A *Campylobacter* integrative and conjugative element with a CRISPR-Cas9 system targeting competing plasmids: a history of plasmid warfare?

Arnoud H.M. van Vliet 1,*, Oliver Charity 2, Mark Reuter 2

1. Department of Pathology and Infectious Diseases, School of Veterinary Medicine, University of Surrey, Guildford, United Kingdom.
2. Quadram Institute Bioscience, Microbes in the Food Chain programme, Norwich, United Kingdom.

* Corresponding author. Mailing address: School of Veterinary Medicine, University of Surrey, Daphne Jackson Road, Guildford GU2 7AL, United Kingdom. Phone +44-1483-684406, E-mail: a.vanvliet@surrey.ac.uk

Running title: CRISPR-Cas on *Campylobacter* mobile elements

Keywords: Campylobacter, mobile genetic elements, plasmids, CRISPR-Cas
ABSTRACT

Microbial genomes are highly adaptable, with mobile genetic elements (MGEs) such as integrative conjugative elements (ICE) mediating the dissemination of new genetic information throughout bacterial populations. This is countered by defence mechanism such as CRISPR-Cas systems, which limit invading MGEs by sequence-specific targeting. Here we report the distribution of the pVir, pTet and PCC42 plasmids and a new 70-129 kb ICE (CampyICE1) in the foodborne bacterial pathogens Campylobacter jejuni and Campylobacter coli. CampyICE1 contains a degenerated Type II-C CRISPR system consisting of a sole Cas9 protein, which is distinct from the previously described Cas9 proteins from C. jejuni and C. coli. CampyICE1 is conserved in structure and gene order, containing blocks of genes predicted to be involved in recombination, regulation, and conjugation. CampyICE1 was detected in 134/5,829 (2.3%) C. jejuni genomes and 92/1,347 (6.8%) C. coli genomes. Similar ICE were detected in a number of non-jejuni/coli Campylobacter species, although these lacked a CRISPR-Cas system. CampyICE1 carries 3 separate short CRISPR spacer arrays containing a combination of 108 unique spacers and 16 spacer variant families. A total of 69 spacers and 10 spacer variant families (63.7%) were predicted to target Campylobacter plasmids. The presence of a functional CampyICE1 Cas9 protein and matching anti-plasmid spacers was associated with the absence of the pVir, pTet and pCC42 plasmids (188/214 genomes, 87.9%), implicating that the CampyICE1-encoded CRISPR-Cas has contributed to the exclusion of competing plasmids. In conclusion, the characteristics of the CRISPR-Cas9 system on CampyICE1 suggests a history of plasmid warfare in Campylobacter.
van Vliet et al. CRISPR-Cas on Campylobacter mobile elements

IMPACT STATEMENT

Understanding pathogen evolution is paramount for enhancing food safety and limiting pathogenic disease in humans and animals. Campylobacter species comprise a group of human and animal pathogens with a remarkable success rate, being the most frequent cause of bacterial food-borne disease in high-income countries. A common theme among Campylobacter evolution is genomic plasticity, and a significant proportion of this plasticity is driven by horizontal gene transfer (HGT) that results in acquisition of complex traits in one evolutionary event. Understanding the mechanisms of transfer of MGEs and how MGEs such as integrative conjugative elements (ICE) exclude other MGEs is fundamental to understanding Campylobacter evolution. CRISPR-Cas9 proteins play a role in bacterial immune systems, mediating the defence against bacteriophage, plasmids, and integrative elements. The use of CRISPR-Cas by a mobile element to fight off competing elements, possibly to the advantage or detriment to their host, also increases our understanding of how important selfish genomic islands undergo co-evolution with bacterial pathogens, and generates insight into the complex warfare between MGEs.

DATA STATEMENT

All genome sequences used in this study are available on the National Center for Biotechnology Information (NCBI) Genome database or in the Campylobacter PubMLST website; the assembly accession numbers (NCBI Genome) or genome ID numbers (Campylobacter PubMLST) are listed in Table S1 (available in the online version of this article). Genome assemblies were quality checked based on N50, L50, genome size and number of contigs. CRISPR Spacer sequences and predicted targets, Cas9 alignments, presence of mobile elements and plasmids are all included in the Supplementary Information.
INTRODUCTION

The genus *Campylobacter* is a member of the Epsilonproteobacteria, and comprises gram-negative bacteria that are commonly found in the intestines of warm-blooded animals. The best studied members are *C. jejuni* and *C. coli*, which are closely related thermophilic species commonly found in birds and animals involved in agriculture, i.e. poultry, cattle and pigs, while they are also found in many wild birds [1, 2]. They jointly represent the most common bacterial human diarrhoeal pathogens in the developed world, with transmission often foodborne via undercooked meat and cross-contamination in kitchen environments [3, 4]. Other related *Campylobacter* species include the recently described *C. hepaticus* found in poultry [5], *C. upsaliensis* which is a zoonotic *Campylobacter* species from dogs and cats [6], and the *C. lari* group consisting of several species isolated from birds and animals connected to coastal environments [7].

Horizontal gene transfer (HGT) plays a major role in the evolution of microbial genomes [8]. Phages and plasmids are contributors to HGT-driven genomic plasticity, with transfer conducted by either transduction or conjugation, or alternatively by natural transformation [9]. One class of mobile genetic elements (MGE) are the integrative and conjugative elements (ICE), which are self-transferable elements that can mediate excision, form a circular intermediate and often encode the genes for the Type IV conjugative pili used to transfer to a new recipient host cell [10, 11]. ICEs often contain genes required for reversible site-specific recombination, conjugation and regulation, but also carry "cargo" genes that may confer antimicrobial resistance, virulence properties or metabolic capabilities to recipient cells [12], as well as addiction modules ensuring stable maintenance within the host cell [13].

Although acquisition of new genetic traits via HGT may have significant benefits for the recipient cell, the newly acquired sequences can also be detrimental to the host. Therefore cells have developed a diverse set of mechanisms to control entry, integration and expression of foreign DNA [14]. One such system is the Clustered Regularly Interspaced Short Palindromic Repeats...
van Vliet et al.  

(CRISPR) and proteins encoded by CRISPR-associated (Cas) genes, which encode the components of an RNA-guided, sequence-specific immune system against invading nucleic acids, often phages, plasmids and other transferable elements [15]. Many CRISPR-Cas systems have the Cas1 and Cas2 proteins mediating spacer acquisition [16] and other Cas proteins involved in expression, maturation/processing and targeting and interference of the foreign DNA or RNA sequences, commonly phages and plasmids [17]. The RNA-guided endonuclease of the Type II CRISPR-Cas system is the Cas9 (Csn1/Csx12) protein, which mediates processing of CRISPR RNAs and subsequent interference with the targets, in combination with a guide RNA called trans-activating CRISPR RNA (tracrRNA) [18].

Early studies using multilocus sequence typing (MLST) indicated a high level of genetic variability in Campylobacter species such as C. jejuni and C. coli [19], and subsequent comparative genomic analyses have shown that this level of genetic variability is achieved by differences in genetic content and high levels of allelic variability [20-22], likely supported by the natural competence of many Campylobacter species. Along with a variety of small plasmids (<10 kb), there are three major classes of 30-60 kb plasmids in C. jejuni and C. coli (pVir, pTet and pCC42) [23-25], although these are of variable size and gene content [26]. There are also four chromosomally located MGEs first identified in C. jejuni RM1221 [27], of which CJIE1 is a Mu-like prophage, CJIE2 and CJIE4 are related temperate prophages [28-31], and CJIE3 is a putative ICE which can contain the Campylobacter Type VI secretion system (T6SS) [32, 33].

In a previous study, we showed that 98% of C. jejuni genomes investigated contained a Type II-C CRISPR-Cas system consisting of cas9-cas1-cas2 genes and a relatively short spacer array (4.9 ± 2.7 spacers, N=1,942 genomes) [34]. In contrast, only 10% of C. coli genomes contained a copy of the C. jejuni CRISPR-Cas system, while genomes from non-agricultural (environmental) C. coli isolates contained a closely related, but separate Type II-C CRISPR-Cas system with the full complement of cas9-cas1-cas2 genes, or an orphan cas9 gene without cas1 or cas2 genes [34].
van Vliet et al.

have expanded this survey of CRISPR-Cas systems in *C. jejuni* and *C. coli*, and show that there is a third, clearly distinct CRISPR-Cas system in both *C. jejuni* and *C. coli*, which is located on a relatively conserved chromosomally located ICE (CampyICE1), and have investigated a possible role of this CRISPR-Cas system in contributing to plasmid competition in *Campylobacter*. 

100
MATERIALS AND METHODS

Identification of CRISPR-Cas systems

A collection of complete and draft genome sequences of *C. jejuni* (N=5,829) and *C. coli* (N=1,347) (Table S1) were obtained from the NCBI Genomes database (http://www.ncbi.nlm.nih.gov/genome/browse/) and the *Campylobacter* pubMLST website (http://pubmlst.org/campylobacter/) [35]. Genome assemblies were quality checked based on N<sub>50</sub>, L<sub>50</sub>, genome size and number of contigs, and have been used previously for studying gene distribution in *Campylobacter* [36, 37]. Genome sequences for non-jejuni/coli *Campylobacter* species such as *C. hepaticus*, *C. lari* group and *C. upsaliensis* were obtained from the NCBI genome database using ncbi-genome-download version 0.2.11 (https://github.com/kblin/ncbi-genome-download/). Genome sequences were annotated with Prokka version 1.13 [38], and the annotation searched for Cas9 orthologs using the *C. jejuni* Cj1523c (Cas9) amino acid sequence using BLASTP, while genome sequences were searched using TBLASTN to identify inactivated copies of *cas9* genes. CRISPR arrays were identified as described previously [34], using the CRISPRfinder software (http://crispr.u-psud.fr/Server/) [39] and the CRISPR Recognition Tool CRT [40], further supported by BLAST searches and manual curation. Conservation of sequences was represented using Weblogo [41].

Prediction of putative targets of CRISPR spacers

A total of 108 unique and 16 variant families of the CampyICE1 CRISPR spacer sequences were used as query on the CRISPRTarget website (http://brownlabtools.otago.ac.nz/CRISPRTarget/crispr_analysis.html) [42], and used to search the Genbank-Phage, Refseq-Plasmid, and Refseq-Viral databases. Only *Campylobacter* targets were included for further analysis. Hits with plasmids from the pVir, pTet and pCC42 families were
recorded. Individual genomes with plasmid-specific spacers and positive for either pVir, pTet or pCC42 were searched for the target sequences of that genome using BLAST.

Analysis of MGE and plasmid distribution

Genome sequences were screened using Abricate (https://github.com/tseemann/abricate) version 0.9.8, with each mobile element/plasmid subdivided into 600 nt fragments used as individual queries, and each 600 nt query sequence was only scored positive with a minimum coverage of 70% and minimum sequence identity of 80%. The CJIE1, CJIE2, CJIE3 and CJIE4 elements were obtained from *C. jejuni* reference strain RM1221 [27]. Nucleotide positions in the RM1221 genome (accession number CP000025) were 207,005-244,247 (CJIE1), 498,503-538,770 (CJIE2), 1,021,082-1,071,873 (CJIE3), and 1,335,703:1,371,932 (CJIE4). The T6SS genes were taken from *C. jejuni* 108 (accession number JX436460). For the CampyICE1 element, genome sequences were screened with the CampyICE1 element from *C. jejuni* strain CCN26 (accession number NZ_FBML01, nucleotide positions contig 11: 109,469-134,196 and reverse strand contig 17: 19,482-78,836), the Clade 1a *C. coli* strain RM1875 (accession number CP007183, nucleotide positions 1,235,330-1,320,414) and the *C. coli* Clade 2 strain C8C3 (accession number FBQX01, nucleotide positions 905,906-996,822). The pCC42 plasmid sequence was obtained from *C. coli* 537360 (accession number CP006703), whereas the pTet (accession number CP000549) and pVir (accession number CP000550) plasmid sequences were obtained from *C. jejuni* 81-176. Other plasmids used were pRM3194 (accession number CP014345), pHELV-1 (accession number CP020479) and pSCJK2-1 (accession number CP038863). Genomes were scored as positive for a mobile element or plasmid if >50% positive for 600 nt queries. Samples scoring between 30-50% were manually inspected for distribution of matches and given a final score. Clinker version 0.0.20 [43] was used to generate comparative gene maps of MGE and plasmids, using the default settings. Table S1 includes the presence/absence information of the pCC42, pTet and pVir plasmids, and the
van Vliet et al. CRISPR-Cas on Campylobacter mobile elements

151 CJIE1, CJIE2, CJIE3 and CJIE4 MGE.
152
153 **Phylogenetic trees**
154 Core genome MLST allelic profiles were generated for the 5,829 *C. jejuni* and 1,347 *C. coli*
155 genomes using a 678 gene set described previously [44]. Allele calling was performed using
156 chewBBACA version 2.6 [45] using the default settings. The phylogenetic trees were generated
157 using GrapeTree version 1.5.0 [46] with the RapidNJ implementation of Neighbor-Joining, and
158 annotated using the standard 7-gene MLST clonal complexes as determined using the MLST
160 Cas9 protein sequences were aligned with MEGA7 using the MUSCLE algorithm with the
161 default settings [47], and phylogenetic trees constructed using the MEGA7 Neighbor-joining
162 option, pairwise deletion and the Jones-Taylor-Thornton (JTT) model, with 500 bootstraps. Trees
163 were visualised using MEGA7 [47] and Figtree version 1.4.2
164 (http://tree.bio.ed.ac.uk/software/figtree/).
RESULTS

Campylobacter jejuni and C. coli contain a third type II-C Cas9-encoding gene

A collection of 5,829 C. jejuni and 1,347 C. coli genomes was searched for the presence of Cas9 orthologs using the C. jejuni NCTC11168 Cj1523c and C. coli 76639 BN865_15240c amino acid sequences, representative of the two type II-C Cas9 proteins previously detected in C. jejuni and C. coli [34]. In addition to the cas9 genes representative of the C. jejuni/agricultural C. coli and the non-agricultural C. coli genomes, a third cas9 gene was detected in 134 (2.3%) of C. jejuni genomes and 92 (6.8%) of C. coli genomes, predicted to encode a 965 amino acid protein, with 4 C. jejuni and 3 C. coli genomes containing an interrupted cas9 gene. This new cas9 gene did not have adjacent cas1 or cas2 genes. Alignment of the predicted new Cas9 proteins from C. jejuni and the C. coli clades with Cas9 proteins from members of the genera Campylobacter and Helicobacter showed that the new Cas9 proteins form a separate cluster (Fig. 1), suggesting these have originated from a more distant common ancestor. Alignment of the additional Cas9 proteins from C. jejuni and the different C. coli genetic clades showed that the three RuvC motifs, the HNH motif and R-rich region were all conserved (Fig. S1).

The novel CRISPR-Cas system is located on an integrative conjugative mobile element

We first looked for the genomic region containing the gene encoding the new Cas9 protein in completed C. jejuni and C. coli genomes. Only two complete C. coli genomes contained the additional cas9 gene; an inactivated copy of the cas9 gene was found on the C. coli RM1875 genome, while a complete copy of the gene was present in C. coli C8C3. The cas9 gene was flanked by a short CRISPR-repetitive region with five to six repeats, similar to the Campylobacter repeat lengths reported previously [34]. Investigation of the surrounding genes showed the downstream presence of a putative Type IV conjugative transfer system, with traG, traN, traL and...
traE genes, as well as a parM gene encoding the chromosome segregation protein ParM, while upstream of cas9, genes annotated as DNA primase, thymidine kinase, XerC tyrosine recombinase, and an integrase were detected, with the integrase flanked by a tRNA-Met gene as integration site (Fig. 2A), thus matching the common components of an ICE [10]. We have named the cas9-containing ICE element CampyICE1.

The C. coli RM1875 and C. coli C8C3 CampyICE1-containing genomic regions were used to search the 134 C. jejuni and 92 C. coli genomes containing the CampyICE1-cas9 gene for additional contigs matching the additional CampyICE1 sequences, and ordered these contigs accordingly. We were able to reconstruct the CampyICE1 genomic regions for 81 C. coli and 133 C. jejuni genomes, annotated these and each showed genetic synteny. The size of the ICE ranged from 70.0-129.3 kb (average 87.7 kb, n=214), and each CampyICE1 region started with a gene encoding a putative integrase (in Genbank often annotated as 30S ribosomal subunit protein), followed by a XerC tyrosine recombinase. There were six relatively conserved blocks of genes downstream, of which the third block ends with the cas9 gene, and the fourth and the fifth blocks contain genes encoding conjugation proteins (Fig. 2A). Finally, the mobile element also contained up to three putative CRISPR arrays, each with at most a few repeats. The conservation of the CampyICE1 gene synteny is shown in Figure 2B using three C. coli and three C. jejuni examples.

Searches of the Genbank sequence database for orthologs of CampyICE1 allowed the identification of a similar element in C. jejuni subsp. doylei, where the element is split into two parts, but lacks the gene block containing the cas9 gene. There were also regions with sequence and CampyICE1 gene structure similarity in C. upsaliensis plasmid pCU110 and C. iguaniorum plasmid pCIG1485E, although both lack the cas9 gene (Fig. S2). Subsequent searches in other Campylobacter spp genomes in the Genbank database allowed the identification of other plasmids and potential ICE elements with similar layouts from diverse Campylobacter species such as C. helveticus, C. insulaenigrae, C. lari and C. subantarcticus, but none of those contained the cas9
van Vliet et al.  

CRISPR-Cas on Campylobacter mobile elements

12

Distribution of CampyICE1 and other mobile elements and linkage to MLST-clonal complexes

To assess whether the distribution of CampyICE1 and other MGEs was linked to specific MLST-types or isolation source, we screened a collection of 5,829 C. jejuni and 1,347 C. coli genomes [36] using BLAST+ for the presence of CampyICE1, CJIE1, CJIE2, CJIE3, CJIE4, the plasmids pVir, pTet, pCC42, and the CJIE3-associated T6SS (Table 1). The CJIE1 element was the most common in C. jejuni, while CJIE4 was the least common of the MGEs from C. jejuni RM1221, although still more common than CampyICE1. In C. coli, the CJIE1, CJIE2 and CJIE3 elements were present in similar fractions, and again much more common than CJIE4 and CampyICE1 (Table 1). There was clear variation within the CJIE1-CJIE4 genetic elements, mostly in length but also in gene content (Fig. S3), with the CJIE3 element differing due to the presence or absence of the T6SS. With regard to the three plasmids, pVir was rare in both C. jejuni and C. coli, while pTet is present in approximately a quarter of the C. jejuni and C. coli genomes. The pCC42 plasmid was relatively rare in C. jejuni, but the most common plasmid in C. coli (Table 1). The plasmids showed more conservation of gene structure and content (Fig. S4), although there were combinations of plasmids and mobile elements that lead to megaplasmids with phage elements or the T6SS [48] which were not separately included in this analysis.

The C. jejuni genomes were clustered in a phylogenetic tree based on a 678 gene core genome (cg)MLST scheme [44], which grouped the genomes mostly according to clonal complexes of the seven-gene MLST for C. jejuni (Fig. 3) and the different C. coli clades (Fig. 4). With the exception of CJIE3 and the associated T6SS in C. jejuni, there was no clear association with specific MLST clonal complexes in either C. jejuni or C. coli. In C. jejuni, CJIE3 without the T6SS was restricted to clonal complexes ST-354 and ST-257, while the CJIE3 with T6SS was mostly found in clonal
complexes ST-464, ST-353, ST-573 and ST-403 (Fig. 3). There was no obvious link between isolation source and any of the MGEs, although it should be noted that the dataset used is biased towards human isolates. Similar to the mobile elements, the pVir, pTet and pCC42 plasmids did not show an association with either MLST clonal complex in *C. jejuni* or *C. coli* clade, or isolation source (Fig. 3, Fig. 4). The specific distribution per genome is provided in Table S1.

The majority of CampyICE1 CRISPR spacers are predicted to target *Campylobacter* plasmids

CRISPR arrays consist of the CRISPR repeats and the individual spacers, which are used to generate the cRNAs used for interference, and the tracrRNA [18]. The layout of the CampyICE1 CRISPR arrays is distinct from most other Type II CRISPR-Cas systems, where the CRISPR array and tracrRNA are often found directly next to the Cas genes. In contrast, the CampyICE1 system does not contain the ubiquitous *cas1* and *cas2* genes, and has a total of three CRISPR arrays spaced over the element (Fig. 2). We were able to identify spacers from 81 *C. coli* and 133 *C. jejuni* CampyICE1 elements. The first array contained 3.0 ± 1.5 spacers (N=197, range 1-6), and also contained a putative tracrRNA in the opposite transcriptional orientation (Fig. 5A), while the second CRISPR array contained 3.1 ± 1.7 spacers (N=208, range 1-10) and lacked a potential tracrRNA. The third CRISPR array is shorter and contained 1.0 ± 0.6 spacers (N=182, range 1-3).

The tracrRNA and repeat sequence are distinct from the previously described *C. jejuni* and *C. coli* CRISPR systems [34], with the changes in the repeat sequence being mirrored in the tracrRNA sequence, thus unlikely to affect functionality (Fig. 5A, 5B). The predicted Protospacer Adjacent Motif (PAM) was 5'-A(C/T)A(C/T) (Fig. 5A), which matches well with the 5'-ACAc PAM-motif described for the *C. jejuni* Cas9 protein [34, 49].

Comparison of the spacers from 214 CampyICE1 elements showed that these consisted of 108 unique spacer sequences, and an additional 40 spacers that were subdivided in 16 variant families, where 2-6 spacers had one or two nucleotide differences to each other and were predicted to match
van Vliet et al. CRISPR-Cas on Campylobacter mobile elements

The same targets (Table S2). The spacers were used to search phage and plasmid databases for putative targets, and a total of 62 unique spacers and eight variant families were predicted to target the Campylobacter plasmids pCC42 (31 unique spacers, two variants), pTet (16 unique spacers, six variants) and pVir (15 unique spacers, see Fig. 5C for an example). Furthermore there were spacers predicted to target the Campylobacter helveticus plasmid pHELV-1 (one unique spacer) and pSCJK2-1 from C. jejuni SCJK2 (six unique spacers, two variants). The pHELV-1 and pSCJK2-1 plasmids were not detected in the 5,829 C. jejuni and 1,347 C. coli genomes used in this study. The predicted targets on the plasmids pCC42, pTet and pVir were plotted against the plasmid maps (Fig. 5D), and showed that targets for pCC42 and pVir were found in multiple genes on these two plasmids, whereas the targets on pTet were limited to two genes, of which YSU_08860 is not universally present on plasmids of the pTet family (Fig. 5D).

Plasmid-mapping CampyICE1 CRISPR spacers are associated with an absence of the corresponding plasmids

To assess whether the CampyICE1 CRISPR-Cas9 system can function to exclude plasmid by using plasmid-mapping spacers, the 226 C. jejuni and C. coli CampyICE1-positive genome assemblies were searched for the presence of plasmid contigs and matches with spacer sequences (Table S3 and Table S4). As one possible escape for CRISPR-Cas9 surveillance could be sequence mutations/changes in the plasmids, we also checked whether the predicted plasmid-matching spacer would recognise any sequence in the genome assemblies (which include plasmid contigs). Of the C. coli assemblies, spacers were detected in 81/92 assemblies, and 56 had no plasmid/spacer matches. Of the 25 assemblies where there were plasmid/spacer matches, three had an inactivated CampyICE1 cas9 gene, and 11 did not have sequences matching the spacer(s) or only partial matches in their genome assembly, suggesting that mutations in the plasmid sequence have made the spacer unusable. This left 11 C. coli assemblies with a functional cas9 gene and spacer matching
the pCC42 plasmid. Similarly, for C. jejuni, spacers were detected in 133/134 genomes, and 109 had no plasmid/spacer matches. Of the 24 assemblies where there were plasmid/spacer matches, two had an inactivated CampyICE1 cas9 gene with frameshifts and stop codons, and seven did not have sequences matching the spacer(s) or only partial matches in their genome assembly. This left 15 C. jejuni assemblies with a functional cas9 gene and spacer matching the pCC42 (seven) and pTet (eight) plasmids. The matching of spacers, CampyICE1 Cas9 status and plasmid presence/absence is given in Figure 6, with more detailed data in Table S3 and Table S4.
DISCUSSION

In the last 25 years, CRISPR-Cas has gone from a relatively obscure repeat system in bacteria to a Nobel Prize winning phenomenon [50]. CRISPR-Cas systems are widespread in prokaryotic organisms, and while early reports predicted them to be a bacterial version of the adaptive immune system against phages, it is now clear that they target a wide variety of MGEs, and can also have a diverse set of alternative functions. Recent studies show that CRISPR-Cas systems are not just located on genomes, but can also be found on MGEs. Type IV and Type I CRISPR-Cas systems have been reported on enterobacterial plasmids [51, 52], and have been predicted to function in competition between plasmids [53]. *Vibrionaceae* species contain a variety of CRISPR-Cas systems associated with putative MGEs and genomic islands [54, 55], although data on their potential role in MGE competition are still lacking. To our knowledge, our study is the first to feature an incomplete Type II-C CRISPR-Cas9 system that is associated with an MGE, and where the majority of spacers matched competing plasmids. We have shown that CampyICE1 is highly conserved in both *C. jejuni* and *C. coli*, that it has up to three short spacer arrays on the ICE, and that the presence of a functional CampyICE1 CRISPR-Cas system and anti-plasmid spacers is associated with the absence of the three targeted plasmid types in *C. jejuni* and *C. coli*.

The Type II-C Cas9 protein encoded on CampyICE1 is closely related to the Cas9 proteins found in other *Campylobacter* and *Helicobacter* species, but clusters separately, suggesting it may have been co-opted from a genomic location in an ancestral *Campylobacteraceae* species. Interestingly, CampyICE1 lacks the *cas1* and *cas2* genes [56], a feature which has also been noted for the hypercompact Cas12j (CasΦ) system found on certain bacteriophages [57]. The Cas12j system much resembles the CampyICE1 Cas9 system described here, as they share a limited CRISPR spacer repertoire [58]. The lack of Cas1 and Cas2 components could mean that the CampyICE1 system is incapable of acquiring new spacers, which is supported by the relative lack of spacer diversity in the 214 genomes containing CampyICE1. However, we cannot exclude that
the CampyICE1 Cas9 may be able to co-opt the Cas1 and Cas2 proteins from the chromosomal version of the CRISPR-Cas system in *C. jejuni* and *C. coli*, although this is speculative. We have previously shown that ~98% of all *C. jejuni* genomes have a CRISPR-Cas system, while this is more limited in *C. coli*, where only ~10% of *C. coli* genomes have a CRISPR-Cas system [34]. Since the diversity in CRISPR spacers is also low in the chromosomal version of CRISPR-Cas of *C. jejuni* and *C. coli* and most spacers cannot (yet) be linked to mobile elements or phages [34, 59-61], it may represent additional or alternative functions for Cas9 in *C. jejuni*, such as control or activity in virulence [62-66]. However, this is not the case for the CampyICE1 CRISPR-Cas9 system, as a majority of spacers can be linked to the three main families of plasmids in *C. jejuni* and *C. coli*: pTet, pVir and pCC42.

In our collection of genomes, 41.7% of *C. coli* and 24.3% of *C. jejuni* genomes are predicted to contain one or more of these three plasmids, in different combinations. The three plasmids do not show signs of incompatibility, as 93 *C. jejuni* and 166 *C. coli* genomes had a combination of two plasmids or all three plasmids together. The role of these plasmids in *C. jejuni* and *C. coli* is still unclear, but they can carry virulence factors and contribute to the dissemination of antibiotic resistance. However, plasmids are not absolutely required for this, and plasmid-free isolates are also common. This is similar for the CJIE elements, where different combinations of the CJIE-elements and CampyICE1 were detected. The different roles of the CJIE-elements in *C. jejuni* and *C. coli* is still not clear, although the T6SS from CJIE3 has been linked with virulence [32, 33, 67, 68], and the DNases of the CJIE1, CJIE2, and CJIE4 elements are associated with reduced biofilm formation and reduced natural transformation [29, 30, 69].

The CRISPR-Cas9 system of the CampyICE1 element has some unique properties, as there are up to three short CRISPR arrays on the mobile element, with the essential tracrRNA not located with the *cas9* gene but located in another CRISPR spacer array on CampyICE1. Although the arrays detected were small, there were still 108 unique spacers and 16 spacer families, with a spacer family
The presence of CampyICE1, functional CRISPR-Cas9 and anti-plasmid spacers was associated with the absence of the competing plasmids targeted, suggesting that CampyICE1 has used its CRISPR-Cas9 system for "plasmid warfare" as a form of incompatibility. The match is not perfect, as there are several examples of a complete CampyICE1 CRISPR-Cas9 system with plasmid-targeting spacers, to which the spacers mapped were present with 100% sequence identity between spacer and predicted plasmid contigs (Table S3, Table S4, Fig. 6). This could potentially mean that the CRISPR-Cas system can prevent acquisition of new plasmids, but for unknown reasons is unable to remove plasmids already present, although this is highly speculative. It also suggests that the CampyICE1 plasmid restriction can be avoided by mutation of the target site disrupting the sequence matching, making the system less functional, especially in a bacterium known for its high levels of genetic variation. We also speculate that DNA modification and transcriptional variation/regulation may play a role in spacer-target discrepancies.

In summary, we have identified a new putative mobile element in C. jejuni and C. coli that contains a degenerated CRISPR-Cas9 system predicted to employ this CRISPR-Cas system to compete with other families of Campylobacter plasmids. We also show that mobile elements and plasmids are semi-randomly distributed within a large set of C. jejuni and C. coli genomes, and display significant levels of genetic variation within the elements. This fits well with the previously described genetic variability of the genus Campylobacter, and adds to the complexity of mobile
elements present within these successful foodborne human pathogens.

**FUNDING INFORMATION**

We gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC) via the BBSRC Institute Strategic Programme Grant BB/J004529/1 (Gut Health and Food Safety), and the BBSRC Doctoral Training Partnership to the Norwich Research Park (BB/M011216/1). The funder did not contribute to the study design, data collection, analysis or interpretation of the data.

**ACKNOWLEDGMENTS**

This publication made use of the PubMLST website (http://pubmlst.org/) developed by Keith Jolley and sited at the University of Oxford. The development of that website was funded by the Wellcome Trust.

**AUTHOR CONTRIBUTIONS**

A.H.M.v.V. conceived the study and study design, performed analysis and wrote the paper; O.C. and M.R. contributed to study design, performed analysis and writing of the paper.

**CONFLICTS OF INTEREST**

The authors declare that there are no conflicts of interest.
van Vliet et al. CRISPR-Cas on Campylobacter mobile elements

REFERENCES


van Vliet et al.  

CRISPR-Cas on *Campylobacter* mobile elements


22.

van Vliet et al. CRISPR-Cas on Campylobacter mobile elements


van Vliet et al.  

CRISPR-Cas on Campylobacter mobile elements


van Vliet et al. CRISPR-Cas on Campylobacter mobile elements


van Vliet et al.  CRISPR-Cas on Campylobacter mobile elements

doi: 10.1038/s41594-021-00632-3


van Vliet et al. CRISPR-Cas on *Campylobacter* mobile elements

van Vliet et al. CRISPR-Cas on *Campylobacter* mobile elements

**Table 1.** Prevalence of chromosomal and extrachromosomal mobile elements in 5,829 *C. jejuni* and 1,347 *C. coli* genome assemblies.

<table>
<thead>
<tr>
<th>Mobile element</th>
<th><em>C. jejuni</em> (N=5,829)</th>
<th><em>C. coli</em> (N=1,347)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chromosomal elements</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CampyICE1</td>
<td>134 (2.3%)</td>
<td>92 (6.8%)</td>
</tr>
<tr>
<td>CJIE1</td>
<td>2,136 (36.6%)</td>
<td>254 (18.9%)</td>
</tr>
<tr>
<td>CJIE2</td>
<td>1,291 (22.1%)</td>
<td>225 (16.7%)</td>
</tr>
<tr>
<td>CJIE3 with T6SS &lt;sup&gt;a&lt;/sup&gt;</td>
<td>1,137 (19.5%)</td>
<td>203 (15.1%)</td>
</tr>
<tr>
<td>CJIE3 without T6SS &lt;sup&gt;b&lt;/sup&gt;</td>
<td>537 (9.2%)</td>
<td>2 (0.1%)</td>
</tr>
<tr>
<td>CJIE4</td>
<td>798 (13.7%)</td>
<td>79 (5.9%)</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCC42</td>
<td>253 (4.3%)</td>
<td>383 (28.4%)</td>
</tr>
<tr>
<td>pTet</td>
<td>1,177 (20.2%)</td>
<td>337 (25.0%)</td>
</tr>
<tr>
<td>pVir</td>
<td>84 (1.4%)</td>
<td>15 (1.1%)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Combined presence of the CJIE3 element and the type VI secretion system

<sup>b</sup> Presence of the CJIE3 element, absence of the Type VI secretion system
van Vliet et al. CRISPR-Cas on Campylobacter mobile elements

LEGENDS TO FIGURES

**Figure 1.** Phylogenetic tree comparing the CampyICE Cas9 proteins with other *Campylobacter* and *Helicobacter* Cas9 proteins. The CampyICE1 Cas9 protein (blue) is distinct from the previously described Cas9 proteins of *C. jejuni* & *C. coli* (red), other *Campylobacter* spp. (green), and selected *Helicobacter* spp. (black). *C. jejuni* subsp. doylei is shown as *C. doylei*. The tree was drawn using the Neighbor-Joining method based on an alignment with the MEGA7 Muscle plugin. Bootstrap values are indicated at branches which scored >95%, based on 500 iterations using MEGA7, using the JTT matrix and pairwise deletion. The scale bar represents the number of amino acid substitutions per site. An alignment of a subset of Cas9 proteins with domain annotation is provided in Figure S1.

**Figure 2.** Structure and genetic conservation of CampyICE1 from *C. jejuni* and *C. coli*. (A) Schematic overview of the gene structure of CampyICE1 from *C. jejuni* and *C. coli*. The relative positions of the three CRISPR arrays and their transcriptional orientation is shown above the blocks of genes. In the CRISPR arrays, repeats are represented by arrowheads, spacers by diamonds, with the ends of the flanking genes shown. The gene category colors are shown to highlight the large proportion of hypothetical proteins with no known function. (B) Graphical comparison of CampyICE1 elements from *C. jejuni* and *C. coli* genomes, presented as output of a comparison of Prokka-generated annotations [38] using Clinker [43]. The colours of the arrows in the figure are used to identify homologous blocks of genes, and are not related to the colours used in part A of the figure.

**Figure 3.** Distribution of mobile elements and plasmids in 5,829 *C. jejuni* genome sequences. The phylogenetic tree was based on core genome MLST. Isolation source category and 7-gene MLST
van Vliet et al.  

CRISPR-Cas on Campylobacter mobile elements

562 information have been included for comparative purposes.

563

Figure 4. Distribution of mobile elements and plasmids in 1,347 C. coli genome sequences. The phylogenetic tree was based on core genome MLST. Isolation source category and 7-gene MLST information have been included for comparative purposes.

567

Figure 5. Characteristics of the CampyICE1 CRISPR spacers, protospacers and tracrRNA, and predicted plasmid targeting by the CampyICE1 CRISPR-Cas9 system. (A) A section of the CRISPR array is shown (center) with the corresponding protospacer (top) with 8 nt flanking sequences which contain the PAM motif at the 3’ end of the protospacer, represented using a sequence logo. The tracrRNA sequence and structure are included below. (B) Comparison of the CRISPR-repeats and predicted tracrRNA part of CampyICE1, C. jejuni and the three C. coli clades. The tracrRNA and CRISPR-repeat show matching changes as indicated by red underlined residues. Asterisks indicate conserved nucleotides, boxes indicate the complementary sequences in CRISPR repeat and tracrRNA. (C) Example of a CampyICE1 CRISPR spacer perfectly matching a segment of the C. jejuni 81-176 pVir plasmid. (D) Schematic representation of the pCC42, pTet and pVir family of plasmids (based on the C. coli 15-537360 pCC42 plasmid and the C. jejuni 81-176 pTet and pVir plasmids), with the locations of plasmid-targeting CampyICE1 spacers indicated. For pTet, the approximate location of target gene YSU_08860 (absent from the C. jejuni 81-176 plasmid) is indicated by the dashed, green colored arrowhead. More information on specific spacers and their targets is provided in Table S2.

583

Figure 6. Low prevalence of pVir, pTet and pCC42 plasmids in CampyICE1-positive C. jejuni and C. coli is associated with CRISPR-spacers targeting these plasmids. The C. jejuni and C. coli isolates have been combined in this graph; specific data per isolate and spacer are available in Table
S3, data for *C. jejuni* and *C. coli* separately are provided in Table S4.
van Vliet et al, Figure 2
van Vliet et al, Figure 3
**C. coli** (n=1,347)

- Source: ruminant; animal; environmental; human (clinical); poultry

**Clade**

- Clade 1a (ST-828)
- Clade 1b (ST-1150)
- Clade 1c (unintrogressed)
- Clade 2
- Clade 3

**IE1**

- IE1 without T6SS

**IE2**

- IE2 with T6SS

**IE3/T6SS**

- IE3 with T6SS; IE3 without T6SS; T6SS without IE3

**IE4**

- IE4 with T6SS

**ICE1**

- ICE1 with T6SS

**pCC42**

- pCC42 with T6SS

**pVir**

- pVir with T6SS

**pTet**

- pTet with T6SS

van Vliet et al, Figure 4
**A**

PROTOSPACE

5'-TATTTTCCaaaaggttttttttgttcasaasaagttGTAAAC-3'

REPEAT

ATGTTTATATGATAAAATaaaaaggttttttttgttcasaasaagttGTTTAGCCTATTTAATGTTTATATGATAAAATagcttataaa-3'

3'-UUGCUAUAUUGCGAGUCCACUCACGAAUUCGGAUAAAUACUCGGAUAAAUAAACAA-A5'

**B**

CRISPR-tracrRNA interaction

CampyICE1

GTTTTAGGCTATTATTAATGTTTT

TTATGATAAAT --

AAAAAAAAATTAGCTAAATAT

AAAGGTTTAGCAC

C.jejuni

GTTTTAGCTTTTATTTAATTGTTT

TTATGATAAAT --

AAATATAAAATGCTAAATAT

AAAGGTTTAGCAC

C.coli CL1

GTTTTAGCTTTTATTTAATTGTTT

TTATGATAAAT --

AAATATAAAATGCTAAATAT

AAAGGTTTAGCAC

C.coli CL2

GTTTTAGCTTTTATTTAATTGTTT

TTATGATAAAT --

AAATATAAAATGCTAAATAT

AAAGGTTTAGCAC

C.coli CL3

GTTTTAGCTTTTATTTAATTGTTT

TTATGATAAAT --

AAATATAAAATGCTAAATAT

AAAGGTTTAGCAC

**C**

Match to: pVir (Cj81-176) position 25057-25028, with: ice048 Spacer

5' -AUUGCAAAAGCUGAGAAAGAUAAACAAAAU- CRISPR spacer RNA

3' CRISPR spacer RNA

5' TTAAAATTATTGCAAAAGCTGAGAAAGATAAACAAAATATTG

PAM motif

**D**

pCC42

26.3 kb

33 ORFs

spacers

pTet

45.0 kb

54 ORFs

spacers

pVir

37.5 kb

54 ORFs

spacers

van Vliet et al, Figure 5
van Vliet et al, Figure 6

The diagram shows the number of CampyCE1-positive genomes across different conditions. Each condition is represented by a bar with different segments indicating the presence or absence of certain plasmids and the presence or absence of Cas9.

- **pTet**: 200 genomes with a red segment indicating the presence of Cas9.
- **pCC42**: 150 genomes with red and blue segments indicating the presence of Cas9 and the absence of a spacer.
- **pVir**: 100 genomes with red and yellow segments indicating the presence of Cas9 and the presence of a spacer.

Legend:
- Red: Cas9 present
- Blue: Cas9 absent
- Yellow: Spacing matches plasmid
- No spacer to plasmid
- Target missing or defective Cas9

The preprint (which was not certified by peer review) is made available under a CC-BY-NC 4.0 International license. It is made available under a perpetuity. It is made available under a preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in this version posted September 13, 2021. doi: bioRxiv preprint