


RESEARCH ARTICLE

Open Access



# *Campylobacter* bacteriophage DA10: an excised temperate bacteriophage targeted by CRISPR-cas

Steven Hooton<sup>1</sup>, Daniela D'Angelantonio<sup>2</sup>, Yang Hu<sup>1</sup>, Phillippa L. Connerton<sup>1</sup>, Giuseppe Aprea<sup>2</sup> and Ian F. Connerton<sup>1\*</sup> 

## Abstract

**Background:** Lytic bacteriophages that infect *Campylobacter* spp. have been utilized to develop therapeutic/decontamination techniques. However, the association of *Campylobacter* spp. and bacteriophages has been the focus of several strands of research aimed at understanding the complex relationships that have developed between predators and prey over evolutionary time. The activities of endogenous temperate bacteriophages have been used to evaluate genomic rearrangements and differential protein expression in host cells, and mechanisms of resistance to bacteriophage infection in campylobacters such as phase variation and CRISPR-mediated immunity.

**Results:** Temperate bacteriophage DA10 represents a novel excised and infective virus capable of replication in a restricted set of *C. jejuni* and *C. coli* hosts. Whole genome sequencing reveals that DA10 (35,379 bp) forms part of a novel group of temperate bacteriophages that have limited distribution among database host genome sequences. Analysis of potential host genomes reveals a robust response against DA10 and DA10-like bacteriophages is driven by CRISPR-mediated immunity with 75% of DA10 ORFs represented as ~ 30 bp spacer sequences in numerous *Campylobacter* Type II-C CRISPR arrays. Several DA10-like homologues have been identified in a small sub-set of *C. jejuni* and *C. coli* genome sequences (ranging from near complete integrated prophage sequences to fragments recognisable in the sequence read archive).

**Conclusions:** A complete intact DA10-like prophage in *C. jejuni* CJ677CC520 provides evidence that the associations between host and DA10-like bacteriophages are long-standing in evolutionary timescales. Extensive nucleotide substitution and loss can be observed in the integrated DA10-like prophage of CJ677CC520 compared to other relatives as observed through pairwise genome comparisons. Examining factors that have limited the population expansion of the prophage, while others appear to have thrived and prospered (Mu-like, CJIE-like, and lytic *Campylobacter* bacteriophages) will assist in identifying the underlying evolutionary processes in the natural environment.

**Keywords:** *Campylobacter*s, *Campylobacter* bacteriophages, Prophage, Lytic phage, CRISPR-mediated immunity, Evolution

\* Correspondence: [ian.connerton@nottingham.ac.uk](mailto:ian.connerton@nottingham.ac.uk)

<sup>1</sup>School of Biosciences, Division of Microbiology Brewing and Biotechnology, University of Nottingham, Sutton Bonington Campus, Leicestershire LE12 5RD, UK

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

*Campylobacter* spp. are recognized as being a major cause of gastrointestinal disease on a global scale [1]. Genomic data representative of ecologically-diverse and widely-distributed species such as *C. jejuni*, *C. coli*, and to a lesser extent other *Campylobacter* spp. has advanced our understanding of the genetic factors driving the worldwide success of these pathogens [2–4]. As with many bacteria *Campylobacter* spp. are known to harbour prophage/prophage remnants, for example the *C. jejuni* integrative elements – CJIE, and Mu-like bacteriophage (phage) sequences [5–7]. It is widely accepted that lysogenic conversion of a host bacterium following integration of a temperate phage genome can alter the virulence profile of the host [8]. In campylobacters the situation is not so clear. Genomic rearrangements mediated by the activity of Mu-like prophages (CJIE1 also known as CMLP-1/CampMu-like phage 1) are recognized as major factors governing host resistance to phage predation [7]. Furthermore, some CJIE1-like elements are associated with dampening the host's ability to acquire exogenous DNA via natural transformation through the activity of an extracellular secreted DNase enzyme [9–11]. A study involving *C. jejuni* encoding homologues of CJIE1 elements indicated a potential role in virulence due to enhanced adherence and invasion of INT407 cells compared to *C. jejuni* lacking these integrated elements [12]. However, a comparable study found no correlation between the presence/absence of CJIE1-like elements and the ability of *C. jejuni* to adhere to and invade HT29 colon cancer cells [13]. A further set of integrated genetic elements (CJIE2, CJIE3, and CJIE4) can also be found embedded in many *C. jejuni* genomes. Whilst CJIE2 and CJIE4 possess hallmark features of prophage/prophage remnants, CJIE3 most likely arose following integration of a plasmid-like mobile element into the host chromosome [14]. Interactions between *Campylobacter*-associated mobile genetic elements can also be observed in the large multidrug resistance plasmid of *C. jejuni* T1–21 [15]. Intriguingly, the 82.7 Kb mega-plasmid contains a 45 Kb Mu-like prophage region that may be a remnant of a transposition event outside of the established lytic and lysogenic lifecycles. The general consensus is that a non-inducible or as yet unidentified mechanism is employed for the activation of CJIE1 and CJIE-like elements [16], although recourse to mitomycin-c has been reported to induce CJIE1 [12, 14]. It has also been shown that CJIE1 and other relatives (CJIE4) have the potential to influence protein expression of their bacterial lysogens [16].

In comparison to other enteric pathogens such as *Escherichia coli* and *Salmonella* spp. [17] there is a distinct lack of temperate phages known to infect, undergo chromosomal integration, and subsequently excise to

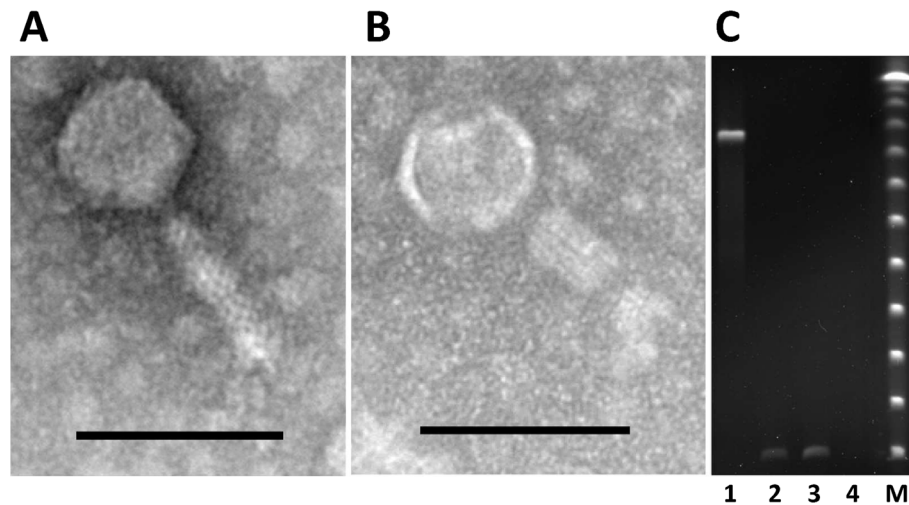
produce nascent infectious virions in *C. jejuni* and related species. However, there are historical reports of the spontaneous induction of temperate phages harboured by *C. fetus* (formerly described as *Vibrio fetus*) [18, 19]. Transmission electron micrographs obtained of phages targeting *C. fetus* indicate a *Siphoviridae* morphology as indicated by the presence of long, non-contractile tail machinery [19, 20]. Multiple mechanisms exist for bacteria to resist phage infection, such as phase variation of cognate receptors, switching from motile to sessile modes of growth, and CRISPR systems. It has been established that *Campylobacter* spp. employ several mechanisms to evade phage, including genetic rearrangement [7], phase variation [21], the use of an alternative flagellin [22], and the acquisition of CRISPR spacers to abrogate phage infection [23, 24]. Alternatively, phages have the capacity to mutate and overcome modifications such as phase-varied cognate receptors, leading to a classical co-evolutionary arms race. Class II/III *Campylobacter* phages (*Eucampyvirinae* Firehammervirus and Fletchervirus) are capable of interfering with host CRISPR systems via interactions between phage-encoded Cas4 protein and the host Type II-C system [23, 25].

*Campylobacter* phage DA10 was isolated from poultry-associated samples during a study performed in 2017 [26]. Analysis of the lytic properties of DA10 indicate a limited host-range against a diverse panel of *Campylobacter* spp. (7/118). Morphologically DA10 has been reported to display classic *Myoviridae* features including a icosahedral head, neck-like structure and a contractile tail [26]. However, here we report DA10 has a novel small genome compared to virulent *Campylobacter* bacteriophages that is related to prophage in the growing number of *Campylobacter* genome sequences available in the database. Despite a distinct lack of inducible prophages in *Campylobacter* genomes, DA10 represents a rare excised and infective phage arising from biological interactions between campylobacters and associated phages. However, extensive representation of DA10 sequences in the CRISPR arrays of diverse campylobacters likely accounts for the restricted host and the progression towards extinction of the infectious virus. Moreover, the degradation of prophage sequences in *Campylobacter* genomes exemplifies the evolutionary pressures exerted by the *Campylobacter* Type II-C CRISPR-Cas system.

## Results

### Characterisation of bacteriophage DA10

Transmission electron microscopy confirmed the morphology of phage DA10 to feature a contractile tail with an icosahedral head (Fig. 1a and b) indicating phage DA10 is a myovirus. The phage tails had a mean length



**Fig. 1** TEMs and PFGE of *Campylobacter* bacteriophage DA10. Panel **a** shows a transmission electron micrograph image of phage DA10. The bar represents 100 nm. Panel **b** shows a transmission electron micrograph image of phage DA10 with a contracted tail. The bar represents 100 nm. Panel **c** shows pulse fields gel electrophoresis of bacteriophage genomic DNAs: lane 1, Class I *Campylobacter* phage 12 DNA; lane 2, Phage DA10 DNA; lane 3, Phage DA10 DNA digested with RNase A; lane 4, Phage DNA DA10 digested with DNase 1; lane M, concatenated lambda phage DNA (48.5 kb) size marker

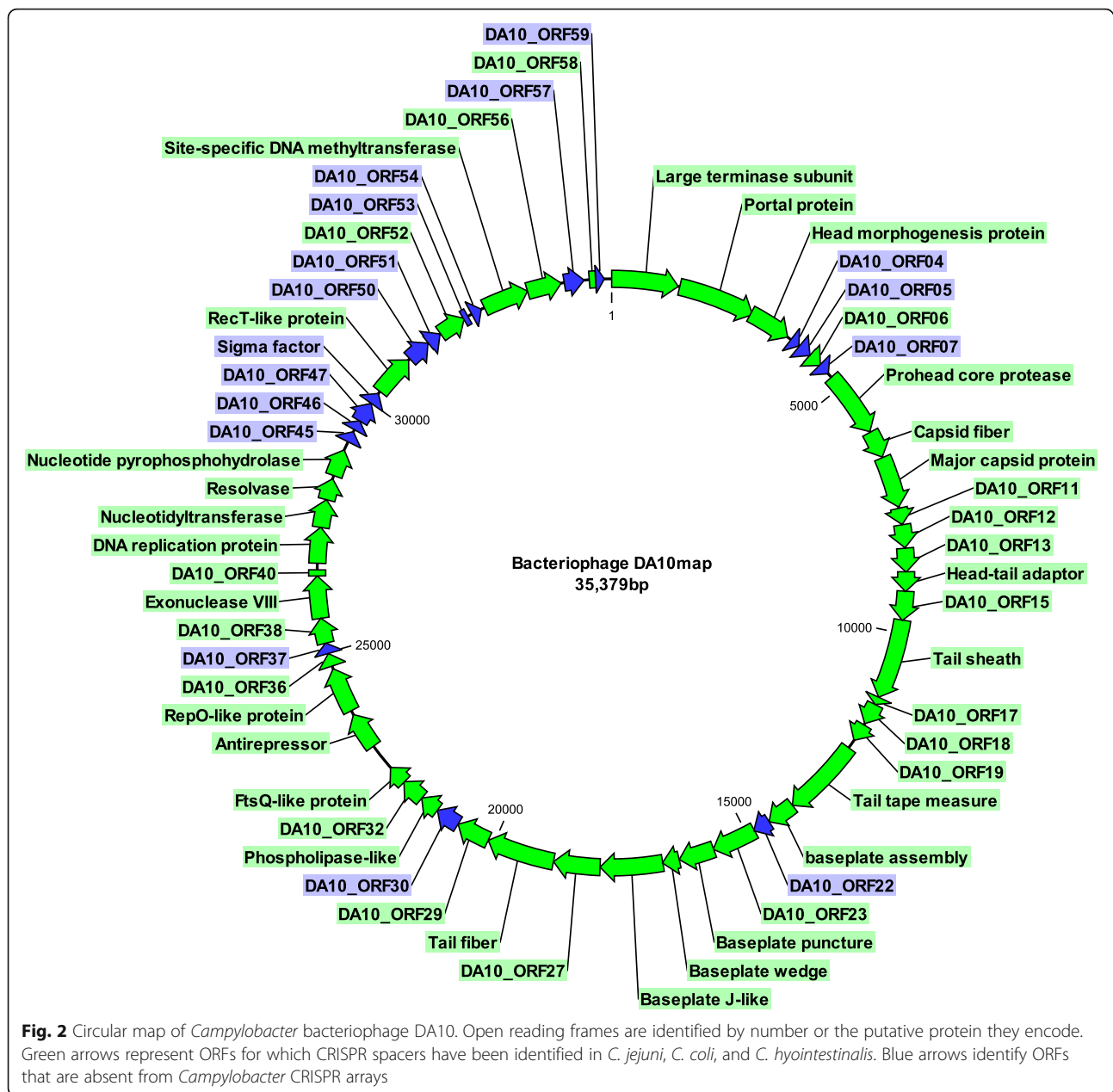
of  $93.5 \pm 3.7$  nm and width of  $21.6 \pm 3.0$  nm. The phage tails were shorter than previous observations of virulent class II/III *Campylobacter* phages (115–148 nm) [27, 28]. The mean head diameter of  $67 \pm 3.7$  nm was also reduced compared to the class II/III *Campylobacter* phages (92–96 nm), suggesting DA10 represents a new class of *Campylobacter* phage. Figure 1c shows PFGE of the D10 genome, which indicates the genome size to be  $\leq 48.5$  kb that represents the lowest molecular weight marker present. This is in contrast to the class I *Campylobacter* bacteriophage 12 DNA on the same gel that is estimated to be 420 kb. The DA10 PFGE band is resistant to RNase digestion but sensitive to DNase1, indicating the phage has a DNA genome (Fig. 1c). Bacteriophage DA10 therefore represents the smallest *Campylobacter* phage DNA genome recorded, which prompted investigation of the genome sequence.

#### Genome features

To further our understanding of phage DA10 whole genome sequencing was performed. Assembly of DA10 sequence reads produced a complete circularly permuted genome of 35,379 bp (Fig. 2) with 600-fold coverage. Many temperate bacteriophages have discrete ends, which prompted us to examine the accumulation of sequence reads that might implicate the presence of terminal repeats. No regions of read accumulation were observed, although these may not be evident due to the tagmentation protocol adopted for DNA sequence library preparation. The genome sequence was therefore deposited in the nucleotide sequence database as a putative circularly permuted genome (Acc. No. MN530981).

Analysis of the nucleotide composition of DA10 indicates a GC content of 27.1% which is slightly below that of *C. jejuni* (~30% GC). BLASTn analysis of the DA10 nucleotide sequence identified a near intact relative of DA10 present as an integrated prophage in *C. jejuni* CJ677CC520 [29]. CJ677CC520 (Acc. No. CP010501) was isolated from human faeces/blood and is a member of sequence type 677 (ST677). The predicted size of the CJ677CC520 prophage is 36,401 bp (base position 446, 910–483,310) and is integrated downstream of a tRNA 2-selenouridine gene. Approximately 20.2 Kb of the phage DA10 genome can be mapped to the prophage region in CJ677CC520 with nucleotide identities ranging from 72 to 96%. Several contigs (5–37 Kb) recently uploaded as part of a study into *Campylobacter* antimicrobial resistance show that DA10-like prophages are present in a small number of *C. jejuni* and *C. coli* chromosomes (using DA10-like anti-repressor protein as an identifier). Of these, six complete circularly permuted DA10-like genomes have been identified (Table 1 shows a pairwise comparison of the nucleotide identities and Table 2 post Mummer alignment of the coding regions). Hallmark features of DA10-like phages include genome sizes ranging from ~33 to 38 Kb in length, a conserved gene set encoding major structural proteins, DNA replication/repair/recombination functions, and DNA methylation. Similar %GC contents are also observed for DA10-like phages with ranges between 26.2% (as observed for the integrated prophage in CJ677CC520) and 27.8% (putative prophage in *C. coli* NC\_C4236).

Although no genes encoding an integrase could be identified in DA10 or DA10-like DNA, a putative



Holliday junction resolvase (DA10\_ORF43) was identified at 20% amino acid identity to bacteriophage T7 endonuclease 1 (HHpred  $p > 99\%$ , expect =  $1.7e-12$ ). Small fragments of the DA10 genome are found in *Campylobacter* spp. chromosomes (*C. jejuni*/*C. coli*), Class II/III *Campylobacter* phage genomes, and CJIE prophage sequences. DA10 and DA10-like phages therefore constitute a unique mosaic genome structure displaying weak, but recognizable associations with *Campylobacter* prophages, prophage remnants, and lytic *Campylobacter* phages. Complete or extended DA10-like prophage sequences are rare when surveying whole genome studies lodged in the NCBI Sequence Read Archive (10 of 29,

736 genomes for *C. jejuni* and 12 of 11,916 genomes for *C. coli*), suggesting that DA10 has a limited ecological distribution in *Campylobacter* populations. The exception to this is a 1.5 Kb region of the DA10 genome (bases 21,460-22,990) that is conserved across multiple *Campylobacter* genomes. In DA10 this region contains genes encoding hypothetical proteins and a FtsQ related protein (HHpred  $p > 96\%$ , expect = 0.015). Internal to this is a 470 bp region conserved in CJIE4 prophage sequences (ANI of 93%): CJIE4-2 (KF751794), CJIE4-3 (KF751795), CJIE4-4 (KF751797) (436/470 bases), CJIE4-1 (KF751793) (391/425 bases) and CJIE4-5 (KF751797A) (394/425 bases). Frequently two copies are

**Table 1** Pairwise comparison of nucleotide identities for DA10 and DA10-like sequences

	DA10	PNUSA C003578	CcNC_C4236	CcNC_C4247	CjFSIS 11,811,828	PNUSA C003324	CJ677 CC520
<b>DA10</b>	*	24,042/ 35,379	23,679/ 35,379	24,064/ 35,379	23,170/ 35,379	26,403/ 35,379	18,052/ 35,379
<b>PNUSA C003578</b>	23,605/ 37,072	*	20,967/ 37,072	20,967/ 37,072	21,273/ 37,072	29,909/ 37,072	19,925/ 37,072
<b>CcNC_C4236</b>	23,520/ 33,831	22,435/ 33,831	*	29,470/ 33,831	22,986/ 33,831	23,927/ 33,831	17,738/ 33,831
<b>CcNC_C4247</b>	20,413/ 33,823	20,262/ 33,823	27,669/ 33,823	*	17,782/ 33,823	21,231/ 33,823	14,885/ 33,823
<b>CjFSIS 11,811,828</b>	23,899/ 33,563	18,783/ 33,563	19,295/ 33,563	19,097/ 33,563	*	22,012/ 33,563	18,704/ 33,563
<b>PNUSA C003324</b>	26,142/ 37,779	27,760/ 37,779	21,504/ 37,779	21,504/ 37,779	22,140/ 37,779	*	19,988/ 37,779
<b>CJ677 CC520</b>	16,146/ 37,405	16,785/ 37,405	14,016/ 37,405	14,217/ 37,405	15,403/ 37,405	16,032/ 37,405	*

present at different genomic locations in *Campylobacter* genome sequences. For example, in *C. jejuni* FDAAR-GOS\_421 (CP023866) this sequence is encoded at ~ 0.87 Mb and ~ 1.69 Mb in the host chromosome.

Examination of the DA10 genome identified a total of 59 open reading frames (ORFs) encoding 44 hypothetical proteins, and 15 ORFs that could be assigned putative functions based on BLASTp analysis of translated amino acid sequences (Fig. 2). All 59 DA10 ORFs are found on the same strand with no putative genes being identified on the opposite strand. For genes that could be assigned putative functions several were associated with the synthesis of phage structural proteins (tail-associated proteins, baseplate assembly, major capsid/head morphogenesis), nucleotide metabolism (sugar phosphate nucleotidyltransferase), and DNA replication/regulation (site-specific methyltransferase, RecT-like, RepO-like, and phage antirepressor). Many of the proteins encoded by DA10 have sequence similarities to a small subset of *C. jejuni* and *C. coli* genome sequences further hinting towards a limited distribution of this phage throughout *Campylobacter* populations.

#### Orthologous relationships

Opening the DA10 genome sequence at the recognisable large terminase subunit gene encoded by DA10\_ORF1 (located at nucleotide positions 1–1284 in Fig. 2). BLASTn analysis indicates the gene encoded by DA10 displays 86% ANI (1087/1264) when aligned against its homologue in CJ677CC520. Several *C. jejuni* and *C. coli* genomes are observed to encode similar proteins ranging from 83 to 99% identity at the amino acid level, with some examples showing two separate proteins separated by an internal stop codon (positions 1–248 and 248–427). Similar large terminase subunits were identified in *Helicobacter* sp. (162/432 amino acid identity) and

*Helicobacter aurati* (167/427 amino acid identity). *Helicobacter*-encoded large terminase subunits contain conserved protein domains associated with the packaging of phage DNA during the virion assembly process. While these domains were not present in the DA10 large terminase protein sequence, numerous alignments to residues associated with Terminase\_3 (pf04466) and Terminase\_6c RNase H-like (pf17289) domains were identified.

DA10\_ORF10 encodes the putative major capsid protein in the DA10 genome (nucleotides 6575–7570). BLASTn analysis of the gene sequence indicates 673/994 (71% ANI) with its homologue encoded in CJ677CC520. Several protein homologues are observed when the translated 331 amino acid sequence is aligned with conserved hypothetical proteins encoded in *C. jejuni* and *C. coli* genomes. Limited homology of the DA10 amino acid sequence is evident when aligned with major capsid proteins of various phages. For example, *Pseudomonas* phages NP3 (AMQ76141), PA5 (APD20704), and KPP22 (BAU20690) all display identical alignments over 28% of the DA10 sequence. Interestingly, the majority of phage-associated BLASTp hits with DA10 major capsid protein are found to be with highly divergent homologues found in *Pseudomonas* phages. No obvious homology at the amino acid level was observed with the major capsid proteins present in Group II/III *Campylobacter* phages or CJIE-like elements.

Recognisable ORFs encoding a putative phage tail fiber protein (DA10\_ORF28; bases 18,813 to 20,126), tail sheath protein (DA10\_ORF16; bases 9749 to 11,296) phage-associated DUF2612 domain-containing protein (DA10\_ORF27; bases 17,914 to 18,810), and a baseplate J-like protein (DA10\_ORF26; bases 16,709 to 17,914) are present in the DA10 genome. A small-subset of *C. jejuni* and *C. coli* possess near-identical homologues of the 272

amino acid phage-associated protein encoded by DA10\_ORF28, and weaker alignments are observed with proteins encoded in *C. lari* and *C. ornithocola*. Amino acids 89–176 display similarity (~36% identity) to a hypothetical protein encoded by several Group III *Campylobacter* phages, and Group II phage CP21. BLASTp hits for short regions of the putative tail fiber sequences from phages infecting *Pectobacterium* spp. and *Erwinia amylovora*. DA10\_ORF27 encodes a 298 amino acid protein containing a DUF2612 domain—that is conserved in several *C. jejuni* and *C. coli* genomes, as well as multiple highly diverged homologues in numerous bacterial species. DUF2612-domains (pfam11041) are widely distributed in phages/prophages associated with *Proteobacteria*. DA10\_ORF16 encodes a protein of 515 amino acids that is predicted to share structural similarity to a R-type phage tail-like bacteriocin contractile sheath (HHpred  $p > 99\%$ , expect = 7.6e-20) that forms a cell penetrative structure with a baseplate, receptor-binding tail fibers and an inner needle-like tube analogous to bacteriophage infection machinery. DA10\_ORF26 encodes a 401 amino acid baseplate J-like protein including the conserved baseplate J family domain (pfam04865) indicating that this ORF encodes an important structural component of the DA10 phage particle. Alignment of the protein sequence with the baseplate J-like protein encoded by CJ677CC520 indicates 251/396 (63%) identity. Sequential nucleotide database searches of the tail fiber region revealed short dispersed regions of similarity, including a 530 bp region (bases 18,425–18,955) (375/530–70% average nucleotide identities - ANI); 131 bp (95/131–73%) and 72 bp regions (54/72–75%) that align to the integrated prophage encoded in CJ677CC520. A further 92 bp region is highly similar to Class II *Campylobacter* phage CP21 (74/92–80%). Adjacent to the baseplate J-like encoding gene are predicted ORFs of related function, notably encoding baseplate wedge (DA10\_ORF25), puncture (DA10\_ORF24) and assembly proteins (DA10\_ORF21).

DA10\_ORF42 encodes a 543 bp sugar phosphate nucleotidyltransferase gene (bases 27,415 to 27,957). Conserved homologues (76% ANI) of DA10\_ORF42 are found to be encoded in Class III *Campylobacter* phage genomes (CP81, CP39, CP8, CPX, CP81, PC5, CP30A, PC14, NCTC 12673, and vB\_CjeM\_Los1). *C. coli* OR12 and an aerotolerant derivative (CP019977) contain two homologues of the 543 bp sugar phosphate nucleotidyltransferase gene encoded by DA10 [30]. Both genes in *C. coli* OR12 display 79% ANI (399/508) over a 508 bp region of the DA10 sequence. The first copy is located at ~1.76 Mb and the second copy at ~1.9 Mb on the *C. coli* OR12 genome. Duplication of this gene in *C. coli* OR12 is interesting as while several other *C. coli* genomes contain an identical gene, OR12 is the only

example observed to possess two copies. Examination of the regions within *C. coli* OR12 highlights that both copies are encoded within intact prophage sequences present in the host genome. However, these integrated prophages in *C. coli* OR12 display very little homology towards DA10-like phage sequences. *C. coli* OR12 possesses another duplication of a gene encoded by DA10 (DA10\_ORF43). This ORF is predicted to encode a resolvase similar to enterobacteria phage T7 that is not present in other *C. coli* sequences. The sequence of DA10 aligns with 87% ANI (214/246) against two copies of this gene in the *C. coli* OR12 genome at ~1.25 and ~1.42 Mb [30].

DA10\_ORF44 (bases 28,439–28,969) was identified as a nucleotide pyrophosphohydrolase (HHpred  $p > 99\%$ , expect = 6.1e-10), which displays sequence identity to Class III *Campylobacter* phage DNA sequences: PC5 (KX229736), CP81 (NC\_042112), CP8 (KF148616), CPX (JN132397), CP30A (JX569801), NCTC 12673 (GU296433), PC14 (KX23633), and vB\_CjeM\_Los1 (NC\_041896). Multiple alignments of the amino acid sequence to a conserved hypothetical protein present in *C. jejuni*/*C. coli* genomes indicate that this phage-associated protein may have prophage associations. The ORF encoded in DA10 encodes 5 bp repeats of 5'-AAAAA-3' at each end suggesting a possible recombination event has allowed for horizontal transfer of this gene.

A putative site-specific DNA methyltransferase is encoded in DA10\_ORF55 (bases 32,866–33,750). Translation of the gene sequence produces a 294 amino acid protein homologous to site-specific DNA methyltransferases encoded in three *C. jejuni* and four *C. coli* genomes. No other matches with *Campylobacter* spp. or *Campylobacter* phage-associated proteins were identified. Alignments (43% identity) with prophage-encoded site-specific (N6\_N4) methyltransferases from *Streptococcus mutans* (WP\_002287175), *S. mutans* NLML9 (EMC100053), and *S. mutans* 24 (EMC47101) signify a possible function for the DA10-encoded protein. However, no conserved protein domains associated with N6\_N4 DNA methylation or similar were identified in the DA10 site-specific DNA methyltransferase. An interesting match is also observed with a putative DNA methylase in *Lactobacillus* phage LBR48 (ADF83450) with 127/282 (45% identity).

#### DNA regulation

DA10\_ORF41 (bases 26,726–27,418) encodes a 230 amino acid AAA-family ATPase protein with homologues observed in *C. jejuni*, *C. coli*, *C. fetus*, and *C. iguaniorum*. BLASTn analysis of the gene sequence indicates that it has very limited distribution in sequences deposited in the database. However, when aligned to its

counterpart encoded in CJ677CC520, a high degree of conservation 94% ANI (652/693) is observed. Translated protein alignments with *C. jejuni* and *C. coli* proteins show a higher degree of conservation (72–96% identities) than those observed for *C. fetus* and *C. ignavium* (41%). Analysis of conserved protein domains (P-Loop NTPase superfamily [cl21455] AAA-family of ATPases) indicate a potential functional role in the replication of DA10 DNA. A single match of limited homology (22% amino acid identity) is evident with a DnaC-like protein encoded in *Streptococcus* phage 73 (KT717083). No alignments with any other phage-encoded proteins are observed at the nucleotide or protein level.

A putative exonuclease VIII-like protein is encoded by DA10\_ORF39 (816 bp, bases 25,668–26,483). At the nucleotide level CJ677CC520 is most similar with 93% ANI (760/816) aligning with DA10\_ORF39. The 271 amino acid exonuclease VIII-like protein has several near-identical homologues (93–99%) encoded by *C. jejuni* and *C. coli* present in the database. A conserved protein domain (PRK09709) indicates the exonuclease VIII-like function of this protein, and numerous weak alignments are observed with similarly annotated proteins in diverse genomes such as *Synechococcus* phage S-CBS1 (ADP06638.1) and Cyanophage KBS-S-2A (AGH57650.1).

Analysis of the putative antirepressor gene encoded by DA10\_ORF34 (bases 22,997–23,713) indicates potential roles in DNA binding and regulation. DA10 antirepressor is highly homologous to antirepressor/BRO family proteins found in *C. jejuni* and *C. coli* (77–88% amino acid identity). Although the DA10 antirepressor lacks some of the canonical conserved KILAC-domain residues (pfam03374) that are often associated with antirepressor proteins, several amino acids are conserved. Immediately downstream of DA10 antirepressor is DA10\_ORF35 (bases 23,798–24,673) that encodes a putative 291 amino acid RepO-like protein. Two single examples of homologues of DA10 RepO-like protein are observed in *C. jejuni* (EAL7532945) and *C. coli* (EAH9758457). Analysis of the amino acid sequence of DA10 RepO-like protein indicates that it is modular in nature with distinctive N- and C-terminal domains. Linking these two protein domains is a sequence motif of 16 amino acids. Residues 165–291 show conservation with several conserved hypothetical proteins in *C. jejuni* and *C. coli*. However, the N-terminal residues (1–164) are distinctly different from that observed in phage DA10. A similar situation is evident when aligning the RepO-like protein of DA10 against the CJ677CC520 prophage. No N-terminal similarity can be observed between the two RepO-like proteins, whereas the C-terminus exhibits 99% identity between amino acid residues 160–291. At the nucleotide level it is apparent that the linker region of DA10\_

ORF35 consists of three short polyA tracts (AAAAAAA-TATAAAAAAAGAAAAAAA) that may be capable of promoting recombination thus leading to the alternative N-terminal protein sequences observed.

#### **Campylobacter CRISPR spacers targeting DA10**

Small fragments of the DA10 genome can be identified as spacer sequences in *Campylobacter* CRISPR arrays (Table 3) that target 43 of the 59 ORFs of DA10 (Fig. 2). The DNA sequences are present as 29–38 bp CRISPR spacers in the Type II-C systems encoded by campylobacters. A spacer targeting DA10\_ORF13 appears more frequently than any other DA10-associated CRISPR spacers identified during this analysis (5'-AACCCC GCTAACTGCATAAAAAATATCCATTTTC-3'). Importantly it appears that a considerable number of DA10 genes can be identified as ~30 bp spacers in *Campylobacter* CRISPR arrays, suggesting a robust and widespread defence system is being deployed in campylobacters to mitigate DA10-like phage infection. In some cases multiple spacer sequences targeting DA10 are observed within the same CRISPR array. For instance, *C. jejuni* R123 (HQ378289) has acquired the widely distributed spacer sequence described above as well as spacers targeting the tail tape measure and baseplate J-like genes. Some spacers are observed to have a single base mismatch due to the presence of an extra nucleotide in the DA10 gene relative to the CRISPR sequence (e.g. cytosine insertion in the *C. jejuni* HCC91 spacer targeting RecT-like recombination/repair gene). Many other 30–35 bp alignments originating from the DA10 genome are present in a wide range of *Campylobacter* genomes, presumably these short alignments map to unannotated CRISPR sequences. It is interesting to note that the genome sequence of CJ677CC520 containing the DA10-like prophage element has an atypical Type II-C CRISPR system configuration with Cas2 and Cas9 present, and Cas1 notably absent (nucleotide positions 1,419,058–1,422,456). The DNA endonuclease function associated with cas1 is critical to the CRISPR adaptation process. Spacer acquisition and adaptation in CRISPR systems arises from Cas1-Cas2 complexes that generate protospacers from foreign DNA [31]. It is possible that a malfunctioning CRISPR system has allowed DA10-like phage infection of CJ677CC520 to establish a lysogenic state. Several examples also exist of CRISPR spacers targeting the DA10 antirepressor gene sequences.

#### **Non-CRISPR targeted DA10 ORFs**

Sixteen ORFs in the DA10 genome are not represented as spacers in *Campylobacter* CRISPR arrays. Analysis of these ORFs suggests that the majority of these are likely derived from host DNA as reflected by matches with

hypothetical proteins encoded in a number of *C. jejuni* and *C. coli* genomes. Notably DA10\_ORF4, ORF5, ORF7, ORF25, ORF22, ORF45, ORF46, ORF47, and ORF48 appear to be conserved hypothetical proteins in this category. DA10\_ORF53 is a 96 bp gene that displays homology to a small fragment encoding a hypothetical protein in *Campylobacter* pVir-like plasmids. For DA10\_ORF50, no significant BLASTn matches were identified in genomes of *Campylobacter* spp. or associated phages. Translation of the nucleotide sequence and subsequent BLASTp analysis identified a single example of a protein homologue (EYI48\_08780) present in a 9471 bp contig originating from a sequence read archive (SRR8592636) of a *C. coli* cattle isolate. Conserved proteins that are found in DA10 and other *Campylobacter* phage genomes also appear to be free from CRISPR-mediated immunity. In this context DA10\_ORF30, ORF37, ORF54, and ORF57 all display conservation with *Campylobacter* phage-associated proteins. DA10\_ORF57 (384 bp) encodes a 127 amino acid protein that is conserved in *Campylobacter* phages CP20, CPt10, CP81, CP8, CPX, CP81, CP30A, PC14, NCTC 12673, PC5, CP39, CP21, and vB\_CjeM\_Los1. A similar observation is made with DA10\_ORF54 with similar proteins encoded in CPt10, CP20, CP39, CP8, CPX, CP81, and vB\_CjeM\_Los1. A further match against *Campylobacter* phages CP39 and PC5 can be found when analysing DA10\_ORF37 (207 bp), which encodes a small 68 amino acid protein. DA10\_ORF30 (474 bp) encodes a 157 amino acid protein that is homologous to counterparts present in *Campylobacter* prophages CJIE4–2 and CJIE4–5.

## Discussion

Phage DA10 and DA10-like phages are present in a limited number of genome sequences associated with their campylobacter hosts. DA10 has been isolated from the environment, has been identified as a *Myoviridae* following TEM analysis, can produce infectious virions following replication in a permissive host, and potentially has a long-standing evolutionary relationship with *Campylobacter* spp. Given the large number of *Campylobacter* spp. genome sequences that have been produced in recent times, and the paucity of DA10-like phage sequences associated with them, it is reasonable to assume that DA10-like phages are rare in nature. DA10-like phages may constitute a novel genus of temperate phages, albeit with extremely limited environmental distribution. Genome sizes ranging from ~33–38 Kb are observed in DA10-like phages with GC contents of 27.1–27.2% for those present in *C. jejuni* whilst for *C. coli* a slightly higher value of 27.7–27.8% is observed. The lowest GC content of 26.3% is found in the integrated prophage of CJ677CC520. It is interesting to note that previous studies of CJ677 strains show that this

sequence type is associated with a higher degree of disease-severity [32]. Increased resistance to normal human serum was reported for CJ677 bacteraemia-associated strains isolated in Finland from 1998 to 2007 [33]. Infection with CJ677 also typically resulted in longer recovery periods following hospitalization [34].

Identification of a small number of DA10-like phages allowed for a limited comparative analysis to be performed. Table 1 (whole genome alignments) and Table 2 (mummer alignments of coding regions) show the results obtained from nucleotide sequences of all DA10-like phages. Pairwise results from whole genome alignments show that DA10 is most similar at the nucleotide level to *C. jejuni* PNUSAC003324 (74.6% ANI), and alignments of the coding regions show comparable results (88.2% ANI). The integrated DA10-like prophage in CJ677CC520 has suffered considerable degradation in respect of nucleotide substitution and loss in comparison with other members of the group. Pairwise alignment of DA10 with CJ677CC520 prophage results in an overall ANI of 51%, however mummer alignment of the intact coding regions shows a higher degree of sequence divergence (29% ANI), implying preservation of gene function is not essential.

It is quite striking that BLASTn analysis of multiple genes encoded by DA10 delivers search results of ~30 bp fragments that are embedded in CRISPR arrays of *C. jejuni* and *C. coli*. Of the 59 ORFs annotated in the DA10 genome, CRISPR spacers targeting ~75% of these (43 ORFs) were identified (Fig. 1 and Table 3). Several *C. jejuni* and *C. coli* CRISPR arrays also harbour multiple spacers that are homologous to DA10 DNA. *C. jejuni* YH003 encodes two 30 bp CRISPR spacers that represent perfect matches to DA10\_ORF16 and DA10\_ORF52. Similarly, *C. coli* BP3183 encodes two identical CRISPR spacers that are homologous to DA10\_ORF5 (30/30) and DA10\_ORF28 (32/32). *C. coli* BP3183 possesses a second CRISPR spacer targeting DA10\_ORF28, which exhibits a single base pair mismatch with its cognate 30 bp sequence in DA10. Other examples are also observed where a CRISPR array harbours two spacers that target DA10 DNA with the caveat that one of the spacers displays multiple mismatches to the target sequence. For *C. jejuni* GB1 this is found within a spacer that targets DA10\_ORF4 (37/40 matches), however a second spacer that is 100% identical to DA10\_ORF52 (32/32 matches) is also present within the host CRISPR array. Multiple CRISPR spacers targeting the genes encoded of a single phage suggests the phage previously had replicative success but has become an active target. There is a bias towards adenine bases in the protospacer adjacent motifs (PAMs) that arise from DA10, which may contribute to the frequency of targeting the low GC contents of DA10 and DA10-related sequences (Fig. 3).



**Table 2** Pairwise comparison of nucleotide identities for DA10 and DA10-like sequences post Mummer alignment of the coding regions

	DA10	PNUSA C003578	CcNC_C4236	CcNC_C4247	CjFSIS 11,811,828	PNUSA C003324	CJ677 CC520
<b>DA10</b>	*	28,824/ 35,379	25,789/ 35,379	25,791/ 35,379	25,635/ 35,379	31,207/ 35,379	10,263/ 35,379
<b>PNUSA C003578</b>	28,932/ 37,072	*	24,273/ 37,072	24,275/ 37,072	23,575/ 37,072	34,654/ 37,072	10,299/ 37,072
<b>CcNC_C4236</b>	25,858/ 33,831	24,284/ 33,831	*	33,616/ 33,831	20,797/ 33,831	26,190/ 33,831	8529/ 33,831
<b>CcNC_C4247</b>	26,066/ 33,823	24,493/ 33,823	33,823/ 33,823	*	21,006/ 33,823	26,405/ 33,823	8536/ 33,823
<b>CjFSIS 11,811,828</b>	25,683/ 33,563	23,494/ 33,563	20,738/ 33,563	20,740/ 33,563	*	24,894/ 33,563	10,308/ 33,563
<b>PNUSA C003324</b>	31,020/ 37,779	34,629/ 37,779	25,651/ 37,779	25,659/ 37,779	24,880/ 37,779	*	11,401/ 37,779
<b>CJ677 CC520</b>	10,327/ 37,405	10,256/ 37,405	8184/ 37,405	8191/ 37,405	10,399/ 37,405	11,396/ 37,405	*

The limited distribution of DA10-like integrated prophages in *C. jejuni* and *C. coli* genomes is testament to a robust CRISPR-mediated defence that targets multiple functional genes in the phage. Circumvention of immunity will be greatly diminished in this scenario as the invading phage genome is unlikely to be able to tolerate the multiple point mutations required to evade CRISPR-mediated responses. In *Pseudomonas aeruginosa* it has been shown that multiple CRISPR spacers targeting a single phage can effectively drive the virion to extinction [35, 36].

While ~75% of the coding regions of the DA10 genome are actively targeted by *Campylobacter* CRISPR systems it is apparent that the remainder of the ORFs are not represented as spacer sequences. Analysis of ORFs that avoid CRISPR-mediated responses indicates that protection from immunity may be afforded for several reasons. Genes that have host homologues (as determined by conservation across a wide number of *C. jejuni/C. coli* genomes) will not be targeted without a detrimental effect on the host, and thus are spared from spacer acquisition due to the deleterious effects of autoimmunity. ORFs that are observed to be conserved between DA10 and other *Campylobacter* phages also appear to be free from CRISPR-mediated responses. This is perhaps unsurprising because homologues encoded in *Campylobacter* phage genomes that have evolved to successfully infect and target *Campylobacter* spp. are unlikely to be CRISPR targets. Phage-mediated circumvention of Type II-C CRISPR-mediated immunity in *C. jejuni* is a documented phenomenon. Class II and Class III *Campylobacter* phages encode conserved Cas4-like proteins that drive acquisition of host-derived CRISPR spacers. Consequently, Class II and III *Campylobacter* phage genomic DNA remains protected from

CRISPR-mediated immunity during the infection cycle [23, 25]. DA10 and DA10-like phages do not encode a CRISPR Cas4-like protein. It is plausible that the distinct lack of ability to control the *Campylobacter* host CRISPR immune response has driven this family of phages to near extinction within the global *Campylobacter* population. In relation to the integrated DA10-like prophage encoded in CJ677CC520, it is reported that genome sequences derived from members of this clade encode atypical CRISPR systems. Deletion of essential *cas* genes will lead to a loss of functionality of the type II-C systems for CJ677-types [32]. For CJ677CC520 that lacks *cas1*, the ability to generate protospacers during phage infection has likely been lost due to absence of the DNA endonuclease function provided by Cas1 protein. Analysis of a Type II-C CRISPR system in *Riemerella anatipestifer* shows that *cas1/cas2* gene knockouts lose the ability to acquire protospacers, with functionality of the system restored upon ectopic restoration of the *cas1/cas2* genes [37]. In the absence of a functional CRISPR-cas system CJ677-types can remain a refuge for the DA10 genome, and kernels from which the phage population can expand when susceptible hosts become available that do not carry DA10-directed CRISPR spacers. However, re-emergence of the phage will depend upon the host maintaining the integrity of the phage genome, which does not appear to be the case for CJ677CC520.

## Conclusions

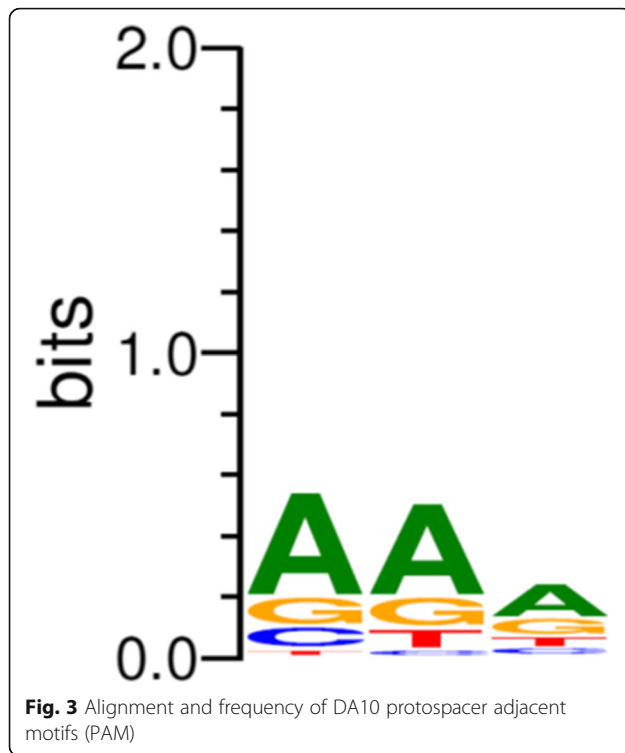
Bacteriophage DA10 has the shortest genome of any *Campylobacter* phage DNA recorded to date and represents a replicative form of a prophage present in the genome of *C. jejuni* CJ677CC520. DA10-like prophage sequences show a restricted distribution in the host

**Table 3** CRISPR spacers targeting *Campylobacter* phage DA10

Host (Acc. No.)	CRISPR spacer sequence (ID)	DA10 target
<i>C. coli</i> BP3183 (CP017871)	CACGTGTGGGAAGGAGAACCTTTAGAGTATAA (32/32)	DA10_ORF01
<i>C. jejuni</i> HSA43 (KR155141)	CCTTTGTCAATTTATGAAAAATTATGTTTCAA (32/32)	DA10_ORF02
<i>C. jejuni</i> HCC31 (KR155133)	AAATTTACAGCACAAAATAAAGATTTTAATATT (33/33)	DA10_ORF03
–	No matches	DA10_ORF04
–	No matches	DA10_ORF05
<i>C. jejuni</i> F041 (EF017333)	TAGATTTAGAAACTTTTAATAAAGTAAAAGA (31/31)	DA10_ORF06
–	No matches	DA10_ORF07
<i>C. jejuni</i> R17 (HQ378296)	ATAATCCTAGCGAGGAAAGCAAACCTGATGATG (33/33)	DA10_ORF08
<i>C. coli</i> YH503 (CP025281)	TGGTTGCTTCGTTCAATCAAAAACAGGTGC (30/30)	DA10_ORF09
<i>C. jejuni</i> NCTC 13268 (LR134497)	AGTGCTACAGCCTTTACCACACTTCAATC (29/29)	DA10_ORF10
<i>C. hyointestinalis</i> LMG9260 (CP015575)	TTCAAAAGATGA-TTTTGATTTACTTTGCAAATCTT (32/36)	DA10_ORF11
<i>C. coli</i> RM4661 (CP007181)	ACCCTGAGTTTAAAGACTTAAGTAAATGTAA (31/31)	DA10_ORF12
<i>C. jejuni</i> YH003 (CP041584)	AGGGGTGATATTGTTATGAGTATTATGCAA (30/30)	DA10_ORF13
<i>C. jejuni</i> F087 (EF017335)	TTAATCCTAGTGAAATTGCAAAATTAACAAG (31/31)	DA10_ORF14
<i>C. jejuni</i> HCC01 (KR155140)	GTTAAAATCTTTGAAACAATAGAATATTTA (30/31)	DA10_ORF15
<i>C. jejuni</i> HCC92 (KR155175)	GAAAATGGATATTTTTATGCAGTTAGCGGGGT (33/33)	DA10_ORF16
<i>C. jejuni</i> HSA25 (KR155156)	GTGGCTAATAATAATAAAGCTAAGACTGATA (31/31)	DA10_ORF17
<i>C. coli</i> 210 (KR155130)	TGTAGCTGGTAAGGTTATAAGAGGTTTATT (30/30)	DA10_ORF18
<i>C. jejuni</i> ATCC 33560 (CP019838)	TTAGTGCTTTATTTGAAAATGAGATGGCAG (30/30)	DA10_ORF19
<i>C. coli</i> RM4661 (CP007181)	TTTTCTTATGAAGTTTTTAAAGGAATTATAC (31/31) GCTTTAGGAAATGCTTTAAACGCTTTGG (29/29)	DA10_ORF20
<i>C. jejuni</i> HCC01 (KR155140)	TATAAAAATATGAAAATAGTATCTTACGCA (30/30)	DA10_ORF21
–	No matches	DA10_ORF22
<i>C. coli</i> 14983A (CP017025)	ATGGAGAAGAGATAAATAAACAAGTTTTAAGC (32/32)	DA10_ORF23
<i>C. coli</i> BP3183 (CP017871)	TCTTATAAACAAAGCGGTAATGAAGGCTTA (30/30)	DA10_ORF24
<i>C. jejuni</i> GB1 (HQ378304)	TGGTATGAGTTAGCAACTTATAAATACTAATGTTATTA (37/40)	DA10_ORF25
<i>C. jejuni</i> F280 (EF017337)	ATTTTGAAGTACTGCTAGAAATGCAAGTTTTA (33/33)	DA10_ORF26
<i>C. coli</i> YH502 (CP018900)	TAGATAACGAGTTATATTCAGAAATACCAAA (31/31)	DA10_ORF27
<i>C. jejuni</i> FORC_046	TGCAGAAAATGGAGATAGACAAGATTTTCCTGT	DA10_ORF28

**Table 3** CRISPR spacers targeting *Campylobacter* phage DA10 (Continued)

Host (Acc. No.)	CRISPR spacer sequence (ID)	DA10 target
(CP017229)	(33/33)	
<i>C. jejuni</i> 183 (KR155181)	TCAAAGTGCTGAAAAAGATAGAACTTACTTA (32/32)	DA10_ORF29
–	No matches	DA10_ORF30
<i>C. jejuni</i> HCC60 (KR155155)	CCTTATAAAAGCGAATATTTCACTGCGTGCG-TTATAC (36/38)	DA10_ORF31
<i>C. jejuni</i> NCTC 12662 (CP019965)	CCATCAAGTCGTGCAATTTAATACACTGGG (30/31)	DA10_ORF32
<i>C. jejuni</i> NCTC 11168 (LS483362)	AGTTAAATTCAACTAATGAACTTTGGAAAA (31/31)	DA10_ORF33
<i>C. jejuni</i> CJ066CC508 (CP0122224)	GTTACTTACACCCTTTAAAACTAATGG (30/30)	DA10_ORF34
<i>C. jejuni</i> YH003 (CP041584)	CTTTATGTAATTTTATAATCAAAGTATAA (30/30)	DA10_ORF35
<i>C. jejuni</i> GB1 (HQ378304)	TCTACAAGAATGAGGATGATGATTTTTACAA (32/32)	DA10_ORF36
–	No matches	DA10_ORF37
<i>C. jejuni</i> NCTC 12660 (CP028910)	AGATGCTTTAACAGTATGTTCTTTAGCGGG (29/30)	DA10_ORF38
<i>C. jejuni</i> HCC04 (KR155185)	CCAAAACAATAGCAAATTTTAAATATTATATA (32/32)	DA10_ORF39
<i>C. jejuni</i> R133 (HQ378258)	ATGAGTAAAAAAGATAGTCAAGAATGGT (28/28)	DA10_ORF40
<i>C. jejuni</i> F007 (EF017341)	AAAAGAGCCTTTAGAAAACAATGTTTTATAATA (31/34)	DA10_ORF41
<i>C. coli</i> RM4661 (CP007181)	ATAATTCCTAATAAAAAAGTAAGTTTTGATG (31/31)	DA10_ORF42
<i>C. jejuni</i> F459 (EF017334)	AAA-TAGTTTATAAAAAATAAGCAAGATGTAGA (31/32)	DA10_ORF43
<i>C. jejuni</i> F226 (EF017339)	TTATTATGATGATAACTTAATAAGTTTATC (31/31)	DA10_ORF44
–	No matches	DA10_ORF45
–	No matches	DA10_ORF46
–	No matches	DA10_ORF47
–	No matches	DA10_ORF48
<i>C. jejuni</i> NCTC 12660 (CP028910)	TTTAAATTTCAAAGATGAGAATATAGCTAA (30/30)	DA10_ORF49
–	No matches	DA10_ORF50
–	No matches	DA10_ORF51
<i>C. jejuni</i> NCTC 13257 (LR134502)	TAAAATAATTTGCAAGGTATATAAAAAATTTGATTTTG (37/38)	DA10_ORF52
–	No matches	DA10_ORF53
–	No matches	DA10_ORF54
<i>C. jejuni</i> R15 (HQ378302)	TTAAAAATAGAAATTTTATAGGTTGCGAAA (30/30)	DA10_ORF55
<i>C. jejuni</i> 80 (KR155173)	AAA-AGAACTATATTAGCAGGTATAAATAATCT (33/34)	DA10_ORF56
–	No matches	DA10_ORF57
<i>C. jejuni</i> NCTC 12664 (CP028912)	GTTGCTAAGAATAAAAATAAGAAACTGGT (29/30)	DA10_ORF58
–	No matches	DA10_ORF59



genomes of *C. jejuni* and *C. coli* available in the sequence read archive compared to previously recognised Mu-like and CJIE-like prophage sequences. DA10 and DA10-like bacteriophages are a target for CRISPR-mediated immunity, which may be a driving factor in their rarity since 75% of the ORFs of DA10 are present as ~30 bp spacer sequences in numerous *Campylobacter* Type II-C CRISPR arrays. Under these circumstances, prophage integration could render the host susceptible to CRISPR-mediated autoimmunity and elimination. CJ677CC520 is among a clade of campylobacters that exhibit atypical CRISPR systems that are deficient *cas* genes, and in the absence of CRISPR-mediated immunity permit prophage integration.

## Methods

### Host strains & bacteriophage DA10 propagation

*Campylobacter jejuni* was routinely grown on blood agar base No.2 plates (BA plates; Oxoid) supplemented with 5% horse blood (TCS Biosciences) under microaerobic conditions in either a modular atmosphere controlled cabinet (5% CO<sub>2</sub>, 5% O<sub>2</sub>, 2% H<sub>2</sub>, 88% N<sub>2</sub>) or in anaerobic jars using gas replacement (7.3% CO<sub>2</sub>, 5.6% O<sub>2</sub>, 3.6% H<sub>2</sub>, 83.5% N<sub>2</sub>) at 42 °C. Bacteriophage DA10 was isolated as described previously [26] and propagated by plate lysis of the host *C. jejuni* GM. *C. jejuni* GM cells were collected from a BA plate and suspended in 10 mM MgSO<sub>4</sub> (Thermo Fisher) to a cell density of approximately 10<sup>7</sup> CFU/ml. A volume of 500 µl of the cell suspension was

added 10<sup>7</sup> PFU DA10 before adding 5 ml of molten NZCYM top agar (0.6%, VWR Leicestershire, UK) and pouring on top of a NZCYM agar plate.

### Pulse fields gel electrophoresis

For PFGE, plugs were prepared as described previously [38]. Briefly a 10 µl suspension of bacteriophage at 10<sup>10</sup> PFU/ml was added to 40 µl of TE buffer (10 mM Tris, 1 mM EDTA [pH 7.5]) and mixed with an equal volume of 1.4% molten low-melting point PFGE agarose in TE buffer and dispensed into plug molds (Bio-Rad). Solidified plugs were washed in 5 ml of lysis buffer (100 mM EDTA, 10 mM Tris [pH 7.2], 1% Sarkosyl [w/v], 0.5 mg of proteinase K; Thermofisher) at 55 °C for 18 h with gentle shaking. The lysis solution was discarded and proteinase K inactivated by washing in 5 ml of 1 mM phenylmethylsulfonyl fluoride 50 mM EDTA, 20 mM Tris [pH 7.2] for 1 h at room temperature, before three successive washes of 20 min in 50 mM EDTA, 20 mM Tris [pH 7.2]. Agarose gels (1% w/v) were loaded with 2 mm plug slices. The gel was run using a Bio-Rad CHEF DR11 system in 0.5× TBE for 18 h at 6 V/cm with a switch time of 30 to 60 s. Lambda concatemers were used as markers (GelSyringe™ New England Biolabs, USA), and the DNA was visualized by staining with ethidium bromide.

### Transmission electron microscopy

Glutaraldehyde-fixed phage suspensions on Pioloform grids were negatively stained with 0.5% uranyl acetate. The specimens were imaged using a Tecnai G12 biotwin TEM, run at 100Kv, with a SIS megaview camera system and Gatan Microscopy Suite software (Gatan Inc).

### Bacteriophage genomic DNA preparation

Genomic DNA was extracted from the DA10 lysate using a modified protocol of the Wizard DNA clean up kit (Promega, Southampton, UK). Phage suspensions were mixed with nuclease solution (10 mg/ml DNase and RNase) and incubated at 37 °C for 30 min. A precipitant solution containing 10% PEG 8000, 1 M NaCl was added at a ratio of 1:2 precipitant to lysate. The solution was mixed by inversion and incubated at 4 °C overnight followed by centrifugation at 10,000 g at 4 °C for 10 min. The supernatant was discarded, and the pellet resuspended in 500 µl of 5 mM MgSO<sub>4</sub> and cleared by centrifugation for 10 s in a microfuge. Residual nucleases were degraded by addition of 10 µl of 0.5 M EDTA pH 8 and proteinase K to a final concentration of 0.1 mg/ml. The purification resin supplied with the Promega Wizard kit was resuspended and 1 ml added to the phage suspension. The mixture was inverted 5 to 6 times and filled into a 3 ml syringe (Part# A809B). A Wizard minicolumn was placed into a reaction tube and the

syringe barrel attached to the column and the resin/lysate mix was pushed into the column until all liquid was forced through the resin. The column was washed with 2 ml of 80% isopropanol and centrifuged at 13,000 g for 2 min to dry the resin. The column was placed in a 1.5 ml microcentrifuge tube and 100 µl of sterile water at 80 °C pipetted into the column. The column was centrifuged at 13,000 g for 1 min to elute the DNA. Residual guanidine contaminants were removed by ethanol precipitation.

### Genome sequencing

Library preparation of DA10 genomic DNA followed the Illumina Nextera™ tagmentation protocol (Illumina, Cambridge, UK). The DA10 genome was assembled de novo using CLC Genomics Workbench Version 9.0.1 (Qiagen Bioinformatics, Aarhus, Denmark) from 2.1 million Illumina MiSeq paired-end sequence reads of 120 to 240 nucleotides.

### DA10 genome analysis

The DA10 nucleotide sequence was initially annotated using a combination of the online tools RAST (<http://rast.nmpdr.org/rast.cgi>) [39], Phaster (<http://phaster.ca/>) [40], and HHpred (<https://toolkit.tuebingen.mpg.de/tools/hhpred>) [41]. Manual curation of DA10 ORFs was performed using a combination of BLASTn and BLASTp to assist in annotation and assigning function to genes when possible [42]. The annotated DA10 DNA sequence was submitted to NCBI (Accession No. MN530981). Average nucleotide identities of DA10 and DA10-like phages were calculated from whole genome alignments using a pairwise BLAST comparative approach, and MUMMER was utilized for coding region alignments (<https://github.com/mummer4/mummer>). Analysis of DA10 ORFs represented as spacers present in *Campylobacter* spp. CRISPR arrays was performed using a combination of BLASTn and CRISPR-Cas++ (<https://crispcas.i2bc.paris-saclay.fr/>) [43, 44]. Protospacer adjacent motif (PAM) sequence alignments were obtained using WebLogo 3 (<http://weblogo.threplusone.com/>) [45, 46].

### Abbreviations

ANI: Average nucleotide identity; BLAST: Basic Local Alignment Search Tool; Cas: CRISPR-associated protein; CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats; CJIE: *Campylobacter jejuni* integrated elements; ORF: Open Reading Frame; PAM: Protospacer Adjacent Motif; PHASTER: Phage Search Tool Enhanced Release; RAST: Rapid Annotation using Subsystem Technology

### Acknowledgments

"Not applicable".

### Authors' contributions

SH and IFC performed the computational analyses and wrote the manuscript. DA, GA and PLC isolated, propagated and characterised the bacteriophage, and provided the biological interpretation of the host range.

YH performed the PFGE and TEM experiments and provided the data interpretation. All authors read and approved the final manuscript.

### Funding

This work was supported by the Biotechnology and Biological Sciences Research Council [grant number BB/P02355X/1], UK. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Availability of data and materials

All data generated or analysed during this study are included in this published article. DNA sequences appear in the National Center for Biotechnology Information (NCBI) nucleotide sequence database accession number MN530981. The bacteriophage is available from DA.

### Ethics approval and consent to participate

"Not applicable".

### Consent for publication

"Not applicable".

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>School of Biosciences, Division of Microbiology Brewing and Biotechnology, University of Nottingham, Sutton Bonington Campus, Leicestershire LE12 5RD, UK. <sup>2</sup>Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale", Teramo, Italy.

Received: 6 December 2019 Accepted: 5 June 2020

Published online: 12 June 2020

### References

1. Kaakoush NO, Castano-Rodriguez N, Mitchell HM, Man SM. Global epidemiology of campylobacter infection. *Clin Microbiol Rev.* 2015;28(3):687–720.
2. Costa D, Iraola G. Pathogenomics of Emerging *Campylobacter* Species. *Clin Microbiol Rev.* 2019;32:4.
3. Meric G, Yahara K, Mageiros L, Pascoe B, Maiden MC, Jolley KA, Sheppard SK. A reference pan-genome approach to comparative bacterial genomics: identification of novel epidemiological markers in pathogenic *Campylobacter*. *PLoS One.* 2014;9(3):e92798.
4. Llarena AK, Taboada E, Rossi M. Whole-genome sequencing in epidemiology of *Campylobacter jejuni* infections. *J Clin Microbiol.* 2017;55(5):1269–75.
5. Parker CT, Quinones B, Miller WG, Horn ST, Mandrell RE. Comparative genomic analysis of *Campylobacter jejuni* strains reveals diversity due to genomic elements similar to those present in *C. jejuni* strain RM1221. *J Clin Microbiol.* 2006;44(11):4125–35.
6. Clark CG, Ng LK. Sequence variability of campylobacter temperate bacteriophages. *BMC Microbiol.* 2008;8:49.
7. Scott AE, Timms AR, Connerton PL, Loc Carrillo C, Adzfa Radzum K, Connerton IF. Genome dynamics of *Campylobacter jejuni* in response to bacteriophage predation. *PLoS Pathog.* 2007;3(8):e119.
8. Brussow H, Canchaya C, Hardt WD. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev.* 2004;68(3):560–602 table of contents.
9. Gaasbeek EJ, Wagenaar JA, Guilhaert MR, Wosten MM, van Putten JP, van der Graaf-van Bloois L, Parker CT, van der Wal FJ. A DNase encoded by integrated element CJIE1 inhibits natural transformation of *Campylobacter jejuni*. *J Bacteriol.* 2009;191(7):2296–306.
10. Gaasbeek EJ, Wagenaar JA, Guilhaert MR, van Putten JP, Parker CT, van der Wal FJ. Nucleases encoded by the integrated elements CJIE2 and CJIE4 inhibit natural transformation of *Campylobacter jejuni*. *J Bacteriol.* 2010;192(4):936–41.
11. Brown HL, Hanman K, Reuter M, Betts RP, van Vliet AH. *Campylobacter jejuni* biofilms contain extracellular DNA and are sensitive to DNase I treatment. *Front Microbiol.* 2015;6:699.
12. Clark CG, Grant CC, Pollari F, Marshall B, Moses J, Tracz DM, Gilmour MW. Effects of the *Campylobacter jejuni* CJIE1 prophage homologs on adherence

- and invasion in culture, patient symptoms, and source of infection. BMC Microbiol. 2012;12:269.
13. Skarp CPA, Akinrinade O, Kaden R, Johansson C, Rautelin H. Accessory genetic content in *Campylobacter jejuni* ST21CC isolates from feces and blood. Int J Med Microbiol. 2017;307(4–5):233–40.
  14. Fouts DE, Mongodin EF, Mandrell RE, Miller WG, Rasko DA, Ravel J, Brinkac LM, DeBoy RT, Parker CT, Daugherty SC, et al. Major structural differences and novel potential virulence mechanisms from the genomes of multiple campylobacter species. PLoS Biol. 2005;3(1):e15.
  15. Marasini D, Fakhr MK. Whole-Genome Sequencing of a *Campylobacter jejuni* Strain Isolated from Retail Chicken Meat Reveals the Presence of a Megaplasmid with Mu-Like Prophage and Multidrug Resistance Genes. Genome Announcements. 2016;4:3.
  16. Clark CG, Chong PM, McCorrister SJ, Simon P, Walker M, Lee DM, Nguy K, Cheng K, Gilmour MW, Westmacott GR. The CJIE1 prophage of *Campylobacter jejuni* affects protein expression in growth media with and without bile salts. BMC Microbiol. 2014;14:70.
  17. Figueroa-Bossi N, Uzzau S, Maloroli D, Bossi L. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in Salmonella. Mol Microbiol. 2001;39(2):260–71.
  18. Firehammer BD, Border M. Isolation of temperate bacteriophages from Vibrio fetus. Am J Vet Res. 1968;29(11):2229–35.
  19. Bryner JH, Ritchie AE, Foley JW, Berman DT. Isolation and characterization of a bacteriophage for Vibrio fetus. J Virol. 1970;6(1):94–9.
  20. Bryner JH, Ritchie AE, Booth GD, Foley JW. Lytic activity of vibrio phages on strains of Vibrio fetus isolated from man and animals. Appl Microbiol. 1973; 26(3):404–9.
  21. Gencay YE, Sorensen MCH, Wenzel CQ, Szymanski CM, Brondsted L. Phase variable expression of a single phage receptor in *Campylobacter jejuni* NCTC12662 influences sensitivity toward several diverse CPS-dependent phages. Front Microbiol. 2018;9:82.
  22. Lis L, Connerton IF. The minor flagellin of *Campylobacter jejuni* (FlaB) confers defensive properties against bacteriophage infection. Front Microbiol. 2016;7:1908.
  23. Hooton SP, Connerton IF. *Campylobacter jejuni* acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. Front Microbiol. 2015;5:744.
  24. Pearson BM, Louwen R, van Baarlen P, van Vliet AH. Differential distribution of type II CRISPR-Cas Systems in Agricultural and Nonagricultural *Campylobacter coli* and *Campylobacter jejuni* isolates correlates with lack of shared environments. Genome Biol Evol. 2015;7(9):2663–79.
  25. Hooton SP, Brathwaite KJ, Connerton IF. The bacteriophage carrier state of *Campylobacter jejuni* features changes in host non-coding RNAs and the Acquisition of new Host-derived CRISPR spacer sequences. Front Microbiol. 2016;7:355.
  26. Aprea G, D'Angelantonio D, Boni A, Connerton P, Connerton I, Scattolini S, Marotta F, Pomilio F, Migliorati G, D'Alterio N, Di Giannatale E. Isolation and morphological characterization of new bacteriophages active against *Campylobacter jejuni*. Am J Clin Microbiol Antimicrobials. 2018;1(1):1004.
  27. Sails AD, Wareing DRA, Bolton FJ, Fox AJ, Curry A. Characterisation of 16 *Campylobacter jejuni* and *C. coli* typing bacteriophages. J Med Microbiol. 1998;47:123–8.
  28. Javed MA, Ackermann HW, Azeredo J, Carvalho CM, Connerton I, Evoy S, Hammerl JA, Hertwig S, Lavigne R, Singh A, Szymanski CM, Timms A, Kropinski AM. A suggested classification for two groups of *Campylobacter* myoviruses. Arch Virology. 2014;159:181–90.
  29. Skarp CP, Akinrinade O, Nilsson AJ, Ellstrom P, Myllykangas S, Rautelin H. Comparative genomics and genome biology of invasive *Campylobacter jejuni*. Sci Rep. 2015;5:17300.
  30. O'Kane PM, Connerton IF. Characterisation of aerotolerant forms of a robust chicken colonizing *Campylobacter coli*. Front Microbiol. 2017;8:513.
  31. Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. Nat Struct Mol Biol. 2014;21(6):528–34.
  32. Kivisto RJ, Kovanen S, Skarp-de Haan A, Schott T, Rahkio M, Rossi M, Hanninen ML. Evolution and comparative genomics of *Campylobacter jejuni* ST-677 clonal complex. Genome Biol Evol. 2014;6(9):2424–38.
  33. Jaakkonen A, Castro H, Hallanvuos S, Ranta J, Rossi M, Isidoro J, Lindstrom M, Hakkinen M. Longitudinal Study of Shiga Toxin-Producing *Escherichia coli* and *Campylobacter jejuni* on Finnish Dairy Farms and in Raw Milk. Appl Environ Microbiol. 2019;85:7.
  34. Feodoroff B, de Haan CP, Ellstrom P, Sarna S, Hanninen ML, Rautelin H. Clonal distribution and virulence of *Campylobacter jejuni* isolates in blood. Emerg Infect Dis. 2013;19(10):1653–5.
  35. van Houte S, Ekroth AK, Broniewski JM, Chabas H, Ashby B, Bondy-Denomy J, Gandon S, Boots M, Paterson S, Buckling A, et al. The diversity-generating benefits of a prokaryotic adaptive immune system. Nature. 2016;532(7599): 385–8.
  36. Hoyland-Kroghsbo NM, Munoz KA, Bassler BL. Temperature, by Controlling Growth Rate, Regulates CRISPR-Cas Activity in *Pseudomonas aeruginosa*. Mbio. 2018;9:6.
  37. He Y, Wang M, Liu M, Huang L, Liu C, Zhang X, Yi H, Cheng A, Zhu D, Yang Q, et al. Cas1 and Cas2 from the type II-CRISPR-Cas system of *Riemerella anatipestifer* are required for spacer acquisition. Front Cell Infect Microbiol. 2018;8:195.
  38. Atterbury RJ, Connerton PL, Dodd CE, Rees CE, Connerton IF. Isolation and characterization of *Campylobacter* bacteriophages from retail poultry. Appl Environ Microbiol. 2003;69:4511–8.
  39. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. The RAST server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
  40. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44(W1):W16–21.
  41. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol. 2018; 430:2237–43.
  42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
  43. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35(Web Server issue):W52–7.
  44. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Neron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018;46(W1): W246–51.
  45. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14(6):1188–90.
  46. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990;18(20):6097–100.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

