



# CRISPR-Cas Systems in *Bacteroides fragilis*, an Important Pathobiont in the Human Gut Microbiome

Mehrdad Tajkarimi<sup>1</sup> and Hannah M. Wexler<sup>1,2,3\*</sup>

<sup>1</sup> Brentwood Biomedical Research Institute, Los Angeles, CA, United States, <sup>2</sup> University of California, Los Angeles, Los Angeles, CA, United States, <sup>3</sup> GLAVA Health Care System, Los Angeles, CA, United States

## OPEN ACCESS

### Edited by:

Scott Norman Peterson,  
Sanford-Burnham Institute for Medical  
Research, United States

### Reviewed by:

Andrei Osterman,  
Sanford Burnham Prebys Medical  
Discovery Institute (SBP),  
United States  
Jozsef Soki,  
University of Szeged, Hungary

### \*Correspondence:

Hannah M. Wexler  
hwexler@ucla.edu

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 23 July 2017

Accepted: 31 October 2017

Published: 23 November 2017

### Citation:

Tajkarimi M and Wexler HM (2017)  
CRISPR-Cas Systems in *Bacteroides*  
*fragilis*, an Important Pathobiont in the  
Human Gut Microbiome.  
Front. Microbiol. 8:2234.  
doi: 10.3389/fmicb.2017.02234

**Background:** While CRISPR-Cas systems have been identified in bacteria from a wide variety of ecological niches, there are no studies to describe CRISPR-Cas elements in *Bacteroides* species, the most prevalent anaerobic bacteria in the lower intestinal tract. Microbes of the genus *Bacteroides* make up ~25% of the total gut microbiome. *Bacteroides fragilis* comprises only 2% of the total *Bacteroides* in the gut, yet causes of >70% of *Bacteroides* infections. The factors causing it to transition from benign resident of the gut microbiome to virulent pathogen are not well understood, but a combination of horizontal gene transfer (HGT) of virulence genes and differential transcription of endogenous genes are clearly involved. The CRISPR-Cas system is a multi-functional system described in prokaryotes that may be involved in control both of HGT and of gene regulation.

**Results:** Clustered regularly interspaced short palindromic repeats (CRISPR) elements in all strains of *B. fragilis* ( $n = 109$ ) with publically available genomes were identified. Three different CRISPR-Cas types, corresponding most closely to Type IB, Type IIIB, and Type IIC, were identified. Thirty-five strains had two CRISPR-Cas types, and three strains included all three CRISPR-Cas types in their respective genomes. The *cas1* gene in the Type IIIB system encoded a reverse-transcriptase/Cas1 fusion protein rarely found in prokaryotes. We identified a short CRISPR (3 DR) with no associated *cas* genes present in most of the isolates; these CRISPRs were found immediately upstream of a *hipA/hipB* operon and we speculate that this element may be involved in regulation of this operon related to formation of persister cells during antimicrobial exposure. Also, blood isolates of *B. fragilis* did not have Type IIC CRISPR-Cas systems and had atypical Type IIIB CRISPR-Cas systems that were lacking adjacent *cas* genes.

**Conclusions:** This is the first systematic report of CRISPR-Cas systems in a wide range of *B. fragilis* strains from a variety of sources. There are four apparent CRISPR-Cas systems in *B. fragilis*—three systems have adjacent *cas* genes. Understanding CRISPR/Cas function in *B. fragilis* will elucidate their role in gene expression, DNA repair and ability to survive exposure to antibiotics. Also, based on their unique CRISPR-Cas arrays, their phylogenetic clustering and their virulence potential, we are proposing that blood isolates of *B. fragilis* be viewed a separate subgroup.

**Keywords:** *Bacteroides*, virulence, CRISPR-Cas system, phage, immune defense, gut microbiome, pathobiome

## INTRODUCTION

Bacteria are constantly exposed to incoming DNA via phages, plasmids, other mobile genetic elements (MGEs) or “naked” nucleic acid from lysed cells. Their ability to incorporate these bits of DNA into their own genetic code, known as horizontal gene transfer (HGT), contributes to the ability of a bacterium to adapt to a wide variety of ecological and environmental pressures, including antibiotics and evolving host niche. Along with this ability to incorporate DNA, bacteria possess immune defense mechanisms to defend themselves against invading DNA or RNA in their immediate environment, including systems to abort infection, specialized enzymes that degrade foreign DNA (restriction/modification or R/M systems), and an adaptive immunity system first described more than a decade ago (Haft et al., 2005) that confers adaptive immunity to invading DNA or RNA (Marraffini, 2015). This system, the CRISPR-Cas system, is defined by clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated (Cas) proteins (Haft et al., 2005).

The specificity of the CRISPR system lies in the short, repetitive (sometimes palindromic) direct repeat sequences (DR) separated by nucleic acid sequences (spacers) previously acquired from invading DNA and a cleavage system that can target incoming DNA based on recognition of those previously encountered DNA sequences (Burstein et al., 2016). The Cas protein machinery that mediates the protection against invading nucleic acid is encoded by gene clusters that are adjacent to and generally upstream of the CRISPR locus (Jore et al., 2011). The CRISPR RNA array (crRNA) that includes the entire DR sequence with the intervening spacers is transcribed (from the leader) as a single transcript and later processed into individual segments that contain one spacer sequence and portions of the DR sequence at both ends. These individual crRNAs are eventually used to direct Cas proteins to destroy incoming invasive genetic elements (including phages, plasmids and other nucleic acid segments) that include sequences complementary to one of the spacer sequences in the CRISPR. Newly acquired spacers are incorporated into CRISPR loci in a directional manner, generally, but not invariably, at repeats adjacent to CRISPR leaders (Rath et al., 2015).

Many surveys reported that a vast number of prokaryotic species possess CRISPR arrays (Louwen et al., 2014; including ~45% of bacteria and ~83% of Archaea; Grissa et al., 2007) although there are entire phyla that lack this system (Burstein et al., 2016). CRISPR-Cas systems were classified into 18 structural families and 24 sequence families (Makarova et al., 2015). While new subtypes are still being identified, the major groups include 5 major types and 16 different subtypes. The basis for these classifications include which *cas* genes are present, the *cas* operon architecture, Cas protein sequences, and the nuclease involved in the degradation of the target DNA (Barrangou, 2015). The function of some of the *cas* genes has been described in detail, while the function(s) of other genes remains unknown (Rath et al., 2015). The most common systems described thus far in prokaryotes are variants of the (1) Class 1 Type I system, (2) the Class 1 Type III system, and (3) the Class 2 Type II system

(Makarova et al., 2015). In general, most strains of the same species contain identical CRISPR-Cas types (Louwen et al., 2014) although there are some variations.

Since the host CRISPR/Cas can acquire and incorporate incoming DNA as spacer DNA, the resultant CRISPR may serve as a molecular record of past exposure to foreign DNA for an individual strain. Alternatively, entire CRISPR loci may be disseminated via HGT. There is evidence that these systems are readily transferred between microbes, potentially even across phylum boundaries (Burstein et al., 2016); this is supported both by the phylogenetic relationship of the *Cas* genes from diverse organisms (Godde and Bickerton, 2006) as well as the fact that DR sequences from diverse organisms can be clustered together into subtypes based on sequence similarities (Kunin et al., 2007).

The function of the CRISPR-Cas system was initially described as protection against invasive nucleic acid. However, the sheer diversity of the systems suggested other functions, and indeed, these elements have now been implicated in regulation of transcription, chromosomal segregation and rearrangement and DNA repair, although most of the processes related to these alternate functions are not completely clear at this point (Stern et al., 2010; Sampson and Weiss, 2014). The potential of regulation via suppressing transcription is also suggested by the broad use of modified Cas9 nuclease precisely to suppress transcription in engineered systems (Barrangou et al., 2007; Sander and Joung, 2014; Mimee et al., 2015; Mougiakos et al., 2016).

The presence of CRISPR-Cas systems are variably associated with virulence (Makarova et al., 2011; Barrangou, 2015) and/or antibiotic resistance in pathogenic bacteria. For example, the presence of a CRISPR-Cas locus correlated inversely with acquired antibiotic resistance (Palmer and Gilmore, 2010) in *Enterococcus faecalis*; a mechanism for CRISPR-Cas loss in this species was identified and the data suggested that antibiotic use inadvertently selects for enterococcal strains with compromised genome defense. Conversely, an *E. faecalis* strain harboring CRISPR-Cas was more virulent than one lacking this element (Bourgoigne et al., 2008). The type of CRISPR element may also be related to the virulence of the bacterium: CRISPR typing was able to discriminate between lineages of *Propionibacterium acnes* with varying degrees of virulence although there were strains belonging to the more invasive lineages that lacked such spacers or a complete CRISPR-Cas system (Marinelli et al., 2012). On the one hand, the ability of a bacterium to incorporate mobile elements bearing pathogenic determinants would argue for a less robust CRISPR-Cas defense system; on the other hand, the presence of multiple CRISPRs systems with a variety of spacers that may indicate previous exposure to these elements, parts of which were incorporated into the bacterial genome.

While CRISPR-Cas systems were identified in bacteria from a wide variety of ecological niches, there are no studies describing CRISPR-Cas elements in *Bacteroides* species, the most prevalent anaerobic bacteria in the lower intestinal tract. The species *Bacteroides fragilis* (BF), an important gut pathobiont, is of particular interest, since its transition from friendly commensal to dangerous threat is not well understood. Outside its colonic niche, *B. fragilis* is an opportunistic pathogen; *B. fragilis* only

accounts for 2% of the total gut *Bacteroides* yet it is the agent of >70% of *Bacteroides* infections. *B. fragilis* is the main cause of anaerobic bacteremia and intraabdominal abscesses, implicated in serious gynecological, soft tissue infections, peritonitis, brain abscess (Wexler, 2014) and surgical site infections (SSIs) following colorectal surgery (Solomkin et al., 2010). Even *B. fragilis* strains isolated as normal gut microbiota possess many genes associated with virulence; as the expression of virulent genes constitutes a form of stress response of pathogenic bacteria during host infection, we would expect differential regulation of the multitude of genes involved in virulence, survival, and host colonization (Louwen et al., 2014).

*Bacteroides fragilis* are also important players in the “hot spot” of HGT between microbes (Kurokawa et al., 2007) and constitute one of the most concentrated reservoirs of resistance genes in the human gut (Salyers et al., 2004). HGT throughout diverse bacterial species has been responsible for the dissemination of both virulence and resistance genes that undermine the usefulness of most antimicrobials (Barlow, 2009). In this gut milieu, resistance genes can move from commensal organisms to potential pathogens, between pathogens, and even between pathogens and probiotics (Capozzi and Spano, 2009). Genetic analysis of *B. fragilis* clinical isolates as well as isolates from the GI tract indicate a high degree of HGT, including resistance genes from very divergent gram-positive bacteria (Husain et al., 2014).

The importance of CRISPR-Cas systems in *B. fragilis* is of particular interest, given the potential of these systems to regulate and/or record HGT, including the acquisition of virulence and antimicrobial resistance genes. CRISPR-Cas systems may also regulate endogenous genes, some of which may be important in the transition of *B. fragilis* from commensal to virulent pathogen. The possible ability of Type I CRISPR-Cas to effect DNA repair would be highly significant for *B. fragilis*, which is very susceptible to the DNA-damaging properties of oxygen exposure. The aim of our study was to investigate and document the occurrence, prevalence and diversity of CRISPR-Cas systems in *B. fragilis* strains with publically available genomes, including strains from a variety of clinical sources and belonging to different phylogenetic clades. With this information, we will explore the capacity of the naturally-occurring CRISPR-Cas systems to regulate gene expression, control HGT (including transmission of antimicrobial resistance and virulence determinants) and control reaction to antibiotic exposure.

## METHODS

### Strains

One hundred nine published sequences of *B. fragilis* as well as the complete genome sequence of *B. fragilis* species UW (a blood isolate reported which clusters phylogenetically with a subset of strains from *B. fragilis*; Salipante et al., 2015) were retrieved from the National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/genome>). The assembly numbers of these complete genomes, the common names of the strains, and the type of infection from which they were isolated are listed

in **Table 1**. Some strains were classified as virulent or multidrug resistant without the source of isolation; these are so noted.

### Phylogenetic Analysis of *B. fragilis* Strains

The J species Web Server (with the kind assistance of Dr. Michael Richter, Ribocoon, Bremen, Germany) was used to construct a matrix of average nucleotide identity between two organisms (ANIb) values among the *B. fragilis* strains tested. The matrix was converted to a three column array using an Excel macro and then converted into a Newick tree file using a routine in R (with the kind assistance of Dr. Luis M. Rodriguez, Kostas Lab, Georgia Institute of Technology, Atlanta, GA). The Newick file was visualized in the Archaeopteryx software tool (Han and Zmasek, 2009).

### Identification of CRISPRs in *B. fragilis*

CRISPR arrays were identified using both the CRT tool (Bland et al., 2007) with default parameters and the CRISPR Repeat Finder program (Grissa et al., 2007). We found the CRT program to be generally more robust than CRISPR Finder in detecting CRISPRs; on the other hand, CRT does not have a filter to limit acceptable spacer homology within a given CRISPRs, leading to incorrectly identifying tandem repeat sequences (perfect and imperfect) as CRISPRs. The CRT program was used with default parameters and the CRISPR Repeat Finder program was used with the following parameters: Min Number of Repeats 2, Min Repeat Length 19, Max Repeat Length 55, Search Window 8, Min Spacer length 18 and Max Spacer Length 55. CRISPRs repeats that had three or more DR sequences were retained for further analysis and then validated using the CRISPTionary program (Grissa et al., 2007).

### Distribution of Spacers among CRISPRs

The CRISPTionary tool (Grissa et al., 2007) was used to analyze spacer distribution and position with individual CRISPRs and to construct a binary file of spacer distribution among the various CRISPR-Cas systems.

### Identification of Protospacer Target Regions

Clustered regularly interspaced short palindromic repeats (CRISPR) targets were identified by submitting the spacer sequences to analysis with the CRISPR Target tool (<http://bioanalysis.otago.ac.nz/CRISPRTarget>) searching the GenBank Phage, RefSeq Bacteria, RefSeq plasmid and RefSeq viral databases. We also searched the Nucleotide collection (nr/nt) at NCBI using the default parameters, both with and without the limit of “NOT Repeat\_Region” added as a query filter. Each matched protospacer was then subjected to a BLAST search (Benson et al., 2014) to find full or partial matches to any other phage, bacteria, plasmid and viruses; annotation of the ten up- and downstream genes flanking the proposed match were recorded.

**TABLE 1** | *Bacteroides fragilis* strains used in this study.

Common name	Clinical source	GenBank accession no.	Assembly number	References
<i>Bacteroides fragilis</i> NCTC 9343	Appy Abscess	CR626927.1	ASM2598v1	NCTC Type Strain
<i>Bacteroides fragilis</i> 638R	Appy Abscess	FQ312004.1	ASM21083v1	Privitera et al., 1979
<i>Bacteroides fragilis</i> HMW 615	Appy Abscess	JH815491.1	Bact_frag_HMW_615_V1	Pumbwe et al., 2007a
<i>Bacteroides fragilis</i> DCMSKEJBY0001B	Blood	JMZ02000001.1	DCMSKEJBY0001B2.0	Sydenham et al., 2015
<i>Bacteroides fragilis</i> DCMOUH0067B	Blood	JPHS01000001.1	DCMOUH0067B_1.0	Sydenham et al., 2015
<i>Bacteroides fragilis</i> HMW 610	Blood	JH815482.1	Bact_frag_HMW_610_V1	Sherwood et al., 2011
<i>Bacteroides fragilis</i> HMW 616	Blood	JH815524.1	Bact_frag_HMW_616_V1	Pumbwe et al., 2007a
<i>Bacteroides fragilis</i> DCMOUH0017B	Blood	JMZ02000001.1	DCMOUH0017B2.0	Sydenham et al., 2015
<i>Bacteroides fragilis</i> DCMOUH0018B	Blood	JMZ02000001.1	DCMOUH0018B2.0	Sydenham et al., 2015
<i>Bacteroides fragilis</i> 894_BFRA	Blood	JUOZ01000315.1	ASM105877v1	Roach et al., 2015
<i>Bacteroides fragilis</i> YCH46	Blood	AP006841.1	ASM992v1	Kuwahara et al., 2004
<i>Bacteroides</i> sp. UW	Blood	JANI01000001.1	ASM78502v1	Kalapila et al., 2013
<i>Bacteroides fragilis</i> DCMOUH0042B	Blood	JPGQ01000001.1	DCMOUH0042B1.0	Sydenham et al., 2015
<i>Bacteroides fragilis</i> DCMOUH0085B	Blood	JPHP01000001.1	DCMOUH0085B_1.0	Sydenham et al., 2015
<i>Bacteroides fragilis</i> 885_BFRA	Blood	JUPJ01000001.1	ASM105875v1	Roach et al., 2015
<i>Bacteroides fragilis</i> 884_BFRA	Blood	JUPK01000001.1	ASM105874v1	Roach et al., 2015
<i>Bacteroides fragilis</i> 3397 T10	ETBF	JGCP01000001.1	ASM59840v1	Science, 2013
<i>Bacteroides fragilis</i> 1007-1-F #4	ETBF	JGDK01000001.1	ASM59854v2	Science, 2013
<i>Bacteroides fragilis</i> 3397 T14	ETBF	JGDO01000001.1	ASM59916v2	Science, 2013
<i>Bacteroides fragilis</i> 1007-1-F #9	ETBF	JGEC01000001.1	ASM59888v1	Science, 2013
<i>Bacteroides fragilis</i> 1007-1-F #8	ETBF	JGCM01000001.1	ASM59826v1	Science, 2013
<i>Bacteroides fragilis</i> 3774 T13	ETBF	JGCR01000001.1	ASM59830v1	Science, 2013
<i>Bacteroides fragilis</i> 3783N1-2	ETBF	JGCS01000001.1	ASM59832v1	Science, 2013
<i>Bacteroides fragilis</i> 3976 T7	ETBF	JGCU01000001.1	ASM59816v1	Science, 2013
<i>Bacteroides fragilis</i> 3986T(B)10	ETBF	JGCV01000001.1	ASM59818v2	Science, 2013
<i>Bacteroides fragilis</i> 3986 N(B)19	ETBF	JGCW01000001.1	ASM59844v1	Science, 2013
<i>Bacteroides fragilis</i> 3988T(B)14	ETBF	JGCY01000001.1	ASM59836v1	Science, 2013
<i>Bacteroides fragilis</i> 3988 T1	ETBF	JGCZ01000001.1	ASM59820v1	Science, 2013
<i>Bacteroides fragilis</i> DS-166	ETBF	JGDD01000001.1	ASM59824v1	Science, 2013
<i>Bacteroides fragilis</i> S36L11	ETBF	JGDJ01000001.1	ASM59912v1	Science, 2013
<i>Bacteroides fragilis</i> 2-F-2 #4	ETBF	JGDM01000001.1	ASM59882v1	Science, 2013
<i>Bacteroides fragilis</i> 3719A10	ETBF	JGDP01000001.1	ASM59884v1	Science, 2013
<i>Bacteroides fragilis</i> 3783N1-8	ETBF	JGDR01000001.1	ASM59860v1	Science, 2013
<i>Bacteroides fragilis</i> J38-1	ETBF	JGDV01000001.1	ASM59864v1	Science, 2013
<i>Bacteroides fragilis</i> S6L3	ETBF	JGDY01000001.1	ASM59922v1	Science, 2013
<i>Bacteroides fragilis</i> S6R6	ETBF	JGDZ01000001.1	ASM59924v1	Science, 2013
<i>Bacteroides fragilis</i> 1009-4-F #10	ETBF	JGED01000001.1	ASM59870v1	Science, 2013
<i>Bacteroides fragilis</i> 1009-4-F #7	ETBF	JGEE01000001.1	ASM59928v2	Science, 2013
<i>Bacteroides fragilis</i> 3397 N3	ETBF	JGEG01000001.1	ASM59892v1	Science, 2013
<i>Bacteroides fragilis</i> 3986 N(B)22	ETBF	JGEI01000001.1	ASM59894v1	Science, 2013
<i>Bacteroides fragilis</i> 3986 T(B)13	ETBF	JGEJ01000001.1	ASM59896v1	Science, 2013
<i>Bacteroides fragilis</i> S36L12	ETBF	JGEP01000001.1	ASM59934v1	Science, 2013
<i>Bacteroides fragilis</i> S36L5	ETBF	JGEQ01000001.1	ASM59902v1	Science, 2013
<i>Bacteroides fragilis</i> S38L3	ETBF	JGEV01000001.1	ASM59876v2	Science, 2013
<i>Bacteroides fragilis</i> I1345	ETBF	JGEW01000001.1	ASM59878v2	Science, 2013
<i>Bacteroides fragilis</i> 3986 N3	ETBF	JGVG01000001.1	ASM60111v1	Science, 2013
<i>Bacteroides fragilis</i> 3725 D9	ETBF	JNHH01000001.1	ASM69968v1	Science, 2013
<i>Bacteroides fragilis</i> 2-F-2 #5	ETBF	JGCN01000001.1	ASM59828v1	Science, 2013
<i>Bacteroides fragilis</i> 34-F-2 #13	ETBF	JGCQ01000001.1	ASM59842v1	Science, 2013
<i>Bacteroides fragilis</i> 3783N2-1	ETBF	JGCT01000001.1	ASM59834v1	Science, 2013
<i>Bacteroides fragilis</i> 3986 T(B)9	ETBF	JGCX01000001.1	ASM59846v1	Science, 2013

(Continued)

TABLE 1 | Continued

Common name	Clinical source	GenBank accession no.	Assembly number	References
<i>Bacteroides fragilis</i> 3996 N(B)6	ETBF	JGDA01000001.1	ASM59822v1	Science, 2013
<i>Bacteroides fragilis</i> 3998T(B)3	ETBF	JGDB01000001.1	ASM59848v1	Science, 2013
<i>Bacteroides fragilis</i> 3998 T(B)4	ETBF	JGDC01000001.1	ASM59838v1	Science, 2013
<i>Bacteroides fragilis</i> DS-208	ETBF	JGDE01000001.1	ASM59850v1	Science, 2013
<i>Bacteroides fragilis</i> DS-71	ETBF	JGDF01000001.1	ASM59908v1	Science, 2013
<i>Bacteroides fragilis</i> Ds-233	ETBF	JGDG01000001.1	ASM59880v1	Science, 2013
<i>Bacteroides fragilis</i> J-143-4	ETBF	JGDH01000001.1	ASM59852v1	Science, 2013
<i>Bacteroides fragilis</i> S13 L11	ETBF	JGDI01000001.1	ASM59910v1	Science, 2013
<i>Bacteroides fragilis</i> 1007-1-F #7	ETBF	JGDL01000001.1	ASM59914v2	Science, 2013
<i>Bacteroides fragilis</i> 3397 N2	ETBF	JGDN01000001.1	ASM59856v1	Science, 2013
<i>Bacteroides fragilis</i> 3725 D9(v)	ETBF	JGDQ01000001.1	ASM59858v1	Science, 2013
<i>Bacteroides fragilis</i> 3976 T8	ETBF	JGDS01000001.1	ASM59918v2	Science, 2013
<i>Bacteroides fragilis</i> 3-F-2	ETBF	JGDT01000001.1	ASM59886v1	Science, 2013
<i>Bacteroides fragilis</i> B1 (UDC16-1)	ETBF	JGDU01000001.1	ASM59862v1	Science, 2013
<i>Bacteroides fragilis</i> Korea 419	ETBF	JGDW01000001.1	ASM59920v1	Science, 2013
<i>Bacteroides fragilis</i> S23 R14	ETBF	JGDX01000001.1	ASM59866v1	Science, 2013
<i>Bacteroides fragilis</i> 1007-1-F #10	ETBF	JGEA01000001.1	ASM59868v2	Science, 2013
<i>Bacteroides fragilis</i> 1007-1-F #3	ETBF	JGEB01000001.1	ASM59926v1	Science, 2013
<i>Bacteroides fragilis</i> 20793-3	ETBF	JGEF01000001.1	ASM59890v1	Science, 2013
<i>Bacteroides fragilis</i> 3719 T6	ETBF	JGEH01000001.1	ASM59872v1	Science, 2013
<i>Bacteroides fragilis</i> A7 UDC12-2	ETBF	JG EK01000001.1	ASM59898v1	Science, 2013
<i>Bacteroides fragilis</i> S23L24	ETBF	JGEL01000001.1	ASM59930v1	Science, 2013
<i>Bacteroides fragilis</i> S24L15	ETBF	JGEM01000001.1	ASM59900v1	Science, 2013
<i>Bacteroides fragilis</i> S24L26	ETBF	JGEN01000001.1	ASM59874v1	Science, 2013
<i>Bacteroides fragilis</i> S24L34	ETBF	JGEO01000001.1	ASM59932v1	Science, 2013
<i>Bacteroides fragilis</i> S38L5	ETBF	JGER01000001.1	ASM59936v1	Science, 2013
<i>Bacteroides fragilis</i> S6L8	ETBF	JGES01000001.1	ASM59938v1	Science, 2013
<i>Bacteroides fragilis</i> S6R5	ETBF	JGET01000001.1	ASM59904v1	Science, 2013
<i>Bacteroides fragilis</i> 3783N1-6	ETBF	JGEU01000001.1	ASM59906v2	Science, 2013
<i>Bacteroides fragilis</i> S6L5	ETBF	JGVC01000001.1	ASM60101v1	Science, 2013
<i>Bacteroides fragilis</i> 1007-1-F #5	ETBF	JGVD01000001.1	ASM60103v1	Science, 2013
<i>Bacteroides fragilis</i> S6R8	ETBF	JGVE01000001.1	ASM60107v2	Science, 2013
<i>Bacteroides fragilis</i> 1007-1-F #6	ETBF	JGVF01000001.1	ASM60109v1	Science, 2013
<i>Bacteroides fragilis</i> S23L17	ETBF	JHEF01000001.1	ASM60105v1	Science, 2013
<i>Bacteroides fragilis</i> 86-5443-2-2	ETBF	LIDS01000001.1	ASM169988v1	Pierce and Bernstein, 2016
<i>Bacteroides fragilis</i> 20656-2-1	ETBF	CM004522.1	ASM169987v1	Pierce and Bernstein, 2016
<i>Bacteroides fragilis</i> 2-078382-3	ETBF	CM004523.1	ASM169986v1	Pierce and Bernstein, 2016
<i>Bacteroides fragilis</i> BOB25	ETBF	CP011073.1	ASM96578v1	Nikitina et al., 2015
<i>Bacteroides fragilis</i> 2-F-2 #7	ETBR	JGCO01000001.1	ASM59814v1	Science, 2013
<i>Bacteroides fragilis</i> KLE1758	Feces	KQ971009.1	ASM158009v1	HMP <sup>a</sup>
<i>Bacteroides fragilis</i> CL03T12C07	Feces	JH724181.1	Bact_frag_CL03T12C07_V1	HMP
<i>Bacteroides fragilis</i> CL05T12C13	Feces	JH724193.1	Bact_frag_CL05T12C13_V1	HMP
<i>Bacteroides fragilis</i> JIM10	Feces	CM004507.1	ASM169269v1	Russia <sup>b</sup>
<i>Bacteroides fragilis</i> CL07T12C05	HMP	JH724215.1	Bact_frag_CL07T12C05_V1	HMP
<i>Bacteroides fragilis</i> 3_1_12	HMP	EQ973245.1	ASM15701v1	HMP
<i>Bacteroides fragilis</i> CL03T00C08	HMP	JH724173.1	Bact_frag_CL03T00C08_V1	HMP
<i>Bacteroides fragilis</i> CL05T00C42	HMP	JH724188.1	Bact_frag_CL05T00C42_V1	HMP
<i>Bacteroides fragilis</i> CL07T00C01	HMP	JH724206.1	Bact_frag_CL07T00C01_V1	HMP
<i>Bacteroides fragilis</i> JCM 11017	Japan	BAIY01000098.1	ASM61342v1	NBRP,Japan <sup>c</sup>
<i>Bacteroides fragilis</i> 4g8B_assembly,	Undernourished Malawian Children	CDQP01000001.1	4g8B_assembly	Kau et al., 2015
<i>Bacteroides fragilis</i> 2d2A_assembly	Undernourished Malawian Children	CDQM01000001.1	2d2A_assembly	Kau et al., 2015

(Continued)

TABLE 1 | Continued

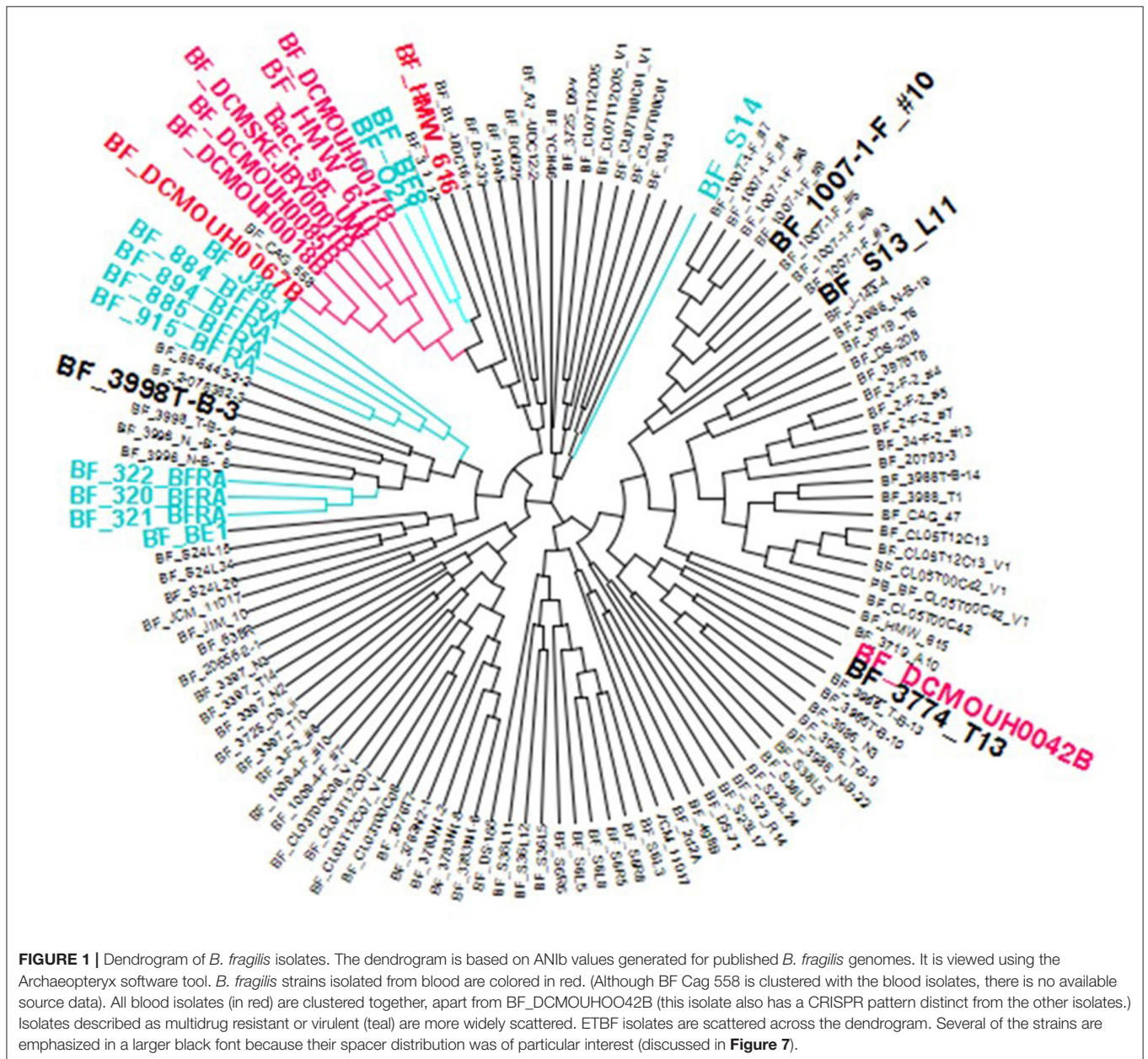
Common name	Clinical source	GenBank accession no.	Assembly number	References
<i>Bacteroides fragilis</i> BF8	ViR/MDR	LGTH01000001.1	ASM169535v1	Soki et al., 2016
<i>Bacteroides fragilis</i> S14	ViR/MDR	CP012706.1	ASM168221v1	Soki <sup>d</sup>
<i>Bacteroides fragilis</i> 322_BFRA	Fluid	JVLP01000024.1	ASM105489v1	Roach et al., 2015
<i>Bacteroides fragilis</i> 321_BFRA	Fluid	JVLQ01000008.1	ASM105633v1	Roach et al., 2015
<i>Bacteroides fragilis</i> 320_BFRA	Fluid	JVLR01000007.1	ASM105486v1	Roach et al., 2015
<i>Bacteroides fragilis</i> O:21	ViR/MDR	KV751174.1	ASM169369v1	Soki et al., 2016
<i>Bacteroides fragilis</i> BFBE1.1	ViR/MDR	LN877293.1	BFBE1.1	Risse et al., 2015

<sup>a</sup>Reference Genome for the Human Microbiome Project (HMP).

<sup>b</sup>Federal Research and Clinical Center of Physical-Chemical Medicine of FMBA of Russia.

<sup>c</sup>National Bioresearch Project, Japan.

<sup>d</sup>Soki et al. (unpublished), Comparative analysis of Division I and II *Bacteroides fragilis* strain genomes.



## Identification of *cas* Genes Associated with Particular CRISPR Repeats

Genes flanking specific CRISPR loci were identified by BLAST analysis of the entire CRISPR sequence against the host genome using RAST-based annotation (Aziz et al., 2008). We have found that the RAST annotation server frequently offered a more robust annotation than GenBank and in this study it afforded an easier way to localize the CRISPR element and to view the adjacent genes. Therefore, all bacterial genomes were uploaded to the RAST server and the adjacent *cas* genes as well as the genomic context of the particular CRISPR repeat array were determined. In some cases, the CRISPR was at the end of a gene scaffold, and the upstream genes could not be identified. The particular *cas* operon identified was assigned to a CRISPR-Cas classification system developed by Makarova (Makarova et al., 2011). Most of the *B. fragilis* genome sequences have not been completely assembled, and sometimes the CRISPR repeat sequence was located near the edge of a contig; in those cases, the adjacent genes could not be identified. In those cases, the presence or absence of *cas* genes somewhere else on the genome was noted. Automatic assembly of contigs for genome assembly is often complicated by the presence of transposase genes, and indeed, we found transposase genes adjacent to the CRISPR repeat element in many instances.

## Other Bioinformatic Tools Used in the Analysis of *B. fragilis* CRISPRs

Consensus sequences and images for the DR sequences were obtained using WEBLOGO (Crooks et al., 2004). Predicted structures of the DR repeat sequence were visualized using the RNA fold program (Ding et al., 2005). Phylogenetic relationships of the consensus DR sequences in the DR database were analyzed using CRISPRmap (Lange et al., 2013). Gene neighborhoods were visualized using tools at the Joint Genome Institute (Markowitz et al., 2012). Venn diagrams were generated using <http://bioinformatics.psb.ugent.be/webtools/Venn/>.

## RESULTS

### All *B. fragilis* Strains with Publically Available Genome Sequences Were Analyzed for CRISPR Repeat Arrays

Deep sequence analysis of the CRISPR system was performed for all of the publicly available genome sequences of *B. fragilis* ( $n = 109$ ) (Table 1). The proportion of “pathogenic” vs. “commensal” strains of *B. fragilis* is artificially inflated, since strains from serious infections are more likely to be chosen for whole genome sequencing. A large proportion of these strains are from studies of enterotoxigenic *B. fragilis* (ETBF), so these strains were disproportionately represented and consequently analyzed. Conversely, only a few strains from “normal” feces or human microbiome projects (where samples are often from healthy patients) have been sequenced.

The *B. fragilis* strains included in our study include several blood isolates from geographically distant locations including Washington State (Schapiro et al., 2004), Seattle (Kalapila

et al., 2013), Denmark (Ank et al., 2015), the United Kingdom (Pumbwe et al., 2007b), and Afghanistan (Sherwood et al., 2011). A previous phylogenetic analysis of several clinical *B. fragilis* isolates indicated that there is a cluster of *B. fragilis* strains that constitute a genomospecies distinct from other *B. fragilis* (Salipante et al., 2015). Two of these isolates, BF HMW 610 and BF HMW 616, were reported and characterized extensively in our laboratory (Pumbwe et al., 2007a,b; Sherwood et al., 2011). The isolates noted as virulent and multidrug resistant (VIR/MDR) include an international cluster of MDR *B. fragilis* isolates from five countries (Soki et al., 2016). Whole genome sequencing of a large number of ETBF isolates has been undertaken and these are also included (Science, 2013). References, when available, for all of the isolates analyzed are listed in Table 1, and include more details about the origin and nature of these isolates.

### Phylogenetic Grouping of *B. fragilis* Strains

A dendrogram based on Average Nucleotide Identity by BLAST (ANIb) values is shown in Figure 1. All but one of the strains isolated from blood cluster together in one group (despite being isolated in geographically distant areas.) Strains associated with enterotoxin related illnesses formed the majority of the strains and were evenly distributed across the tree as were “commensal” strains isolated as part of the human microbiome project.

### CRISPR-Cas Systems in *B. fragilis*

Three types of CRISPR-Cas systems (and a fourth array without associated *cas* genes) are variably distributed in *B. fragilis* strains (Table 2) and most closely match types Class 1 Type IB, Class 1 Type IIIB, and Class 2 Type IIC (Makarova et al., 2015; Figure 2); each was in a highly conserved gene neighborhood (Figure 3). One hundred strains of *B. fragilis* from those studied had one or more CRISPR-Cas systems; 9 strains had no discernable CRISPRs (Figure 4). If the Orphan CRISPR-Cas system was excluded from the analysis, 84 strains of *B. fragilis* had Type IB, Type IIIB, or Type IIC CRISPR systems. Three strains (S38L3, S38L5, and S14) harbored all three types. In most bacteria, CRISPR systems are restricted to one or possibly two types but occasional strains harboring three types of CRISPR systems were seen in *Streptococcus* and *Clostridium* species (Louwen et al., 2014). In *Streptococcus thermophiles*, for example, three different major CRISPR-Cas systems are present and function independently in crRNA biogenesis (Carte et al., 2014).

Our final data set of *B. fragilis* CRISPR-Cas arrays, reflecting manual curation, is listed in Supplementary Table 1. The CRISPR repeat arrays are written with the oldest spacer at the top of the array; therefore the adjacent *cas* genes would be located proximal to the newest spacer, at the bottom of the array. The length and sequence of repeats and the length of spacers are generally well conserved within a CRISPR locus, but may vary between CRISPRs in the same or different genomes (Rath et al., 2015); CRISPR sequences may vary in both DR and spacer sequences even among strains more than 99% at the DNA level (Kunin et al., 2007).

**TABLE 2** | Distribution of CRISPR-Cas Systems in *B. fragilis*.

Type IB	Adjacent cas genes	Length <sup>a</sup>	Type IIIB	Adjacent cas genes	RT-cas <sup>b</sup>	Length <sup>a</sup>	Type IIC	Adjacent cas genes	Length <sup>a</sup>	Orphan
Consensus length DR: 29			Consensus length DR: 35			Consensus length DR: 47			Consensus length DR: 24	
Average length spacer: 37			Average length spacer: 35			Average length spacer: 29			Average length spacer: 39	
<i>Bacteroides</i> sp. UW	IB	25	<i>Bacteroides</i> sp. UW	Alternate <sup>f</sup>	No	5	BF 1009 4 F 10	IIC	31,8,14	BF 1007 1 F 10
BF 1007 1 F 10	IB	15	BF 3976t7	ND <sup>c</sup>		5	BF 1009 4 F 7	IIC	8,7,14,31	BF 1007 1 F 3
BF 1007 1 F 3	IB	14	BF 3976t8	IIIB		5	BF 2 078382 3	IIC	19	BF 1007 1 F 4
BF 1007 1 F 4	IB	14	BF 1007 1 F 10	IIIB		15	BF 2 F 2 4	IIC	6	BF 1007 1 F 5
BF 1007 1 F 5	IB	14	BF 1007 1 F 3	IIIB		15	BF 2 F 2 5	IIC	4,5	BF 1007 1 F 6
BF 1007 1 F 6	IB	14	BF 1007 1 F 4	IIIB		15	BF 20656 2 1	IIC	15	BF 1007 1 F 7
BF 1007 1 F 7	IB	14	BF 1007 1 F 5	IIIB		15	BF 3 F 2	IIC	15,15,15	BF 1007 1 F 8
BF 1007 1 F 8	IB	6,5,(12)	BF 1007 1 F 6	IIIB		15	BF 320	IIC	5	BF 1007 1 F 9
BF 1007 1 F 9	IB	14	BF 1007 1 F 7	IIIB		15	BF 321	IIC	7,6	BF 1009 4 F 10
BF 1009 4 F 10	Truncated IB	9	BF 1007 1 F 8	ND <sup>c</sup>		15	BF 322	IIC	5	BF 1009 4 F 7
BF 1009 4 F 7	Truncated IB	9	BF 1007 1 F 9	IIIB		15	BF 34 F 2 13	IIC	6	BF 2 F 2 4
BF 3 F 2 #6	Truncated IB	8	BF 3397 N2	IIIB		34	BF 3719 T6	IIC	15	BF 2 F 2 5
BF 320	IB	20	BF 3397 N3	IIIB		35	BF 3725 D9 ii	IIC	32	BF 2 F 2 7
BF 321	IB	20	BF 3397 T10	ND <sup>c</sup>		28	BF 3774 T13	IIC	14	BF 20793 3
BF 322	IB	20	BF 3397 T14	IIIB		35	BF 3783N1 2	IIC	13	BF 3 1 12
BF 3998T B 3	Truncated IB	20,1b, inferred	BF 3719A10	IIIB		31	BF 3783N1 6	IIC	13	BF 3 F 2 #6
BF 3998T B 4	ND <sup>a</sup>	3,1	BF 3774 T13	ND <sup>c</sup>		11	BF 3783N1 8	IIC	13	BF 3397 N2
BF DCMOUH0017B	Odd genes <sup>d</sup>	67	BF 3783N1 2	ND <sup>c</sup>		4	BF 3783N2 1	ND <sup>a</sup>	13	BF 3397 T14
BF DCMSKEJBY0001B	IB	25	BF 3783N1 6	IIIB		6	BF 3976T7	ND <sup>a</sup>	13	BF 34 F 2 13
BF HMW 610	IB	8,2 <sup>a</sup>	BF 3783N1 8	IIIB		8	BF 3986 N B 19	IIC	14	BF 3774 T13
BF KLE1758	Truncated IB	4,2	BF 3783N2 1	ND <sup>c</sup>		5	BF 3996 N B 6	IIC	9,5,3	BF 3783N1 6
BF Korea 419	Truncated IB	8	BF 3986 N B 22	IIIB		7	BF 3998T B 4	IIC	5,5	BF 3783N1 8
BF NCTC 9343	Truncated IB	8	BF 3986 N3	IIIB		15	BF 3998T B 3	IIC	5	BF 3783N2 1
BF S13 L11	ND <sup>a</sup>	14	BF 3986T B 13	IIIB		7	BF 638R	IIC	29	BF 3976T7
BF S14	Truncated IB	8	BF 3986T B 9	IIIB		7	BF 86 5443 2 2	IIC	9	BF 3986 N B 22
BF S38L5	Truncated IB	8	BF 3986T B 10	ND <sup>c</sup>		4	BF BE1 1	IIC	11	BF 3986 N3
BF YCH46	IB	8	BF 3988 T1	ND <sup>c</sup>		7	BF DCMOUH0042B	IIC	4	BF 3986T B 13
BF s38L3	Truncated IB	8	BF 3988T B 14	ND <sup>c</sup>		4,24	BF DS 208	IIC	9	BF 3986T B 9
BF 3996 NB6	IB	20	BF 884	ND <sup>c</sup>		8	BF DS 71	IIC	19	BF 3986T B 10
			BF 885	ND <sup>c</sup>		17	BF KLE1758	IIC	3	BF 3988 T1
			BF 894	IIIB		17	BF NCTC 9343	IIC	26	BF 3988T B 14
			DCMOUH0018B	IIIB		4,7	BF S14	IIC	24	BF A7 UDC12 2

(Continued)



TABLE 2 | Continued

Type IB	Adjacent cas genes	Length <sup>a</sup>	Type IIIB	Adjacent cas genes	RT-cas <sup>b</sup>	Length <sup>a</sup>	Type IIC	Adjacent cas genes	Length <sup>a</sup>	Orphan
Consensus length DR: 29			Consensus length DR: 35			Consensus length DR: 47			Consensus length DR: 24	
Average length spacer: 37			Average length spacer: 35			Average length spacer: 29			Average length spacer: 39	
			BF DCMOUH0042B	IIIB		17	BF S23 R14	IIC	9,10,4	BF B1 UDC16 1
			BF DCMOUH0067B	Alternate <sup>f</sup>	No	4	BF S23L17	IIC	14	BF BE1 1
			BF-DCMSKEJBY0001B	Alternate <sup>f</sup>	No	5	BF S23L24	IIC	14	BF BF8
			BF DS 166	ND <sup>c</sup>		12,5	BF S24L15	IIC	8	BF BOB25
			BF DS 71	ND <sup>c</sup>		1	BF S24L26	IIC	8	BF CL03T00C08
			BF HMW 610	Alternate <sup>f</sup>	No	2	BF S24L34	IIC	8	BF CL03T12C07
			BF J38 1	ND <sup>c</sup>		17	BF S38L3	IIC	19	BF CL07T00C01
			BF Korea 419	IIIB		15	BF S38L5	IIC	19	BF CL07T12C05
			BF S13 L11	ND <sup>c</sup>		5,9				BF DCMOUH0018B
			BF S14	IIIB		22				BF DCMOUH0042B
			BF S23 R14	ND <sup>c</sup>		21				BF DS 166
			BF S23L17	IIIB		21				BF DS 71
			BF S23L24	IIIB		21				BF HMW 616
			BF S36L11	ND <sup>c</sup>		24,3				BF I1345
			BF S36L12	IIIB		24,3				BF JCM 11017
			BF S36L5	IIIB		23,3				BF JIM10
			BF S38L3	IIIB		29				BF KLE1758
			BF S38L5	ND <sup>c</sup>		29				BF Korea 419
			BF S6L3	Alternate <sup>f</sup>	YES	1,3				BF NCTC 9343
			BF S6L5	ND <sup>c</sup>		1,3				BF O:21
			BF S6L8	ND <sup>c</sup>		2,4				BF S14
			BF S6R5	ND <sup>c</sup>		1,3				BF S23L17
			BF s6R6	IIIB		1,3				BF S23L24
			BF s6R8	ND <sup>c</sup>		1,3				BF S23 R14
			BF-S6L5	Alternate <sup>f</sup>	YES					BF S24L15
			BF-S6L8	Alternate <sup>f</sup>	YES					BF S24L26
			BF-S6R5	Alternate <sup>f</sup>	YES					BF S24L34
			BF-S6R6	Alternate <sup>f</sup>	YES					BF S36L11
			BF-S6R8	Alternate <sup>f</sup>	YES					BF S36L12
			BF-S36L11	Alternate <sup>f</sup>	YES					BF S36L5
			BF-S36L12	Alternate <sup>f</sup>	YES					BF S38L3
			BF-S36L5	Alternate <sup>f</sup>	YES					BF S38L5

<sup>a</sup>Because many of the genomes are not yet assembled, the same CRISPR array may have been identified in different contigs, the length refers to the number of spacers in a particular CRISPR array.

<sup>b</sup>RT-cas: Gene coding for Reverse-transcriptase Cas1 fusion protein.

<sup>c</sup>ND: The contig on which the repeat array was found was very short and no adjacent cas genes could be identified.

<sup>d</sup>Has cas2, cas1, cas 4a and cas 7 only; missing the effector cas genes.

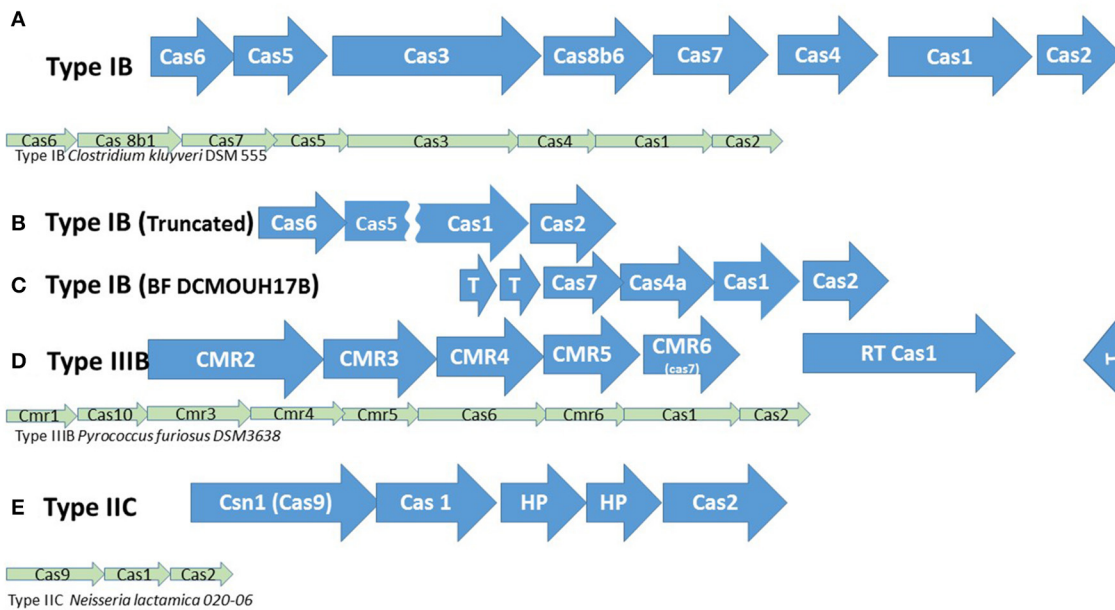
<sup>e</sup>BF HMW 610 has large segments of N's in the midst of what appears to be a long, continuous repeat array.

<sup>f</sup>Alternate gene neighborhood in the midst of polysaccharide and other metabolic genes.

## Class 1 Type IB CRISP-Cas Systems

The closest match for the Type I Cas system is the canonic Type IB found in *Clostridium kluveri* (Makarova et al., 2015;

Figure 2A) although the gene order is different. Using the traditional annotation servers, we found genes coding for Cas2, Cas1, Cas3, Cas4a, and Cas6 (TM1814). Three additional



**FIGURE 2** | CRISPR-Cas systems in *B. fragilis*. Cas genes found in the respective systems are listed. The canonic arrangement of the closest matching type according to Makarova et al. (2015) is represented by the smaller green arrows below the *B. fragilis* cas operon. *T*, Transposase gene. **(A) Type IB**: The traditional annotation servers identified genes coding for *cas2*, *cas1*, *cas3*, *cas4a*, and *cas6* (TM1814). Three additional genes, classified by all the publicly available annotation sites as hypothetical proteins, have been classified as *cas5*, *cas7*, and *cas8b6* genes by Makarova et al. (2015); their sequences are very divergent. **(B) Type IB, truncated**: These truncated CRISPR-Cas systems were located in the same neighborhood as the complete Type IB CRISPR-Cas systems. The *cas* genes include: *cas2*, truncated *cas1* (252 vs. 1014 bp), truncated *cas5* (381 vs. 564 bp) and *cas6*. The *cas4*, *cas7*, *cas8b6*, and *cas3* genes were missing in those strains, as were the 5' and 3' ends of *cas1* and *cas5*, respectively. They occur in strains BF 1009 4 F 10, BF 1009 4 F 7, BF 3 F 2-6, BF 3998T B 3, BF KLE1758, BF Korea 419, BF NCTC 9343, BF S14, BF S38L5 and BF s38L3. **(C) Type IB, BF DCMOUH17B**: This array has only some of the genes in the typical Type IB systems. **(D) Type IIIB**: The BF *cas1* gene in the Type IIIB system codes for a reverse transcriptase-Cas1 fusion protein. The other genes present, *cmr2-6* are typical of Type IIIB CRISPR-Cas systems, although in the canonic operon, the order is somewhat different. **(E) Type IIC**: The BF Type IIC system includes the canonic *cas2*, *cas1* and *csn1* (*cas9*) genes. Two additional genes coding for hypothetical proteins of unknown function are situated between *cas2* and *cas1*.

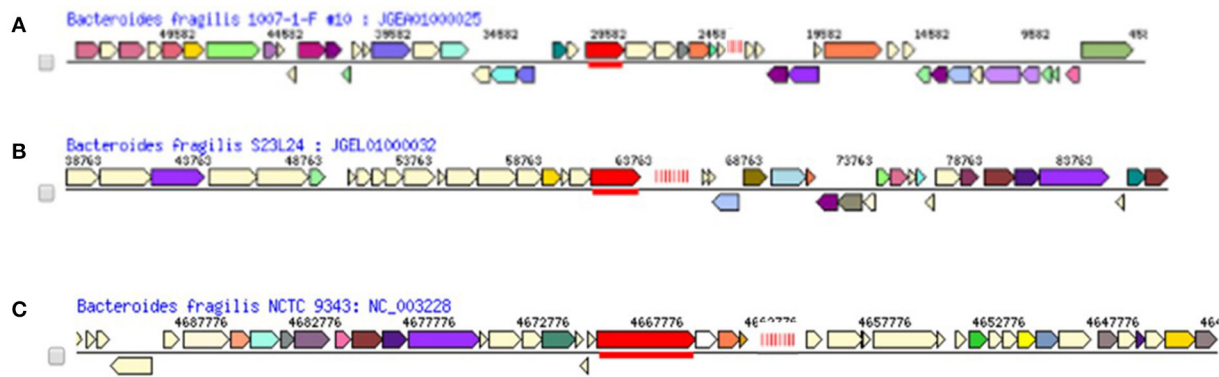
genes, classified by all the publicly available annotation sites as hypothetical proteins, were classified as *cas5*, *cas7*, and *cas8b6* genes by Makarova et al. (2015) (Figure 2A); their sequences are very divergent (K. Makarova, personal communication). We did find annotated genes coding for Cas5 and Cas8 in *Bacteroides oleiciplenus* YIT 12058, but their sequences did not have any homologs in *B. fragilis* isolates.

The Type IB *cas1* gene is highly conserved among *B. fragilis* strains. It is otherwise most closely aligned with *cas1* genes from other anaerobes, including *Parabacteroides* sp., *Finexgoldia magna*, *Clostridium tetani*, and *Clostridioides difficile* (data not shown). The Type IB *cas3* gene is also highly conserved among *B. fragilis* strains. Its closest matches were in other *Prevotella*, *Porphyromonas*, *Chitinophaga*, and *Spirosoma* species. It was highly divergent from other *cas3* genes in the model organism GenBank databank; its closest match was with *Clostridioides difficile* (score 126, E-value: 1 e-28).

Twenty-nine strains of *B. fragilis* had Type IB CRISPR-Cas systems and were found in a highly conserved gene neighborhood (Figure 3A). These CRISPR-Cas systems could be divided into two main groups. In the first group, either a full set of Type IB *cas* operon genes could be identified upstream of the CRISPR repeat, or a portion of the operon was seen but was cut off because it was at the end of the sequenced contig;

in those cases, the presence of a full TYPE IB *cas* operon was inferred. The second group had CRISPR repeats with 8 or fewer spacers and only a few *cas* genes (*cas2*, truncated *cas1* (252 vs. 1014 bp), truncated *cas5* (381 vs. 564 bp) and *cas6*) were adjacent to the CRISPR repeat. The *cas4*, *cas7*, *cas8b6*, and *cas3* genes were missing in those strains, as were the 5' and 3' ends of *cas1* and *cas5*, respectively. The deletion of the *cas* genes, including *cas3*, and the truncation of the *cas1* and *cas5* probably occurred as a single deletion event, presumably after these spacers had been acquired, since the absence of the acquisition module probably crippled the ability of the strain to assimilate new spacer sequences. These truncated CRISPR-Cas systems, occurring in strains scattered across the dendrogram, were located in the same neighborhood as the complete Type IB CRISPR-Cas systems. All of the strains with the truncated CRISPR-Cas IB system have either Type IIIB or Type IIC systems, or both, so it is possible that the associated CRISPR-arrays could still use proteins encoded by those genes, despite their lack of the full set of Type IB *cas* genes.

Three blood isolates (*Bacteroides* sp. UW, BF DCMSKEJBY0001B and BF HMW610) had typical Type IB CRISPR-Cas systems (with adjacent *cas* genes including *cas3*) with relatively long repeat arrays (the array for HMW 610 was split into several segments because of N's in the sequencing results [data not shown]). The predicted Cas3 protein sequence



**FIGURE 3 |** Conserved gene neighborhoods of *B. fragilis* CRISPR-Cas arrays. Proteins coded for by genes are listed below. Vertical red lines represent repeat array. Associated *cas* genes products are in bold. Representative strain is shown with genes surrounding CRISPR-Cas array. The neighborhoods are highly conserved within each CRISPR-Cas type. The most characteristic *cas* gene for each group (i.e., A: *cas 3*, B: *RT-cas 1*-fusion and C: *cas 9*) are underlined.

**(A)** Conserved gene neighborhood of Type IB CRISPR-Cas system in *B. fragilis*. **Upstream genes:** Hypothetical protein;hypothetical protein;Manganese transport protein MntH;Exodeoxyribonuclease III; hypothetical protein; hypothetical protein; hypothetical protein; Translation elongation factor LepA;putative Na<sup>+</sup>/H<sup>+</sup> exchange protein;Na<sup>+</sup>/H<sup>+</sup> antiporter NhaA type;hypothetical protein;DNA recombination protein RmuC; Methionine aminopeptidase; **CRISPR-associated protein, TM1814 family (*cas 6*); hypothetical protein (*cas 5*);CRISPR-associated helicase *cas 3* (underlined); hypothetical protein; hypothetical protein;CRISPR-associated RecB family exonuclease *cas 4a*;CRISPR-associated protein *cas 1*;CRISPR-associated protein *cas 2*; **CRISPR REPEAT ARRAY**; **Downstream genes:** Hypothetical protein;hypothetical protein;RNA pseudouridylate synthase BT0642; RNA methyltransferase, TrmA family;hypothetical protein;Pyruvate,phosphate dikinase;hypothetical protein; hypothetical protein;Thiamin-phosphate pyrophosphorylase;Sulfur carrier protein adenylyltransferase ThiF;Thiazole biosynthesis protein ThiH; hypothetical protein;Thiamin biosynthesis protein ThiC;Thiazole biosynthesis protein ThiG;Thiamin-phosphate pyrophosphorylase; Sulfur carrier protein ThiS; Superoxide dismutase [Fe];ATP-dependent DNA helicase UvrD/PcrA;Carboxynorspermidine decarboxylase, putative;hypothetical protein;**

**(B)** Conserved gene Neighborhood of Type IIIB CRISPR-Cas system in *B. fragilis*. **Upstream genes:** Two-component system sensor histidine kinase; Two-component system response regulator; Outer membrane protein assembly factor YaeT precursor; ABC transporter, permease; Probable ABC transporter permease; putative ABC transporter permease; putative ABC transporter permease; ABC transporter, permease protein; ABC transporter, permease protein; ABC transporter ATP-binding protein YvcR; Thiol:disulfide interchange protein; *M. jannaschii* predicted coding region MJ0978; protein of unknown function DUF88; hypothetical protein; hypothetical protein; hypothetical protein; hypothetical protein; **CRISPR-associated RAMP Cmr2; CRISPR-associated RAMP Cmr3; CRISPR-associated RAMP Cmr4; CRISPR-associated RAMP Cmr5; CRISPR-associated RAMP Cmr6; Retron-type RNA-directed DNA polymerase RT-Cas1 fusion protein (underlined). CRISPR Repeat Array. Downstream genes:** ISNCY family transposase; ISNCY family transposase; Aminotransferase class II, serine palmitoyltransferase like; Transcription regulator [contains diacylglycerol kinase catalytic domain]; Aspartyl-tRNA synthetase; hypothetical protein; N-carbamoylputrescine amidase/Aliphatic amidase AmiE Argmatine deiminase;Ferredoxin domain containing protein; YbbL ABC transporter ATP-binding protein; YbbM seven transmembrane helix protein; hypothetical protein; Possible glyoxylase family protein (Lactoylglutathione lyase); hypothetical protein; Hypothetical protein YbgI; Hypothetical protein; RND efflux system, outer membrane lipoprotein CmeC; RND efflux system, membrane fusion protein CmeA; RND efflux system, inner membrane transporter CmeB. **In contrast, Gene Neighborhood of TYPE IIIB CRISPR Repeat Array in Blood Isolates (absence of *cas* genes) (not shown): Upstream Genes:** Lipopolysaccharide biosynthesis protein; hypothetical protein;UDP-N-acetylglucosamine 2-epimerase; glycosyl transferase, group 1 family protein; hypothetical protein; mannosyltransferase B; UDP-N-acetylglucosamine 4,6-dehydratase; UDP-N-acetylglucosamine 2-epimerase; dTDP-4-dehydrohannose reductase; Glycosyl transferase, group 1 precursor; UDP-glucose 4-epimerase; Undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase; TYPE IIIB REPEAT ARRAY in BF BLOOD ISOLATES; Downstream genes: Putative non-specific DNA-binding protein; hypothetical protein; Na<sup>+</sup>/H<sup>+</sup>-dicarboxylate symporter; 6-phosphogluconate dehydrogenase, decarboxylating; Glucose-6-phosphate 1-dehydrogenase; 6-phosphogluconolactonase, eukaryotic type; hypothetical protein;

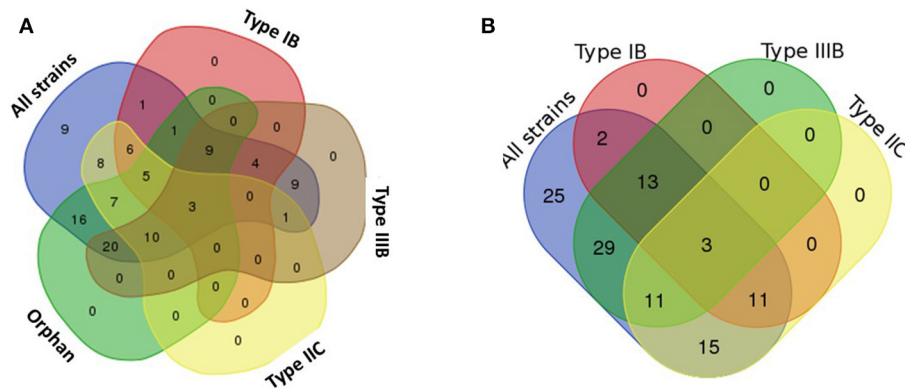
**(C)** Gene Neighborhood of Type IIC CRISPR-Cas system in *B. fragilis*. **Upstream gene neighborhood:** Conserved hypothetical protein;putative transmembrane protein;putative transmembrane DNA mismatch repair-like protein;conserved hypothetical protein;putative urocanate hydratase;putative formimidoyltransferase-cyclodeaminase;putative imidazolonepropionase;putative formiminotransferase-cyclodeaminase;putative histidine ammonia-lyase;putative TetR transcriptional regulator;putative outer membrane efflux protein;putative membrane fusion protein transporter;putative transmembrane Acr-type transport protein;conserved hypothetical protein;putative transmembrane protein;conserved hypothetical protein;putative transmembrane polysaccharide modification protein;hypothetical protein;hypothetical protein;hypothetical protein; ***csn1(cas9)* (underlined);conserved hypothetical protein (pseudogene);cas1;cas2;** **CRISPR REPEAT ARRAY. Downstream:** Putative transmembrane protein;conserved hypothetical protein;putative transmembrane MotA/TolQ/ExbB proton channel family protein;conserved hypothetical protein;conserved hypothetical protein;putative TonB-family outer membrane receptor protein;conserved hypothetical protein;putative TPR-repeat family protein;putative ATP-binding component of ABC transporter;putative transmembrane protein;putative GntR family transcriptional regulator;conserved hypothetical protein;conserved hypothetical protein;putative ABC transporter transmembrane component.

of these three isolates cluster as a separate group in a phylogenetic analysis of *Cas3* in *B. fragilis* as does *Cas 6* and most of the other *Cas* proteins.

### Class 1 Type IIIB CRISPR-Cas Systems

Fifty-six strains had Type IIIB CRISPR-Cas systems. The Type IIIB CRISPR-Cas system had a *cas* operon upstream of the

repeat segment consisting of *Cas1*, *CMR6*, *CMR5*, *CMR4*, *CMR3*, and *CMR2*; a ISNCY (i.e., Not Classified Yet) transposase was frequently located at the opposite end of these CRISPRs (Figure 2B). The *Cas1* gene in Type IIIB CRISPRs was often not annotated as *Cas1* in either RAST or NCBI annotation but as a retron-type DNA polymerase or reverse transcriptase (RT). This *cas1* gene (2271 bp) encodes a 756 aa protein that is more than



**FIGURE 4 |** Venn diagram of distribution of CRISPR-Cas systems among strains of *B. fragilis*. The Venn diagrams in **Figure 4** include CRISPR-Cas repeats of a given type, whether full or truncated; whether or not there was a full set of adjacent *cas* genes is detailed in **Table 2**. **(A) Distribution of all CRISPR arrays in *B. fragilis*.** Orphan Type IB Type IIC Type IIIB-3 Strains: BF S38L3, BF S38L5, BF S14. Orphan Type IB Type IIIB-9 Strains: BF 1007 1 F 4, BF 1007 1 F 3, BF Korea 419, BF 1007 1 F 6, BF 1007 1 F 7, BF 1007 1 F 9, BF 1007 1 F 8, BF 1007 1 F 10, BF 1007 1 F 5. Orphan Type IB Type IIC-5 Strains: BF 1009 4 F 7, BF KLE1758, BF 3-F-2 #6, BF NCTC 9343, BF 1009 4 F 10. Orphan Type IIC Type IIIB-10 Strains: BF 3783N1 8, BF 3783N2 1, BF 3976T7, BF DS 71, BF S23L17, BF S23L24, BF S23 R14, BF 3783N1 6, BF DCMOUH0042B, BF 3774 T13. Type IB Type IIIB-4 Strains: BF DCMSKEJBY0001B, BF S13 L11, BF HMW 610 *Bacteroides* sp. UW. Type IB Type IIC-6 Strains: BF 3996 N B 6, BF 321, BF 322, BF 3998T B 3, BF 320, BF 3998 T B 4. Orphan Type IB-1 Strain: BF YCH46. Type IIC Type IIIB-1 Strain: BF 3783N1 2. Orphan Type IIIB-20 Strains: BF 3986T B 10, BF DCMOUH0018B, BF S6L8, BF S36L12, BF 3986 T B 9, BF S6L3, BF DS 166, BF 3988T B 14, BF S6R8, BF S6L5, BF S36L5, BF 3986 N B 22, BF 3988 T1, BF S6R6, BF 3397 N2, BF S36L11, BF 3397 T14, BF 3986 T B 13, BF S6R5, BF 3986 N3. Orphan Type IIC-7 Strains: BF 34 F 2 13, BF 2 F 2 4, BF 2 F 2 5, BF S24L26, BF S24L34, BF S24L15, BF BE1 1. Type IB-1 Strains: BF DCMOUH0017B. Type IIIB-9 Strains: BF 3397 T10, BF 3976 T8, BF DCMOUH0067B, BF J38 1, BF 894, BF 885, BF 884, BF 3719A10, BF 3397 N3. Type IIC-8 Strains: BF 3719 T6, BF 3725 D9 ii, BF 2 078382 3, BF 20656 2 1, BF DS 208, BF 3986 N B 19, BF 638R, BF 86 5443 2 2. Orphan-16 Strains: BF BOB25, BF A7 UDC12 2, BF CL07T00C01, BF CL03T00C08, BF HMW 616, BF 2 F 2 7, BF 3 1 12, BF CL07T12C05, BF CL03T12C07, BF JCM 11017, BF JIM10, BF, BF8, BF I1345, BF 20793 3, BF B1 UDC16 1, BF O: 21. **No CRISPR-Cas- 9 Strains:** BF CL05T00C42, BF HMW 615, BF DCMOUH0085B, BF Ds 233, BF 3725 D9, BF CL05T12C13, BF J 143 4, BF 4g8B, BF 2d2A. **(B) Distribution of CRISPR-Cas systems excluding orphan CRISPR arrays.** Type IB Type IIC Type IIIB-3 Strains: BF S38L3, BF S38L5, BF S14. Type IB Type IIIB-13 Strains: BF 1007 1 F 4, BF DCMSKEJBY0001B, BF 1007 1 F 3, BF Korea 419, BF S13 L11, BF 1007 1 F 6, BF 1007 1 F 7, BF HMW 610 *Bacteroides* sp. UW, BF 1007 1 F 9, BF 1007 1 F 8, BF 1007 1 F 10, BF 1007 1 F 5. Type IB Type IIC-11 Strains: BF 3996 N B 6, BF 321, BF 1009 4 F 7, BF KLE1758, BF 3-F-2 #6, BF 322, BF 3998T B 3, BF 320, BF NCTC 9343, BF 1009 4 F 10, BF 3998 T B 4. Type IIC Type IIIB-11 Strains: BF 3783N1 8, BF 3783N2 1, BF 3976T7, BF DS 71, BF S23L17, BF 3783N1 2, BF S23L24, BF S23 R14, BF 3783N1 6, BF DCMOUH0042B, BF 3774 T13. Type IB-2 Strains: BF YCH46, BF DCMOUH0017B. Type IIIB-29 Strains: BF 3397 T10, BF 3986T B 10, BF DCMOUH0018B, BF 3976 T8, BF DCMOUH0067B, BF S6L8, BF J38 1, BF S36L12, BF 3986 T B 9, BF S6L3, BF DS 166, BF 3988T B 14, BF 894, BF S6R8, BF S6L5, BF S36L5, BF 3986 N B 22, BF 885, BF 3988 T1, BF 884, BF S6R6, BF 3397 N2, BF 3719A10, BF S36L11, BF 3397 N3, BF 3397 T14, BF 3986 T B 13, BF S6R5, BF 3986 N3. Type IIC-15 Strains: BF 34 F 2 13, BF 3719 T6, BF 3725 D9 ii, BF 2 078382 3, BF 20656 2 1, BF DS 208, BF 3986 N B 19, BF 2 F 2 4, BF 638R, BF 2 F 2 5, BF S24L26, BF 86 5443 2 2, BF S24L34, BF S24L15, BF BE1 1. **No CRISPR-Cas (with adjacent *cas* genes)-25 Strains:** BF CL05T00C42, BF BOB25, BF A7 UDC12 2, BF CL07T00C01, BF CL03T00C08, BF HMW 615, BF HMW 616, BF 2 F 2 7, BF 3 1 12, BF DCMOUH0085B, BF CL07T12C05, BF CL03T12C07, BF JCM 11017, BF Ds 233, BF JIM10, BF, BF8, BF 3725 D9, BF CL05T12C13, BF J 143 4, BF I1345, BF 4g8B, BF 2d2A, BF 20793 3, BF B1 UDC16 1, BF O: 21.

twice as long as the 243 aa Cas1 protein encoded by the *cas1* gene in the Class 1 Type IB CRISPRs. These proteins contain both an RT-like superfamily domain and a Cas1\_I-II-II superfamily domain (the amino acid sequence of RT-like Cas1, when present, is almost completely conserved among *B. fragilis* strains).

The RT-Cas protein is highly conserved within strains of *B. fragilis*, but very divergent from other RT-Cas 1 proteins. When compared to the landmark organisms (proteomes from 27 genomes spanning a wide taxonomic range) at NCBI, the closest matches were *Clostridiodes difficile* 630, fission yeast and *Sulfolobus acidocaldirius* (data not shown). If a wider search of the NR database was done and *Bacteroides* species eliminated, the closest matches were *Culturomica*, *Candidatus*, *Sulphuromonas*, and *Porphyromonas* species.

The *B. fragilis* Type IIIB CRISPR-Cas systems occur in highly conserved gene neighborhoods in the strains in which they are found (**Figure 3B**); these neighborhoods include a high percentage of genes involved in efflux processes. Frequently, a transposase gene was located just downstream of the CRISPR

repeat region in those strains with adjacent Type IIIB CRISPR-Cas operons. In some cases, the contig ended just downstream of the *RT-cas1* gene; this type of sequencing result is frequently due to the presence of a transposase gene that causes breaks in the automatic scaffolding assembly of genome sequences. Transposase genes are often part of MGEs but other genes typical of mobile elements were not seen in these genomic neighborhoods. It is conceivable that the transposase gene has some function in the formation of the adjacent CRISPR (including transfer via HGT) but we could not find any mention of an association of this kind in the literature to date. The implications of the proximity of the CRISPR to the particular genomic neighborhood, if any, are unknown.

Some of the strains had Type IIIB CRISPR repeat arrays without adjacent *cas* genes in a different gene neighborhood; these included *B. fragilis* blood isolates as well as several ETBF isolates (in the ETBF isolates there were two type IIIB CRISPR arrays, one in the conserved neighborhood with adjacent *cas* genes and one in the same alternate neighborhood as the arrays

in the *B. fragilis* blood isolates). These arrangements are discussed further below (virulence association with CRISPR-Cas types).

### The Type IIIB *B. fragilis* CRISPR/Cas System with the Cas1-RT Fusion Protein Are Not Common among Bacteria

A GenBank search for proteins with both RT and Cas1 domains revealed that only 8% of bacterial Type IIIB elements have a Cas1/reverse transcriptase fusion protein, and these were most prevalent among cyanobacteria (Silas et al., 2016). The limited phylogenetic distribution of Cas1-RT and its association with only one CRISPR type suggests that there are a small number of origins of these RT-Cas1 fusions (Makarova et al., 2006; Silas et al., 2016). Recent data supports the notion that this protein may provide an efficient method to facilitate acquisition of spacers directly from RNA (Silas et al., 2016). Since other Type III CRISPR systems target RNA for degradation, these RT-associated CRISPR-Cas systems would effectively generate adaptive immunity against RNA parasites.

Further, this mechanism could target highly transcribed regions at both the DNA and RNA levels and thus serve in a regulatory capacity. The ability to target RNA substrates without targeting the bacterial chromosome would be a useful method for CRISPR regulation of endogenous genes (Sampson and Weiss, 2014). In *B. fragilis*, the conserved gene neighborhood of the Type IIIB CRISPRs included efflux operons, genes involved in cell wall biosynthesis and division, and/or genes involved in iron transport and storage.

### Class 2 Type IIC CRISPR-Cas Systems

Forty strains had Type IIC systems that contained the *cas2*, *cas1*, and *csn1* (*cas9*) genes (Figure 2C). For most of the strains, there are additional ORFs between *cas1* and *cas9* that encode two hypothetical proteins annotated as putative DNA binding proteins in a cluster with Type 1 Restriction-Modification systems. These hypothetical proteins have a conserved domain annotated as RhuM; this is a group of proteins implicated in virulence/pathogenicity (RhuM, of unknown function, is encoded in the SPI-3 pathogenicity island in *Salmonella typhimurium*). Fourteen percent (46/320) of the spacers could be identified in phage sequences (Supplementary Table 2C).

Csn1 (Cas9) proteins in the Type IIC systems were highly conserved among strains of *B. fragilis* (>99% identity). The closest homologs (52–70% identity) were found in other species including *Parabacteroides*, other *Bacteroidetes*, *Coprobacter*, *Capnocytophaga*, and *Prevotella* species. Csn1 (Cas9) belongs to the COG3513 - CRISPR/Cas system Type II associated protein, and contains McrA/HNH and RuvC-like nuclease domains.

Many of the Type IIC CRISPR-Cas arrays were found on contigs that ended just downstream of the CRISPR repeat, so not much information could be gleaned about the downstream neighborhood. The upstream gene neighborhood, when it could be discerned, was highly conserved (Figure 3C).

### Short CRISPR Repeats with 3 DRs and No Adjacent *cas* Genes (i.e., Orphan CRISPRs) Were Found Upstream of the *hipB* Gene in 71 Strains of *B. fragilis*

The entire sequence was completely conserved in these isolates, which originated from a variety of different sources, including normal feces, human microbiome project isolates, ETBF isolates, blood and virulent/MDR isolates. The *hipAB* operon is thought to be important in development of persister cells and multidrug tolerance in chronic infections due to *E. coli* and other bacteria (Day et al., 2004; Correia et al., 2006; Schumacher et al., 2015). HipA inhibits protein synthesis resulting in inhibition of cell growth and leading to multidrug tolerance by driving the cells into reversible dormancy and resulting in the production of persister cells (Schumacher et al., 2015). HipB binds HipA and acts as a transcriptional repressor of the *hipBA* operon (Hansen et al., 2012) and regulation of this operon is a key factor in controlling persister formation. Recently, high persister mutations were found in *E. coli* in which the mutation was not in the active site but rather interfered with higher order HipA-HipB promoter complexes that occluded the active site, thereby “unleashing” HipA to effect multidrug tolerance (Schumacher et al., 2015).

Tight regulation of *hipAB* is important since dormancy is only desirable if bacterial viability is threatened. Indeed, in global transcriptome studies of several *B. fragilis* strains we found that the *hipA* homologs were transcribed but at a very low level (manuscript submitted for publication). The ubiquitous proximity of this short orphan CRISPR to the *hipAB* operon in *B. fragilis* and the lack of adjacent *cas* genes raises the possibility that this “orphan” CRISPR is involved in regulation of *hipAB* but experimental analysis is necessary to determine whether this is the case. If confirmed, this would constitute a novel mode of regulation and could be important in the recalcitrance of chronic *B. fragilis* infections to antibiotic treatment. We did not find any significant repeat region upstream of *hipAB* in bacteria with genes phylogenetically related to *B. fragilis hipA* (*E. coli* NEB5A\_07695, *E. coli* 0104: H4 str. 2011c-3493, *Shewanella oneidensis* MR-1, *Streptomyces coelicolor* and *Myxococcus xanthus*). Orphan CRISPRs (i.e., without associated *cas* genes) have been implicated in regulation in other bacteria (Sampson and Weiss, 2014). In some cases, the spacers within the CRISPR target endogenous genes and there are no associated *cas* genes. Whether the lack of adjacent *cas* genes explains why the chromosome itself is not targeted (Stern et al., 2010), or whether these CRISPRs may use CAS proteins from other locations and still be involved in regulation of the targeted genes or gene transcripts remains unclear.

### Different CRISPR/Cas Subtypes Have Associated DR Sequence Subtypes and May Be Associated with Specific Functions in the Cell

This agrees with patterns seen in other bacteria (Kunin et al., 2007; Makarova et al., 2015) and is consistent with the notion

that different sets of genes are needed to recognize, bind and process the different repeat types; the differences are probably related to the particular fold structure assumed by the DR sequence. Consensus sequences of the repeat sequence for the four CRISPR-Cas systems, their placement within the database of CRISPR direct repeat sequences, and their predicted associated RNA structure are shown in **Figures 5** and **6**. **Figure 5** shows the placement of the four consensus repeat sequences within the entire Direct Repeat database as well as the nearest phylogenetic neighboring DRs. Notably, the four CRISPR consensus DRs were located across the entire DR spectrum, with phylogenetic neighbors from very distantly related species. More details about the phylogenetic placement of the DRs can be found in the legend to **Figure 5**.

The length of the consensus repeat sequence differed among the CRISPR-Cas types; Type IB (29 bp), Type IIIB (35 bp), Type IIC (47 bp), and Orphan (24 bp). This is in agreement with the finding that unusually long repeats (up to 48 bp) are exclusively present in Type IIC systems, especially in the *Bacteroidetes* phylum (Chylinski et al., 2014); the average CRISPR repeat in this system among all prokaryotes is 36 bp (Chylinski et al., 2013). The predicted fold structures of the DR sequences in each of the different CRISPR-Cas types was distinctive (**Figure 6**).

## Identification of Protospacers and Spacer Distribution within Each of the CRISPR Types Are Shown in Table 2

The search itself is somewhat problematic because the spacer sequences are so short. We used both the CRISPR Target program and NCBI BLAST with specified parameters to search for protospacer matches. Occasionally, a particular gene was identified to which the spacer had significant homology but, more often, other CRISPR arrays were identified. We repeated the search using the NCBI BLAST engine but adding a filter to screen out Repeat Arrays; we were able to identify protospacer matches for some additional spacer sequences but more often the match was to non-related nucleotide sequences.

We found that 7% of the spacer sequences were significantly homologous to a *Bacteroides* phage sequence. The majority of protospacers could not be assigned to a specific sequence (phage, plasmid or other bacterial gene or even inter-gene sequences) and most were identified in CRISPR regions in other *B. fragilis* strains. These numbers are consistent with other studies of prokaryotic CRISPR-Cas arrays; for example, in *Riemerella* strains, only 13/153 (8%) of spacers were homologous with a phage or plasmid target (Zhu et al., 2016).

In closely related strains, spacer identity and placement was frequently highly conserved. Spacers common to multiple strains can be due to either clonal dispersion of the CRISPR element through heredity or to HGT of these CRISPRs. The identified spacers, along with the strains in which they are found, their placement in those strains, and their predicted protospacer target are listed in Supplementary Tables 2A–C. A visual representation of their distribution is shown in Supplementary Tables 3A–C (binary distribution tables in Excel format).

One hundred thirty three unique spacers were found in the Type IB CRISPR-Cas systems. Five of the spacer targets matched, imperfectly but across the entire spacer length, to a phage sequence; these spacers were all found in an atypical, extremely long (67 spacers!) CRISPR repeat array in the blood isolate BF DCMOUH0017B. Notably, this array had an incomplete set of adjacent *cas* genes (only *cas1*, *cas2*, and *cas4*) and two transposases (one IS21 type) immediately upstream of *cas1*. *Cas1*, *Cas2*, and often *Cas4* comprise the adaptation module that is responsible for new spacer insertion (Makarova et al., 2015). The very long length of this CRISPR repeat array and multiple spacers matching *B. fragilis* phage protospacers is consistent with a CRISPR-Cas system that can acquire new spacers but cannot defend against the invading nucleic acid.

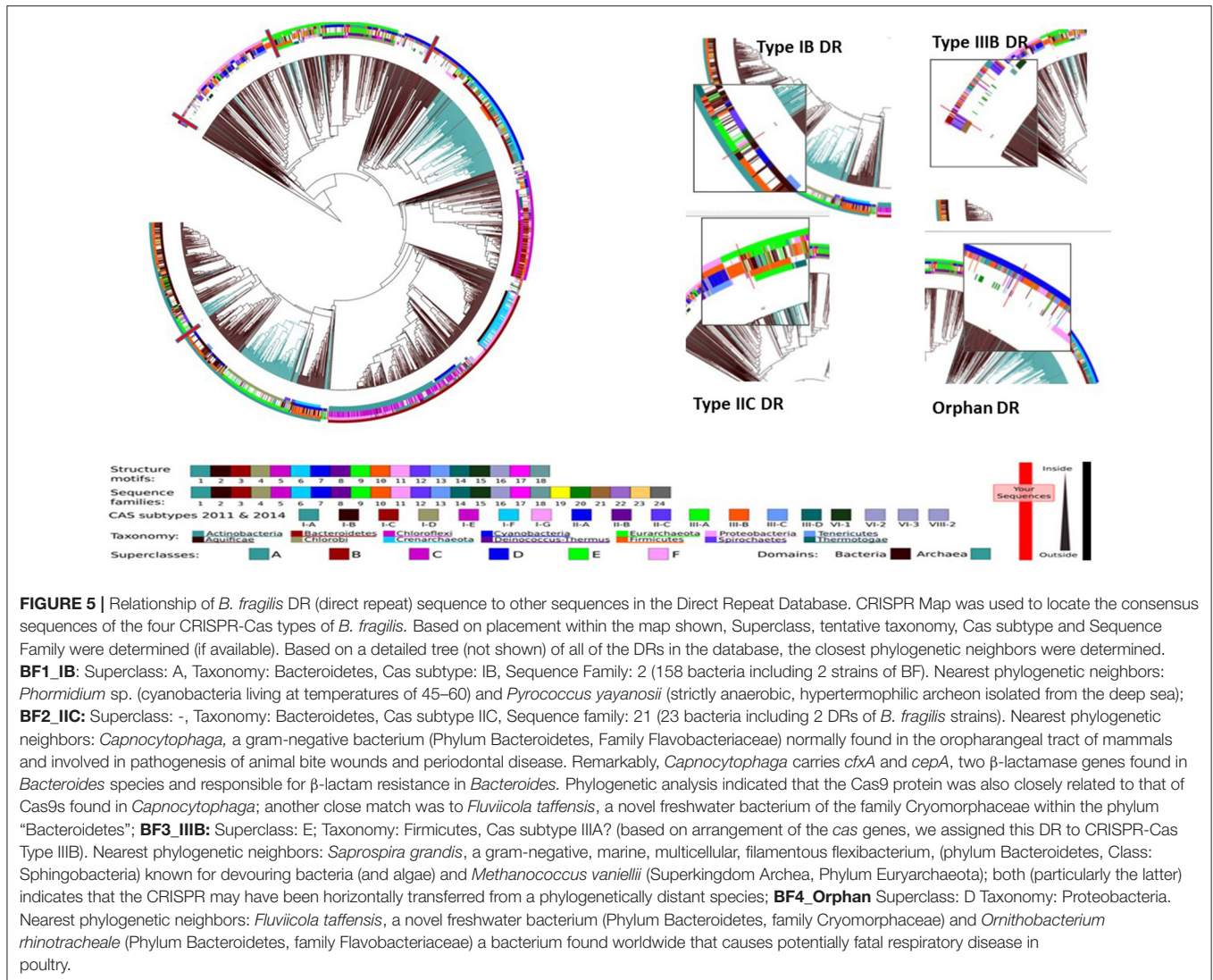
Type IIC systems contained significantly higher portion of spacers with crRNA homologous to a phage sequence than were found in the other systems (87% of spacers matching a phage sequence belonged to Type IIC CRISPR-Cas systems, 13% belonged to Type IB systems).

Spacer distribution in the Type IIIB CRISPR-Cas systems is shown in Supplementary Table 2B and graphically in Supplementary Table 3B. There were 252 unique spacers in the Type IIIB system, and the distribution was very broad. Strains with close phylogenetic relationship had identical spacers (e.g., BF 1007 F-2 thru F-10; BF 320, BF 321, and BF 322). Fifteen of the spacers (15/252, 6%) had matches with prokaryotic genes; no matches with phage or plasmid genes were identified and most matches were with other CRISPR regions (Supplementary Table 2B).

Type III CRISPR-Cas systems are associated with degradation of phage DNA in *Staphylococcus epidermidis* (Jiang et al., 2016). However, in *Streptococcus pyogenes*, there is a strong bias of the Type II system for acquisition of spacers matching viral protospacers (Heler et al., 2015). In our study, as in others, the DR repeat structures and associated *cas* genes of the two types were very distinct. Apparently, more than one CRISPR-Cas systems with the associated DR repeat structure can be used for acquisition and/or degradation of phage genetic material.

## Leader Sequences in CRISPR-Cas Systems

A 193 bp leader sequence was identified in the Type IB system of BF YCH 46 using CRISPRleader version 3.0 (Alkhnabashi et al., 2016) (kindly analyzed for us by Dr. Omer Alkhnabashi). The sequence was located between the repeat array and the *cas* genes and was extremely conserved among the Type IB CRISPR-Cas systems identified in *B. fragilis*. (Leader sequence: (5'-TGTTATTGTGAATTATCAATGGTAAAGTAACGGAAACGCTCTATGACATGTTGATATATAGATGTTTAGTACCTATGTCGCTAACCTATGTTTTTATATTATTCTTGATCGACACATTATTTCTAAAGAAAAGCAATTTTCTGCAAAGCATTGCTTATTACTAAGTAATTGCGTTGATTGATGGGTAGA-3')). Strains BF HMW610, *Bacteroides* sp. UW, BF DCMSKEJBY0001B and BF DCMOUH0017B, all phylogenetically related BF blood isolates, had slightly divergent Type IB leader sequences (data not shown). The promoter sequences for *B. fragilis* are not completely established and no

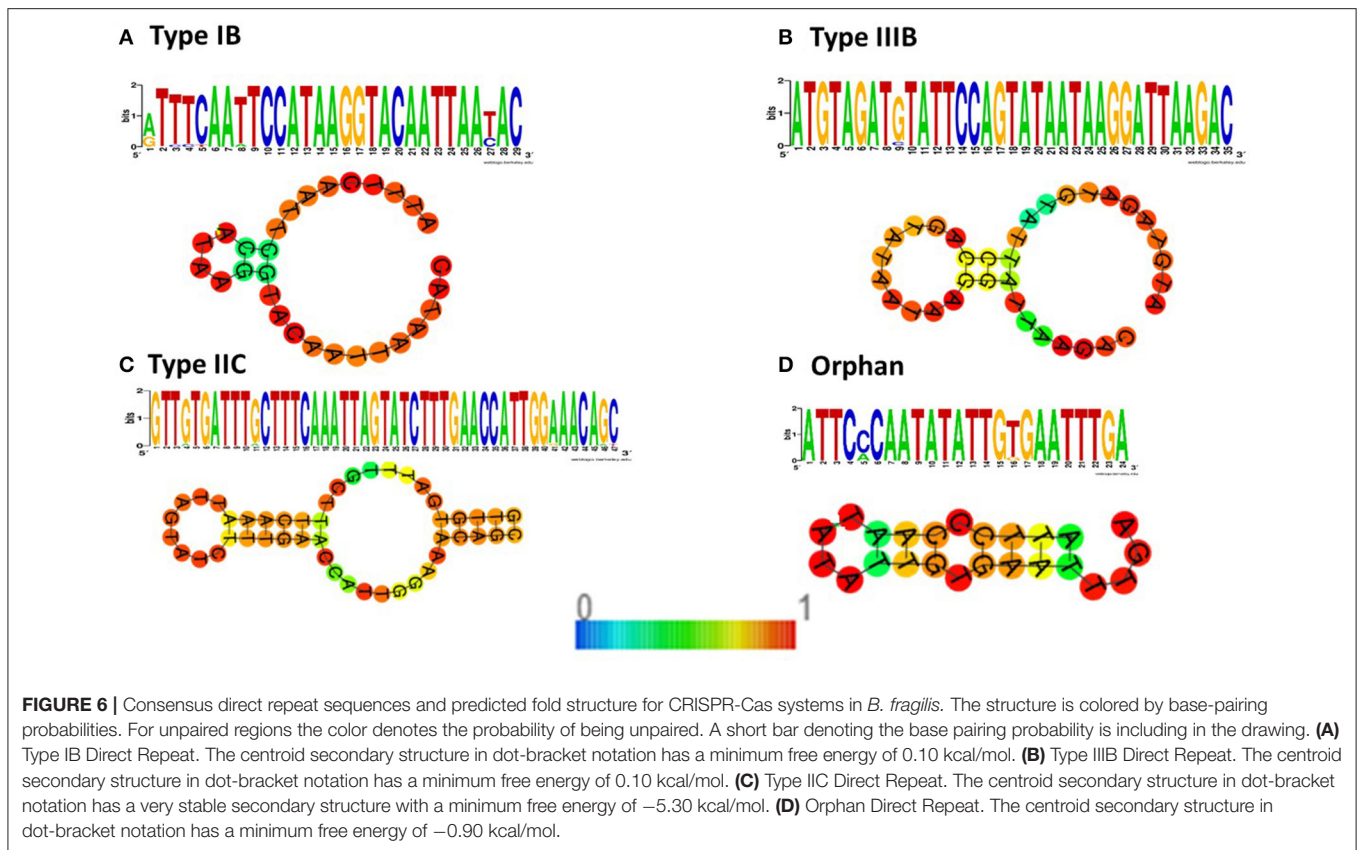


obvious promoter sequence was found on the leader sequence (Bayley et al., 2000). While no specific leader sequence could be definitively assigned in the Type IIB or Type IIC systems using the CRISPRleader program, the sequences upstream of the CRISPR repeat array (and downstream of the *cas* operon) were highly conserved, respectively, in each of those two CRISPR-Cas types. As was found for the Type IB leader, strains BF HMW610, *Bacteroides* sp. UW, BF DCMSKEJBY0001B and BF DCMOUH0017B had slightly divergent leader sequences for both Types IIB and IIC systems.

## No “Typical” CRISPR Systems Could Be Identified on Mobile Elements

Phylogenetic analysis indicates that CRISPR-Cas arrays have undergone extensive HGT, possibly on megaplasmids, as very similar *cas* genes are found in distantly related organisms (Godde and Bickerton, 2006). CRISPR-Cas arrays were found on a variety of MGEs (Sorek et al., 2008) including *Clostridium butyricum* megaplasmids (Iacobino et al., 2013) and viruses

attacking *Cyanobacter* (Chenard et al., 2016). We examined a variety of mobile elements in *B. fragilis* for the presence of CRISPR arrays. Ten plasmids and/or conjugative transposons for which there is partial or complete sequence were examined by CRT and did not have any discernable CRISPRs. These included: pBF9343, CTnHybL, pHAG88, p610 88, BOB\_25 PAO, CTn86, CTn341, CTnDot, and CTnPgl-a (note: many of the conjugative transposons in *B. fragilis* were only partially sequenced). Additionally, no CRISPRs were detected in the large horizontally transferred chromosomal fragments recently described (Husain et al., 2017). We also examined available sequences for 14 *B. fragilis* plasmids and did not find any CRISPR elements contained within their sequences. The examined plasmids (and their Genbank accession numbers) were: 2-078382-3, pBFP53, complete sequence, (CM004523.1); DCMOUH0042B, pBFU42e contig 1, (JPGQ0100001.1); pBFU42e, insertion sequence ISBf13 (complete), (KJ417513.1); pBF69566b, (KJ830768.1); pBF69566a, (KJ830769.1); 20656-2-1, pBI143, (LIDU01000064.1); IS4351, R plasmid encoding



macrolide B resistance, (M17124.1); JIM 10, unnamed plasmid, (MBRB01000037.1); JIM 10, unnamed plasmid, (NZ\_CM004507.1) pBFP35, complete sequence, (NC\_011073.1); pBFUK1, complete sequence, (NC\_019534.1); 2-078382-3, pBFP53, (NZ\_CM004523.1); DCMOUH0042B, pBFU42e contig 1, (NZ\_JPGQ01000001.1); pBI143, complete sequence, (U30316.1). No CRISPRs could be identified on the two sequenced BF phages (B40-8 and B124-14). However, using the more relaxed parameters in the CRISPR finder program, a potential CRISPR repeat was identified in pBF9343 but was not associated with adjacent *cas* genes. It should be noted that many of the CRISPRs that were identified on the genome had adjacent genes (e.g., transposase, mobility (*mob*) genes) that are typically found in mobile elements, so it is certainly possible that these CRISPRs are MGE-associated but that the exact MGE is not defined.

### Transposase Genes Were Frequently Found Adjacent to CRISPR Systems

Since the CAS proteins themselves possess transposase activity, it is not clear why it would be advantageous to have another transposase gene in close proximity. The presence of these genes, which are ubiquitous on mobile elements, might suggest that these CRISPRs are (or were in the past) contained within a putative MGE. There are ancestral innate immune systems that were formed from transposon-like elements containing *cas1* and *cas2*, eventually using the terminal inverted repeats

characteristic of the transposon to form the ancestral CRISPR repeats that were then duplicated by *cas1* (*cas1* also functioned in addition of spacers) (Nuñez et al., 2015). The molecular mechanism of CRISPR spacer integration is similar to that of both retroviral integration and DNA transposition that are mediated by integrases/transposases (Rath et al., 2015). Thus, it is possible that this adjacent transposase has some sort of function with the CRISPR function itself; however, that is purely speculative.

### Association of Virulence in *B. fragilis* Strains with Specific CRISPR-Cas Systems

The presence of CRISPR-Cas systems are variably associated with virulence (Makarova et al., 2011; Barrangou, 2015) and/or antibiotic resistance in pathogenic bacteria. On the one hand, the ability of a bacterium to incorporate mobile elements bearing pathogenic determinants would argue for a less robust CRISPR-Cas defense system; on the other hand, the presence of multiple CRISPRs systems with a variety of spacers may indicate previous exposure to these elements, parts of which were incorporated into the bacterial genome.

The *B. fragilis* strains isolated from blood (Table 1) that clustered into a tight phylogenetic group (Figure 1) had distinctive properties in their CRISPR-Cas content. None of these strains, with the exception of BF DCMOUH0042B (which clearly has a different lineage) had CRISPR-Cas systems belonging to the



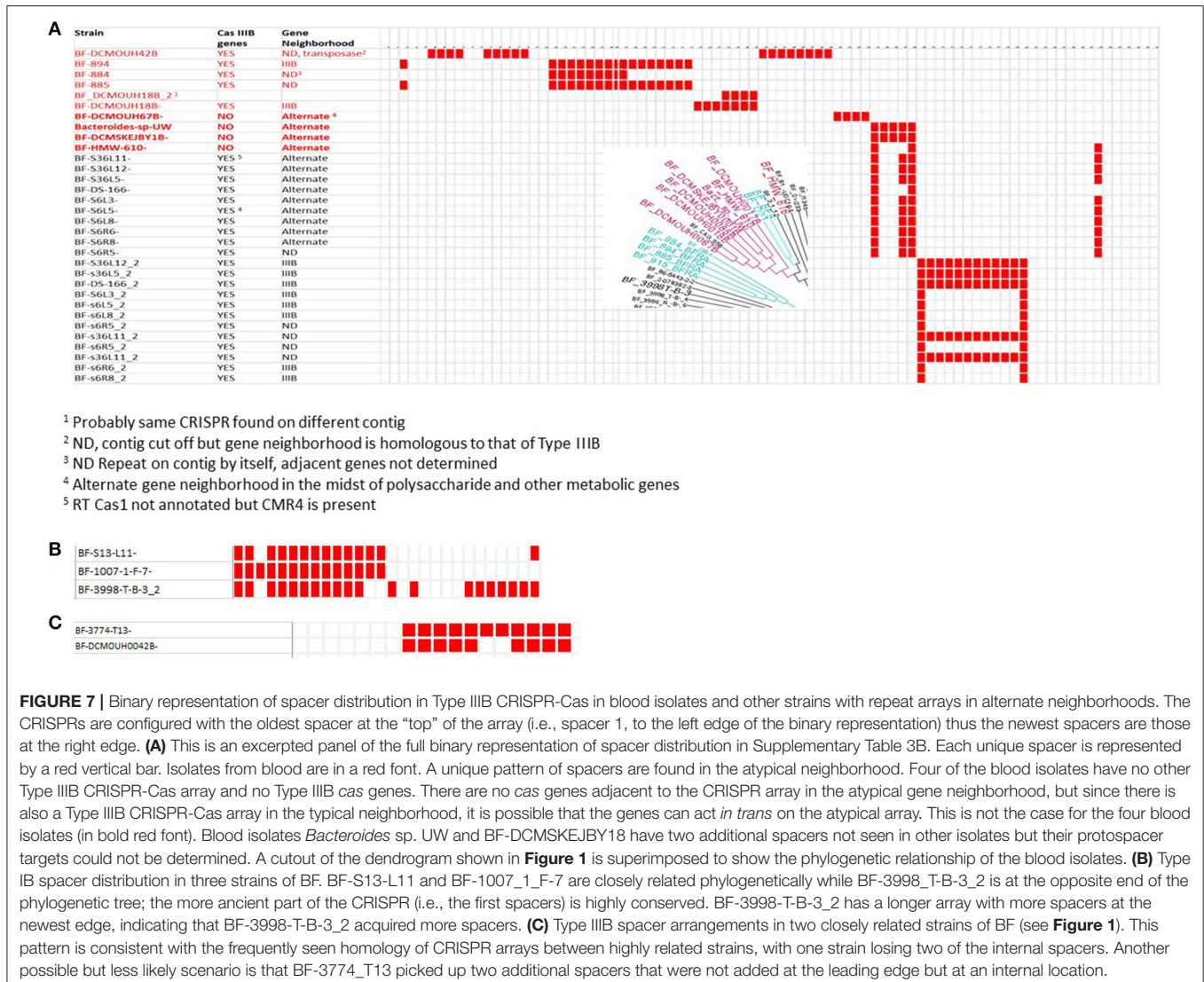
Type IIC CRISPR-Cas system. The distinct lack of the Type IIC CRISPR-Cas system for the majority of the blood isolates might suggest that the ability to incorporate phage DNA (which often carries antimicrobial resistance or virulence genes) is a beneficial trait for these virulent strains.

Further, the while these isolates did contain Type IIIB CRISPR repeat arrays, they differed from the typical Type IIIB array in two ways: (1) they were located in a completely different gene neighborhood (**Figure 3B**) and (2) had no adjacent *cas* genes. Interestingly, several of the ETBF isolates contained two sets of Type IIIB CRISPR-Cas repeat arrays: one in the typical Type IIIB gene neighborhood with a full set of adjacent Type IIIB *cas* genes and one (without adjacent *cas* genes) in the same neighborhood as the Type IIIB blood isolate CRISPR arrays.

No significant target sequences for the blood isolate spacers were identified, except for one spacer in BF DCMOUH0018B with significant homology to a region carrying the *B. fragilis* insertion sequence IS1168 and *nimB* (nitroimidazole resistance)

gene. The particular spacers carried by these isolates, despite the lack of Type IIIB *cas* genes, could be due either to them acquiring them at some point via an intact Type IIIB CRISPR-Cas system that they subsequently lost (perhaps during the transfer to an alternate gene neighborhood) or by HGT of the repeat array (without the *cas* genes) from another strain that did contain those spacers.

**Figure 7A** is an excerpted panel of the full binary representation of spacer distribution in Supplementary Table 3B and highlights the spacer distribution of the atypical Type IIIB CRISPR-Cas arrays in blood isolates (red font) as well as in those isolates that have both typical and atypical Type IIIB systems. Each unique spacer is represented by a red vertical bar. Four of the blood isolates have no other Type IIIB CRISPR-Cas array and no Type IIIB *cas* genes anywhere on the genome. Some of the ETBF isolates also have an array containing these same spacers, and also in the same “atypical” neighborhood. But these isolates *also* have another Type IIIB array, in the traditional



neighborhood with adjacent *cas* genes; thus, it is possible that the genes can act *in trans* on the atypical array. Again, this is not the case for isolates *Bacteroides* sp. UW, BF-DCMSKEJBY18, BF HMW 610, and BF-DCMOUH0067B. Blood isolates *Bacteroides* sp. UW and BF-DCMSKEJBY18 have two additional spacers not seen in other isolates. It is not clear how these atypical systems appeared but it is tempting to speculate that the atypical array broke away from the original CRISPR-Cas array and that the very virulent blood isolates somehow discarded the *cas* genes portion of the original array. At this point we have no evidence for that speculation nor a reason that it would be beneficial for those isolates to discard that array or adjacent genes. **Figures 7B** and **C** illustrate examples of spacer pattern diversions between closely related strains and spacer conservation between distant strains. **Figure 7B** shows distribution of two closely related strains and one distant strain and **Figure 7C** shows Type IIIB spacer arrangements in two closely related strains of BF and is consistent with a pattern in which spacers can get lost.

## The Plasticity of the *B. fragilis* Genome Is Balanced by Multiple Systems to Avoid Invading DNA Elements

The *B. fragilis* genome is very “plastic,” due both to its ability to incorporate pathogenicity islands from other *B. fragilis* and foreign genes via HGT as well as its ability to simply turn specific genes on or off as needed. Combined, these traits allow *B. fragilis* to adapt to new nutrition pathways, utilize specific efflux pumps to rid the cell of toxic substrates, and display new surface epitopes—taken together, allowing them to change from friendly commensal to dangerous threat (Wexler, 2007). We previously demonstrated transfer of a mobile element that contained multiple resistance genes of aerobic origin clustered within a conjugative transposon (Husain et al., 2014).

This ability of *B. fragilis* to easily incorporate foreign genes is intriguing since *B. fragilis* also possesses strong DNA restriction modification (DNA/RM) systems to degrade “non-self” DNA. We previously demonstrated horizontal transfer of mobile elements bearing alternate variants of these genes from a multidrug resistant *B. fragilis* isolate (HMW 615) to *B. fragilis* 638R (Husain et al., 2017). These systems are located in shufflons with invertible promoters that can be turned off or on, so that the bacterium can control whether the incoming DNA will be degraded (Patrick et al., 2010). If the system is turned “off,” the incoming DNA can survive degradation. External or internal signals that regulate these systems have not yet been described. Now we have shown that in addition to the DNA/RM systems, there are abundant CRISPR elements in *B. fragilis* with adjacent *cas* genes, suggesting that these CRISPRs are indeed active as a bacterial defense system. We had previously suggested that predation by bacteriophages is an evolutionary driving force for generation of variable polysaccharide and R-M systems in *B. fragilis* (Patrick et al., 2010) and that this system can be diversified by HGT (Husain et al., 2017). In an interesting twist, *B. fragilis* also has a Type IIC CRISPR system with a high

proportion of crRNAs with homology to *B. fragilis* phages; it is likely that this CRISPR system is largely directed to protect against invading bacteriophage (but these systems are lacking in the most virulent *B. fragilis* strains).

Thus, at least two unique systems are in place in *B. fragilis* to control and degrade incoming DNA, despite the well-established role of *B. fragilis* as a “resistance reservoir” (Salys et al., 2004; Coyne et al., 2014) and its known ability to incorporate both other *B. fragilis* elements as well as “foreign” genes (Husain et al., 2014). It seems likely that an intricate balance of these two evolutionary forces is responsible for the remarkable adaptability of *B. fragilis* to the changing nutritional availability, immune forces and competitive organisms in the complex environment of the gut microbiome.

## CONCLUSIONS

In this study, we identified and described CRISPR-Cas systems in *B. fragilis* as a first step for further exploration of the roles and mode of function of *B. fragilis* CRISPRs. We now need to determine whether the *B. fragilis* CRISPRs are functional in bacterial immunity and/or in regulation and whether they can signal acquisition of virulence determinants. Aside from these associations, there is growing evidence that CRISPR-Cas systems may have alternative roles that allow the bacterium to survive host defenses and replicate (Sampson and Weiss, 2014). For example, the Cas1 protein of the CRISPR-Cas system of *E. coli* may play a role in DNA repair (Babu et al., 2011). Since DNA damage to bacteria can be the result of specific host defenses during infection (e.g., the production of radical nitrogen and oxygen species), it would be beneficial to the bacterium to possess redundancy in its DNA repair capability (Sampson and Weiss, 2014). An additional way to avoid or repair DNA damage due to radical oxygen species would obviously be of great benefit to the anaerobic *B. fragilis* as well. In addition to these functions, there are reports of CRISPR-Cas involvement in resistance to stress, pathogenicity and regulation of biofilm formation (Barrangou, 2015).

We found that the most virulent strains of *B. fragilis*, the blood isolates, did not have Type IIC CRISPR-Cas systems, which may suggest that they remain capable of incorporating resistance and virulence factors that may be transferred via phages. Also, they do not appear to have functional Type IIIB systems, although they retain the CRISPR repeat array characteristic of that system, but in a distinct gene neighborhood. Thus, it appears that for these most virulent blood isolates, there is a benefit to being able to acquire genes from phages and possibly other mobile elements, which is completely consistent with the evidence that many antimicrobial resistance genes and other virulence genes are mobilized via HGT. Further analysis of the spacer acquisition pattern will help to determine how these spacers were acquired and help to elucidate the extent to which strains evolve by vertical evolution and/or horizontal gene transmission. Experimental molecular manipulation to determine the functions of the various CRISPR arrays will lead to a more comprehensive understanding of their functions in bacterial defense, gene

regulation, virulence and commensalism of this important gut pathobiont.

## AUTHOR CONTRIBUTIONS

MT contributed to the design, performed bioinformatic analyses, assisted in the interpretation of results and in writing the manuscript. HW designed the study, interpreted the results and wrote the manuscript. All authors have read and approved the final version of the manuscript.

## FUNDING

These studies were supported by grants to HW from the Department of Veteran's Affairs (BX000563-01A1) and the National Institutes of Health (R21 AI109545-01).

## REFERENCES

- Alkhnbashi, O. S., Shah, S. A., Garrett, R. A., Saunders, S. J., Costa, F., and Backofen, R. (2016). Characterizing leader sequences of CRISPR loci. *Bioinformatics* 32, i576–i585. doi: 10.1093/bioinformatics/btw454
- Ank, N., Sydenham, T. V., Iversen, L. H., Justesen, U. S., and Wang, M. (2015). Characterisation of a multidrug-resistant *Bacteroides fragilis* isolate recovered from blood of a patient in Denmark using whole-genome sequencing. *Int. J. Antimicrob. Agents* 46, 117–120. doi: 10.1016/j.ijantimicag.2015.02.024
- Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75
- Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., et al. (2011). A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol. Microbiol.* 79, 484–502. doi: 10.1111/j.1365-2958.2010.07465.x
- Barlow, M. (2009). What antimicrobial resistance has taught us about horizontal gene transfer. *Methods Mol. Biol.* 532, 397–411. doi: 10.1007/978-1-60327-853-9\_23
- Barrangou, R. (2015). Diversity of CRISPR-Cas immune systems and molecular machines. *Genome Biol.* 16, 247. doi: 10.1186/s13059-015-0816-9
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi: 10.1126/science.1138140
- Bayley, D. P., Rocha, E. R., and Smith, C. J. (2000). Analysis of *cepA* and other *Bacteroides fragilis* genes reveals a unique promoter structure. *FEMS Microbiol. Lett.* 193, 149–154. doi: 10.1111/j.1574-6968.2000.tb09417.x
- Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2014). GenBank. *Nucleic Acids Res.* 42, D32–D37. doi: 10.1093/nar/gkt1030
- Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., et al. (2007). CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209. doi: 10.1186/1471-2105-8-209
- Bourgogne, A., Garsin, D. A., Qin, X., Singh, K. V., Sillanpaa, J., Yerrapragada, S., et al. (2008). Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome Biol.* 9:R110. doi: 10.1186/gb-2008-9-7-r110
- Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., et al. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* 7:10613. doi: 10.1038/ncomms10613
- Capozzi, V., and Spano, G. (2009). Horizontal gene transfer in the gut: is it a risk? *Food Res. Int.* 42, 1501–1502. doi: 10.1016/j.foodres.2009.08.001
- Carte, J., Christopher, R. T., Smith, J. T., Olson, S., Barrangou, R., Moineau, S., et al. (2014). The three major types of CRISPR-Cas systems function independently

## ACKNOWLEDGMENTS

We wish to thank and acknowledge the kind assistance of Dr. Michael Richter (Ribocon, Bremen, Germany) in constructing a matrix of average nucleotide identity between two organisms (ANIb) values using the J species Web Server and of Dr. Luis M. Rodriguez (Kostas Lab, Georgia Institute of Technology, Atlanta, Georgia) in using an R routine to convert this data into a Newick tree file.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02234/full#supplementary-material>

- in CRISPR RNA biogenesis in *Streptococcus thermophilus*. *Mol. Microbiol.* 93, 98–112. doi: 10.1111/mmi.12644
- Chénard, C., Wirth, J. F., and Suttle, C. A. (2016). Viruses infecting a freshwater filamentous *Cyanobacterium* (Nostoc sp.) encode a functional CRISPR array and a proteobacterial DNA polymerase B. *MBio* 7:e00667-16. doi: 10.1128/mBio.00667-16
- Chylinski, K., Le Rhun, A., and Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* 10, 726–737. doi: 10.4161/rna.24321
- Chylinski, K., Makarova, K. S., Charpentier, E., and Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* 42, 6091–6105. doi: 10.1093/nar/gku241
- Correia, F. F., D'Onofrio, A., Rejtar, T., Li, L., Karger, B. L., Makarova, K., et al. (2006). Kinase activity of overexpressed HipA is required for growth arrest and multidrug tolerance in *Escherichia coli*. *J. Bacteriol.* 188, 8360–8367. doi: 10.1128/JB.01237-06
- Coyne, M. J., Zitomersky, N. L., McGuire, A. M., Earl, A. M., and Comstock, L. E. (2014). Evidence of extensive DNA transfer between Bacteroidales species within the human gut. *MBio* 5, e01305–e01314. doi: 10.1128/mBio.01305-14
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Day, B. J., van Heeckeren, A. M., Min, E., and Velsor, L. W. (2004). Role for cystic fibrosis transmembrane conductance regulator protein in a glutathione response to bronchopulmonary pseudomonas infection. *Infect. Immun.* 72, 2045–2051. doi: 10.1128/IAI.72.4.2045-2051.2004
- Ding, Y., Chan, C. Y., and Lawrence, C. E. (2005). RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* 11, 1157–1166. doi: 10.1261/rna.2500605
- Godde, J. S., and Bickerton, A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* 62, 718–729. doi: 10.1007/s00239-005-0223-z
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360
- Haft, D. H., Selengut, J., Mongodin, E. F., and Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* 1:e60. doi: 10.1371/journal.pcbi.0010060
- Han, M. V., and Zmasek, C. M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* 10:356. doi: 10.1186/1471-2105-10-356
- Hansen, S., Vulić, M., Min, J., Yen, T. J., Schumacher, M. A., Brennan, R. G., et al. (2012). Regulation of the *Escherichia coli* HipBA toxin-antitoxin system by proteolysis. *PLoS ONE* 7:e39185. doi: 10.1371/annotation/e608601c-eadd-4c11-adb2-7b605aba9c44

- Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D., et al. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* 519, 199–202. doi: 10.1038/nature14245
- Husain, F., Tang, K., Yaligara, V., Boente, R., Blakely, G., Patrick, S., et al. (2017). Novel large-scale chromosomal transfer in *Bacteroides fragilis* contributes to its pan-genome and rapid environmental adaptation. *Microb. Genom.* 4:e000136. doi: 10.1099/mgen.0.000136
- Husain, F., Veeranagouda, Y., Boente, R., Tang, K., Mulato, G., and Wexler, H. M. (2014). The Ellis Island effect: a novel mobile element in a multi-drug resistant *Bacteroides fragilis* clinical isolate includes a mosaic of resistance genes from Gram-positive bacteria. *Mob. Genet. Elem.* 4, E29801–E29812. doi: 10.4161/mge.29801
- Iacobino, A., Scalfaro, C., and Franciosa, G. (2013). Structure and genetic content of the megaplasmids of neurotoxicogenic *Clostridium butyricum* type E strains from Italy. *PLoS ONE* 8:e71324. doi: 10.1371/journal.pone.0071324
- Jiang, W., Samai, P., and Marraffini, L. A. (2016). Degradation of phage transcripts by CRISPR-associated RNases enables type III CRISPR-Cas immunity. *Cell* 164, 710–721. doi: 10.1016/j.cell.2015.12.053
- Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* 18, 529–536. doi: 10.1038/nsmb.2019
- Kalapila, A., Pergam, S., Pottinger, P., Butler-Wu, S., Whimbe, E., and Duchin, J. (2013). Multidrug-resistant *Bacteroides fragilis*–Seattle, Washington, 2013. *MMWR Morb. Mortal. Wkly. Rep.* 62, 694–696.
- Kau, A. L., Planer, J. D., Liu, J., Rao, S., Yatsunenko, T., Trehan, I., et al. (2015). Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. *Sci. Transl. Med.* 7:276ra224. doi: 10.1126/scitranslmed.aaa4877
- Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* 8:R61. doi: 10.1186/gb-2007-8-4-r61
- Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., et al. (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* 14, 169–181. doi: 10.1093/dnares/dsm018
- Kuwahara, T., Yamashita, A., Hirakawa, H., Nakayama, H., Toh, H., Okada, N., et al. (2004). Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14919–14924. doi: 10.1073/pnas.0404172101
- Lange, S. J., Alkhnbashi, O. S., Rose, D., Will, S., and Backofen, R. (2013). CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucleic Acids Res.* 41, 8034–8044. doi: 10.1093/nar/gkt606
- Louwen, R., Staals, R. H., Endtz, H. P., van Baaren, P., and van der Oost, J. (2014). The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiol. Mol. Biol. Rev.* 78, 74–88. doi: 10.1128/MMBR.00039-13
- Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., and Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* 1:7. doi: 10.1186/1745-6150-1-7
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569
- Marinelli, L. J., Fitz-Gibbon, S., Hayes, C., Bowman, C., Inkeles, M., Loncaric, A., et al. (2012). Propionibacterium acnes bacteriophages display limited genetic diversity and broad killing activity against bacterial skin isolates. *MBio* 3:12. doi: 10.1128/mBio.00279-12
- Markowitz, V., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012). IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* 40, 22. doi: 10.1093/nar/gkr1044
- Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature* 526, 55–61. doi: 10.1038/nature15386
- Mimee, M., Tucker, A. C., Voigt, C. A., and Lu, T. K. (2015). Programming a human commensal bacterium, *Bacteroides thetaiotaomicron*, to sense and respond to stimuli in the murine gut microbiota. *Cell Syst.* 1, 62–71. doi: 10.1016/j.cels.2015.06.001
- Mougiakos, I., Bosma, E. F., de Vos, W. M., van Kranenburg, R., and van der Oost, J. (2016). Next generation prokaryotic engineering: the CRISPR-Cas toolkit. *Trends Biotechnol.* 34, 575–587. doi: 10.1016/j.tibtech.2016.02.004
- Nikitina, A. S., Kharlampieva, D. D., Babenko, V. V., Shirokov, D. A., Vakhitova, M. T., Manolov, A. I., et al. (2015). Complete genome sequence of an enterotoxigenic *Bacteroides fragilis* clinical isolate. *Genome Announc.* 3:e00450–15. doi: 10.1128/genomeA.00450-15
- Núñez, J. K., Lee, A. S., Engelman, A., and Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519, 193–198. doi: 10.1038/nature14237
- Palmer, K. L., and Gilmore, M. S. (2010). Multidrug-resistant enterococci lack CRISPR-cas. *MBio* 1:10. doi: 10.1128/mBio.00227-10
- Patrick, S., Blakely, G. W., Houston, S., Moore, J., Abratt, V. R., Bertalan, M., et al. (2010). Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of *Bacteroides fragilis*. *Microbiology* 156(Pt 11), 3255–3269. doi: 10.1099/mic.0.042978-0
- Pierce, J. V., and Bernstein, H. D. (2016). Genomic diversity of enterotoxigenic strains of *Bacteroides fragilis*. *PLoS ONE* 11:e0158171. doi: 10.1371/journal.pone.0158171
- Privitera, G. F., Dublanchet, A. F., and Sebald, M. (1979). Transfer of multiple antibiotic resistance between subspecies of *Bacteroides fragilis*. *J. Infect. Dis.* 139, 97–101. doi: 10.1093/infdis/139.1.97
- Pumbwe, L., Chang, A., Smith, R. L., and Wexler, H. M. (2007a). BmeRABC5 is a multidrug efflux system that can confer metronidazole resistance in *Bacteroides fragilis*. *Microb. Drug Resist.* 13, 96–101. doi: 10.1089/mdr.2007.719
- Pumbwe, L., Wareham, D. W., Aduse-Opoku, J., Brazier, J. S., and Wexler, H. M. (2007b). Genetic analysis of mechanisms of multidrug resistance in a clinical isolate of *Bacteroides fragilis*. *Clin. Microbiol. Infect.* 13, 183–189. doi: 10.1111/j.1469-0691.2006.01620.x
- Rath, D., Amlinger, L., Rath, A., and Lundgren, M. (2015). The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie* 117, 119–128. doi: 10.1016/j.biochi.2015.03.025
- Risse, J., Thomson, M., Patrick, S., Blakely, G., Koutsovoulos, G., Blaxter, M., et al. (2015). A single chromosome assembly of *Bacteroides fragilis* strain BE1 from Illumina and MinION nanopore sequencing data. *Gigascience* 4, 60. doi: 10.1186/s13742-015-0101-6
- Roach, D. J., Burton, J. J., Lee, C., Stackhouse, B., Butler-Wu, S. M., Cookson, B. T., et al. (2015). A year of infection in the intensive care unit: prospective whole genome sequencing of bacterial clinical isolates reveals cryptic transmissions and novel microbiota. *PLoS Genet.* 11:e1005413. doi: 10.1371/journal.pgen.1005413
- Salipante, S. J., Kalapila, A., Pottinger, P. S., Hoogestraat, D. R., Cummings, L., Duchin, J. S., et al. (2015). Characterization of a multidrug-resistant, novel *Bacteroides* genomospecies. *Emerg. Infect. Dis.* 21, 95–98. doi: 10.3201/eid2101.140662
- Salyers, A. A., Gupta, A., and Wang, Y. (2004). Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends Microbiol.* 12, 412–416. doi: 10.1016/j.tim.2004.07.004
- Sampson, T., and Weiss, D. (2014). CRISPR-Cas systems: new players in gene regulation and bacterial physiology. *Front. Cell. Infect. Microbiol.* 4:37. doi: 10.3389/fcimb.2014.00037
- Sander, J. D., and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355. doi: 10.1038/nbt.2842
- Schapiro, J. M., Gupta, R., Stefansson, E., Fang, F. C., and Limaye, A. P. (2004). Isolation of metronidazole-resistant *Bacteroides fragilis* carrying the nimA nitroreductase gene from a patient in Washington State. *J. Clin. Microbiol.* 42, 4127–4129. doi: 10.1128/JCM.42.9.4127-4129.2004
- Schumacher, M. A., Balani, P., Min, J., Chinnam, N. B., Hansen, S., Vulić, M., et al. (2015). HipBA-promoter structures reveal the basis of heritable multidrug tolerance. *Nature* 524, 59–64. doi: 10.1038/nature14662
- Science, U.o.M.S.o.M.I.f.G. (2013). *GSCID Genomic Sequencing Center for Infectious Disease [Online]*. NCBI. Available online at: [http://gscid.igs.umaryland.edu/wp.php?wp=comparative\\_genomics\\_of\\_human\\_mucosal\\_isolates\\_of\\_the\\_enterotoxigenic\\_bacteroides\\_fragilis\\_group](http://gscid.igs.umaryland.edu/wp.php?wp=comparative_genomics_of_human_mucosal_isolates_of_the_enterotoxigenic_bacteroides_fragilis_group).
- Sherwood, J., Fraser, S., Citron, D., Wexler, H., Blakely, G., Jobling, K., et al. (2011). Multi-drug resistant *Bacteroides fragilis* recovered from blood and severe leg wounds caused by an improvised explosive device (IED) in Afghanistan. *Anaerobe* 17, 152–155. doi: 10.1016/j.anaerobe.2011.02.007

- Silas, S., Mohr, G., Sidote, D. J., Markham, L. M., Sanchez-Amat, A., Bhaya, D., et al. (2016). Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* 351:aad4234. doi: 10.1126/science.aad4234
- Soki, J., Hedberg, M., Patrick, S., Balint, B., Herczeg, R., Nagy, I., et al. (2016). Emergence and evolution of an international cluster of MDR *Bacteroides fragilis* isolates. *J. Antimicrob. Chemother.* 71, 2441–2448. doi: 10.1093/jac/dkw175
- Solomkin, J. S., Mazuski, J. E., Bradley, J. S., Rodvold, K. A., Goldstein, E. J., Baron, E. J., et al. (2010). Diagnosis and management of complicated intra-abdominal infection in adults and children: guidelines by the Surgical Infection Society and the Infectious Diseases Society of America. *Surg. Infect. (Larchmt)* 11, 79–109. doi: 10.1089/sur.2009.9930
- Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* 6, 181–186. doi: 10.1038/nrmicro1793
- Stern, A., Keren, L., Wurtzel, O., Amitai, G., and Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* 26, 335–340. doi: 10.1016/j.tig.2010.05.008
- Sydenham, T. V., Soki, J., Hasman, H., Wang, M., Justesen, U. S., and Esgai (2015). Identification of antimicrobial resistance genes in multidrug-resistant clinical *Bacteroides fragilis* isolates by whole genome shotgun sequencing. *Anaerobe* 31, 59–64. doi: 10.1016/j.anaerobe.2014.10.009
- Wexler, H. M. (2007). Bacteroides—the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* 20, 593–621. doi: 10.1128/CMR.00008-07
- Wexler, H. M. (2014). “The genus Bacteroides,” in *The Prokaryotes: Other major lineages of Bacteria and the Archaea*, eds E. Y. Rosenberg, E. F. DeLong, F. Thompson, S. Lory, and E. Stackebrandt (Berlin; Heidelberg: Springer Verlag), 459–484.
- Zhu, D. K., Yang, X. Q., He, Y., Zhou, W. S., Song, X. H., Wang, J. B., et al. (2016). Comparative genomic analysis identifies structural features of CRISPR-Cas systems in *Riemerella anatipestifer*. *BMC Genomics* 17:689. doi: 10.1186/s12864-016-3040-4

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AO and handling Editor declared their shared affiliation.

Copyright © 2017 Tajkarimi and Wexler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.