Genomic analyses of *Bacteroides fragilis*: subdivisions one and two represent distinct species

Jamie English[1*], Fiona Newberry[2*], Lesley Hoyles[2], Sheila Patrick[1,3], Linda Stewart[1]

Jamie English – ORCiD 0000-0001-7491-1366

Fiona Newberry – ORCiD 0000-0002-7253-6950

Lesley Hoyles – ORCiD 0000-0002-6418-342X

Sheila Patrick – ORCiD 0000-0003-3230-1986

Linda Stewart – ORCiD 0000-0002-2400-6162

[1]Institute of Global Food Security, School of Biological Sciences, Queen's University, Belfast

[2]School of Science and Technology, Nottingham Trent University

[3]Wellcome Wolfson Institute for Experimental Medicine, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast, BT9 7BL

Corresponding authors: Linda Stewart, l.stewart@qub.ac.uk; Lesley Hoyles, lesley.hoyles@ntu.ac.uk

* Joint first authors

Keywords: antimicrobial resistance, taxonomy, *Bacteroidaceae*.

Abbreviations: AMR, antimicrobial resistance; ANI, average nucleotide identity; BLAST; basic local alignment search tool; CARD, Comprehensive Antibiotic Resistance Database; CLIMB, cloud infrastructure for microbial bioinformatics; ETBF, enterotoxigenic *B. fragilis*; GTDB, Genome Taxonomy Database; HGT, horizontal gene transfer; KO, KEGG orthology; PCA, principal component analysis; RGI, Resistance Gene Identifier.

Abstract

Introduction. Bacteroides fragilis is a Gram-negative anaerobe that is a member of the human gastrointestinal microbiota and is frequently found as an extra-intestinal opportunistic pathogen. B. fragilis comprises two distinct groups – division I and II – characterised by the presence/absence of genes [cepA and ccrA (cfiA), respectively] that confer resistance to β-lactam antibiotics by either serine or metallo-β-lactamase production. No large-scale analyses of publicly available B. fragilis sequence data have been undertaken, and the resistome of the species remains poorly defined.

Hypothesis/Gap Statement. Reclassification of division I and II B. fragilis as two distinct species has been proposed but additional evidence is required.

Aims. To investigate the genomic diversity of GenBank B. fragilis genomes and establish the prevalence of division I and II strains among publicly available B. fragilis genomes, and to generate further evidence to demonstrate that B. fragilis division I and II strains represent distinct genomospecies.

Methodology. High-quality (n=377) genomes listed as Bacteroides fragilis in GenBank were included in pangenome and functional analyses. Genome data were also subject to resistome profiling using The Comprehensive Antibiotic Resistance Database.

Results. Average nucleotide identity and phylogenetic analyses showed B. fragilis divisions I and II represent distinct species: B. fragilis sensu stricto (n = 275 genomes) and B. fragilis A (n = 102 genomes; Genome Taxonomy Database designation), respectively. Exploration of the pangenome of B. fragilis sensu stricto and B. fragilis A revealed separation of the two species at the core and accessory gene levels.

Conclusion. The findings indicate that B. fragilis A, previously referred to as division II B. fragilis, is an individual species and distinct from B. fragilis sensu stricto. The B. fragilis pangenome analysis supported previous genomic, phylogenetic and resistome screening analyses collectively reinforcing that divisions I and II are two separate species. In addition, it was confirmed that differences in the accessory genes of B. fragilis divisions I and II are primarily associated with carbohydrate metabolism and suggests that differences other than antimicrobial resistance could also be used to distinguish between these two species.

Impact statement

Bacteroides fragilis is an opportunistic pathogen that poses a major risk to public health due to its capacity to cause anaerobic infections in extraintestinal sites. In addition, *B. fragilis* clinical isolates possess some of the highest levels of antimicrobial resistance genes among anaerobes. Concerningly, multidrug-resistant *B. fragilis* clinical isolates have become increasingly reported over the past decades and represent a challenge in treating infections caused by this bacterium. *B. fragilis* divisions I and II were distinguished based on the presence/absence of β-lactam antimicrobial resistance genes. The *B. fragilis* pangenome was also interrogated, with findings indicating that *B. fragilis sensu stricto* (division I) and *B. fragilis* A (division II) also possess noticeable differences in carbohydrate-metabolising gene composition. This suggests that *B. fragilis* is continuously adapting to accommodate the degradation of certain carbohydrates.

Data summary

Supplementary material detailing all genome data included in this study is available from figshare (10.6084/m9.figshare.23516403, 10.6084/m9.figshare.24077736)**.**

Introduction

The composition and function of the gut microbiota are increasingly appreciated as factors influencing human health and disease [1, 2]. A reduced number of colonising members of the phylum *Bacteroidota* has been associated with gut-localised and systemic diseases such as rheumatoid arthritis [3–8]. The phylum *Bacteroidota* can be divided into six classes (*Bacteroidia*, *Chitinophagia*, *Flavobacteriia*, *Sphingobacteriia*, *Saprospiria* and *Cytophagia*) [9, 10]. Of the *Bacteroidota* present within the human large intestine, bacteria belonging to the order *Bacteroidales* are among the most prevalent and represent almost half of the entire bacterial populations localised to this microbially dense region of the gastrointestinal tract [11]. At the species level, *Bacteroides* spp. represent ~25 % of all anaerobes present in the large intestine. *Bacteroides caccae* prevents invasion of enteric pathogens through its ability to colonise the luminal mucosa of the intestine, whilst species such as *Bacteroides thetaiotaomicron* and *Bacteroides ovatus* have roles in the breakdown of many indigestible polysaccharides that in turn supply the host with up to 15 % of daily metabolic requirements [12–16].

Bacteroides fragilis represents an estimated 2 % of all gut *Bacteroides* spp. in colonised individuals [17, 18]. Although the abundance of *B. fragilis* in the colon is 10- to 100-fold less than other intestinal *Bacteroidales* (including *B. thetaiotaomicron*, *Phocaeicola vulgatus* and *Parabacteroides distasonis*) that are present at 10^{10} per gram dry weight of faeces, *B. fragilis* is an important contributor to the development of an effective immune system and maintenance of an anti-inflammatory environment within the intestinal lumen [13, 18, 19]. Enterotoxigenic *B. fragilis* (ETBF) secretes a zinc-metalloprotease toxin, Bft, that exists in three isoforms (Bft1, Bft2 and Bft3), each of which can disrupt intestinal barrier permeability through cleavage of E-cadherin, an intercellular adhesion protein also involved in tumour suppression [20–22]. Although the Bft protein is associated with diarrhoea, inflammatory bowel disease and colon cancer, it has been reported that up to 67 % of individuals who are colonised by ETBF are asymptomatic [23]. This may be due to asymptomatic individuals harbouring a greater number of non-toxigenic *B. fragilis* strains that utilise type-6 secretion systems to limit intestinal colonisation by ETBF [24].

B. fragilis is the most common cause of Gram-negative anaerobic infection and accounts for 60 % or more of clinical isolates. These infections arise due to a loss of integrity of the intestinal epithelium and are potentially lethal. The precise nature of *B. fragilis* virulence remains to be resolved; however, a combination of within- and between-strain surface polysaccharide diversity, multiple extracellular enzymes targeting host components, outer membrane vesicle production, iron scavenging mechanisms and oxygen tolerance likely contribute to multifactorial virulence. Interestingly, the *B. fragilis* enterotoxin is not an essential virulence determinant; it is absent in, for example, 80 % or more

of blood culture isolates (reviewed in [25]). Infections caused by *B. fragilis* are typically treated with multiple antibiotics, including metronidazole, chloramphenicol, carbapenems and β-lactam agents administered in combination with β-lactamase inhibitors [26, 27]. An increase in the prevalence of antimicrobial resistance (AMR) genes and resistance mechanisms encoded by *B. fragilis* has occurred globally in recent years [19, 28–31] along with reports of multidrug-resistant isolates [27]. The chromosomally encoded cephalosporinase genes *cepA* and *cfiA* (*ccrA*) have been used to separate *B. fragilis* into two divisions: I and II, respectively [26, 32, 33]. *cepA* encodes a class 2e cephalosporinase (β-lactamase) that confers resistance to commonly administered β-lactam antibiotics; *cepA*$^{+}$ strains remain susceptible to treatment with cephamycins, carbapenems and β-lactamase inhibitor combinations [32, 34]. The *cfiA* gene encodes a metallo-β-lactamase and is a greater threat to public health due to its ability to hydrolyse carbapenems and resist β-lactamase inhibitors that are commonly administered to treat anaerobic infections [35–37].

In addition to *cepA* and *cfiA*, *B. fragilis* divisions I and II can be differentiated based on *recA* (a ubiquitous protein involved in DNA repair and homologous recombination) and *glnA* (a glutamine synthetase encoding an enzyme associated with nitrogen metabolism and ammonia assimilation) gene sequences [38, 39]. Despite the phenotypically homogenous appearance of *B. fragilis* isolates, 65-70 % intergroup and 80-90 % intragroup similarities have been confirmed between division I and II *B. fragilis* strains by DNA-DNA hybridisation experiments [33, 40, 41]. Furthermore, the application of species delimitation methods, including genome BLAST distance phylogeny (GBDP) [10] and average nucleotide identity (ANI) [42], has facilitated recent whole-genome sequencing studies that continue to propose that division I and II *B. fragilis* are two distinct species [43, 44]. Interestingly, it was recently highlighted that genetic differences between division I and II go beyond AMR genes, with the core and accessory genomes between these subspecies displaying considerable amounts of genetic diversity [44]. Nonetheless, the proposed reclassification of division I and II *B. fragilis* as two distinct species is yet to be approved by the *International Journal of Systemic and Evolutionary Microbiology* and reinforces that additional evidence is required for this to occur.

The present study aimed to investigate the genomic diversity of GenBank *B. fragilis* genomes, to establish the prevalence of division I and II strains among publicly available *B. fragilis* genomes, and to generate further evidence to demonstrate that *B. fragilis* division I and II strains represent distinct genomospecies.

Methods

Identification of *B. fragilis* genomes used in this study. Bioinformatics analyses were done using the cloud infrastructure for microbial bioinformatics (CLIMB) [45] and HPC facilities of Nottingham Trent

University. Non-redundant genomes (*n* = 187) listed as '*Bacteroides fragilis*' were downloaded from GenBank during 2020, with an updated dataset created on 25 August 2022 (**Supplementary Table 1**). Completeness and contamination of the 418 genomes were assessed using CheckM2 v0.1.3 [46]. Average nucleotide identity (ANI) analysis was done with all GenBank genomes >90 % complete and with <5 % contamination [47] (*n* = 379) using fastANI v1.33 [48] against 111 representative *Bacteroides* genomes (**Supplementary Table 2**) from the Genome Taxonomy Database (GTDB) Release 07-RS207 (8th April 2022) [49, 50]. A 95 % ANI threshold was set to assign species affiliation, as recommended by Jain *et al*. (2018) [48], and similarly applied by Tortoli *et al*. (2019) [51]. Strains with <95 % genomic sequence similarity to *B. fragilis* NCTC 9343^{T} were not considered *B. fragilis sensu stricto*. FastANI results were summarised and visualised using R (tidyverse v1.3.1; reshape2 v1.4.4; gplots 3.1.3). The '*Bacteroides fragilis*' genomes were annotated using Bakta v1.4.2 (database release 3.1) [52]. Phylogenetic analysis of the genomes was carried out using PhyloPhlAn v3.0.58 [53], to confirm species affiliations. The tree was visualised using iToL v6.6 [54] and annotated using iToL and Adobe Illustrator.

Phylogenetic analyses of 16S rRNA gene sequences encoded within genomes. barrnap v0.9 was used to identify ribosomal RNA genes within genome sequences. All 16S rRNA gene sequences >1300 nt identified were used to generate a multiple-sequence alignment (Clustal Omeaga v.1.2.2) in Geneious Prime 2023.0.1. Unrooted neighbour-joining (Jukes-Cantor; 100 bootstrap replications) and maximum-likelihood (PhyML 3.3.20180214; substitution model JC69; 100 bootstrap replications) phylogenetic trees were generated from the alignment. Trees were visualized and annotated using iToL v6.6 and Adobe Illustrator. Alignment, similarity matrix and newick files generated from these analyses are available from figshare as Supplementary Material.

Characterisation of AMR genes among the genomes. The Resistance Gene Identifier (RGI) v6.0.0 [Comprehensive Antibiotic Resistance Database (CARD) v3.2.4] was used to identify AMR genes encoded within *B. fragilis* and *B. fragilis* A genomes [55]. Data for strict and perfect matches were extracted from the .txt output files and visualised using R (tidyverse v1.3.1; ggtree v3.4.1; aplot v0.1.8) with a phylogenetic tree generated for the 377 genomes using PhyloPhlAn v3.0.58.

Analysis of pangenome. Panaroo (v.1.3.0) was used to generate a pangenome and core genome alignment of all isolates (default settings; -a core, --remove-invalid-gene, --clean-mode strict, --threshold 0.98) [56]. Principal component analysis (PCA) was undertaken with the accessory genes (present in 5-95 %) of isolates using a binary gene presence/absence file in R Studio (v. 4.1.2 with FactoMineR (v.2.6) and factoextra (v.1.0.7) [57, 58]. A core single nucleotide polymorphism (SNP) maximum likelihood tree was generated using IQTree (v.1.16.10, maximum bootstrap: 1000, default settings) and best fit model determined using ModelFinder [59]. The core genome alignment output

from Panaroo was input to snp-sites (v.2.5.1; default settings) [60]. The genomes were clustered according to hierarchical Bayesian clustering algorithm using fastbaps [61].

Functional analysis of pangenome. The pan reference genome fasta file generated from Panaroo was input to eggnog mapper server (accessed: 31/10/2022; default settings; [56, 62]). The KEGG orthology (KO) terms assigned to genes within the accessory genome were retained and duplicate KO terms across multiple genes were collated. A KO table of the occurrence of each KO term within *B. fragilis* and *B. fragilis* A isolates was generated and input to FuncTree for visualisation [63]. Wilcoxon test with Benjamini-Hochberg adjustment was used to determine the KO values that were significantly different (adjusted *P* value <0.05) between both groups.

Results and Discussion

Confirmation of identities of genomes included in this study

Of the genomes listed on NCBI GenBank as '*Bacteroides fragilis*' (*n* = 418), 379 were considered to be of high quality [<5 % contamination, > 90 % complete; criteria of [47] after CheckM2 analysis (**Supplementary Table 1**)]. ANI analysis showed 275 of these genomes belonged to *B. fragilis* and 102 genomes belonged to *B. fragilis* A (**Supplementary Table 3**; **Supplementary Figure 1**), based on comparison with GTDB reference genomes (**Supplementary Table 2**). One genome (accession GCA_019583405) that represented a novel species within the genus *Bacteroides* (<95 % ANI with the representative genome of *B. fragilis* A, assembly GCF_002849695) and one (accession GCA_000699685) that belonged to *B. ovatus* (>97 % ANI with the reference genome, assembly GCF_001314995) were excluded from further analyses (**Supplementary Table 3**; **Supplementary Figure 1**). Phylogenetic analysis of the 377 genomes with GTDB reference genomes confirmed the affiliations of the 275 and 102 genomes with *B. fragilis* and *B. fragilis* A, respectively (**Figure 1**). This supports recent work by Wallace and colleagues who also confirmed that division II (i.e. *cfiA*positive) *B. fragilis* genomes share <95 % ANI with the *B. fragilis* type strain NCTC 9343^{T}, and ultimately do not meet the threshold required for species-level identification [44, 48, 64].

16S rRNA gene sequence-based analyses

Among the 377 high-quality *B. fragilis* (A) genomes, 231 (170 *B. fragilis* – division I; 61 *B. fragilis* A – division II) encoded 16S rRNA genes that were ≥80 % complete (length range 1302–1586 nt; mean 1519 ± 30 nt; median 1525 nt). Our dataset included a mixture of publicly available complete and draft genomes, with (unsurprisingly) many draft genomes not encoding any or encoding only truncated 16S rRNA gene sequences. It was common for genomes to encode more than one almost-complete copy

of the 16S rRNA gene (copy number range 1–8; mean 2 ± 2; median 1). The genome of *B. fragilis* NCTC 9343^T encoded six copies of the 16S rRNA gene, sharing 100 % similarity with one another. *B. fragilis sensu stricto* (division I) 16S rRNA gene sequences shared between 95.89 and 100 % similarity with those of *B. fragilis* NCTC 9343^T, while *B. fragilis* A (division II) 16S rRNA gene sequences shared between 94.89 and 97.92 % similarity with those of *B. fragilis* NCTC 9343^T. There was no significant difference (P = 0.13, unpaired Student's t test) in the number of copies of the 16S rRNA gene encoded by *B. fragilis* and *B. fragilis* A genomes. Phylogenetic analyses of the 16S rRNA gene sequences showed they clustered according to division with high (≥90 %) bootstrap support (**Supplementary Figure 2** and **Supplementary Figure 3**). Given the wide range of sequence divergence among 16S rRNA gene sequences from *B. fragilis* and *B. fragilis* A genomes (as noted above, but also refer to the similarity matrix available as Supplementary Material), we recommend that alternative genes – such as *recA* and *glnA* [38, 39] – be used to distinguish between these bacteria.

AMR genes encoded in *B. fragilis* genomes

The 275 *B. fragilis* and 102 *B. fragilis* A genomes were analysed using RGI with the most-recent release of CARD. All authentic *B. fragilis* genomes were predicted to encode variants of *cepA*, a β-lactamase-encoding gene conferring resistance to cephalosporin antibiotics [26, 29, 65]. Division II *B. fragilis* genomes were characterised by the presence of variants of the AMR gene *cfiA* (also referred to as *ccrA*) [26, 66, 67]. Of the authentic (division I) *B. fragilis* genomes subject to resistome screening, 100 % generated both 'perfect' and 'strict' hits, as described by Alcock *et al*. [55], for the presence of *cepA*, which confers resistance towards penicillins and cephalosporins [68]. In addition, all genomes with <95 % ANI to *B. fragilis* NCTC 9343^T were confirmed to encode *ccrA/cfiA*, as expected (**Figure 2**). It has been reported previously that these AMR genes are present in different regions of division I and II genomes, as confirmed by analysis between *B. fragilis* NCTC 9343^T, which acts as a reference genome for the identification of division I strains, and *B. fragilis* IHMA_4, that while not included in this study is a division II *B. fragilis* strain due to the presence of the *cfiA* gene [44]. Therefore, our AMR-based analysis complements findings from previous studies to demonstrate that the *B. fragilis sensu stricto* genomes belonged to division I *B. fragilis* and the *B. fragilis* A genomes belonged to division II *B. fragilis,* as confirmed by the presence of *cepA* and *cfiA* genes as well as phylogenetic clustering (**Figure 2**). Aside from *cepA* and *cfiA*, variants of the *cfxA* gene were identified in 23 of the *B. fragilis sensu stricto* genomes; this also confers antibiotic resistance through the expression of β-lactamases. For example, CARD analysis confirmed 14 and 9 hits for the presence of *cfxA2* and *cfxA3* AMR genes, respectively. Similar to *cepA* and *cfiA*, *cfxA* genes also encode a class A cephalosporinase and, as a trio, these genes are primarily responsible for β-lactamase expression among *Bacteroides* species [34, 69, 70]. Nonetheless, it is the class B metallo-β-lactamase that enables the hydrolysis of carbapenems and

poses the greatest threat given the reliance on these antibiotics to treat multidrug-resistant infections [37, 71]. Despite being considered as the largest of the β-lactamase families, AMR genes encoding the OXA class-D β-lactamases were relatively scarce among genomes investigated, with 2 and 8 hits being generated for *B. fragilis sensu stricto* and *B. fragilis* A, respectively. This suggests that OXA AMR genes are not utilised as frequently by *B. fragilis* to confer resistance to β-lactam antibiotics unlike pathogenic bacteria including *Acinetobacter baumannnii*, *Pseudomonas aeruginosa* and *Klebsiella pneumoniae,* where these genes are found in greater abundance [72, 73].

It was confirmed by CARD analysis that a total of 31 different AMR genes were encoded by the authentic *B. fragilis* genomes (**Figure 2**), while out of 102 *B. fragilis* A genomes, perfect and strict hits were generated for the presence of 23 different AMR genes. AMR genes including *adeF* and tetracycline resistance gene variants were among the most common AMR genes detected among both *B. fragilis sensu stricto* and *B. fragilis* A genomes. Specifically, a total of 538 and 204 hits were generated among *B. fragilis sensu stricto* (*n* = 275) and *B. fragilis* A (*n* = 102) genomes for the presence of *adeF* that confers resistance to fluoroquinolone and tetracycline antibiotics by acting as an efflux pump component [74]. This AMR gene has been detected previously among *Bacteroides* clinical isolates as well as being prevalent among other gut-associated bacteria including *Akkermansia muciniphila* and the pathogen *Acinetobacter baumannii* [75–78]. Tetracycline resistance gene variants were also detected among *B. fragilis sensu stricto* and *B. fragilis* A genomes, including *tetB*, *tetC*, *tetM*, *tetQ* and *tetX*. Of these, *tetQ* was the most prevalent with 203 and 93 hits generated among *B. fragilis sensu stricto* and *B. fragilis* A genomes, respectively, potentially mediating resistance by protecting ribosomal proteins of encoding strains from antibiotic activity [79, 80]. Both *adeF* and *tetQ* have been reported previously as the most abundant AMR genes present in the gut during metagenomic analysis, with the prevalence of the latter of these two AMR genes thought to have almost tripled in the last decades among *Bacteroides* isolates [79, 81]. The presence of the *tetQ* on mobile genetic elements, such as conjugative transposons which are transferred at increased frequency on exposure to low concentrations of tetracycline, is likely to facilitate the spread of this AMR gene via horizontal gene transfer (HGT) among *Bacteroides* species [82]. This has been reinforced by earlier studies that demonstrate the genetic homology between *tetQ* genes present in *Bacteroides* species, including *B. fragilis* [81]. In addition, *tetQ* was the most abundant AMR gene detected among *Bacteroidota* present in the faecal microbiota of animals treated with oxytetracycline [83]. The high prevalence of *tetQ* among *B. fragilis* may also facilitate the dissemination of this AMR gene to fellow intestinal colonisers that also act as clinically important opportunistic pathogens. An example of this would be the Gram-positive bacterium *Enterococcus faecalis*, which has the ability to acquire the *tetQ* from *B. fragilis* and

reinforces the concern that the spread of AMR genes among gut bacteria poses to public health [84, 85].

Resistome screening also revealed perfect and strict hits for the presence of *nim* genes in *B. fragilis sensu stricto* and *B. fragilis* A genomes. Out of the 11 *nim* genes that have been identified to date, six were detected among the *B. fragilis* genomes investigated, namely *nimA*, *nimB*, *nimD*, *nimE*, *nimG* and *nimJ* (**Figure 2**). Of these, *nimB* and *nimG* were restricted to *B. fragilis* A and *B. fragilis sensu stricto*, respectively, while the other *nim* gene variants were detected among both genomospecies. *nim* genes are, however, more prevalent among *B. fragilis* A genomes in contrast to *B. fragilis sensu stricto* and suggests that these strains previously considered as division II *B. fragilis* have a greater capacity to acquire these AMR genes and ultimately facilitate resistance to the antimicrobial agent metronidazole, which is commonly administered to treat and prevent anaerobic infections [86]. These AMR genes are a growing concern and have been identified among *B. fragilis* clinical isolates in recent studies [87–90]. While the nitroimidazole reductase enzyme that is encoded by *nim* genes is responsible for contributing to reduced metronidazole susceptibility in encoding strains, by inhibiting the formation of toxic nitroso residues, metronidazole resistance can occur in the absence of these AMR genes and indicates that other mechanisms can confer metronidazole resistance [88, 91]. Such unrelated *nim* gene mechanisms include overexpression of multidrug efflux pumps and the DNA repair protein, RecA, as well as ferrous iron transporter deficiency [92–94]. Nonetheless, given that *nim* genes are typically accompanied by upstream insertion sequence elements which contain the *B. fragilis* consensus promoter sequence, it is likely that these genes are spread throughout bacterial communities. Furthermore, the fact that metronidazole resistance can be induced in *nim+* strains, also reinforces that even if *nim+ B. fragilis* are not initially resistant to this antimicrobial, exposure to sub-lethal concentrations may encourage an increase in resistant strains within the gut and make treating *B. fragilis* infections more challenging [95]. Continued resistome screening of clinical *B. fragilis* isolates is therefore encouraged on a regular basis to help monitor the changes in *nim* gene prevalence and tackle the burden posed by antimicrobial-resistant microbes.

AMR genes associated with resistance to the glycopeptide antibiotic vancomycin, used to treat infections by Gram-positive pathogens by acting as an inhibitor of cell wall synthesis, were prevalent among *B. fragilis sensu stricto* and *B. fragilis* A genomes, with 274 and 100 hits being generated, respectively (**Figure 2**). Given the presence of an outer membrane in Gram-negative bacteria, glycopeptides are unable to interact with the bacterial cell wall component peptidoglycan and are therefore not the antibiotic of choice when treating infections caused by Gram-negative bacteria. The *vanT* gene in the *vanG* cluster was the most common vancomycin resistance gene detected among *B. fragilis sensu stricto* and *B. fragilis* A genomes. In enterococci, vancomycin resistance gene clusters are

involved in the expression of membrane-associated enzymes that lead to the synthesis of peptidoglycan precursors with reduced compatibility to vancomycin, thereby aiding resistance against this drug [96]. While *B. fragilis* possesses an outer membrane that will limit the interaction of glycopeptide antibiotics, such as vancomycin, with the cell wall and intracellular environment, it is likely that encoding strains have potentially acquired these genes from fellow intestinal colonisers and possibly act as an additional mechanism of resistance. Given that AMR genes can be transferred between Gram-positive and Gram-negative species, the high prevalence of *van* genes among *B. fragilis sensu stricto* and *B. fragilis* A represents a risk for their dissemination to susceptible bacteria and ultimately reduce the efficacy of this drug in treating infections caused by Gram-positive bacteria [97].

Other AMR genes with lower prevalence include *erm* gene variants, particularly *ermF*, with 52 and 42 hits for this gene being generated among *B. fragilis sensu stricto* and *B. fragilis* A genomes, respectively. The *erm* genes are responsible for counteracting the inhibitory activity of erythromycin on protein synthesis through the expression of a methylase that facilitates modification of the 50S ribosomal subunit that acts as the target site for this antibiotic. For instance, the role of *ermF* in erythromycin resistance has been reported previously in the bird pathogen and fellow member of the phylum *Bacteroidota, Riemerella anatipestifer* [98], while other studies have confirmed the high prevalence of this gene in environmental *B. fragilis* isolates, including those from hospital wastewater [99]. Furthermore, the high prevalence of such AMR genes among isolates from these environments is likely to aid their dissemination among bacteria, particularly due to their association with mobile genetic elements and the sub-lethal antibiotic concentrations present in wastewater that select for resistant strains [100]. The AMR gene *mef(En2)* encodes an efflux pump that also confers resistance to macrolide antibiotics such as erythromycin and clindamycin, and was predicted to be present among *B. fragilis sensu stricto* and *B. fragilis* A genomes, with 45 and 17 hits being generated, respectively. Of the *B. fragilis* A genomes predicted to be *mef(En2)*+, genome GCA_014639005 was central to a previous publication that also proposed *cfiA+ B. fragilis* as a distinct genomospecies [101]. This genome was included in the current study, with findings from resistome screening analysis supporting those made previously that also detected the presence of both *cfiA* and *mef(En2)* in this strain.

Of the *B. fragilis sensu stricto* strains, genome GCA_000601055 (*B. fragilis* S23L17) was predicted to carry the most antibiotic resistance genes, with 18 hits being generated for the presence of AMR genes that include *aadS*, *adeF*, *cepA*, *ermF* and *tet* gene variants. *aadS* is not likely to be relevant as *Bacteroides* are intrinsically resistant to aminoglycoside antibiotics; however, it may contribute to the pool of horizontally transmissible resistance genes with the gut microbiota. This strain has been confirmed in previous studies to express a type-6 secretion system that is likely to facilitate modulation of the surrounding environment, while others reported the presence of a CRISPR-Cas system within its

genetic architecture that may also contribute to antibiotic resistance [102, 103]. Despite generating a smaller number of hits for the presence of AMR genes, three *B. fragilis* A genomes, namely GCA_000297695 (*B. fragilis* strain HMW610), GCA_001693695 (*B. fragilis* strain O:21) and GCA_001695355 (*B. fragilis* strain BF8) were predicted to harbour 11 AMR genes that include *ccrA*, *cfiA14*, *ermF* as well as *nim* and *tet* gene variants, among others. Of these, *B. fragilis* strains O:21 and BF8 were central to a previous study by Sóki and colleagues who sequenced both genomes and confirmed the multidrug-resistant properties of these strains due to the presence of AMR genes, all of which were also detected in the current study [104]. Although these genomes are predicted to harbour fewer AMR genes than the individual *B. fragilis sensu stricto* genome, the presence of genes that help confer resistance towards commonly administered antibiotics such as carbapenems and metronidazole make monitoring the prevalence of *B. fragilis* A strains a top priority for the benefit of public health.

Although resistome screening analysis in the current study has determined the type and abundance of AMR genes among publicly available *B. fragilis sensu stricto* and *B. fragilis* A genomes, it is noteworthy that the presence of AMR genes may not confer phenotypic resistance. For example, the *tetX* AMR gene that was detected in 31 and 8 *B. fragilis sensu stricto* and *B. fragilis* A genomes, respectively, was initially identified in *Bacteroides* spp. and yet did not confer resistance to the host strain [105]. However, transfer of the *B. fragilis* associated transposons, Tn4351 and Tn4400, that harbour the *tetX* gene led to tetracycline resistance in aerobically grown *Escherichia coli* [106, 107]. This is likely due to the fact that the TetX protein requires the presence of oxygen to transform tetracycline antibiotics, which is relatively scarce in the anaerobic mucosa of the gut where *B. fragilis* exists [105]. Such findings suggest that *B. fragilis sensu stricto* and *B. fragilis* A act as reservoirs for silent AMR genes that have the capacity to become incorporated into clinically relevant pathogens via the frequent HGT that occurs in the gut [108]. It is therefore important that the resistome of intestinal bacteria, including *B. fragilis*, is closely monitored in future studies even if strains lack phenotypic resistance. Ultimately, this would facilitate our understanding of the silent AMR genes that are present among bacterial populations and prevent the threat of their dissemination via HGT being underestimated.

Pangenome analysis of division I and division II *B. fragilis* genomes

Panaroo analysis revealed a total of 24,451 genes in the pangenome of 377 genomes. The core genome accounted for 8.8 % (present in 99-100 % of isolates) of the total pangenome and contained 2,175 genes (**Table 1**). The majority of genes were identified within relatively few isolates, as noted previously with non-clinical pangenome studies [109, 110]. Compared to pathogenic bacteria, the core

pangenome of *B. fragilis* was found to be smaller [111–113]. The core genome of 4,401 *E. coli* isolates was reported to be 53 % of the total gene count (128,193 genes). Additionally, the core genome of *Staphylococcus aureus* was 75 % of the total pangenome (21,133 genes) [114]. *Bifidobacterium longum,* a commensal intestinal microbe, has also exhibited a small core genome (3.2 %) similar to *B. fragilis* [115]. The small core genome observed in this study suggests that the core housekeeping genes necessary for basic survival are conserved between both *B. fragilis sensu stricto* and *B. fragilis A*, as noted with *Bifidobacterium longum.*

Generation of a PCA revealed that 15.1 % of variation was explained by Dimension 1 and 4.8 % was explained by Dimension 2 (**Figure 3**).

A clear division between the accessory genes of *B. fragilis* and *B. fragilis* A was observed, suggesting functional differences existed between the two groups of bacteria. The top 49 accessory genes contributing to the variation in dimensions 1 and 2 were present in all *B. fragilis* division isolates (**Supplementary table 4**). Within the accessory genome, there were 49 genes present in all *B. fragilis sensu stricto* isolates and 42 genes present in all *B. fragilis* A isolates; however, the absence of these genes from a division does not infer the gene and its function are missing from the other division. It is important to be aware of the sequence identity cut-offs used during pangenome analysis. A core SNP maximum likelihood phylogenetic tree was generated using IQTree with GTR+F model according to Bayesian information criteria (**Figure 4**). *B. fragilis sensu stricto* and *B. fragilis* A isolates formed two distinct monophyletic clades, as seen in the accessory gene-based PCA (**Figure 3**). According to fastbaps, *B. fragilis sensu stricto* and *B. fragilis* A formed two clusters (outer ring, **Figure 4**).

Functional analysis of pangenome

The majority of KO values within the accessory genome were assigned to metabolism, specifically carbohydrate metabolism (**Figure 5; Supplementary table 5**). Of the 825 KO values, 213 were significantly (adjusted *P* value <0.05) different between *B. fragilis sensu stricto* and *B. fragilis* A (**Supplementary table 6**). Several KO values were found in either only *B. fragilis* or *B. fragilis* A genomes. The majority of these were hydrolases or transporter proteins (**Table 2**). Additionally, the significant KO values appeared to be involved in glycan biosynthesis/metabolism, metabolism of cofactors/vitamins and carbohydrate metabolism (**Figure 5**). The diversity of capsular polysaccharide biosynthesis loci within the *B. fragilis* pangenome is reflected in observed capsular antigenic diversity between clonal isolates, with more than 30 divergent microcapsule biosynthesis operons identified [116, 117]. A recent study explored the pangenome of *B. ovatus* and *B. xylanisolvens* and revealed only 17.5 % (2,264 genes) were shared among the selected strains, a similar core genome sized observed during this study. Several key components of *Bacteroidota* polysaccharide metabolism (2 classes of

core polysaccharide utilization loci, SusC/D homologs and degradative CAZymes) were heavily represented in the accessory genome and not common to all strains [118]. Members of the genus *Bacteroides* are well-known polysaccharide degraders and can adapt to changes in available dietary fibres [119, 120]. For example, *B. thetaiotaomicron, B. ovatus,* and *B. cellulosilyticus* encode over 250 CAZymes that target nearly all commonly available dietary polysaccharides. Although no specific gene subsets within the accessory genome were explored in this study, it is possible that the main diversification between *B. fragilis sensu stricto* and *B. fragilis* A is due to genes involved in polysaccharide metabolism. A recent study revealed constant adaptation of *B. fragilis* within the intestinal microbiome is a common feature of within-person evolution [121]. Therefore, the variation within the accessory genome and large number of genes present in single isolates could be due to the adaptation of *B. fragilis* to fill specific carbohydrate degradation niches within individual microbiomes.

Conclusion

Here, we confirm that 275/377 genomes listed as *B. fragilis* on the NCBI public database are *B. fragilis sensu stricto* and share >=95 % ANI with *B. fragilis* NCTC 9343^{T}. Of the remaining genomes with <95 % ANI, 102 were assigned as *B. fragilis* A by the GTDB. Findings from fastANI analyses were reinforced by phylogenetic analyses and emphasised the importance of investigating the identities of publicly available genomes. These findings indicate that *B. fragilis* A, previously referred to as division II *B. fragilis*, is an individual species and distinct from *B. fragilis sensu stricto*. Whether this divergence is the result of barriers to HGT or occupation of micro-environments in different gut locations remains to be determined. Furthermore, it has yet to be confirmed whether individuals are colonised with *B. fragilis sensu stricto*, *B. fragilis* A or both simultaneously, and therefore highlights an avenue for future investigation.

Resistome screening, facilitated by CARD, confirmed that all *B. fragilis sensu stricto* genomes encoded *cepA*, an AMR gene that was absent in all *B. fragilis* A genomes analysed in the present study. In contrast, all *B. fragilis A* genomes encoded *ccrA*, an AMR gene that encodes a different class of β-lactamase that was absent from all *B. fragilis sensu stricto* genomes. This supports findings from previous studies that distinguished division I and II *B. fragili*s based on the presence or absence of these AMR genes in the genomic architecture of *B. fragilis* strains. The AMR gene *adeF*, which leads to the expression of an efflux pump component, was among the most prevalent resistance genes predicted during resistome screening analysis among *B. fragilis sensu stricto* and *B. fragilis* A genomes and suggests that this may be an important mechanism in conferring resistance. Additionally, AMR genes predicted to confer resistance to tetracycline were also abundant among *B. fragilis sensu stricto*

and *B. fragilis* A, with *tetQ* being the most frequently detected *tet* gene variant among all genomes investigated and reinforces that tetracycline should no longer be considered in treating *B. fragilis* infections. Resistome screening analysis from the current study also emphasises the concern regarding metronidazole resistance by determining the prevalence of *nim* genes among publicly available genomes. Given that the treatment of *B. fragilis* infections is often dependent on metronidazole administration, the prevalence of *nim* genes among clinically isolates should be closely monitored in the future.

Exploration of the pangenome of *B. fragilis sensu stricto* and *B. fragilis* A revealed separation of the two groups at the core and accessory genome level, confirming separation of two subdivisions into two species. This separation was confirmed by phylogenetic analysis of the core genome and PCA of the accessory genome. Significant functional differences were observed between both groups, mainly in genes associated with amino acid, carbohydrate, and glycan metabolism. While this study did not explore specific gene subsets, future studies should aim to identify mobile DNA signatures in the accessory genes and intergenomic recombination between species in core genes to determine if there are hot spots for genome transfer within each group. Importantly, this study adds to the growing body of evidence that *B. fragilis* A, previously referred to as division II *B. fragilis*, should be considered a distinct species of *Bacteroides*. To ensure that the clinical association with the potential for lethal infection arising from these bacteria remains easily memorable, while enabling understanding of the different antimicrobial susceptibilities, we propose that in a formal nomenclature change Division II *Bacteroides fragilis* A is renamed *Bacteroides fragila*. Compilation of the taxonomic details necessary for a formal proposal are ongoing.

References

1. **Sekirov I, Russell SL, Antunes LCM, Finlay BB**. Gut microbiota in health and disease. *Physiol Rev* 2010;90:859–904.

2. **Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ**. Dysbiosis of the gut microbiota in disease. *Microb Ecol Health Dis* 2015;26:26191.

3. **Sze MA, Schloss PD**. Looking for a signal in the noise: revisiting obesity and the microbiome. *mBio* 2016;7:e01018-16.

4. **Khan I, Ullah N, Zha L, Bai Y, Khan A, *et al.*** Alteration of gut microbiota in inflammatory bowel disease (ibd): cause or consequence? IBD treatment targeting the gut microbiome. *Pathogens* 2019;8:126.

5. **Ho JT, Chan GC, Li JC**. Systemic effects of gut microbiota and its relationship with disease and modulation. *BMC Immunol* 2015;16:21.

6. **Kasselman LJ, Vernice NA, DeLeon J, Reiss AB**. The gut microbiome and elevated cardiovascular risk in obesity and autoimmunity. *Atherosclerosis* 2018;271:203–213.

7. **Wang H, Ong E, Kao JY, Sun D, He Y**. Reverse microbiomics: a new reverse dysbiosis analysis strategy and its usage in prediction of autoantigens and virulent factors in dysbiotic gut microbiomes from rheumatoid arthritis patients. *Front Microbiol* 2021;12:633732.

8. **Oren A, Garrity GM**. Valid publication of the names of forty-two phyla of prokaryotes. *Int J Syst Evol Microbiol*;71. Epub ahead of print October 2021. DOI: 10.1099/ijsem.0.005056.

9. **Hahnke RL, Meier-Kolthoff JP, García-López M, Mukherjee S, Huntemann M, *et al.*** Genome-based taxonomic classification of *Bacteroidetes*. *Front Microbiol* 2016;7:2003.

10. **García-López M, Meier-Kolthoff JP, Tindall BJ, Gronow S, Woyke T, *et al.*** Analysis of 1,000 type-strain genomes improves taxonomic classification of *Bacteroidetes*. *Front Microbiol* 2019;10:2083.

11. **Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, *et al.*** The long-term stability of the human gut microbiota. *Science* 2013;341:1237439.

12. **Salyers AA, West SE, Vercellotti JR, Wilkins TD**. Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. *Appl Environ Microbiol* 1977;34:529–533.

13. **Salyers AA**. *Bacteroides* of the human lower intestinal tract. *Annu Rev Microbiol* 1984;38:293–313.

14. **Croucher SC, Houston AP, Bayliss CE, Turner RJ**. Bacterial populations associated with different regions of the human colon wall. *Appl Environ Microbiol* 1983;45:1025–1033.

15. **Guarner F, Malagelada J-R**. Gut flora in health and disease. *Lancet* 2003;361:512–519.

16. **Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, *et al.*** A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* 2003;299:2074–2076.

17. **Huang J, Lee S, Mazmanian S**. The human commensal *Bacteroides fragilis* binds intestinal mucin. *Anaerobe*;17. Epub ahead of print August 2011. DOI: 10.1016/j.anaerobe.2011.05.017.

18. **Wexler HM**. The Genus *Bacteroides*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (editors). *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. Berlin, Heidelberg: Springer. pp. 459–484.

19. **Wexler HM**. *Bacteroides*: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* 2007;20:593–621.

20. **Franco AA, Mundy LM, Trucksis M, Wu S, Kaper JB, *et al.*** Cloning and characterization of the *Bacteroides fragilis* metalloprotease toxin gene. *Infect Immun* 1997;65:1007–1013.

21. **Wu S, Lim K-C, Huang J, Saidi RF, Sears CL**. *Bacteroides fragilis* enterotoxin cleaves the zonula adherens protein, E-cadherin. *Proc Natl Acad Sci U S A* 1998;95:14979–14984.

22. **Sears CL**. Enterotoxigenic *Bacteroides fragilis*: a rogue among symbiotes. *Clin Microbiol Rev* 2009;22:349–369, Table of Contents.

23. **Boleij A, Hechenbleikner EM, Goodwin AC, Badani R, Stein EM, *et al.*** The *Bacteroides fragilis* toxin gene is prevalent in the colon mucosa of colorectal cancer patients. *Clin Infect Dis* 2015;60:208–215.

24. **Hecht AL, Casterline BW, Earley ZM, Goo YA, Goodlett DR, *et al.*** Strain competition restricts colonization of an enteric pathogen and prevents colitis. *EMBO reports* 2016;17:1281–1291.

25. **Patrick S**. A tale of two habitats: *Bacteroides fragilis*, a lethal pathogen and resident in the human gastrointestinal microbiome. *Microbiology (Reading)*;168. Epub ahead of print April 2022. DOI: 10.1099/mic.0.001156.

26. **Rashidan M, Azimirad M, Alebouyeh M, Ghobakhlou M, Asadzadeh Aghdaei H, *et al.*** Detection of *B. fragilis* group and diversity of bft enterotoxin and antibiotic resistance markers *cepA*, *cfiA* and *nim* among intestinal *Bacteroides fragilis* strains in patients with inflammatory bowel disease. *Anaerobe* 2018;50:93–100.

27. **Sherwood JE, Fraser S, Citron DM, Wexler H, Blakely G, *et al.*** Multi-drug resistant *Bacteroides fragilis* recovered from blood and severe leg wounds caused by an improvised explosive device (IED) in Afghanistan. *Anaerobe* 2011;17:152–155.

28. **Nagy E, Urbán E, Bacteria CEN on behalf of the ESG on AR in A**. Antimicrobial susceptibility of *Bacteroides fragilis* group isolates in Europe: 20 years of experience. *Clinical Microbiology and Infection* 2011;17:371–379.

29. **Kierzkowska M, Majewska A, Szymanek-Majchrzak K, Sawicka-Grzelak A, Mlynarczyk A, *et al.*** The presence of antibiotic resistance genes and *bft* genes as well as antibiotic susceptibility testing of *Bacteroides fragilis* strains isolated from inpatients of the Infant Jesus Teaching Hospital, Warsaw during 2007-2012. *Anaerobe* 2019;56:109–115.

30. **Hashimoto T, Hashinaga K, Komiya K, Hiramatsu K**. Prevalence of antimicrobial resistant genes in *Bacteroides* spp. isolated in Oita Prefecture, Japan. *Journal of Infection and Chemotherapy* 2023;29:284–288.

31. **Wang Y, Guo B, Gao X, Wen J, Wang Z, *et al.*** High prevalence of *cfiA* positive *Bacteroides fragilis* isolates collected at a teaching hospital in Hohhot, China. *Anaerobe* 2023;79:102691.

32. **Nagy E, Becker S, Sóki J, Urbán E, Kostrzewa M**. Differentiation of division I (*cfiA*-negative) and division II (*cfiA*-positive) *Bacteroides fragilis* strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J Med Microbiol* 2011;60:1584–1590.

33. **Gutacker M, Valsangiacomo C, Piffaretti J-C**. Identification of two genetic groups in *Bacteroides fragilis* by multilocus enzyme electrophoresis: distribution of antibiotic resistance (*cfiA*, *cepA*) and enterotoxin (*bft*) encoding genes. *Microbiology (Reading)* 2000;146 (Pt 5):1241–1254.

34. **Parker AC, Smith CJ**. Genetic and biochemical analysis of a novel Ambler class A beta-lactamase responsible for cefoxitin resistance in *Bacteroides* species. *Antimicrob Agents Chemother* 1993;37:1028–1036.

35. **Edwards R**. Resistance to beta-lactam antibiotics in *Bacteroides* spp. *J Med Microbiol* 1997;46:979–986.

36. **Hansen KCM, Schwensen SAF, Henriksen DP, Justesen US, Sydenham TV**. Antimicrobial resistance in the *Bacteroides fragilis* group in faecal samples from patients receiving broad-spectrum antibiotics. *Anaerobe* 2017;47:79–85.

37. **Yekani M, Rezaee MA, Beheshtirouy S, Baghi HB, Bazmani A, *et al.*** Carbapenem resistance in *Bacteroides fragilis*: A review of molecular mechanisms. *Anaerobe* 2022;76:102606.

38. **Karlin S, Weinstock GM, Brendel V**. Bacterial classifications derived from *recA* protein sequence comparisons. *J Bacteriol* 1995;177:6881–6893.

39. **Gutacker M, Valsangiacomo C, Bernasconi MV, Piffaretti J-C**. *recA* and *glnA* sequences separate the *Bacteroides fragilis* population into two genetic divisions associated with the antibiotic resistance genotypes *cepA* and *cfiA*. *J Med Microbiol* 2002;51:123–130.

40. **Johnson JL**. Taxonomy of the *Bacteroides*. *International Journal of Systematic and Evolutionary Microbiology* 1978;28:245–256.

41. **Johnson JL, Ault DA**. Taxonomy of the *Bacteroides*: II. Correlation of phenotypic characteristics with deoxyribonucleic acid homology groupings for *Bacteroides fragilis* and other saccharolytic *Bacteroides* species. *International Journal of Systematic Bacteriology* 1978;28:257–268.

42. **Boyanova L, Kolarov R, Mitov I**. Recent evolution of antibiotic resistance in the anaerobes as compared to previous decades. *Anaerobe* 2015;31:4–10.

43. **Jean S, Wallace MJ, Dantas G, Burnham C-AD**. Time for some group therapy: update on identification, antimicrobial resistance, taxonomy, and clinical significance of the *Bacteroides fragilis* Group. *Journal of Clinical Microbiology* 2022;60:e02361-20.

44. **Wallace MJ, Jean S, Wallace MA, Burnham C-AD, Dantas G**. Comparative genomics of *Bacteroides fragilis* group isolates reveals species-dependent resistance mechanisms and validates clinical tools for resistance prediction. *mBio* 2022;13:e03603-21.

45. **Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, *et al.*** CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microb Genom* 2016;2:e000086.

46. **Chklovski A, Parks DH, Woodcroft BJ, Tyson GW**. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. 2022;2022.07.11.499243.

47. **Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, *et al.*** Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017;35:725–731.

48. **Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S**. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.

49. **Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, *et al.*** A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;36:996–1004.

50. **Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, *et al.*** A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 2020;38:1079–1086.

51. **Tortoli E, Meehan CJ, Grottola A, Fregni Serpini G, Fabio A, *et al.*** Genome-based taxonomic revision detects a number of synonymous taxa in the genus *Mycobacterium*. *Infect Genet Evol* 2019;75:103983.

52. **Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, *et al.*** Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom* 2021;7:000685.

53. **Segata N, Börnigen D, Morgan XC, Huttenhower C**. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* 2013;4:2304.

54. **Letunic I, Bork P**. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.

55. **Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, *et al.*** CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;48:D517–D525.

56. **Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, *et al.*** Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 2020;21:180.

57. **Lê S, Josse J, Husson F**. FactoMineR: An R Package for multivariate analysis. *Journal of Statistical Software* 2008;25:1–18.

58. **Kassambara A, Mundt F**. factoextra: extract and visualize the results of multivariate data analyses. https://CRAN.R-project.org/package=factoextra (2020, accessed 28 February 2023).

59. **Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS**. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.

60. **Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, *et al.*** SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics* 2016;2:e000056.

61. **Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J**. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Research* 2019;47:5539–5549.

62. **Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, *et al.*** eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 2019;47:D309–D314.

63. **Uchiyama T, Irie M, Mori H, Kurokawa K, Yamada T**. FuncTree: functional analysis and visualization for large-scale omics data. *PLOS ONE* 2015;10:e0126967.

64. **Richter M, Rosselló-Móra R**. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 2009;106:19126–19131.

65. **Colney Z, Antony B, Kanthaje S**. Genotyping of multi drug resistant *Bacteroides fragilis* group of clinical isolates from mangalore, south India. *Indian Journal of Medical Microbiology* 2021;39:19–23.

66. **Ferløv-Schwensen SA, Sydenham TV, Hansen KCM, Hoegh SV, Justesen US**. Prevalence of antimicrobial resistance and the *cfiA* resistance gene in Danish *Bacteroides fragilis* group isolates since 1973. *Int J Antimicrob Agents* 2017;50:552–556.

67. **Jeverica S, Sóki J, Premru MM, Nagy E, Papst L**. High prevalence of division II (*cfiA* positive) isolates among blood stream *Bacteroides fragilis* in Slovenia as determined by MALDI-TOF MS. *Anaerobe* 2019;58:30–34.

68. **Nakano V, Nascimento e Silva A do, Merino VRC, Wexler HM, Avila-Campos MJ**. Antimicrobial resistance and prevalence of resistance genes in intestinal *Bacteroidales* strains. *Clinics (Sao Paulo)* 2011;66:543–547.

69. **Rogers MB, Parker AC, Smith CJ**. Cloning and characterization of the endogenous cephalosporinase gene, *cepA*, from *Bacteroides fragilis* reveals a new subgroup of Ambler class A beta-lactamases. *Antimicrob Agents Chemother* 1993;37:2391–2400.

70. **Thompson JS, Malamy MH**. Sequencing the gene for an imipenem-cefoxitin-hydrolyzing enzyme (CfiA) from *Bacteroides fragilis* TAL2480 reveals strong similarity between CfiA and *Bacillus cereus* beta-lactamase II. *J Bacteriol* 1990;172:2584–2593.

71. **Bush K**. Past and present perspectives on β-Lactamases. *Antimicrob Agents Chemother* 2018;62:e01076-18.

72. **Yoon E-J, Jeong SH**. Class D β-lactamases. *Journal of Antimicrobial Chemotherapy* 2021;76:836–864.

73. **Pandey D, Singhal N, Kumar M**. Investigating the OXA variants of ESKAPE pathogens. *Antibiotics (Basel)* 2021;10:1539.

74. **Coyne S, Rosenfeld N, Lambert T, Courvalin P, Périchon B**. Overexpression of resistance-nodulation-cell division pump AdeFGH confers multidrug resistance in *Acinetobacter baumannii*. *Antimicrob Agents Chemother* 2010;54:4389–4393.

75. **Matos J, Matos I, Calha M, Santos P, Duarte I, *et al.*** Insights from *Bacteroides* species in children with type 1 diabetes. *Microorganisms* 2021;9:1436.

76. **Filardi R, Gargari G, Mora D, Arioli S**. Characterization of antibiotic-resistance traits in *Akkermansia muciniphila* strains of human origin. *Sci Rep* 2022;12:19426.

77. **Kaviani R, Pouladi I, Niakan M, Mirnejad R**. Molecular detection of *adeFG* efflux pump genes and their contribution to antibiotic resistance in *Acinetobacter baumannii* clinical isolates. *Rep Biochem Mol Biol* 2020;8:413–418.

78. **Ketter PM, Yu J-J, Guentzel MN, May HC, Gupta R, *et al.*** *Acinetobacter baumannii* Gastrointestinal colonization is facilitated by secretory IgA which is reductively dissociated by bacterial thioredoxin A. *mBio* 2018;9:e01298-18.

79. **Wu L, Xie X, Li Y, Liang T, Zhong H, *et al.*** Metagenomics-based analysis of the age-related cumulative effect of antibiotic resistance genes in gut microbiota. *Antibiotics (Basel)* 2021;10:1006.

80. **Veloo ACM, Baas WH, Haan FJ, Coco J, Rossen JW**. Prevalence of antimicrobial resistance genes in *Bacteroides* spp. and *Prevotella* spp. Dutch clinical isolates. *Clin Microbiol Infect* 2019;25:1156.e9-1156.e13.

81. **Shoemaker NB, Vlamakis H, Hayes K, Salyers AA**. Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. *Appl Environ Microbiol* 2001;67:561–568.

82. **Patrick S**. *Bacteroides*. In: Tang Y-W, Sussman M, Liu D, Poxton I, Schwartzman J (editors). *Molecular Medical Microbiology*. London: Academic Press; 2015. pp. 917–944.

83. **Ghanbari M, Klose V, Crispie F, Cotter PD**. The dynamics of the antibiotic resistome in the feces of freshly weaned pigs following therapeutic administration of oxytetracycline. *Sci Rep* 2019;9:4062.

84. **Leng Z, Riley DE, Berger RE, Krieger JN, Roberts MC**. Distribution and mobility of the tetracycline resistance determinant *tetQ*. *Journal of Antimicrobial Chemotherapy* 1997;40:551–559.

85. **Sava IG, Heikens E, Huebner J**. Pathogenesis and immunity in enterococcal infections. *Clin Microbiol Infect* 2010;16:533–540.

86. **Alauzet C, Lozniewski A, Marchandin H**. Metronidazole resistance and *nim* genes in anaerobes: A review. *Anaerobe* 2019;55:40–53.

87. **Vishwanath S, Shenoy PA, Chawla K**. Antimicrobial resistance profile and *nim* Gene detection among *Bacteroides fragilis* group isolates in a university hospital in South India. *J Glob Infect Dis* 2019;11:59–62.

88. **Sethi S, Shukla R, Bala K, Gautam V, Angrup A, *et al.*** Emerging metronidazole resistance in *Bacteroides* spp. and its association with the *nim* gene: a study from North India. *Journal of Global Antimicrobial Resistance* 2019;16:210–214.

89. **Jahan L, Biswas R**. Molecular study on metronidazole resistance in *Bacteroides fragilis* group isolates from a South Indian tertiary care center. *Anaerobe* 2023;102692.

90. **Kouhsari E, Mohammadzadeh N, Kashanizadeh MG, Saghafi MM, Hallajzadeh M, *et al.*** Antimicrobial resistance, prevalence of resistance genes, and molecular characterization in intestinal *Bacteroides fragilis* group isolates. *APMIS* 2019;127:454–461.

91. **Akhi MT, Ghotaslou R, Alizadeh N, Yekani M, Beheshtirouy S, *et al.*** *nim* gene-independent metronidazole-resistant *Bacteroides fragilis* in surgical site infections. *GMS Hyg Infect Control* 2017;12:Doc13.

92. **Pumbwe L, Glass D, Wexler HM**. Efflux pump overexpression in multiple-antibiotic-resistant mutants of *Bacteroides fragilis*. *Antimicrob Agents Chemother* 2006;50:3150–3153.

93. **Veeranagouda Y, Husain F, Boente R, Moore J, Smith CJ, *et al.*** Deficiency of the ferrous iron transporter FeoAB is linked with metronidazole resistance in *Bacteroides fragilis*. *J Antimicrob Chemother* 2014;69:2634–2643.

94. **Steffens LS, Nicholson S, Paul LV, Nord CE, Patrick S, *et al.*** *Bacteroides fragilis* RecA protein overexpression causes resistance to metronidazole. *Res Microbiol* 2010;161:346–354.

95. **Löfmark S, Fang H, Hedberg M, Edlund C**. Inducible metronidazole resistance and *nim* genes in clinical *Bacteroides fragilis* group isolates. *Antimicrob Agents Chemother* 2005;49:1253–1256.

96. **Aqib AI, Alsayeqh AF**. Vancomycin drug resistance, an emerging threat to animal and public health. *Front Vet Sci* 2022;9:1010728.

97. **Courvalin P**. Transfer of antibiotic resistance genes between Gram-positive and Gram-negative bacteria. *Antimicrob Agents Chemother* 1994;38:1447–1451.

98. **Xing L, Yu H, Qi J, Jiang P, Sun B, *et al.*** *ErmF* and *ereD* are responsible for erythromycin resistance in *Riemerella anatipestifer*. *PLoS One* 2015;10:e0131078.

99. **Niestępski S, Harnisz M, Korzeniewska E, Aguilera-Arreola MaG, Contreras-Rodríguez A, *et al.*** The emergence of antimicrobial resistance in environmental strains of the *Bacteroides fragilis* group. *Environment International* 2019;124:408–419.

100. **Uluseker C, Kaster KM, Thorsen K, Basiry D, Shobana S, *et al.*** A review on occurrence and spread of antibiotic resistance in wastewaters and in wastewater treatment plants: mechanisms and perspectives. *Frontiers in Microbiology*;12. https://www.frontiersin.org/articles/10.3389/fmicb.2021.717809 (2021, accessed 9 March 2023).

101. **Valdezate S, Cobo F, Monzón S, Medina-Pascual MJ, Zaballos Á, *et al.*** Genomic background and phylogeny of *cfiA*-positive *Bacteroides fragilis* strains resistant to meropenem-EDTA. *Antibiotics (Basel)* 2021;10:304.

102. **Coyne MJ, Roelofs KG, Comstock LE**. Type VI secretion systems of human gut *Bacteroidales* segregate into three genetic architectures, two of which are contained on mobile genetic elements. *BMC Genomics* 2016;17:58.

103. **Tajkarimi M, Wexler HM**. CRISPR-Cas systems in *Bacteroides fragilis*, an important pathobiont in the human gut microbiome. *Front Microbiol* 2017;8:2234.

104. **Sóki J, Hedberg M, Patrick S, Bálint B, Herczeg R, *et al.*** Emergence and evolution of an international cluster of MDR *Bacteroides fragilis* isolates. *Journal of Antimicrobial Chemotherapy* 2016;71:2441–2448.

105. **Yang W, Moore IF, Koteva KP, Bareich DC, Hughes DW, *et al.*** TetX Is a flavin-dependent monooxygenase conferring resistance to tetracycline antibiotics. *Journal of Biological Chemistry* 2004;279:52346–52352.

106. **Park BH, Levy SB**. The cryptic tetracycline resistance determinant on Tn4400 mediates tetracycline degradation as well as tetracycline efflux. *Antimicrob Agents Chemother* 1988;32:1797–1800.

107. **Speer BS, Salyers AA**. Characterization of a novel tetracycline resistance that functions only in aerobically grown *Escherichia coli*. *J Bacteriol* 1988;170:1423–1429.

108. **Deekshit VK, Srikumar S**. 'To be, or not to be'—The dilemma of 'silent' antimicrobial resistance genes in bacteria. *Journal of Applied Microbiology* 2022;133:2902–2914.

109. **Odamaki T, Bottacini F, Kato K, Mitsuyama E, Yoshida K, *et al.*** Genomic diversity and distribution of *Bifidobacterium longum* subsp. *longum* across the human lifespan. *Sci Rep* 2018;8:85.

110. **Tomida S, Nguyen L, Chiu B-H, Liu J, Sodergren E, *et al.*** Pan-genome and comparative genome analyses of *Propionibacterium acnes* reveal its genomic diversity in the healthy and diseased human skin microbiome. *mBio* 2013;4:e00003-00013.

111. **Deng X, Phillippy AM, Li Z, Salzberg SL, Zhang W**. Probing the pan-genome of *Listeria monocytogenes*: new insights into intraspecific niche expansion and genomic diversification. *BMC Genomics* 2010;11:500.

112. **Salipante SJ, Roach DJ, Kitzman JO, Snyder MW, Stackhouse B, *et al.*** Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. *Genome Res* 2015;25:119–128.

113. **Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, *et al.*** Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc Natl Acad Sci U S A* 2015;112:E3574-3581.

114. **Park S-C, Lee K, Kim YO, Won S, Chun J**. Large-scale genomics reveals the genetic characteristics of seven species and importance of phylogenetic distance for estimating pan-genome size. *Front Microbiol* 2019;10:834.

115. **Albert K, Rani A, Sela DA**. Comparative pangenomics of the mammalian gut commensal *Bifidobacterium longum*. *Microorganisms* 2019;8:7.

116. **Patrick S, Blakely GW, Houston S, Moore J, Abratt VR, *et al.*** Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of *Bacteroides fragilis*. *Microbiology (Reading)* 2010;156:3255–3269.

117. **Husain F, Tang K, Veeranagouda Y, Boente R, Patrick S, *et al.*** Novel large-scale chromosomal transfer in *Bacteroides fragilis* contributes to its pan-genome and rapid environmental adaptation. *Microbial Genomics* 2017;3:e000136.

118. **Pudlo NA, Urs K, Crawford R, Pirani A, Atherly T, *et al.*** Phenotypic and genomic diversification in complex carbohydrate-degrading human gut bacteria. *mSystems* 2022;7:e0094721.

119. **Porter NT, Martens EC**. The critical roles of polysaccharides in gut microbial ecology and physiology. *Annu Rev Microbiol* 2017;71:349–369.

120. **El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B**. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol* 2013;11:497–504.

121. **Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, *et al.*** Adaptive evolution within gut microbiomes of healthy people. *Cell Host Microbe* 2019;25:656-667.e8.

Author statements

Authors and contributors

Data curation: JE, LH. Investigation: JE, LH, FN. Formal analysis: JE, LH, FN. Methodology: LH, FN. Supervision: LH, SP, LS. Visualization: LH, FN. Writing – original draft: JE, LH, FN, SP, LS. Writing – reviewing and editing: all authors.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Funding information

This work received no specific grant from any funding agency.

Acknowledgements

High-performance computing resources used in this study were supported by the Research Contingency Fund of the Department of Biosciences, Nottingham Trent University.

Table 1: Summary statistics generated from Panaroo pangenome analysis of 377 *B. fragilis* genomes

Pangenome component*	**Present in strains**	**No. of genes**	**Proportion of genes (%)**
Core	99 % <= strains <= 100 %	2,175	8.8
Soft core	95 % <= strains <99 %	517	2.1
Shell	15 % <= strains <95 %	2,519	10.3
Cloud	0 % <= strains <15 %	19,240	78.6
Total	0 % <= strains <= 100 %	24,451	100

* The accessory genome comprises shell and cloud pangenome components.

Table 2: Overview of KO IDs that were found exclusively in either *B. fragilis sensu stricto* or *B. fragilis A*. KEGG description, BRITE description, adjusted *P* value (Benjamini-Hochberg) and count in species also shown.

KO	Adjusted *P* value	Count in *B. fragilis*	Count in *B. fragilis* A	KEGG description	BRITE description
K08998	7.67×10^{-81}	275	0	Unknown	Unknown
K08717	7.67×10^{-81}	275	0	Urea transporter (utp)	Transporters
K07267	7.67×10^{-81}	275	0	Porin (oprB)	Transporters
K05989	2.90×10^{-80}	549	0	Alpha-L-rhamnosidase (ramA)	Hydrolases
K03498	2.90×10^{-80}	276	0	trk/ktr system potassium uptake protein	Transporters
K03551	9.35×10^{-80}	274	0	Holliday junction DNA helicase RuvB	DNA repair and recombination
K01424	1.09×10^{-78}	273	0	L-asparaginase (ansA,ansB)	Hydrolases
K18369	1.10×10^{-72}	267	0	Alcohol dehydrogenase (adh2)	Oxidoreductases
K03648	1.10×10^{-72}	267	0	Uracil-DNA glycosylase	DNA repair and recombination
K05520	7.67×10^{-81}	0	102	Protease I (pfpI)	Peptidases and inhibitors
K00865	7.67×10^{-81}	0	102	Glycerate 2-kinase (garK)	Transferases

FIGURE LEGENDS

Figure 1. Phylogenetic tree showing relationships of '*Bacteroides fragilis*' genomes with members of the genus *Bacteroides*. Taxonomic information based on GTDB annotations. Most (*n* = 275) genomes (shown in yellow) were affiliated with *B. fragilis sensu stricto*, with the remainder (*n* = 102; shown in green) affiliated with *Bacteroides fragilis* A. The tree was created using PhyloPhlAn. Scale bar, average number of amino acid substitutions per site.

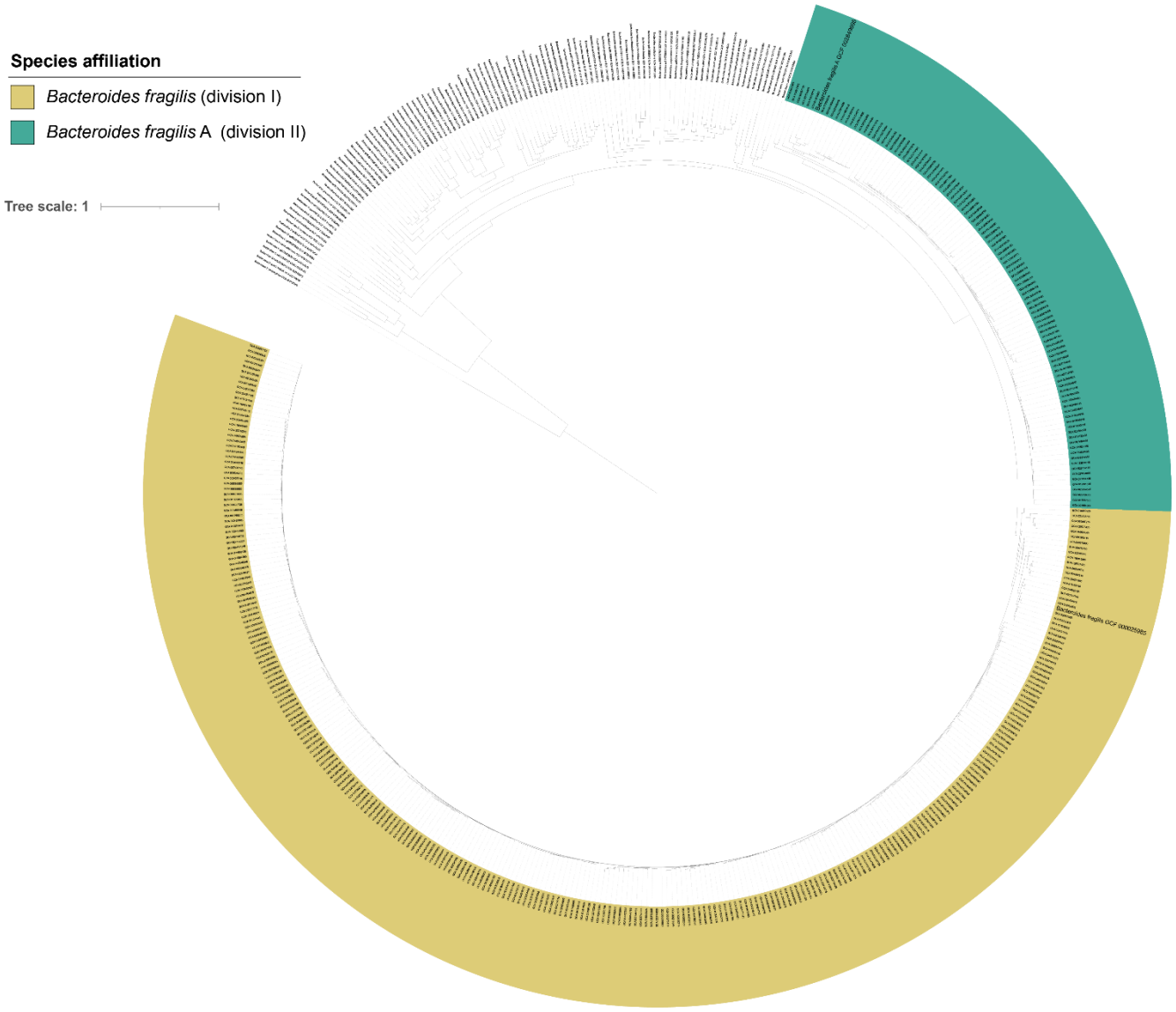
Figure 2. AMR genes predicted to be encoded in *B. fragilis* (Division I; *n* = 275) and *B. fragilis* A (Division II; *n* = 102) genomes. The phylogenetic tree was generated using PhyloPhlAn, and rooted at the midpoint. Strict CARD match, not identical to a CARD entry but the bit score of the matched sequence is greater than the curated BLASTP bit score cut-off; perfect CARD match, 100 % identical to the reference CARD sequence along its entire length. Loose matches are not shown to avoid presenting false positives based on sequences with low homology and bit scores below CARD BLASTP cut-off recommendations.

Figure 3. PCA of the accessory genome (genes present in 5-95%) of all *Bacteroides fragilis sensu stricto* (Division I; *n* = 275 and *Bacteroides fragilis* A (Division II; *n* = 102) genomes.

Figure 4. *B. fragilis sensu stricto* (Division I; *n* = 275) and *B. fragilis A* (Division II; *n* = 102) core SNP maximum likelihood tree generated from the core genome alignment. The inner ring shows classification (*B. fragilis sensu stricto* or *B. fragilis A*) and outer ring shows the designated fastbaps cluster (Cluster 1 or Cluster 2). The phylogenetic tree was generated with IQTree and iToL. The scale bar represents the average number of SNPs per site.

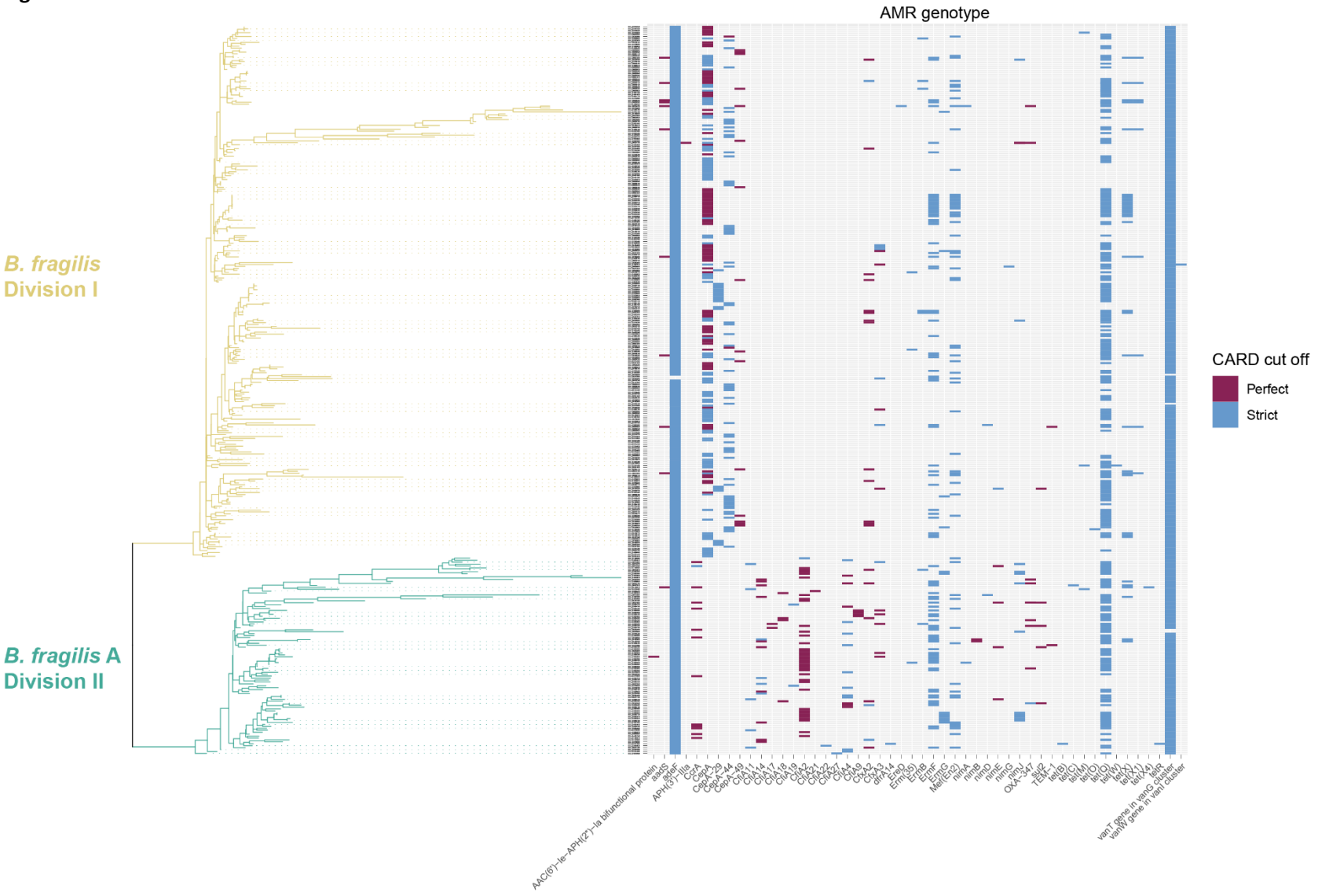
Figure 5. Accessory gene-based functional map of *B. fragilis sensu stricto* and *B. fragilis* A. The figure was generated from eggnog mapper server output using the Panaroo pangenome reference fasta file. The KOs associated with the accessory genome were retained and KO table input to FuncTree2 for visualisation. Significant KO values (adjusted *P* value <0.05; Benjamini-Hochberg) were determined using Wilcoxon test. Each ring of the circular dendrogram represents a different functional layer of the KEGG functional hierarchy (inner ring to outer ring: Biological Category, Biological Process, KEGG Pathway, KEGG Module; see labels). The module coverage of each functional layer is represented by the size of the circle and coloured according to Biological Category (e.g. all layers associated with Metabolism have yellow-coloured circles). The columns within the circle show the total of each KO value associated with *B. fragilis* (yellow columns) or *B. fragilis* A (green columns) with 100 % stacking. The significant KO values have been annotated in the outer ring of the circle and show the location within the functional hierarchy. See Supplementary Material for KOs that could not be assigned to a pathway.

826 **Figure 1**



827

Figure 2


AMR genotype
B. fragilis
Division I
B. fragilis A
Division II
CARD cut off
Perfect
Strict
AAC(6')-Ie-APH(2")-Ia bifunctional protein
aadS
aphF
APH(3")-IIIa
CepA
CepA-29
CepA-44
CepA-49
CfiA11
CfiA14
CfiA17
CfiA18
CfiA19
CfiA2
CfiA21
CfiA22
CfiA27
CfiA4
CfiA9
CfxA2
CfxA3
dfrA14
EreD
Erm(35)
ErmB
ErmF
ErmG
Mef(En2)
ninA
ninB
ninD
ninE
ninG
ninJ
OXA-347
sul2
TEM-1
tet(B)
tet(G)
tet(M)
tet(O)
tet(Q)
tet(W)
tet(X)
tet(X14)
tetR
vanT gene in vanG cluster
vanW gene in vanI cluster

829

Figure 3

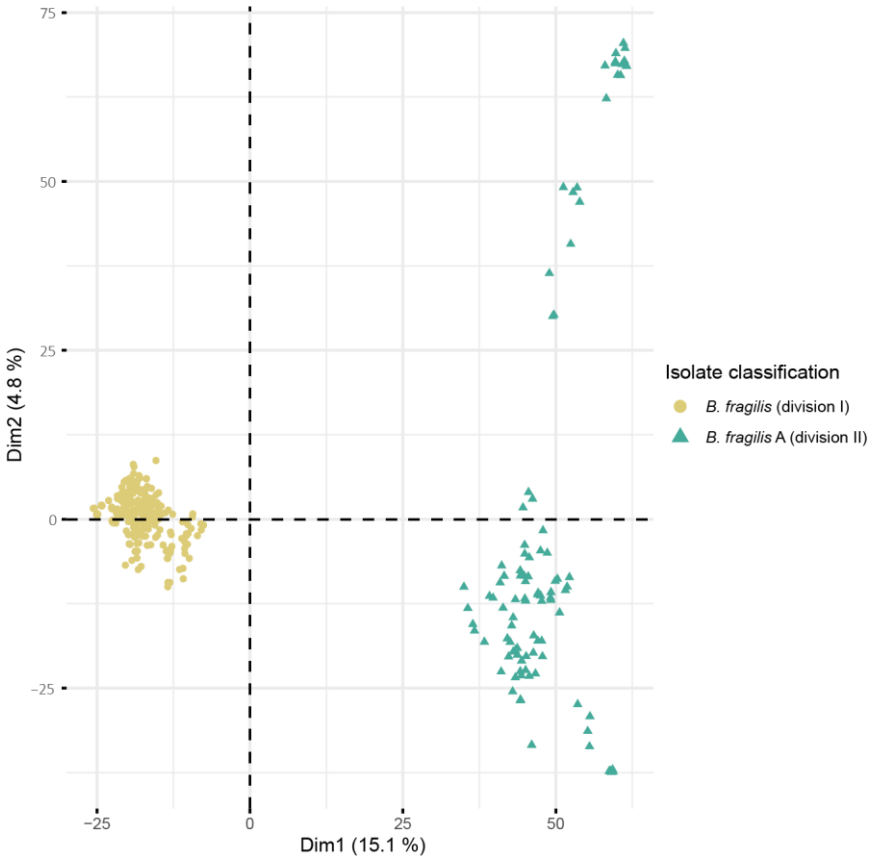

75
50
25
0
−25
Dim2 (4.8 %)
−25
0
25
50
Dim1 (15.1 %)
Isolate classification
B. fragilis (division I)
B. fragilis A (division II)

Figure 4

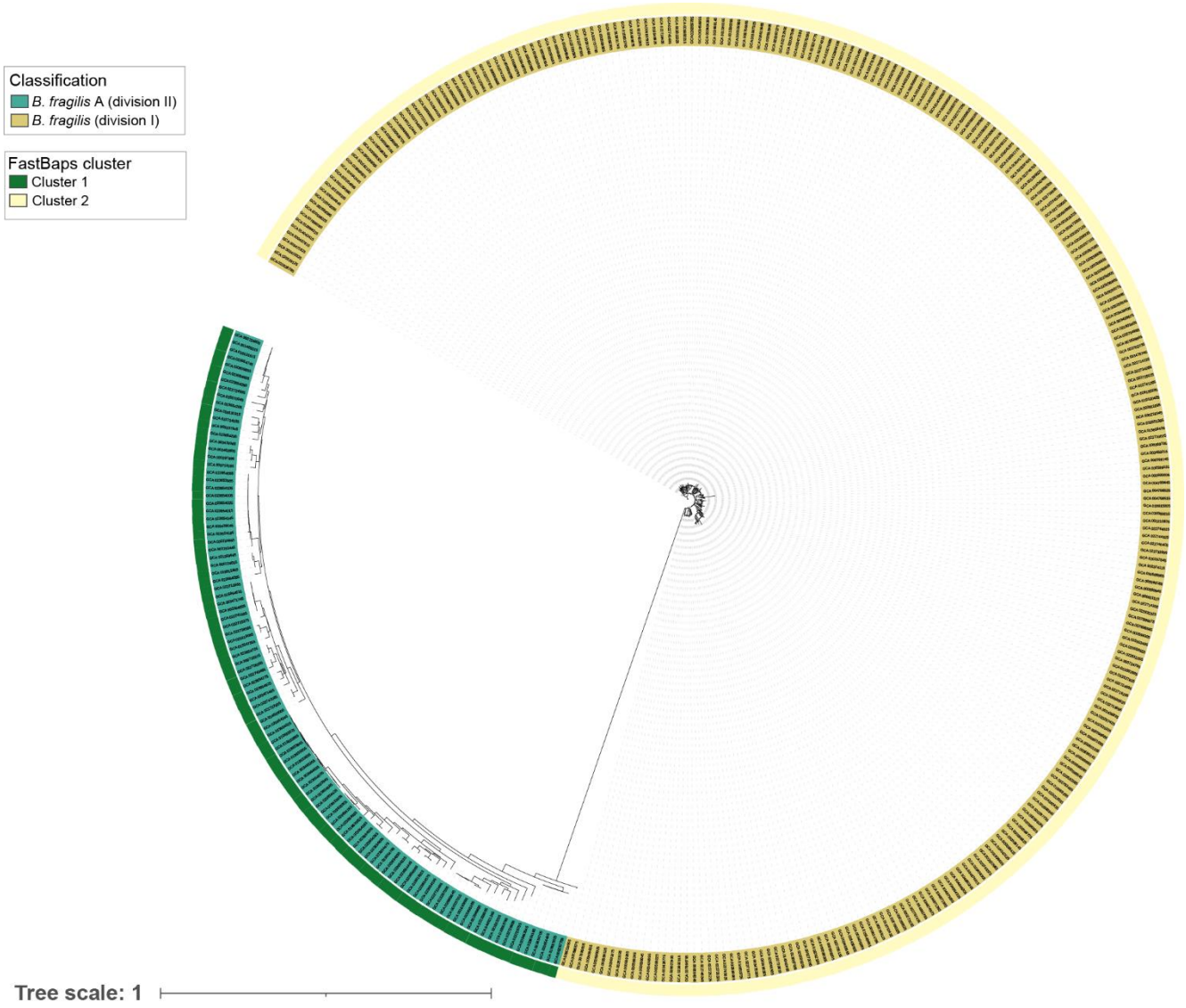
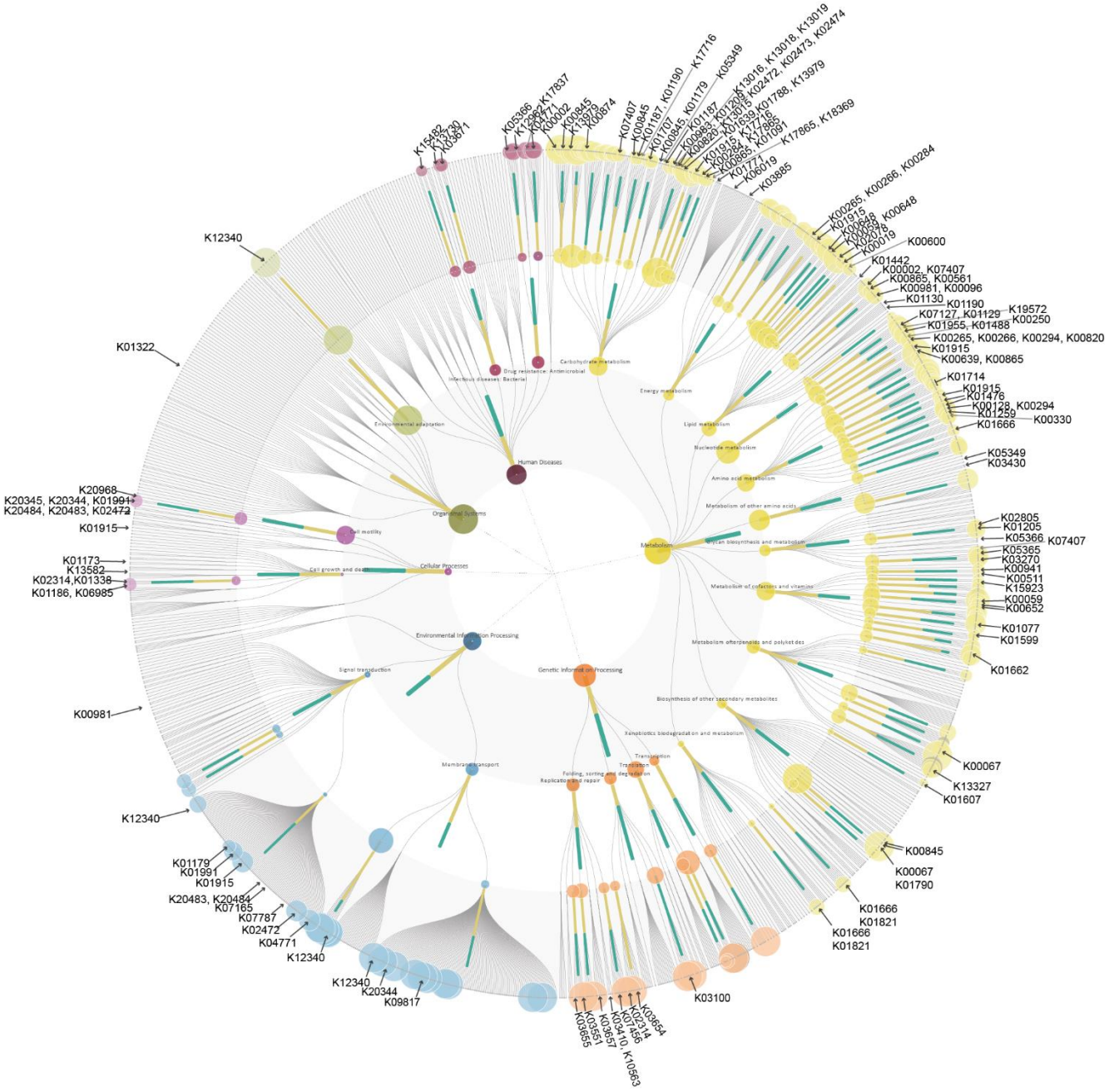
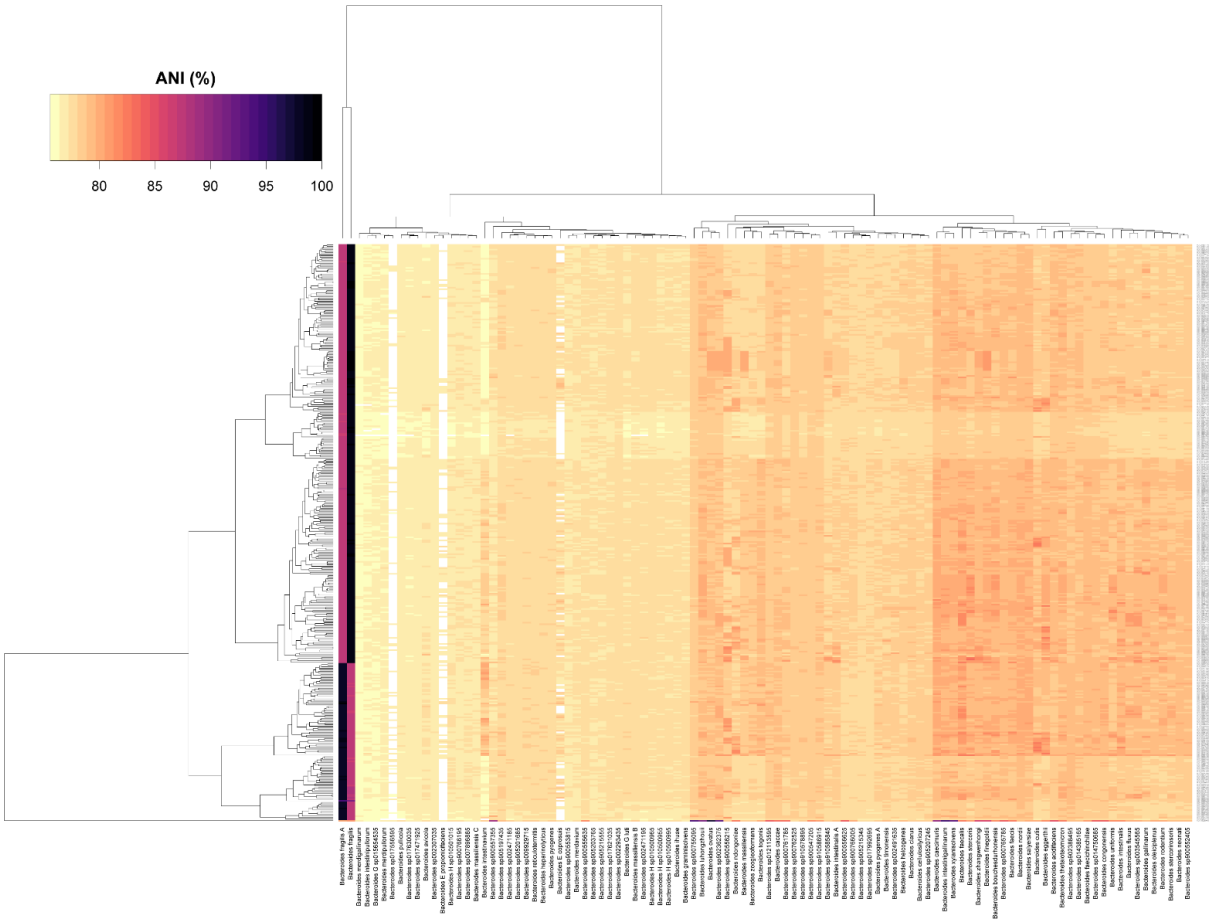

Classification
B. fragilis A (division II)
B. fragilis (division I)
FastBaps cluster
Cluster 1
Cluster 2
Tree scale: 1

Figure 5

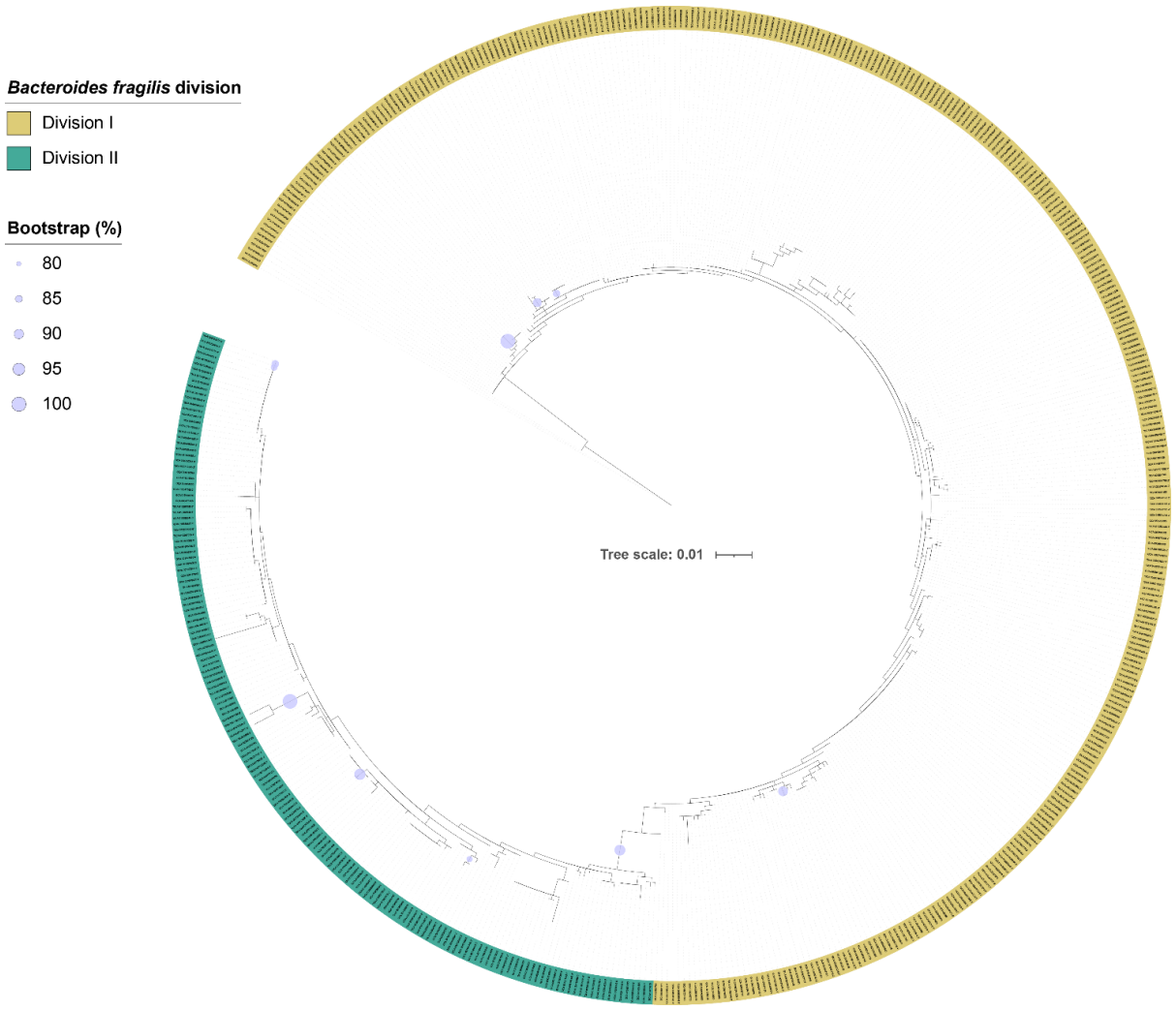

K15482, K03170, K03671
K05366, K12262, K17837
K04771
K00002
K01845
K13979
K00874
K07407
K00845
K01187, K01190
K17716
K01179
K05349
K13016, K13018, K13019
K02472, K02473, K02474
K01788, K13979
K17865, K18369
K03885
K00265, K00266, K00284
K01915
K00648
K00639
K00978
K00019
K00600
K01442
K00002, K07407
K00865, K00561
K00981, K00096
K01130
K01190
K19572
K07127, K01129
K00250
K01955, K01488
K00265, K00266, K00294, K00820
K01915
K00639, K00865
K01714
K01915
K01476
K00128, K00294
K01259
K00330
K01666
K05349
K03430
K02805
K01205
K05366, K07407
K05365
K03270
K00941
K00511
K15923
K00059
K00652
K01077
K01599
K01662
K00067
K13327
K01607
K00845
K00067
K01790
K01666
K01821
K01666
K01821
K03100
K03654
K02314
K07456
K03410, K10563
K03657
K03551
K03655
K12340
K20344
K09817
K12340
K12340
K07787
K02472
K04771
K12340
K01179
K01991
K01915
K20483, K20484
K07165
K00981
K12340
K01322
K20968
K20345, K20344, K01991
K20484, K20483, K02472
K01915
K01173
K13582
K02314,K01338
K01186, K06985
Human Diseases
Organismal Systems
Metabolism
Genetic Information Processing
Environmental Information Processing
Cellular Processes
Environmental adaptation
Drug resistance: Antimicrobial
Infectious diseases: Bacterial
Carbohydrate metabolism
Energy metabolism
Lipid metabolism
Nucleotide metabolism
Amino acid metabolism
Metabolism of other amino acids
Glycan biosynthesis and metabolism
Metabolism of cofactors and vitamins
Metabolism of terpenoids and polyketides
Biosynthesis of other secondary metabolites
Xenobiotics biodegradation and metabolism
Translation
Folding, sorting and degradation
Replication and repair
Membrane transport
Signal transduction
Cell motility
Cell growth and death

SUPPLEMENTARY FIGURES ASSOCIATED WITH ENGLISH *ET AL.*

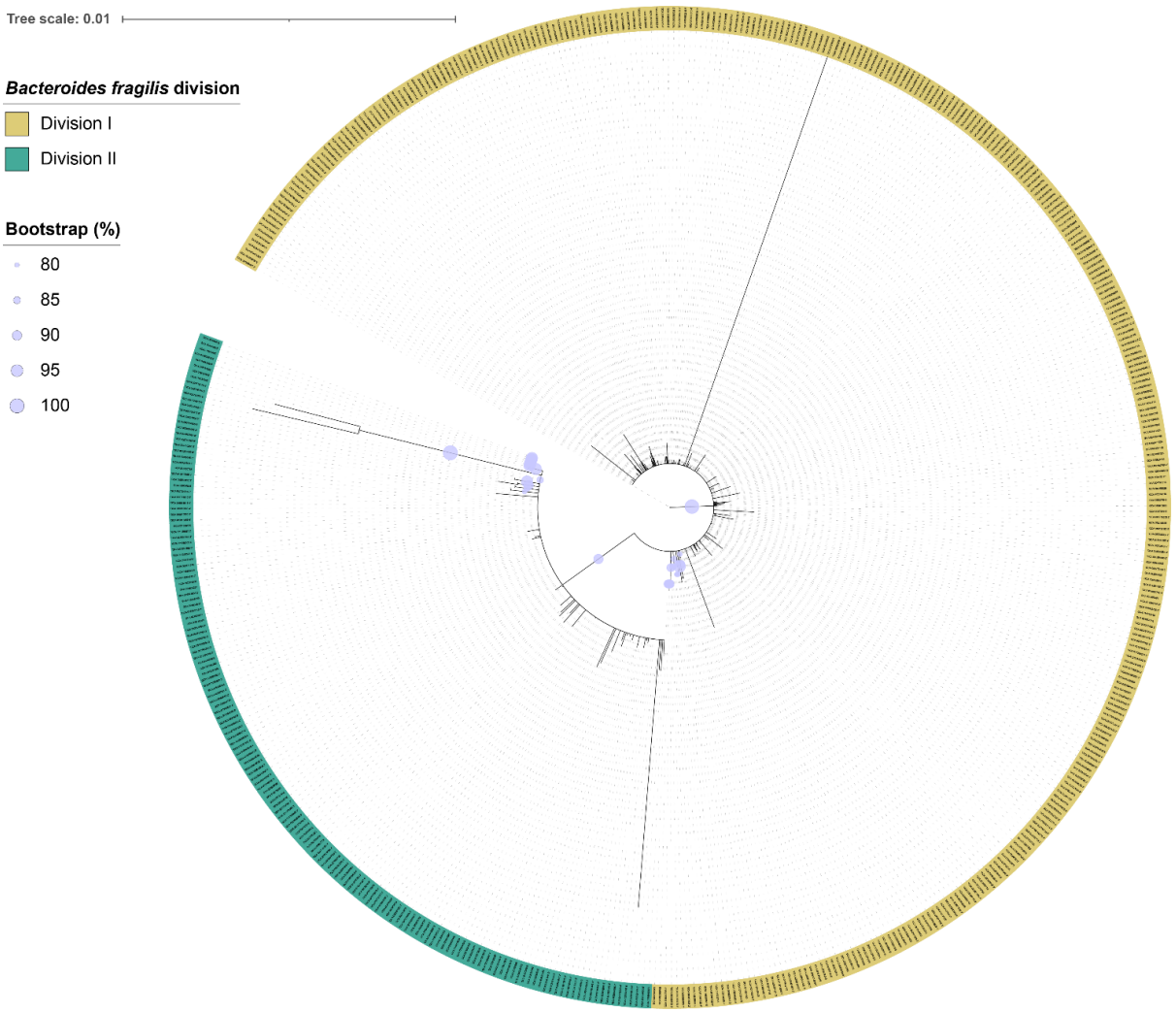
Genomic analyses of *Bacteroides fragilis*: subdivisions one and two represent distinct species



Supplementary Figure 1. Bidirectional clustered heatmap showing results from an ANI analysis of high-quality '*Bacteroides fragilis*' genomes (*n* = 379) downloaded from NCBI GenBank. Genomes were subject to an all-versus-all fastANI analysis along with *Bacteroides* spp. reference genomes (**Supplementary Table 2**) to confirm species identities. Most genomes clustered with the reference genomes of *B. fragilis* or *B. fragilis* A (*n* = 275 and *n* = 102 genomes, respectively).



Supplementary Figure 2. Maximum-likelihood tree showing the phylogenetic relationship between 16S rRNA gene sequences encoded within 170 *B. fragilis* (division I) and 61 *B. fragilis* A (division II) genomes. Some genomes encoded more than one copy of the 16S rRNA gene. The tree was generated from a multiple-sequence alignment of 522 16S rRNA gene sequences. Bootstrap values (represented by circles, size relative to a percentage of 100 replications) are shown at nodes. Scale bar, average number of nucleotide substitutions per position.



Supplementary Figure 3. Neighbour-joining tree showing the similarity between 16S rRNA gene sequences encoded within 170 *B. fragilis* (division I) and 61 *B. fragilis* A (division II) genomes. Some genomes encoded more than one copy of the 16S rRNA gene. The tree was generated from a multiple-sequence alignment of 522 16S rRNA gene sequences. Bootstrap values (represented by circles, size relative to a percentage of 100 replications) are shown at nodes.