https://doi.org/10.1099/acmi.0.000599.v2.1

iThenticate report

https://doi.org/10.1099/acmi.0.000599.v2.2

Author response to reviewers to Version 1

Dear Editor,

We would like to thank the reviewers for taking the time to evaluate our manuscript and provide valuable feedback and suggestions. Please find below the point-by-point response to the reviewers' comments are in red. We have also modified the original manuscript by incorporating the reviewers' comments and suggestions.

Reviewer 1 comments:

Reviewer 1 Comments to Author: This manuscript describes the crispr loci of a strain of Serratia associated with the mosquito gut. There are two type I crispr systems and 3 associated CRISPR loci. The genes are shown to be transcribed more highly during exponential than log phase. Perhaps the most interesting aspect of this paper is the observation that crispr 1 seems to be mostly targeting plasmids while 2.1 and 2.2 mostly target phage. Given the small number of spacers and the numbers that don't match well to any target, this is a bit hard to interpret the significance of this.

Overall, this is a short, descriptive paper that represents a rather incremental contribution to the literature. One suggestion to strengthen it without too much extra work would be to analyse the crisprs of the highly related Ag1 system to determine whether they are related to the Ag2 crisprs, and whether there is a similar bias in spacer origin.

We appreciate the reviewer's comments on the overall significance of this work and highlighting how this study represents a contribution to the literature. We analysed two mosquito symbiont isolates and found a total of two CRISPR systems. Our analysis of the CRISPR system in the two genomes Ag1 and Ag2 did not find any CRISPR array in the Ag2. Hence, we think that the Serratia Ag2 genome lacks a CRISPR system. We agree with the reviewer's intention that it will be interesting to observe the closely related systems in these two genomes isolated from the same mosquito species. However, our efforts did not yield any significant results concerning CRISPR array in the Ag2 genome and hence, we speculate that the Ag2 might not have acquired the CRISPR array.

Specific points:

1. In figure 1, indicate that there is a fused Cas2-Cas3 gene for type I-F crispr

We have modified the figure 1 and attached the new figure.

2. P3 line 53. The term "small" here is not clear - small in relation to what? They are large in comparison with most bacterial immune systems.

"small" refers to small RNAs; i.e. the CRISPR-RNAs/crRNAs

3. P7 line 166 - please check this sentence as it looks like there's a problem.

We have modified the sentence and now it reads "Of the remaining 23 spacer targets, only one matched to a spacer in CRISPR1 and 22 spacers matched to a phage target".

4. The RT-PCR analyses seem to have been carried out only once. It would be preferable to repeat this to provide more confidence in the data and conclusions drawn.

The RT-PCR experiments were run independently twice (different bacterial cultures), and on one of these, the PCR step was performed twice on the same cDNA. The gel below is from the independent experiment from the figure shown in the paper. Densitometry was not performed as the difference between log and stationary phase was clear.

Reviewer 2 Comments to Author: This study describes the identification of two CRISPR-Cas systems in Serratia Sp. Ag1 and investigates the origin of the spacers in each array.

The introductory literature analysis provides an overview of the relevant CRISPR subtypes allowing the non-expert reader to understand the context of the paper findings.

Overall, the materials and methods section is concisely written, potentially contributing to the loss of some detail. Though the reasoning for the below points may be obvious to a CRISPR-focused scientist, they may not be to a general audience. The figures and tables are sensible representations of the underlying data, though figure 3 may contain small typos. Throughout the results section there are a few points to address, but overall, the methods are likely sound, assuming proper replicates of the RT-PCR experiments were undertaken.

In terms of discussion of results, it is my view that the findings of the paper could explored further. Whilst there is discussion of the concordance of the results of this study and that of others looking at the E. coli type I-E system, there is no discussion of how this work's Type I-F results fit into the field, for example. Further, the repressor H-NS is mentioned, but there is no effort to put this into context of the results presented here. Is there a H-NS homologue in Serratia, for example?

We thank reviewer for helpful comments and suggestions on our manuscript. We have now corrected the manuscript considering the suggestions of reviewer. As suggested, we have added following discussion points (line number 206) about the Type I-F system. "Our results on the expression analyses of Type I-F *cas*gene shows continuous expression even in the stationary phase. This result is similar to what has been reported in a phytopathogen *Pectobacterium atrosepticum*showing expression of cas protein in both exponential and stationary growth stages (57, 58)". We have also performed the sequence analysis between E. coli H-NS and homologous sequence in Serratia Sp. Ag1 genome. The results shows that there is 77% similarity between *Serratia*transcriptional regulator gene and E. coli H-NS gene (see below the alignment).

H-NS ATGAGCGAAGCACTTAAAATTCTGAACAACATCCGTACTCTTCGTGCGCAGGCAAGAGAA

transcriptional ATGAGCGAAGCATTAAAGATTTTGAACAACATCCGTACTCTACGTGCACAGGCTAGAGAA

********** * ** *** ****************** ***** ***** ******

H-NS TGTACACTTGAAACGCTGGAAGAAATGCTGGAAAAATTAGAAGTTGTTGTTAACGAACGT

transcriptional TGCAGCTTGGAAACACTGGAAGAGATGCTTGAGAAATTGGAAGTTGTTGTTAACGAGCGT

** * * ***** ******** ***** ** ***** ***************** ***

H-NS CGCGAAGAAGAAAGCGCGGCTGCTGCTGAAGTTGAAGAGCGCACTCGTAAACTGCAGCAA

transcriptional CGTGATGAAGACAGCCAAGCTCAAGCAGAAATTGAAGAGCGTACTCGCAAACTGCAACAA

** ** ***** *** *** ** *** ********** ***** ******** ***

H-NS TATCGCGAAATGCTGATCGCTGACGGTATTGACCCGAACGAACTGCTGAATAGCCTTGCT

transcriptional TATCGTGAAATGCTGATTGCTGATGGTATTGATCCAAACGAATTGCTGCAAACAATGGCT

***** ********** ***** ******** ** ****** ***** * * * ***

H-NS GCCGTTAAATCTGGCACCAAAGCTAAACGTGCTCAGCGTCCGGCAAAATATAGCTACGTT

transcriptional GCTACTAAAGCCGCTGGCAAAGCAAAACGTGCTGCGCGCCCAGCTAAATACCAATATAAA

** **** * * ****** ********* *** ** ** ***** **

H-NS GACGAAAACGGCGAAACTAAAACCTGGACTGGCCAAGGCCGTACTCCAGCTGTAATCAAA

transcriptional GATGAAAACGGCGAAATGAAAACCTGGACTGGCCAGGGCCGTACCCCAGCTGTGATTAAA

** ************ ********** ******** ******** ** ***

H-NS AAAGCAATGGATGAGCAAGGTAAATCCCTCGACGATTTCCTGATCAAGCAATAA

transcriptional AAAGCTCTCGAAGAGCAGGGAAAATCCTTAGACGATTTCCTG----------

***** * ** ***** ** ****** * **********

The authors should double check their grammar and conventions (I.e., ensuring italicisation of gene and species names) throughout to improve readability.

Thanks for suggestions and we have thoroughly checked the grammar and conventions. We have also italicized wherever necessary.

Specific points to address:

Line 20: reword to ensure that it isn't implied, as it currently is, that Serratia Sp. Ag1 is a novel bacterium. It is the discovery of the CRISPR-Cas systems that is novel.

We have now rephrased the sentence which is now reads "we characterized two novel CRISPR-Cas systems in *Serratia*Sp. Ag1, that was isolated from the gut of an *Anopheles gambiae*mosquito."

Line 59: Introduce the concept of a leader sequence.

We have included the definition of leader sequence (line number 59). Now the sentence reads as follows "…..being found closest to the leader sequence, which is directly upstream of first repeat containing regulatory elements necessary for adaptation."

Line 81: typo. Correct from "has" to have.

Corrected!

Line 89: should read "between the gut" to improve sentence structure.

Corrected!

Paragraph starting line 111: How many repetitions of the RT-PCR experiments were undertaken? Was densitometry of the gel undertaken? If not, why? This would allow quantitative comparison between log and stationary growth, and statistical analysis of the results between replicates.

We have now added following information at the end of section RT-PCR expression analysis. "The RT-PCR experiments were run independently twice (different bacterial cultures), and on one of these, the PCR step was performed twice on the same cDNA. The gel below is from the independent experiment from the figure shown in the paper. Densitometry was not performed as the difference between log and stationary phase was clear."

Line 120-121; why were different numbers of cycles used for the control/ cas genes? Perhaps this is a lack of experience on my part.The 16S rRNA gene is much more highly expressed than the *cas*genes. If we had done higher cycles of 16S, the brightness of the bands would be completely saturated.

Line 126: What settings were used when CRISPR-Finder was used?

There are not 'settings' on CRISPR-Finder – we used the default.

Line 128: What settings were used when CRISPR-Target was used?

We used the default except were noted

Line 129: Is there any precedent for the ~70-75% matching nucleotides to identify spacer sequences?

There is no precedent for using ~70-75% matching nucleotide. We have used default setting initially then increased the stringency level to 85% (see line number 171).

Line 131; how similar were the "20 similar sequences from different species"?

We do not have a record of this; BLAST-p was used and the top 20 hits belonging to different species were selected. The bacteria selected appear in the phylogenetic tree (Figure 2)

Line 132; What settings were used within MEGA7? I am not familiar with the software, but one would expect additional information on the settings used would be required for replication.

The default settings for ClustalW protein alignment analysis were used (including a gap penalty of 10)..

Line 142; Cluster 2 and Cluster 4 are not introduced in the text. What is the significance of this assignation?

This is assigned based on the publication by Kunin et al 2017, which has now been cited.

Line 148; as mentioned above, the details of the matches to other bacterial species Cas3 homologues should be provided.

We do not have a record of this; BLAST-p was used and the top 20 hits belonging to different species were selected. The bacteria selected appear in the phylogenetic tree (Figure 2)

Line 163; clarification of text. "…could identify, most targets were" would read better as "…could identify, with most identified targets…"

Corrected as suggested.

Line 165; remove "we reduced" so the sentence reads "…27/32 for CRISPR2 arrays), the number of hits reduced significantly" to improve clarity.

Corrected as suggested.

Line 174; provide references.

We added references here.

Line 186; STEC not defined

We have now defined STEC.

Line 187-189; Again, this could be my lack of specialist knowledge, but I don't see why the presence of a phage record in NCBI suggests that acquisition events occurred recently? Please clarify.

We agree with reviewer's suggestion. We have now deleted that sentence.

Line 191-194: Several typos; "Discrete" to "A discrete". "spaces" to "spacers". Remove "the", it is unnecessary. Provide references for the Enterobacteriaceae data.

Corrected

Line 200-201: What value is this sentence adding as it is? What is the context in Serratia? Is there any context for the Type I-F system?

The sentence is providing the mechanistic basis of cas3 repression in E. coli. Since these studies are not directly linked to Type I-F system, we have now deleted the sentence. "In the *E. coli*system, the transcription repressor, H-NS, is responsible for repressing *cas3*expression (54, 55)."

Line 216: reword to "recurrent encounters between phages and symbiotic bacteria"

Done!

Line 217: reword to "demonstrated that phage infection can alter bacterial levels" if appropriate.

Done!

Line 219: Interact, not interaction.

Done!

In Figure 3, the X axis reads CRISPR1, CRISPR2, and CRISPR3. What is CRISPR3? Is this a typo where the CRISPR 2 field should read CRISPR 2.1 and CRISPR 3 read CRISPR 2.2? If not, please explain the labelling more comprehensively in the figure legend.

We have now modified the figure to correct this mistake and also, we have corrected the figure legend.

Line 434 typo; "closes BLAST-p" ought to read "closest"

We have now corrected the typo.

Line 439 in the Fig 3 figure legend, there is a typo. "CRISP loci" should read CRISPR loci

We have corrected this typo in the figure legend.

In table 2 Some target species names are marked with an asterix. What is the significance of this? It seems to relate to fields where the target species is listed as phage, but the organism suggests the sequence is that of a plasmid? Please address this in the table legend or resolve.

We have now added footnote to explain the asterisks.

Also in table 2, the nucleotide ID is listed as a fraction. It would be more helpful if this were a percentage.

We have now added an extra column with percentage.

VERSION 1

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000599.v1.5

Giuseppina Mariano; University of Surrey, UNITED KINGDOM

Date report received: 31 May 2023
Recommendation: Minor Amendment

Comments: The work presented is clear and the arguments well formed. This study would be a valuable contribution to the existing literature.

Reviewer 2 recommendation and comments

https://doi.org/10.1099/acmi.0.000599.v1.3

Anonymous.

Date report received: 31 May 2023
Recommendation: Minor Amendment

Comments: This study describes the identification of two CRISPR-Cas systems in Serratia Sp. Ag1 and investigates the origin of the spacers in each array. The introductory literature analysis provides an overview of the relevant CRISPR subtypes allowing the non-expert reader to understand the context of the paper findings. Overall, the materials and methods section is concisely written, potentially contributing to the loss of some detail. Though the reasoning for the below points may be obvious to a CRISPR-focused scientist, they may not be to a general audience. The figures and tables are sensible representations of the underlying data, though figure 3 may contain small typos. Throughout the results section there are a few points to address, but overall, the methods are likely sound, assuming proper replicates of the RT-PCR experiments were undertaken. In terms of discussion of results, it is my view that the findings of the paper could explored further. Whilst there is discussion of the concordance of the results of this study and that of others looking at the E. coli type I-E system, there is no discussion of how this work's Type I-F results fit into the field, for example. Further, the repressor H-NS is mentioned, but there is no effort to put this into context of the results presented here. Is there a H-NS homologue in Serratia, for example? The authors should double check their grammar and conventions (I.e., ensuring italicisation of gene and species names) throughout to improve readability. Specific points to address: Line 20: reword to ensure that it isn't implied, as it currently is, that Serratia Sp. Ag1 is a novel bacterium. It is the discovery of the CRISPR-Cas systems that is novel. Line 59: Introduce the concept of a leader sequence Line 81: typo. Correct from "has" to have. Line 89: should read "between the gut" to improve sentence structure Paragraph starting line 111: How many repetitions of the RT-PCR experiments were undertaken? Was densitometry of the gel undertaken? If not, why? This would allow quantitative comparison between log and stationary growth, and statistical analysis of the results between replicates. Line 120-121; why were different numbers of cycles used for the control/ cas genes? Perhaps this is a lack of experience on my part. Line 126: What settings were used when CRISPR-Finder was used? Line 128: What settings were used when CRISPR-Target was used? Line 129: Is there any precedent for the ~70-75% matching nucleotides to identify spacer sequences? Line 131; how similar were the "20 similar sequences from different species"? Line 132; What settings were used within MEGA7? I am not familiar with the software, but one would expect additional information on the settings used would be required for replication. Line 142; Cluster 2 and Cluster 4 are not introduced in the text. What is the significance of this assignation? Line 148; as mentioned above, the details of the matches to other bacterial species Cas3 homologues should be provided. Line 163; clarification of text. "…could identify, most targets were" would read better as "…could identify, with most identified targets…" Line 165; remove "we reduced" so the sentence reads "…27/32 for CRISPR2 arrays), the number of hits reduced significantly" to improve clarity. Line 174; provide references. Line 186; STEC not defined Line 187-189; Again, this could be my lack of specialist knowledge, but I don't see why the presence of a phage record in NCBI suggests that acquisition events occurred recently? Please clarify. Line 191-194: Several typos; "Discrete" to "A discrete". "spaces" to "spacers". Remove "the", it is unnecessary. Provide references for the Enterobacteriaceae data. Line 200-201: What value is this sentence adding as it is? What is the context in Serratia? Is there any context for the Type I-F system? Line 216: reword to "recurrent encounters between phages and symbiotic bacteria" Line 217: reword to "demonstrated that phage infection can alter bacterial levels" if appropriate. Line 219: Interact, not interaction. In Figure 3, the X axis reads CRISPR1, CRISPR2, and CRISPR3. What is CRISPR3? Is this a typo where the CRISPR 2 field should read CRISPR 2.1 and CRISPR 3 read CRISPR 2.2? If not, please explain the labelling more comprehensively in the figure legend. Line 434 typo; "closes BLAST-p" ought to read "closest" Line 439 in the Fig 3 figure legend, there is a typo. "CRISP loci" should read CRISPR loci In table 2 Some target species names are marked with an asterix. What is the significance of this? It seems to relate to fields where the target species is listed as phage, but the organism suggests the sequence is that of a plasmid? Please address this in the table legend or resolve. Also in table 2, the nucleotide ID is listed as a fraction. It would be more helpful if this were a percentage.

Please rate the manuscript for methodological rigour
Satisfactory

Please rate the quality of the presentation and structure of the manuscript
Satisfactory

To what extent are the conclusions supported by the data?
Partially support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

Reviewer 1 recommendation and comments

https://doi.org/10.1099/acmi.0.000599.v1.4

Malcolm White; University of St Andrews School of Biology, School of Biology, North Haugh, St Andrews, UNITED KINGDOM

Date report received: 24 April 2023
Recommendation: Minor Amendment

Comments: This manuscript describes the crispr loci of a strain of Serratia associated with the mosquito gut. There are two type I crispr systems and 3 associated CRISPR loci. The genes are shown to be transcribed more highly during exponential than log phase. Perhaps the most interesting aspect of this paper is the observation that crispr 1 seems to be mostly targeting plasmids while 2.1 and 2.2 mostly target phage. Given the small number of spacers and the numbers that don't match well to any target, this is a bit hard to interpret the significance of this. Overall, this is a short, descriptive paper that represents a rather incremental contribution to the literature. One suggestion to strengthen it without too much extra work would be to analyse the crisprs of the highly related Ag1 system to determine whether they are related to the Ag2 crisprs, and whether there is a similar bias in spacer origin. Specific points: 1. In figure 1, indicate that there is a fused Cas2-Cas3 gene for type I-F crispr 2. P3 line 53. The term "small" here is not clear - small in relation to what? They are large in comparison with most bacterial immune systems. 3. P7 line 166 - please check this sentence as it looks like there's a problem. 4. The RT-PCR analyses seem to have been carried out only once. It would be preferable to repeat this to provide more confidence in the data and conclusions drawn. Malcolm White

Please rate the manuscript for methodological rigour
Satisfactory

Please rate the quality of the presentation and structure of the manuscript
Good

To what extent are the conclusions supported by the data?
Partially support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

SciScore report

https://doi.org/10.1099/acmi.0.000599.v1.1

iThenticate report

https://doi.org/10.1099/acmi.0.000599.v1.2