Comprehensive Comparative Genomic revels: *Bacteroides fragilis* is a reservoir of antibiotic resistance genes in the gut microbiota

Short Title: *Bacteroides fragilis* is a reservoir of antibiotic resistance genes

Eliane Evanovich[1], Patricia Jeanne de Souza Mendonça Mattos[1], João Farias Guerreiro[1]

Filiation:

*Laboratório de Genética Humana e Médica, Instituto de Ciências Biológicas - Universidade Federal do Pará, Pará, Brasil.

E-mail address for correspondence: lianevanovich@gmail.com

Co-authors e-mail addresses: patjeanne@gmail.com, joao.guerreiro53@gmail.com

Mailing address: Laboratório de Genética Humana e Médica - Universidade Federal do Pará, Rua Augusto Corrêa, 1 - Guamá, Belém - PA, 66075-110.

ORCID ID: https://orcid.org/0000-0001-7076-5621

Abstract

Bacteroides fragilis are commensal bacteria of the gut microbiota of mammals and may cause severe infection in a susceptible host. Treatment can be cumbersome if multidrug resistant strains are present in the affected tissue. The principal aim of this study was to provide new insights into the genomic properties of *B. fragilis* through different approaches in comparative genomics. Results revealed that the pan-genome is opened, and an intense exchange of genetic material reinforces this inference. The Don complex, responsible for extraintestinal adaptation, is present in all strains, suggesting a crucial role for *B. fragilis* adaptation. CRISPR-Cas system is at 76% of the samples, but it apparently has low accuracy against prophage. Multidrug resistance genes are in 80% of strains. Conjugative transposons and integrative and conjugative elements (ICE) are the main spreaders of genes for antimicrobial resistance. We also reported evidence for horizontal gene transfer (HGT) of antimicrobial resistance genes among the *B. fragilis* strains and Bacteriodales. At least 398 genes are under positive selection, including genes for antimicrobial resistance and transport of toxins and nutrients.

Keywords: *Bacteroides fragilis*, Comparative genomics, Genes for antimicrobial resistance, Horizontal gene transfer.

1. Introduction

Bacteroides fragilis is an obligatory anaerobic Gram-negative bacillus. It is usually commensal in mammals and is critical for immune development and intestinal mucosal integrity [1, 2, 3]. It may become pathogenic in a vulnerable host, being able to invade intra- and extra-abdominal tissues, including blood [4–9]. Recent evidence suggests that enterotoxigenic *B. fragilis* (ETBF) is associated with colorectal cancer development [8-11].

ETBF can produce fragilysin (known as BFT), an endotoxin from the zinc-dependent metalloproteinase family that cleaves E-cadherin in colonocytes [12-16]. The interaction between BFT and intestinal epithelial cells triggers an inflammatory response that increases apoptosis and favors extraintestinal invasion [17]. The *bftP* gene encodes the BFT and has three alleles: *bftP-1*, *bftP-2*, and *bftP-3* [18]. It is on a pathogenic island termed BfPAI, along with genes for a hypothetical protein and metalloproteinase II (*mpII*) [13, 14, 18–21]. BfPAI of some ETBFs is in the CTn86 transposon [13-14].

Antimicrobial multidrug resistance (AMR) genes are widespread among the B. fragilis strains and hamper antimicrobial therapy [22]. Integrative and conjugative elements (ICE) disseminate AMR genes because they can mobilize plasmids, genomic islands, and other non-conjugative elements [23]. In addition, their genes encode proteins for excision, conjugation (type IV secretion system-T4SS), and transferring DNA sequences [24].

Other genes that improve *B. fragilis* adaptation to intestinal and extraintestinal environments belong to Type VI secretion system (T6SS), capsular lipopolysaccharide biosynthesis locus (LPS locus), and Don PUL complex. T6SS is a protein complex that acts during competition, inhibiting or killing other bacteria or eukaryotic cells [25-27]. The LPS locus has 19 orthologous genes (OG) that control the expression of three capsular polysaccharides (polysaccharides A, B, and C) and may trigger the formation of intra-abdominal abscesses [28–32]. Don PUL complex enables extraintestinal adaptation through uptake of N-linked glycans from fluids [33].

Despite the approaches described above, few studies have investigated the comparative genomics of *B. fragilis*. Thus, this study provides insights into the genomic properties of these bacteria. For this purpose, we analysed 183 strains by applying comparative genomics tools. Some insights obtained were 1) Most genes are related to metabolic processes, mainly carbohydrates and amino acids - a frequent pattern of bacterial gut microbiota; 2) The Don PUL is the only system common to all strains, suggesting that alternative resource capture is decisive for their survival in the gut

microbiota; and 3) We also reported evidence of HGT within order Bacteroidales. Hence, this research extends our knowledge of the genomic evolution of B. fragilis.

2. Materials and Methods

Sample

Samples used were available at the National Center for Biotechnology Information - NCBI (https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/414/). Genome lengths range from 4.87 to 7.60 megabases (complete, scaffold, and contig), and the minimal coverage was 15X [34] The strains used and their sources are in S1 Table.

Pan-genome Analysis

Bacterial Pan Genome Analysis tool version 1.3 (BPGA) was used to identify core, accessory, and unique genomes [35]. Gene functions were based on Clusters of Orthologous Groups of proteins (COG) and the Kyoto Encyclopedia of Genes and Genomes (KEGG). Orthologous cluster analysis (default setting) was performed by Usearch9.2.64 [36] under a cut-off value of 95% to generate high-quality clusters [37]. Multiple sequence comparison by log-expectation (MUSCLE) was used to generate alignments and phylogenies [36]. Gnuplot 5.0 was used for plotting graphs [38].

Prediction of CRISPR and Resistance genes

CRISPR-Cas sequences were estimated by CRISPRRCasFinder (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) [39]. Antimicrobial resistance genes were obtained from CARD (Comprehensive Antibiotic Resistance Database) (http://arpcard.mcmaster.ca/). CARD was performed under the Resistance Gene Identifier option - RGI (Perfect match equal to 100) [40]. Results with a ‘strict match’ between 95.0 and 99.9 were also analyzed under the Basic Local Alignment Search Tool- BLAST, as suggested by Jia et al. [40].

Prediction of Mobilome

PhiSpy software was used to predict putative prophages, and ICEFinder was performed to identify ICE and IME (Integrative mobilizable elements) [41-42]. These analyses were restricted to 19 complete genomes because the prophage, ICE, and IME regions are usually long and would be cut into short sequences, such as contigs and scaffolds.

Candidate proteins were searched against the CARD (BLASTp) [43], Virulence factor database (VFDB) [http://www.mgc.ac.cn/VFs/search_VFs.htm; blastp/Protein sequences from VFDB full dataset (setB)] [44], and NCBI server (BLASTp) (https://blast.ncbi.nlm.nih.gov/) [45]. The score values significant were bit-score equal to or higher than 50; E-value ≤ 0.0, and an identity equal or higher than 30% [46]. Mauve under ProgressiveMauve (default setting) [47–48] was used to compare syntenic regions. MobileElementFinder (V1.0.3) (https://cge.cbs.dtu.dk/services/MobileElementFinder/) was performed to predict mobile genetic elements (MGEs) associated with antibiotic and virulence factors (Minimum alignment coverage = 95; minimum sequence identity-90%, maximum truncation-30 bp) [49–50].

Analysis of Molecular Evolution

Genetic Algorithm Recombination Detection (GARD) was used to detect recombination. Random effect likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR), and Single-likelihood ancestor counting (SLAC) methods implemented by HyPhy 2.5 under the MG94xREV codon model were used to identify sites under positive selection [50–55]. Evidence of episodic positive selection was verified by the mixed-effects model of evolution (MEME) and Branch-site unrestricted statistical test for episodic diversification (BUSTED). The score values were significant whether it is less than or equal to 0.05 (BUSTED, SLAC, FEL, and MEME models), and for the posterior probability if it is ≥ 95% (FUBAR).

3. Results and Discussions

3.1 Pan-genome

Pan-genome is opened and has about 4232 protein-coding genes, 814 of which belong to the core genome, 3455 are in the set of accessory genomes, and 37 are in the unique genome, on average. Most of which are related to the metabolism of carbohydrates and amino acids (KEGG). Zou et al. [57] described the same pattern in other bacteria of the gut microbiota. According to COG distribution, most of the genes of the flexible gene pool (accessory and unique) are in the 'Information storage and processing' category [Translation, ribosomal structure and biogenesis (J); RNA processing and modification (A); Transcription (K); Replication, recombination and repair (L); Chromatin structure and dynamics (B)]. Analysis of USEARCH results indicated that many flexible genes encode truncated proteins. These genes might have been acquired through HGT because they are often are in mobile element regions.

Orthologous genes distributed in their COG and KEGG general categories are in Additional files, S1 Fig. COG distribution in specific categories is in S2 Fig.

Fig 1. Distribution of orthologous genes into KEGG categories. (a) Frequency of core (in blue), accessory (in red), and unique (green) genes. The arrows point to the table corresponding to the graph bars. (c) Table showing the genes supposedly related to infectious diseases and their features.

KEGG analysis also displayed genes potentially related to human infections, mainly in the flexible genome. These genes often encode proteins with bacterial metabolic functions but may be related to virulence in specific bacteria. For example, the molecular chaperone DNAK protein, encoded by the *dnaK* gene, contributes to stress tolerance and cell growth, but it is also a virulence factor in *Mycobacterium tuberculosis* [58–61]. Fig 1 indicates the distribution of pan-genome across KEGG categories and a table that highlights genes associated with infectious diseases.

Don PUL operon is in all strains, in this way belongs to the core genome. It encodes proteins that enable the deglycosylate N-linked glycans of the gut mucus layer, transferrin, and other glycoproteins [33]. Therefore, the operon is essential for *B. fragilis* adaptation to intra- and extra-intestinal environments. Other genes related to adaptation and virulence are in the accessory genome, such as *bftP*, perforin, LPS locus, and *T6SS* genes.

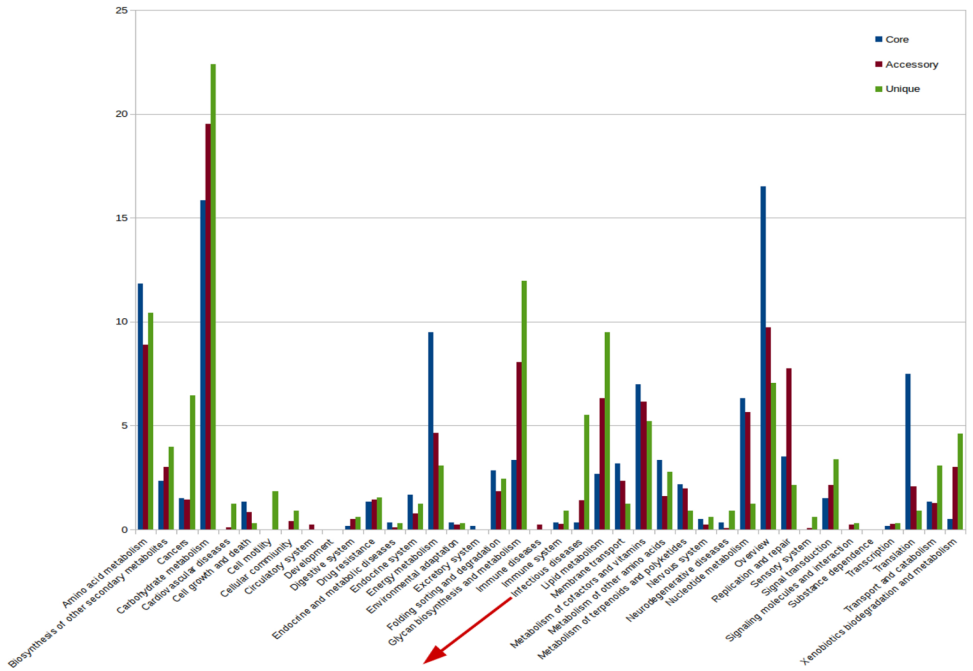
Forty strains have *bftP-1* allele while *bftP-2* is in eight strains (20793-3, 2-078382-3, 86-5443-2-2, BOB25, CL07T00C01, CL07T12C05, CM13 and DS71). No strain has *bftP-3*. BfPAI is absent in 47 strains described as ETBF [62], suggesting that it was not sequenced or annotated in these strains. Pan-genome prediction and the genomic location of BfPAI are in S2 Table.

The perforin gene is in 80 strains, both ETBFs and NTBFs (non-enterotoxigenic *B. fragilis*). The gene encodes a nonfunctional protein in five strains (1007-1-F #8, 3783N1-2, 3783N2-1, 3976T7, and S13 L11) because of the absence of 139 amino acids from the terminal portion of the protein, which includes the MACPF domain (YGTHVLTDITLGG)- critical for protein activity [61]. Perforin [membrane attack complex/perforin (MACPF) domain] is a bacteriocin that can rupture the plasma membrane of a cell host and might be a virulence factor [63]. It is found mainly in eukaryotes and some bacteria, including *Bacteroides spp*. [64].

LPS locus was described in 638R and NCTC9343 strains [31]. We also reported it in thirteen strains [DCMOUH0042B, 3719 A10, 3986 N3, 3986 T (B) 13, 3986 T (B) 9, AD126T_1B, AD126T_2B, ATCC 25285, BFR_KZ09, CL04T03C20, HMW 615, KLE1257, and KLET1758]. Most of them

were isolated from extraintestinal infections (abscess, blood, or tissue). The capsular polysaccharide complex is one of the major virulence factors related to abscess formation [29, 31], and this might explain the source of the strains.

a)



b)

Gene	Definition	Function	Description	Pathogen	Kegg	Pathway	Genomes	ID%	Bit-score	E-value
InlA	Internalin/Ig-like domain	Adhesion	Bacterial invasion of epithelial cells	*Listeria monocytogenes*	K13730	ko05100	3725 D9 ii	47.9	49	2.21E-05
sdhA	Succinate dehydrogenase / Fumarate reductase	Tricarboxylic acid cycle	Legionellosis	*Legionella pneumophila* *Legionella longbeachae*	K00239	ko05134	Core genome	30	141	5.36E-33
rpoN	RNA polymerase sigma-54 factor	Biofilm formation	*Vibrio cholerae* infection	*Vibrio cholerae*	K03092	ko05111	Acessory genome (178)	92.5	905	5.32E-263
dltA	D-alanine--poly(phosphoribitol) ligase subunit 1	Antimicrobial activity	*Staphylococcus aureus* infection	*Staphylococcus aureus*	K03367	ko05150	Acessory genome (178)	31.3	199	1.55E-50
dnaK	Molecular chaperone DnaK	Stress response	Tuberculosis bacterial	*Mycobacterium tuberculosis*	K04043	ko05152	Core genome	95.6	1193	0.0
ACTB_G1	Actin beta/gamma 1	Actin polymerization	Shigellosis	*Shigella dysenteriae* *Shigella flexneri* *Shigella boydii* *Shigella sonnei*	K05692	ko05131	AD135F_2B	92.5	905	5.32E-263
ACTB_G1	Actin beta/gamma 1	Actin polymerization	*Vibrio cholerae* infection	*Vibrio cholerae*	K05692	ko05110	AD135F_2B	97.6	258	5.34E-69
prtC	U32 family peptidase	Cell signaling	Epithelial cell signaling in *H. pylori* infection	*Helicobacter pylori*	K08303	ko05120	3725 D9 ii	95.1	815	6.27E-236
tolC	TolC family protein (BF_RS08835)	Secretion	Pertussis	*Bordetella pertussis*	K12340	ko05133	Core genome	80	728	8.55E-210

Fig 1. Distribution of orthologous genes into KEGG categories. (a) Frequency of core (in blue), accessory (in red), and unique (green) genes. The arrows point to the table corresponding to the graph bars. (c) Table showing the genes supposedly related to infectious diseases and their features.

Fifteen complete genomes have GA3 T6SS genetic architecture [27], except the genomes of DCMOUH0067B, DCMSKEJBY0001B, FDAARGOS_763, Q1F2 strains, which belong to the carbapenem-resistant *B. fragilis* group [63]. This finding suggests T6SS is nonessential to *B. fragilis*.

3.2 Prediction of CRISPR-Cas system

Results revealed that 76,4% of the genomes have a CRISPR-Cas system. Fifty-eight strains have two CRISPR-Cas types in their genomes. Three types are only in six strains (AF26-6, BFR_KZ08, S14, S38L3, S38L5, and TM08-15). CRISPR-Cas system is absent in most NTBFs, whereas the IIIB type is the most common among ETBFs, corroborating Tajkarimi and Wexler [64]. Monophyletic groups often share the same CRISPR-Cas system. A phylogenetic tree displaying the CRISPR-Cas type of each strain is in S3 Fig.

3.3 Antimicrobial Resistance Genes

The most frequent antimicrobial resistance genes in *B. fragilis* are *cepA* [Class A beta-lactamase (EC:3.5.2.6)] and *tetQ* (Tetracycline-resistant ribosomal protection protein). Only 40 strains presented a 'perfect match' (100%) for *cepA*, but the results showed high identity values in 158 strains (range from 98 to 99%, ARO: 3003559). Thus, the functionality of *cepA* gene is probable.

Thirty-six strains have the *tetQ* gene with a 'perfect match', and 125 strains presented a gene with a 'strict match' (ARO: 3000191), but likely functional, according to the comparison to the protein described by Lépine et al. [65]. *tetQ* is the most abundant gene for antimicrobials in the human gut microbiota, as reported in a metagenomics analysis [66]. Therefore, the spread of the gene seems advantageous.

ermF gene [23S rRNA (Adenine (2058)-N (6))-methyltransferase (EC: 2.1.1.-)] is in 43 strains with a 'strict match' (identity from 98 to 99.9%), but their binding sites - essential for protein functionality [67] are unchanged. This gene confers multidrug resistance (macrolide, lincosamide, and streptogramin B) [68] and has a clinical impact on treatment. Antimicrobial resistance genes present in the strains are in S3 Table.

aadS [Aminoglycoside 6-nucleotidyltransferase (EC: 2.7.1.95)], *aphA* [Aminoglycoside 3'-phosphotransferase (EC: 2.7.1.95)], *ccrA* [Subclass B1 beta-lactamase (EC 3.5.2.6)], *OXA-347* [Beta-lactamase class D OXA-209 (EC: 3.5.2.6)], *SUL2* [Dihydropteroate synthase type-2 (EC

2.5.1.15)], and *TEM-1* [Beta-lactamase class A TEM (EC: 3.5.2.6)] genes had 'perfect matches', but they were restricted to a few strains.

3.4 Mobilome

a) Analysis of Prophages

Results indicated the complete genomes have two (BOB25 strain) to fifteen (DCMOUH0067B and FDAASGOS_763 strains) prophage regions. Most virulence-related proteins in these regions had low scores, except for EF-Tu, which had higher values [VFDB database = VFG046475 (F7308_0636) Tu translation elongation factor (EF-Tu) *(Francisella sp*. TX077308) ID = 64%; bit-score = 540, and E-value = e-154]. EF-Tu protein performs various functions in bacteria, including catalyzing the binding of aminoacyl-tRNA to the ribosome, forming a biofilm, and decreasing the host immune response [69]. It is unclear whether it plays a role related to virulence in *B. fragilis*. We found no antimicrobial resistance genes in the prophage region (S4 Table).

b) ICEFinder Analysis

The size of ICEs varies from 24.6 kilobase pairs (kb) to 193 kb, while EMIs are smaller (5 kb to 76.8 kb). Genes for antimicrobial resistance, mainly *tetQ*, were identified in ICEs and IMEs. Complete genomes have at least one putative ICE with T4SS, except those from BE1 and Q1F2 with only putative IMEs. S5 Table indicates the genomic location of the ICEs, IMEs and the antibiotic resistance genes found in them.

ICEFinder, Mauve, and NCBI BLAST search (Bit-score = 70088; e-value = 1.020e + 05; ID = 98%) analyses supported the hypothesis of an HGT (horizontal gene transfer) event between the genomic regions of *B. fragilis DCMOUH0085B* (position 2993358 to 3044318) and one from *Butyricimonas faecalis* H184 (Accession no. CP032819.1). The two regions also share the genes for antimicrobial resistance, *AADE* (aminoglycoside 6-adenylyltransferase gene), *ermB* [rRNA adenine N-6-methyltransferase gene], *(APH (3') - IIIA* {ANT-like pseudogene (9) ANT-like aminoglycoside (9) aminoglycoside nucleotidyltransferase family gene]. HGT among Bacteroidales species was also reported by Coyne et al. [70] and García-Bayona et al. [71].

The results of the Mauve and ICEFinder analyses are in Fig 2 and Table 1.

Fig 2. Insertion elements found in the strains. Frequency of insertion elements inferred by MobileElementFinder. The green bar corresponds to the insertion elements with an antimicrobial resistance gene, the red bar denotes insertion elements without antimicrobial resistance gene, and the gray bar shows the untransferred plasmid (repUS2).

Table 1. Features of the insertion elements.

Strain	Accession Number	IS Type	IS Family	Position in Contigs	ID (%)
2_1_16	GG705215.1	ISBthe1- *nimB*	IS4	39206-40521	98.9
320_BFRA	JVLR01000059.1	ISBf8 (Tn4555)- *cfxA*	IS4	1-1663	99.7
885_BFRA	JUPJ01000031.1	IS4351 (Tn4351)- *tetX*	IS30	2258-3424	99.9
894_BFRA	JUOZ01000029.1	IS4351 (Tn4351)- *tetX*	IS30	2258-3424	99.9
915_BFRA	JUOA01000729.1	IS4351 (Tn4351)- *tetX*	IS30	3147-3947	100
BF8 Consensus_4	LGTH01000004.1	IS4351 (Tn4351)- *tetX*	IS30	924279-925445	99.9
	LGTH01000004.1	IS4351 (Tn4351)- *ermF*	IS30	922085-923239	100
BF8 Consensus_1	LGTH01000001.1	IS1168- *nimB*		612426-612920	100
BF8 Consensus_2	LGTH01000002.1	IS942- *cfiA*	IS4	1220392-1221141	99.1
CL07T00C01	AGXM01000001.1	IS4351 (Tn4351)- *ermF*	IS30	582395-583195	100

A genomic region from the CL03T12C07 strain (position 3350878 to 3378618) is a putative ICE with T4SS that includes a CtnDOT but its *ermF* gene is unfunctional. CtnDOT is a transposon widespread among *Bacteroides spp.* and contains the genes for antimicrobial resistance, *ermF*, *tetQ*, *tetX* (flavin-dependent monooxygenase gene) and *tetX1* (flavin-dependent monooxygenase gene) [72-74].

We also identified in DCMSKEJBY0001B strain a putative IME without identified DR that harbours CTn341. This conjugative transposon is typical of *B. fragilis* and carries the *tetQ* gene. CTn341 (or CTn341-like) may also be found in other Bacteroidates, such as *Bacteroides sp.* PHL 2737, *B. xylanisolvens XB1A*, and *Phocaeicola vulgatus* VIC01is. Thus, CTn341 is apparently the major disseminator of the *tetQ* gene within Bacteroidales (S4 Fig).

c) Conjugative Transposon

Among the conjugative transposons previously described in *Bacteroides spp.*, we reported CTnDOT (aforementioned), CTn86, and CTn341.

We have identified a CTn86-like (Accession no. AY372755.1) in thirteen ETBF strains (2-F-2#5, 3397 N3, 078320–1, 1001285H_161024_D4, AM31-13AC, BFR_KZ09, CL07T00C01, CL07T12C05, CL05T12C13, HMW 615, HMW 616, TL139C_1B, and TL139C_1B2). The *PER* gene, which encodes an ABC permease transporter protein, was excluded from these strains, possibly because of transposon insertion, as described by Franco [13]. CTn86 was also identified in 20793–3, 86-5443-2-2, and BOB25 strains [13, 75–76].

CTn341-like is in 20656-2-1, AF14-14AC, and AF14-26 strains. It does not have all CTn341 genes and attachment sites (*attL* and *attR*) [77]. It has genes for transfer (*traG*, *traI*, *traJ*, and *traM* genes), integration (*int* gene), mobilization (*mobA*, *mobB*, and *mobC*), and excision (*exc* gene) [78–80], besides *tetQ* (S5 Fig).

c) Insertion Elements (IS)

ISBf5 (IS1182 family) is the insertion element most common in *B. fragilis* genomes. It has 1830 bp and carries one transposase gene [81].

ISBf1 (IS21 family) [82] is in some strains, but its *cepA* gene is incomplete in eight strains (885_BFR, AD135F_2B, AF32-10, Am47-7, COR2-248-WT-1, HCK-B3, HAP130N_2B, TM08 -15) and absent in AD135F_1B, AD135F_3B, ATCC 25285, and Ds-233 strains. However, they have in their genomes a *cepA* functional with the insertion of IS1224 in their promoter region that increases the *cepA* expression [82–83].

Functional resistance genes were identified in ISBthe1 (*nimB* gene, metronidazole), ISBf8 (*cfxA* gene), and IS4351 (*tetX* gene). ISBf8 and IS4351 are within the non-conjugative transposons (Tn4555 and Tn4351, respectively), and for that reason, they cannot spread their genes.

Therefore, the spread of genes for antibiotic resistance in *B. fragilis* depends on conjugative transposons, as suggested by Whittle et al. [84], Yan, Hall and Jiang [85], and Johnson and Grossman [23], and of ICE with T4SS, as we observed in our results. IS frequency and its genes for antimicrobial resistance are in Fig 3 and the genomic location of the insertion elements is shown in S6 Table.

Fig 3. Comparison between two putative ICEs with T4SS from *Bacteroides fragilis* DCMOUH0085B and *Butyricimonas faecalis* H184 genomes. In (a) shown the conserved synteny

between these genomic regions; and b) indicates their features in DCMOUH0085B and H184. Antimicrobial resistance genes and their localization in the genomes are indicated in the figure.

3.5 Analysis of Molecular Evolution

a) Molecular Evolution of Antibiotic Resistance Genes

Results indicated evidence of positive selection in the *cepA* (beta-lactamase gene) and *tetQ* genes. Codon 88 of the *cepA* gene is under positive selection (FUBAR method; Posterior posterior = 0.9878). The substitutions found do not contain any putative ribosome-binding sites indicated by Rogers et al. [86].

tetQ alignment displayed 116 variable sites. FUBAR method inferred three codon sites under positive selection: 35 (posterior probability = 0.9676), 237 (posterior probability = 0.9669), and 426 (posterior probability = 0.9623) while MEME suggested episodic positive selection at 354 codon (p-value = 0.003), and 550 (p-value = 0.0464). Comparative analysis with the *tetQ* protein indicated that the four GDP-GTP binding sites are intact, as described by Lépine et al. [65].

b) Molecular Evolution in Pan-genome

At least 396 genes are under positive selection (episodic or pervasive). Most of them encode membrane proteins, such as the adenosine triphosphate (atp)-binding cassette (abc) superfamily (*macB* genes), resistance nodulation-division (*rnD*) family (*macA, mdtA, mdtE*, and *mexA* genes), and tonb-dependent receptor (*susC* and *susD* genes).

ATP-binding and RND proteins are efflux pumps that facilitate the elimination of toxins and are related to antimicrobial resistance phenotypes [87–90].

TonB-dependent receptors contribute to the transport and nutrient absorption [90]. Recently, they have been the subject of research for developing vaccines against Gram-negative bacteria because of their cell multiple functions [92–93].

Seven paralogous genes of the Colicin I receptor (*cirA* genes) also had several sites under positive selection. These proteins belong to the TonB-dependent copper receptor family (IPR010100) and perform the transport of nutrients across the periplasmic space [94–95]. Colicins in *Escherichia coli* are toxins that act in intraspecific competition [96–97], but their role in *B. fragilis* is unclear. Yoshizaki et al. [90] compared the genomes of 638R, NCTC9343, and YCH46 and found 52 genes under positive selection. The greater diversity used in our work might explain the difference between the results. The results of Hyphy are in S7 Table.

4. Conclusions

HGT is an event of extreme evolutionary relevance for *B. fragilis*. The *tetQ* gene spread was mainly by ICE with T4SS and conjugative transposons, whereas the *cepA* gene was an ancient spread through a mobile element that lost its function. We can describe HGT events not yet reported in previous research. Nonetheless, greater availability of complete genomes is required for the verification and validation of these results. Besides antimicrobial resistance genes, other genes related to efflux pumps facilitating the exit of substances from the cells; transport and absorption of nutrients; and alternative carbon acquisition (DON PUL) might provide adaptive advantages for *B. fragilis*. There is no evidence that *EF-Tu*, *dnaK,* or colicin genes are associated with virulence and need further research to clarify their functions in *B. fragilis*.

Abbreviation:

AMR: Antimicrobial multidrug resistance

BFT: Fragilysin

CRISPR: Clustered regularly interspaced short palindromic repeat

ETBF: Enterotoxigenic *Bacteroides fragilis*

HGT: Horizontal gene transfer

ICE: Integrative and conjugative elements

IME: Integrative mobilizable elements

NTBF: Non-enterotoxigenic *B. fragilis*

Supporting information

S1 Table. *Bacteroides fragilis* strains used in this study.

S2 Table. Pan-genome prediction of 183 *B. fragilis* strains.

S3 Table. Antibiotic resistance genes identified by CARD.

S4 Table. Results obtained by PhiSpy compared to the VFDB database.

S5 Table. Results obtained by ICEFinder compared to NCBI BLAST and CARD databases.

S6 Table. Results obtained by MobileElementFinder.

S7 Table. Results obtained by Hyphy software.

S1 Fig. Pan-genome prediction.

S2 Fig. COG distribution of genes found in the core, accessory and unique genomes.

S3 Fig. Pan-genome tree indicating the features of each strain.

S4 Fig. Comparison among the CTn341 from DCMSKEJBY0001B, *Phocaeicola vulgatus* VIC01, *Bacteroides sp.* PHL 2737, and *B. xylanisolvens* XB1A.

S5 Fig. Comparison among the CTn341 from AF14-14AC, AF14-26 and 20656-2-1 strains.

References

[1] S. M. Finegold, “Overview of clinically important anaerobes”, *Clin Microbiol Rev*, vol. 20, pp. 205-207, 1995.

[2] J. L. Round and S. K. Mazmanian, “Inducible Foxp3þ regulatory T-cell development by a commensal bacterium of the intestinal microbiota”, *Proc Natl Acad Sci U S A*, vol. 107, pp. 12204-12209, 2010.

[3] D. Qu, F. Sun, S. Feng et al., "Protective effects of Bacteroides fragilis against lipopolysaccharide-induced systemic inflammation and their potential functional genes", Food Funct, vol. 12, pp. 1015-1025. 2022.

[4] C.L. Sears, "Enterotoxigenic *Bacteroides fragilis*: a Rogue among Symbiotes", *Clin Microbiol Rev*, vol. 22, 349-69, 2009.

[5] H. W. Wexler, "*Bacteroides*: the Good, the Bad, and the Nitty-Gritty", *Clin Microbiol Rev,* vol. 20, pp. 593-621, 2007.

[6] N. Anka, T. V. Sydenhamb, L. H. Iversenc et al., "Characterisation of a multidrug-resistant *Bacteroides fragilis* isolate recovered from blood of a patient in Denmark using whole-genome sequencing", *Int J Antimicrob Agents* vol. 46, pp. 117-120, 2015.

[7] J. Papaparaskevas, A. Katsandri, A. Pantazatou et al., "Epidemiological characteristics of infections caused by *Bacteroides*, *Prevotella* and *Fusobacterium* species: A prospective observational study", *Anaerobe,* vol. 17, pp. 113-117, 2011.

[8] E. Valguarnera and J. B. Wardenburg, "Good gone bad: One toxin away from disease for *Bacteroides fragilis", J Mol Biol*, vol. 432, pp. 765-785, 2020.

[9] H. Elsaghir and A. K. R. Reddivari (2021) *Bacteroides fragilis*, https://www.ncbi.nlm.nih.gov/books/NBK553032/ (March 18, 2021)

[10] M. Vacante, R. Ciuni, F. Basile and A. Biondi, "Gut Microbiota and Colorectal Cancer Development: A Closer Look to the Adenoma-Carcinoma Sequence", *Biomedicines*, vol. 8pp. 489, 2020.

[11] W. T. Cheng, H. K. Kantilal and F. Davamani, "The Mechanism of *Bacteroides fragilis* toxin Contributes to Colon Cancer Formation", *Malays J Med Sci*, vol. 27, pp. 9–21, 2020.

[12] J. S. Moncrief, A. J. Duncan, R. L. Wright et al., "Molecular characterization of the fragilysin pathogenicity islet of enterotoxigenic *Bacteroides fragilis"*, *Infect Immun*, vol. 66, pp. 1735-1739, 1998.

435 [13] A. A. Franco, "The *Bacteroides fragilis* pathogenicity island is contained in a putative novel
436 conjugative transposon", *J Bacteriol*, vol. 186, pp. 6077-6092, 2004.
437
438 [14] S. A Shiryaev, A. E. Aleshin and N. Muranaka, "Structural and functional diversity of
439 metalloproteinases encoded by the *Bacteroides fragilis* pathogenicity island", *FEBS J*. vol. 281, pp.
440 2487-2502, 2014.
441
442 [15] J. Herrou, V. M. Choi, J. M. Wardenburg and S. Crosson, "Activation Mechanism of the
443 *Bacteroides fragilis* Cysteine Peptidase, Fragipain", *Biochemistry,* vol. 55, pp. 4077–4084, 2016.
444
445 [16] G. Bamias and F. Cominelli, "Cytokines and intestinal inflammation", *Curr Opin*
446 *Gastroenterol*,vol. 32, pp. 437-442, 2016.
447
448 [17] S. H. Ko, D. J. Rho, J. I. Jeon et al., "*Bacteroides fragilis* Enterotoxin Upregulates Heme
449 Oxygenase-1 in Intestinal Epithelial Cells via a Mitogen-Activated Protein Kinase- and NF-
450 kappaB-Dependent Pathway, Leading to Modulation of Apoptosis", *Infect Immun*, vol. 84, pp.
451 2541–2554, 2016.
452
453 [18] A. A. Franco, L. M. Mundy, M. Trucksis et al., "Cloning and characterization of the
454 *Bacteroides fragilis* metalloprotease toxin gene", *Infect. Immun*, vol. 65, pp. 1007-1013, 1997.
455
456 [19] N. Kato, C. X. Liu, H. Kato et al, "A new subtype of the metalloprotease toxin gene and the
457 incidence of the three bft subtypes among *Bacteroides fragilis* isolates in Japan", *FEMS Microbiol*
458 *Lett*, vol. 182, pp; 171-176, 2000
459
460 [20] C. L. Sears, "The toxins of *Bacteroides fragilis", Toxicon*, vol. 39:1737-1746, 2001.
461
462 [21] A. A. Franco, R. K. Cheng, A. Goodman and C. L. Sears CL, "Modulation of bft expression by
463 the *Bacteroides fragilis* pathogenicity island and its flanking region", *Mol Microbiol*, vol. 45, pp.
464 1067-77, 2002.
465

[22] S. Niestępski, M. Harnisz, E. Korzeniewska et al., "*Bacteroides spp*. - clinical significance, antibiotic resistance and identification methods", *Postepy Mikrobiol*, vol. 56, pp. 67–76, 2017.

[23] C. M. Johnson and A. D. Grossman, "Integrative and Conjugative Elements (ICEs): What They Do and How They Work", *Annu Rev Genet*, vol. 49, pp. 577-601, 2015.

[24] J. Guglielmini, L. Quintais, M. P. Garcillán-Barcia et al., "The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation", *PLoS Genet*, vol. 7, pp. e1002222, 2011.

[25] A. B. Russell, R. D, Hood, N. K. Bui et al., "Type VI secretion delivers bacteriolytic effectors to target cells", *Nature,* vol. 475, pp. 343–347, 2011.

[26] B. T. Ho, T. G, Dong and J. J. Mekalanos, "A view to a kill: the bacterial type VI secretion system", *Cell Host Microbe*, vol. 15, pp. 9–21, 2014.

[27] M. Chatzidaki-Livanis, N. Geva-Zatorsky and L. E. Comstock, "*Bacteroides fragilis* type VI secretion systems use novel effector and immunity proteins to antagonize human gut Bacteroidales species", *Proc Natl Acad Sci U S A,* vol. 113, pp. 3627–3632, 2016.

[28] R. A. Adegbola, "Bacterial adhesion and pathogenicity", *Afr J Med Med Sci,* vol. 17, pp. 63-69, 1988.

[29] A. B. Onderdonk, D. L. Kasper, R. L. Cisneros, et al., "The capsular polysaccharide of *Bacteroides fragilis* as a virulence factor: comparison of the pathogenic potential of encapsulated and unencapsulated strains", *J Infect Dis*, vol. 136, pp. 82-89, 1977.

[30] M. J. Coyne, A. O. Tzianabos, B. C. Mallory, et al., "Polysaccharide Biosynthesis Locus Required for Virulence of *Bacteroides fragilis", Infect Immun,* vol. 69, pp. 4342–4350, 2001.

[31] L. E. Comstock, M. J. Coyne, A. O. Tzianabos, et al., "Interstrain Variation of the Polysaccharide B Biosynthesis Locus of *Bacteroides fragilis*: Characterization of the Region from Strain 638R", *J Bacteriol,* vol. 181, pp. 6192–6196, 1999.

[32] L. K. Vestby, T. Grønseth, R. Simm et al., "Bacterial Biofilm and its Role in the Pathogenesis of Disease", *Antibiotics*, vol. 2, pp. 59, 2020.

[33] Y. Cao, E. R, Rocha and C. J. Smith, "Efficient utilization of complex N-linked glycans is a selective advantage for *Bacteroides fragilis* in extraintestinal infections", *Proc Natl Acad Sci U S A,* vol. 111, pp. 12901–12906, 2014.

[34] D. Sims, I. Sudbery, N. E. Ilott et al., "Sequencing depth and coverage: key considerations in genomic analyses", *Nat Rev Genet*,vol. 15, pp. 121-132, 2014.

[35] N. M. Chaudhari, V. K. Gupta and C. Dutta, "BPGA- an ultra-fast pan-genome analysis pipeline", *Sci Rep*, vol. 6, pp. 24373, 2016.

[36] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput", *Nucleic Acids Res*, vol. 32, pp. 1792–1797, 2004.

[37] R. C. Edgar, R. C, "Search and clustering orders of magnitude faster than blast", Bioinformatics, vol. 26, pp. 2460–2461, 2010.

[38] T. Williams and C. Kelley (2018) Gnuplot 4.6.6, http://www.gnuplot.info (March, 18, 2021).

[39] D. Couvin, A. Bernheim, C. Claire Toffano-Nioche et al., "CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins", *Nucleic Acids Res*, vol. 46, pp. W246–W251, 2018.

[40] B. Jia, A. Raphenya, B. Alcock et al., "CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database", *Nucleic Acids Res*, vol. 45, pp. D566-D573, 2017.

[41] S. Akhter, R. K. Aziz and R.A. Edwards, "PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies", *Nucleic Acids Res,* vol. 40, pp. e126, 2012.

[42] M. Liu, X. Li, Y. Xie et al., "ICEberg 2.0: an updated database of bacterial integrative and conjugative elements", *Nucleic Acids Res*, vol. 47, pp. D660–D665, 2019.

[43] B. P. Alcock, A. R. Raphenya, T. T. Y. Lau et al., "CARD 2020: Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance Database", *Nucleic Acids Res*, vol. 48, pp. D517-D525, 2020.

[44] L. Chen, J. Yang, J. Yu J et al., "VFDB: a reference database for bacterial virulence factors", *Nucleic Acids Res*, vol. 33, pp. D325-8, 2005.

[45] S. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: A new generation of protein database search programs", *Nucleic Acids Res,* vol. 25, pp. 3389–3402, 1997.

[46] W. R. Pearson, "An introduction to sequence similarity ("homology") searching", *Curr Protoc Bioinform*, vol. 3, pp. 3.1, 2013.

[47] A. E. Darling, B. Mau, F. R. Blattner et al., "Mauve: multiple alignment of conserved genomic sequence with rearrangements", *Genome Res*, vol. 14, pp. 1394–1403, 2004.

[48] A. E. Darling, B. Mau and N. T Perna, "ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement", *PLoS One*, vol. 5, pp. e11147, 2010.

555 [49] M. H. K. Johansson, V. Bortolaia, S. Tansirichaiya et al., "Detection of mobile genetic
556 elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web
557 tool: MobileElementFinder", *J Antimicrob Chemother*, vol. 76, pp. 101–109, 2021.
558
559 [50] S. Guindon, J. F. Dufayard and V. Lefort et al., "New Algorithms and Methods to Estimate
560 Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0", *Syst Biol*, vol. 59,
561 pp. 307-321, 2010.
562
563 [51] S. L. Kosakovsky Pond, D. Posada, M. B. Gravenor et al., "Automated Phylogenetic Detection
564 of Recombination Using a Genetic Algorithm", *Mol Biol Evol*, vol. 23, pp. 1891-1901, 2006.
565
566 [52] S. L. Kosakovsky Pond and S. D. W. Frost, "Not So Different After All: A Comparison of
567 Methods for Detecting Amino Acid Sites Under Selection", *Mol Biol Evol*, vol. 22, pp. 1208-1222,
568 2005.
569
570 [53] B. Murrell, S. Moola, A. Mabona et al., "FUBAR: a fast, unconstrained bayesian
571 approximation for inferring selection", *Mol Biol Evol*, vol. 30, pp. 1196-1205, 2013.
572
573 [54] S. L. Kosakovsky Pond, A. F. Y. Poon, R. Velazquez et al., "HyPhy 2.5—A Customizable
574 Platform for Evolutionary Hypothesis Testing Using Phylogenies", *Mol Biol Evol*, vol. 37, pp. 295-
575 299, 2020.
576
577 [55] B. Murrell, J. O. Wertheim, S. Moola et al., "Detecting Individual Sites Subject to Episodic
578 Diversifying Selection", *PLoS Genetics*, vol. 8, pp. e1002764, 2012.
579
580 [56] B. Murrell, S. Weaver, M. D. Smith et al., "Gene-Wide Identification of Episodic Selection",
581 *Mol Biol Evol,* vol. 32, pp. 1365-1371, 2015.
582
583 [57] Y. Zou, W. Xue, G. Luo et al. "1,520 reference genomes from cultivated human gut bacteria
584 enable functional microbiome analyses", *Nat Biotechnol*, vol. 37, pp. 179–185, 2019.
585

[58] T. Tomoyasu, A. Tabata, H. Imaki et al., "Role of *Streptococcus intermedius* DnaK chaperone system in stress tolerance and pathogenicity", *Cell Stress Chaperon*, vol. 17, pp. 41–55, 2011.

[59] V. K. Singh, S. Utaida, L. S. Jackson et al., "Role for *dnaK* locus in tolerance of multiple stresses in *Staphylococcus aureus", Microbiology*, vol. 153, pp. 3162–3173, 2007.

[61] K. R. Fourie and H L. Wilson, "Understanding GroEL and DnaK Stress Response Proteins as Antigens for Bacterial Diseases", *Vaccines,* vol. 8, pp. 773, 2020.

[62] 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.

[63] M. Chatzidaki-Livanis, M. J. Coyne and L. E. Comstock, "An antimicrobial protein of the gut symbiont Bacteroides fragilis with a MACPF domain of host immune proteins", *Mol Microbiol*, vol. 94, pp. 1361–1374, 2914.

[64] Q. Xu, P. Abdubek, T. Astakhova et al., "Structure of a membrane-attack complex/perforin (MACPF) family protein from the human gut symbiont *Bacteroides thetaiotaomicron*", *Acta Crystallogr Sect F Struct Biol Cryst Commun*, vol. 66, pp. 1297-305, 2010.

[63] S. Valdezate, F. Cobo, S. M. et al., "Genomic Background and Phylogeny of *cfiA*-Positive *Bacteroides fragilis* Strains Resistant to Meropenem-EDTA", *Antibiotics (Basel)*, vol. 10, pp. 304, 2021.

[64] M. Tajkarimi and H. W. Wexler, "CRISPR-Cas Systems in *Bacteroides fragilis*, an Important Pathobiont in the Human Gut Microbiome", *Front Microbiol,* vol. 23, pp. 2234, 2017.

[65] G. Lépine, J. M. Lacroix and C.G Walker, "A Progulske-Fox. Sequencing of a *tet(Q)* gene isolated from *Bacteroides fragilis* 1126", *Antimicrob Agents Chemother*, 37:2037-2041, 2013.

617 [66] Y. Hu, X. Yang, J. Qin et al., "Metagenome-wide analysis of antibiotic resistance genes in a
618 large cohort of human gut microbiota", *Nat Commun*, vol. 4, pp. 2151, 2013.
619
620 [67] M. C. Halula, S. Manning and F. L. Macrina, "Nucleotide sequence of *ermFU*, a macrolide-
621 lincosamide-streptogramin (MLS) resistance gene encoding an RNA methylase from the conjugal
622 element of *Bacteroides fragilis* V503", *Nucleic Acids Res,* vol. 19, pp. 3453, 1991.
623
624 [68] E. K. Park, H. Kim and H. J. Jin, "Phylogenetic analysis of rRNA methyltransferases, Erm and
625 KsgA, as related to antibiotic resistance", *FEMS Microbiol Lett*, vol. 309, pp. 151–162, 2010.
626
627 [69] K. L. Harvey, V. M. Jarocki, I. G. Charles et al., "The Diverse Functional Roles of Elongation
628 Factor Tu (EF-Tu) in Microbial Pathogenesis", *Front Microbiol*, vol. 10:2351, 2019.
629
630 [70] M. J. Coyne, N. L. Zitomersky, A. M. McGuire et al., "Evidence of Extensive DNA Transfer
631 between Bacteroidales Species within the Human Gut", *mBio*, vol. 5, pp. e01305-14, 2014.
632
633 [71] L. García-Bayona, M. J. Coyne and L. E. Comstock, "Mobile Type VI secretion system loci of
634 the gut Bacteroidales display extensive intra-ecosystem transfer, multi-species spread and
635 geographical clustering", *PLoS Genet*, vol. 17, pp. e1009541, 2021.
636
637 [72] G. Whittle, B. D. Hund, N. B. Shoemaker et al., "Characterization of the 13-kilobase ermF
638 region of the Bacteroides conjugative transposon CtnDOT", *Appl Environ Microbiol*, vol. 67, pp.
639 3488-95, 2001.
640
641 [73] N. B. Shoemaker, R. D. Barber and A. A. Salyers, "Cloning and characterization of a
642 Bacteroides conjugal tetracycline-erythromycin resistance element by using a shuttle cosmid
643 vector", *J Bacteriol*, vol. 171, pp. 1294–1302, 1989.
644
645 [74] F. Delavat, R. Miyazaki, N. Carraro et al., "The hidden life of integrative and conjugative
646 elements", *FEMS Microbiology Reviews*, vol. 41, pp. 512–537, 2017.
647

[75] J. V. Pierce and H. D. Bernstein, "Genomic Diversity of Enterotoxigenic Strains of *Bacteroides fragilis",* PLoS One, vol. 11, pp. e0158171, 2016.

[76] A. S. Nikitina, D. D. Kharlampieva, V. V. Babenko et al., "Complete Genome Sequence of an Enterotoxigenic *Bacteroides fragilis* Clinical Isolate", *Genome Announc*, vol. 3, pp. e00450-15, 2015.

[77] G. Guédon, V. Libante, C. Coluzzi et al., "The Obscure World of Integrative and Mobilizable Elements, Highly Widespread Elements that Pirate Bacterial Conjugative Systems", *Genes (Basel)* , vol. 8, pp. 337, 2017.

[78] G. T. Bonheyo, B. D. Hund, N. B. Shoemaker et al., "Transfer region of a *Bacteroides* conjugative transposon contains regulatory as well as structural genes", *Plasmid*, vol. 46, pp. 202–209, 2001.

[79] N. García, G. Gutiérrez, M. Lorenzo et al., "Gene Context and DNA rearrangements in the carbapenemase locus of division II strains of *Bacteroides fragilis", Antimicrob Agents Chemother,* vol. 53, pp. 2677-8, 2009.

[80] A. Haggoud, G. Reysset, H. Azeddoug et al., "Nucleotide sequence analysis of two 5-nitroimidazole resistance determinants from *Bacteroides* strains and of a new insertion sequence upstream of the two genes", *Antimicrob Agents Chemother*, vol. 38, pp. 1047-51, 1994.

[81] T. Kuwahara, A. Yamashita, H. Hirakawa et al., "Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation", *Proc Natl Acad Sci U S A,, vol.* 101, pp. 14919-14924, 2004.

[82] M. B. Rogers, T. K. Bennett, C. M. Payne et al., "Insertional activation of *CEPA* leads to high-level beta-lactamase expression in *Bacteroides fragilis* clinical isolates", *J Bacteriol*, vol. 176, pp. 4376–4384, 1994.

679 [83] J. Vandecraen, M. Chandler, A. Aertsen et al., “The impact of insertion sequences on bacterial
680 genome plasticity and adaptability”, *Crit Rev Microbiol*, vol. 43, pp. 709-730, 2017.
681
682 [84] G. Whittle, N. B. Shoemaker and A. A. Salyers, “The role of *Bacteroides* conjugative
683 transposons in the dissemination of antibiotic resistance genes”, *Cell Mol Life Sci,* vol. 59, pp.
684 2044-54, 2002.
685
686 [85] W. Yan, A. B. Hall and X. Jiang X, "Bacteroidales species in the human gut are a reservoir of
687 antibiotic resistance genes regulated by invertible promoters", npj Biofilms Microbiomes, vol. 8,
688 2022.
689
690 [86] M. B. Rogers, A. C. Parker, C. J. Smith, “Cloning and characterization of the endogenous
691 cephalosporinase gene, *CEPA*, from *Bacteroides fragilis* reveals a new subgroup of Ambler class A
692 beta-lactamases”, *Antimicrob Agents Chemother*, vol. 37, pp. 2391-400, 1993.
693
694 [87] X. Z. Li and H. Nikaido, “Efflux-mediated drug resistance in bacteria”, *Drugs*, vol. 64, pp.
695 159-204, 2004.
696
697 [88] A. L. Pimenta, K. Racher, L. Jamieson et al., Mutations in *HlyD*, part of the type 1 translocator
698 for hemolysin secretion, affect the folding of the secreted toxin, *J. Bacteriol*, vol. 187, pp. 7471-80,
699 2005.
700
701 [89] L. J. Piddock, “Multidrug-resistance efflux pumps - not just for resistance”, *Nat Rev*
702 *Microbiol,* vol. 4, pp. 629-36, 2006.
703
704 [90] A. L. Davidson, E. Dassa, C. Orelle et al., “Structure, function, and evolution of bacterial ATP-
705 binding cassette systems”, *Microbiol Mol Biol Rev,* vol. 72, pp. 317-364, 2008.
706
707 [91] S. Yoshizaki, T. Umemura, K. Tanaka et al., “Genome-wide evidence of positive selection in
708 *Bacteroides fragilis”, Computational Biology and Chemistry*, vol. 52, pp. 43-50, 2014.
709

[92] J. Wang J, K. Xiong, Q. Pan et al., Application of TonB-Dependent Transporters in Vaccine Development of Gram-Negative Bacteria", *Front Cell Infect Microbiol*, vol. 10, pp. 589115, 2021.

[93] E. B. Bettin, J. Dorneles, A. S. Heckttheuer et al., "TonB-dependent receptor epitopes expressed in M. bovis BCG induced significant protection in the hamster model of leptospirosis", *Appl Microbiol Biotechnol,* vol. 106, pp. 173-184, 2022.

[94] D. P. Chimento, A. K. Mohanty, R. J. Kadner et al., "Substrate-induced transmembrane signaling in the cobalamin transporter Btu", *Nat Struct Biol*, vol. 10, pp. 394-401, 2003.

[95] A. R. Reeves, J. N. D'Elia, J. Frias et al., "A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch", *J. Bacteriol*, vol. 178, pp. 823-30, 1996.

[96] E. Cascales, S. K. Buchanan, D. Duché et al., "Colicin biology", *Microbiol Mol Biol Rev,* vol. 71, pp. 158–229, 2007.

[97] F. Calcuttawala, A. Pal, P. Nath et al., "Structural and functional insights into colicin: a new paradigm in drug discovery", *Arch Microbiol,* vol. 20, pp. 37, 2021.