BMC Genomics

RESEARCH

Open Access

Genomic alterations in *Bacteroides fragilis* favor adaptation in colorectal cancer microenvironment

Hao Yang[1†], Yu Gan[1†], Shenghai Jiang[1], Xianchang Zhu[1], Yang Xia[4], Dengmei Gong[1], Xianrang Xie[1], Yao Gong[1], Yi Zhang[1,2], Qian Lei[1], Maijian Wang[3,5*] and Jida Li[1,2,6*]

Abstract

Background The occurrence and development of colorectal cancer (CRC) is an incredibly long process that involves continuous changes in the tumor microenvironment. These constant changes may ultimately result in genetic alterations and changes in the metabolic processes of some symbiotic bacteria as a way to adapt to the changing environment. Patients with CRC exhibit an altered abundance of *Bacteroides fragilis* (*B. fragilis*) as indicated by several studies. To better understand the genomic characteristics and virulence spectrum of *B. fragilis* strains in tumor tissues, *B. fragilis* strains were isolated from tumor and paracancerous tissues of CRC patients.

Methods The isolates were identified using 16 S rRNA sequencing, morphological analysis, physiological and biochemical characterization and PCR, and they were then subjected to whole genome sequencing (WGS) analysis.

Results A strain of *B. fragilis* enterotoxin (BFT) bft1-producing ZY0302 and a non-enterotoxin-producing *B. fragilis* ZY0804 were isolated from cancerous and paraneoplastic tissues, respectively. Analysis based on the core and nonessential genes showed that the genomic profiles of the isolates, ZY0302 and ZY0804, differed from those of *B. fragilis* from other tissue sources. This core and the co-evolution of non-essential genes may be the result of their adaptation to fluctuations in the tumor microenvironment and enhancing their survival. In addition, the ZY0302 and ZY0804 genomes underwent extensive horizontal gene transfer and varying degrees of genomic rearrangements, inversions, insertions, and deletion events, which may favor the enhancement of bacteria's ability to adapt to environmental changes. For instance, the virulence factors, such as the capsular biosynthesis gene clusters and components of the type IV secretion system, acquired through horizontal gene transfer, may facilitated *B. fragilis* in evading immune responses and managing oxidative stress. Moreover, our analysis revealed that multiple virulence factors identified in the isolates were mainly involved in bacterial adhesion and colonization, oxidative stress, iron acquisition, and immune evasion. This observation is worth noting given that enzymes such as neuraminidase, lipase,

[†]Hao Yang and Yu Gan contributed equally to this work.

*Correspondence:
Maijian Wang
864205468@qq.com
Jida Li
lijida198485@163.com

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

hemolysin, protease, and phosphatase, along with genes responsible for LPS biosynthesis, which are recognized for their association with the virulence of *B. fragilis*, were prevalent among the isolates.

Conclusions In summary, it is our assertion that the alterations observed in both core and nonessential genes of *B. fragilis*, which have been isolated from tissues of colorectal cancer patients, along with significant instances of horizontal gene transfer to the genome, are likely intended to enhance adaptation to the evolving conditions of the tumor microenvironment. This study may provide new insights into the interaction between *B. fragilis* and the CRC microenvironment.

Keywords Colorectal cancer, *Bacteroides fragilis*, Complete genome, Tumor microenvironment, Virulence factor

Introduction

As one of the most common cancer types, colorectal cancer (CRC) is recognized as the third most diagnosed cancer and also ranks as the second leading cause of cancer-related deaths, highlighting its significant threat to human health [1]. Currently, the pathogenesis of CRC remains obscure as its occurrence is complex and involves multiple factors such as genetics, environment, diet and lifestyle. As the key metabolic and immunomodulatory factors of the host, intestinal microorganisms participate in crucial regulatory roles where they mediate environmental factors and the occurrence and development of CRC [2]. The development of sequencing technology has provided evidence to show that the microbial diversity in tumor tissues of patients with CRC is altered and that this change may be an important factor in promoting the development of CRC [3, 4].

B. fragilis is a commensal Gram-negative facultative anaerobic bacterium found mainly in the colon, accounting for approximately 1% of the intestinal microbiota [5]. *B. fragilis* is also an opportunistic pathogen that has been strongly associated with the development of acute diarrhea, inflammatory bowel disease (IBD), and CRC, and it is one of the major causative agents of anaerobic sepsis [6]. *B. fragilis* is classified into enterotoxin-producing *B. fragilis* (ETBF) and non-enterotoxin-producing *B. fragilis* (NTBF), depending on its ability to secrete *B. fragilis* enterotoxin (BFT) [5]. BFT is a ~ 20-kDa zinc metalloprotease [7], which cleaves E-calmodulin, disrupts inter-epithelial cellular barriers [8], and induces *c-Myc* expression and cellular proliferation [9], as well as activating helper T cell 17 (Th17) response and NF-κB signaling pathway [10], thereby promoting intestinal inflammation and colorectal carcinogenesis. In addition, ETBF induces cellular DNA damage, thereby increasing the risk of cancer development [11]. Recent studies have also shown that ETBF and its generated BFT are associated with high levels of CpG island methylation in CRC and may promote colorectal tumorigenesis through the serrated tumor pathway [12, 13]. NTBF, however, has been recognized as a potential probiotic that promotes the production of short-chain fatty acids [14] and inhibits the growth of other pathogenic microorganisms [15]. Additionally, it contributes to the diversity of intestinal flora and prevents intestinal epithelial damage, which in turn inhibits intestinal inflammation and inflammation-associated colorectal cancer [16, 17]. Polysaccharide A (PSA) is a major effector of NTBF, and TLR2 signaling plays a crucial role in mediating the protective effect against colitis-associated CRC [17].

Related studies have shown that NTBF strains isolated from different tissues may exhibit distinct biological characteristics. For example, the use of proteomics indicates that NTBF strains from patients with polyps are different from NTBF strains isolated from patients without polyps. NTBF strains from patients with polyps are rich in LPS biosynthetic genes, which may increase their ability to activate the immune system and cause inflammation [18]. Various changes may occur during the development and progression of CRC such as the growth of the transformed tissues and the secretion of various tissue factors and cellular metabolites. Consequently, the local environment may change, making it difficult for some commensal bacteria to survive. As a means of adaptation to the altered microenvironment, commensal bacteria may switch on new genetic and metabolic programs [18]. A research study found that *B. fragilis* possesses numerous DNA inversion systems distributed across its genome. This genetic feature allows the organism to selectively express a diverse array of cell surface structural proteins, which in turn enhances its ability to adapt to varying nutritional environments, assists in evading the host immune response, and boosts its potential for colonization and infection [19, 20]. In conclusion, *B. fragilis* can adapt to tissue-specific and microenvironment changes through genomic changes.

Researchers have long speculated about the role of gut bacteria in the development of CRC, and the isolation and culture methods of target bacteria have made the exploration of the etiology of CRC easier and achievable [21]. Koch's postulates have long proposed that microbial culture is the gold standard for detecting bacteria, but most bacteria in the colon are "unculturable", which makes it a great challenge to further explore the pathogenic mechanism of intestinal pathogens [21]. The isolation of bacteria from clinical specimens is usually affected

by many factors, as such studies have pointed out that the ideal specimen (tissue or fluid) for the isolation of bacteria should be obtained during surgery [22]. When isolating and culturing *B. fragilis*, proper anaerobic collection containers and methods must be ensured. *B. fragilis* is an obligate anaerobe and requires specialized transport conditions, such as dedicated anaerobic transport media [22]. The anaerobic nature of *B. fragilis* requires specialized environments for its isolation and culture, which makes research challenging. Therefore, the primary goal of this study is to successfully and effectively isolate *B. fragilis* from CRC tumors and paracancerous tissues and culture it.

At present, *B. fragilis* used in CRC research is obtained from skin tissue after a traumatic infection, fecal samples, patients with septicemia, or from other species like piglets; however, isolates from tumor sites of patients are rarely reported. To narrow the gap in infrequent reports pertaining to isolates from tumor sites, the aim of this study was to isolate *B. fragilis* from tumors and paracancerous tissues of CRC patients. The whole genome sequences of these isolates were analyzed to investigate their evolutionary features, as well as the virulence genes of the isolates. The aim is to provide a better understanding of the genomic characteristics and virulence profiles in CRC tumor tissues.

Materials and methods

Specimen collection

Samples were collected from tumor tissues and paracancerous tissues of CRC patients from the Affiliate Hospital of Zunyi Medical University (Guizhou, China), and informed and signed consent was obtained from all patients. After collection into sterile sampling bags, the samples were stored in ice boxes and immediately sent to the laboratory for testing. This study was approved by the Medical Ethics Review Committee (ZunYi ethical examination No. [2023] 1–014).

Isolation and purification of *B. fragilis*

According to the method reported in literature [23, 24], the specific scheme of bacterial isolation is as follows: Tissue samples were obtained from the center of the tumor and paracancerous tissue using an aseptic surgical knife cauterized by an alcohol lamp. The tissue sections were inoculated into Brain Heart Infusion Broth (BHI Broth) medium with an aseptic inoculation loop, and the anaerobic culture was incubated at 37℃ for 18–24 h. Then the bacterial solution was inoculated into *Bacteroides Bile Esculin* (BBE) blood plates and anaerobically incubated at 37℃ for 24–48 h. The colony morphology, color, size and surface properties were studied and observed. At least three single colonies (slightly convex morphology, smooth and shiny surface, neat edge, brown or off-white, and black surroundings) were selected from each plate and inoculated into the CDC anaerobic blood plate. The same colony was inoculated into three separate plates and incubated under different conditions, including anaerobic, microaerobic, and aerobic environments at 37 °C for 24–48 h. Observed the growth status of the colonies in the three environments, selected the colonies that grow only in anaerobic environment, inoculated them on BBE blood plates and incubated them anaerobically for 24–48 h. After that, picked a single colony and inoculated it into 10 mL of BHI Broth medium, and then incubated it anaerobically at 37℃ for 18–24 h to get the purified strains. Blood plates used for isolation contained 5% sheep blood, vitamin K1 (0.5 mg/L), and hemin (5 mg/L). Each purified strain was made into a glycerol stocks using 0.25 mL sterile glycerol, 0.75 mL sterile PBS and 0.5 mL bacterial solution mixed thoroughly and preserved at -20℃ and liquid nitrogen.

Identification of target strains

Morphological, physiological and biochemical characteristics

After anaerobic incubation at 37℃ for 24 h in BHI Broth medium, the isolated strains were then inoculated on CDC anaerobic blood plate containing 5% sheep blood, vitamin K1 (0.5 mg/L), and hemin (5 mg/L), as well as BBE blood plates. These plates were then incubated at 37℃ for 24–48 h, and their morphological characteristics were studied and observed. The isolates were stained with Gram staining solution (HuanKai Microbial, Guangdong, China), and the shape and size of the bacteria were observed using an optical microscope (OLYMPUS BX-53, Tokyo, Japan). The physiological and biochemical characteristics of the isolated strains were evaluated using microbial microbiochemical identification tubes (HuanKai Microbial, Guangdong, China).

DNA extraction, PCR amplification and phylogenetic analysis

The isolated strains were inoculated in BHI Broth medium and anaerobically incubated at 37℃ for 18–24 h. Genomic DNA was extracted utilizing Ezup column bacteriological DNA extraction kit (Sangon Biotech, Shanghai, China). To identify the species of isolated strains, the sequences of 16 S ribosomal RNA gene (*16 S rRNA*) were amplified using universal primers, 27 F (5′-AGAGTTTG ATCATGGCTCAG-3′) and 1492R (5’-GGTTACCTTG TTACGACTT-3’). The final PCR master mix was 50 μL, including 5 μL 10 × PCR Buffer (Mg2 + Plus), 0.5 μL r Taq (250U, 5U/mL), 2 μL dNTP Mixture (2.5 mM each), 0.5 μL each of upstream and downstream primers (10 μM), 2.5 μL genomic DNA, and 39 μL ddH2O. PCR amplification was carried out using BIO-RAD T100TM Thermal Cycler, the samples were pre-denatured at 95℃ for 5 min, denatured at 95℃ for 30 s, annealed at 55℃, extended for 1 min at 72℃ for 30 cycles, and finally extended for

10 min at 72℃. After the completion of the process, 5 μL of the PCR product was used for 1% agarose gel electrophoresis to confirm the fragment amplified by PCR. Subsequently, the PCR product was sent to Sangon Biotech (Shanghai) Co., Ltd for sequencing. The sequencing results were analyzed using BLAST alignment by the National Center for Biotechnology Information (NCBI). According to the BLAST results, 16 S rRNA sequences which were closely related to the sequences in this study were retrieved from GenBank for phylogenetic analysis. The phylogenetic tree was constructed by the maximum likelihood method by MEGA11 software, and 1000 Bootstrap repeated samples were used to evaluate the reliability and statistical significance of the constructed phylogenetic tree.

After the 16 S rRNA results confirmed that the strain was *B. fragilis*, the BFT, bft-1, bft-2, and bft-3 gene sequences of the strains were amplified by PCR using specific primers, and the type of the isolated strain was determined by 1% agarose gel electrophoresis. The primer sequence is shown in Table 1. The final PCR master mix and conditions employed were similar to those used to amplify the 16 S rRNA sequence.

Genome sequencing and assembly

Genomic DNA was extracted using the Ezup Column Bacterial Genomic DNA Extraction Kit (Sangon Biotech). The concentration of the DNA samples was detected using a Qubit fluorescence quantifier; the integrity of the DNA samples was detected by 1% agarose gel electrophoresis. The DNA samples were ultrasonically interrupted using a Covaris instrument, and fragment selection was carried out with an Agencourt AMPure XP-Medium kit to make the sample bands concentrated around 300–400 bp; the amount of purified DNA samples was detected using the kit Qubit dsDNA HS Assay Kit 500 assays. Double-stranded DNA end repair was then performed and an A base was added to the 3' end. The DNA was ligated to the junction and the ligation product was amplified. The amplification products were screened for fragments with the reagent Agencourt AMPure XP-Medium. PCR products were detected with an Agilent 2100 Bioanalyzer. After denaturing the PCR products to single-stranded forms, single stranded cyclization of the single strands was performed and then the uncyclized linear DNA molecules were digested to obtain the final library. The fragment size and concentration of the library were detected using an Agilent 2100 Bioanalyzer (Agilent DNA 1000 Reagents). The single-stranded circular DNA molecules were replicated by ring rolling to form a DNA nanoball (DNB) containing more than 300 copies. The resulting DNBs were spiked into mesh pores on the chip using high-density DNA nano-chip technology. Sequencing was performed by co-probe anchored polymerization (cPAS). The *B. fragilis* strain S1-BF-GZ0804 and S2-BF-GZ0302 genomes were sequenced using a Pacbio sequelII and DNBSEQ platform at the Beijing Genomics Institute (BGI, Shenzhen, China). Four SMRT cells Zero-Mode Waveguide arrays of sequencing were used by the PacBio platform to generate the subreads set. PacBio subreads of length < 1 kb were removed. The program Canu was then used for self-correction (Parameters: estn = 24, npruseGrid = 0, corOvlMemory = 4). Draft genomic unitigs, which are uncontested groups of fragments, were assembled using the Canu a high quality corrected circular consensus sequence subreads set. GATK (https://www.broadinstitute.org/gatk/) was used to make single-base corrections to improve the accuracy of the genome sequences (Parameters: -cluster 2 -window 5 -stand_call_conf 50 -stand_emit_conf 10.0 -dcov 200MQ0 > = 4).

Table 1 Primer sequences of *B. fragilis* toxin genotypes

Target gene	Primer sequence 5'- 3'
BFT	F: GGATACATCAGCTGGGTTGTAG R: GCGAACTCGGTTTATGCAGT
bft-1	F: TCTTTTGAATTATCCGTATGCTC R: CTTGGGATAATAAAATCTTAGGGATG
bft-2	F: ATTTTTAGCGATTCTATACATGTTCTC R: GGGCATATATTGGGTGCTAGG
bft-3	F: TGGATCATCCGCATGGTTA R: TTTGGGCATATCTTGGCTCA

Genome component prediction

Gene prediction was performed on the S1-BF-GZ0804 and S2-BF-GZ0302 genome assembly by utilizing glimmer3 (http://www.cbcb.umd.edu/software/glimmer/) with Hidden Markov models. tRNA, rRNA and sRNAs recognition made use of tRNAscan-SE (Parameters:–Spec_ tag(BAOG)–o *. tRNA–f *. tRNA. structure) [25], RNAmmer (Parameters:–s Species–m Type–gff *. rRNA. gff–f *.rRNA. fq), and the Rfam (Parameters:–p blastn–W 7–e 1–v 10000–b 10000–m 8–i subfile–o *. blast. m8) database. The tandem repeats annotation was obtained using the Tandem Repeat Finder (Parameters: 2 7 7 80 10 50 2000 -d -h) (http://tandem.bu.edu/trf/trf.html) and the minisatellite DNA and microsatellite DNA were selected based on the number and length of repeat units. Then the Genomic Island Suite of Tools (GIST) was used for genomicis lands analysis(http://www5.esu.edu/cpsc/bioinfo/software/GIST/) employing the IslandPath-DIOMB, SIGI-HMM, and IslandPicker methods. Insertion Sequence (IS) components were analyzed using ISfinder (database update: 2023-11-10) [26]. Finally, prophage regions were predicted using the PHAge Search Tool (PHAST) web server (http://phast.wishartlab.com/) and CRISPR identification using CRISPRFinder.

Gene annotation and protein classification

The best Blast hit was abstracted using the Blast alignment tool for function annotation. Seven databases including KEGG (Kyoto Encyclopedia of Genes and Genomes), COG (Clusters of Orthologous Groups), NR (Non-Redundant Protein Database databases), Swiss-Prot [27], GO (Gene Ontology), TrEMBL, and EggNOG were used for general function annotation. Four databases for the detection of pathogenicity and drug resistance analysis were also utilized. Virulence factors and resistance genes were identified based on the core dataset in VFDB (Virulence Factors of Pathogenic Bacteria), CARD (The Comprehensive Antibiotic Resistance Database) database, as well as CAZy (Carbohydrate-Active enZYmes Database). Type III secretion system effector proteins were detected using EffectiveT3.

Comparative genomics and phylogenetic analysis

The analysis of nucleic acid synteny was conducted utilizing MUMmer and BLAST to examine the core and pan genes of S1-BF-GZ0804, S2-BF-GZ0302, and a reference strain. The genome sequences of the target bacteria were sequenced in accordance with the genome sequences of the reference bacteria, based on the comparative results obtained from MUMmer (parameter: -b 200 -c 65 --extend -l 20). Subsequently, the protein sets of the target and reference bacteria were compared using BLASTp, with the most favorable comparison result selected for each protein in the library to establish protein pairs (besthit). Ultimately, only those protein pairs that were consistent across both comparisons were retained, with the consistency value for each pair calculated as the average of the consistency values from the two comparisons. Protein gene sets of target and reference strains were clustered by the CD-HIT (parameter: -c 0.5 -n 3 -p 1 -g 1 -d 0) rapid clustering of similar proteins software. The threshold of pairwise identity was set at 50% with a 0.7 length difference cutoff in amino acid. The final gene set of the clustering was used as the Pan gene set. The sequences identified in each sample within the cluster, as per the extracted clustering results, constituted the Core gene set. The gene set unique to each sample was referred to as the Specific gene set. Furthermore, the gene set obtained by excluding both the Core and Specific gene sets from the Pan gene set was termed the Dispensable gene set. The gene family was established using target genes and reference strains, employing multiple software tools in the following manner: protein sequence alignment was performed using BLAST, redundancy was removed through the Solar tool, and gene family clustering was conducted on the alignment results utilizing Hcluster_sg software. The phylogenetic tree was constructed by the TreeBeSTutilizing the NJ method (parameter: treebest nj -b 1000).

Results

Morphological, physiological and biochemical identification

Through isolation and purification, we obtained two strains of *B. fragilis* from tumor and paracancerous tissues, namely ZY0302 and ZY0804. Both strains were considered Gram-negative as they stained red after Gram staining. When viewed under an optical microscope (100× oil microscope), individual bacteria were visualized as short rods with blunt round ends that did not form spores (Fig. 1A and D). After being cultured on the CDC anaerobic blood plate for 48 h, the two strains formed milky white colonies with similar shapes and sizes (Fig. 1B and E). On the BBE blood plates, the colonies were slightly convex, smooth, and neat. The ZY0302 colonies were slightly larger and the color was brown, while the ZY0804 colonies were grayish white. They both could decompose bile aesculin, turning the culture medium black, a black halo around the colonies was also present (Fig. 1C and F).

Physiological and biochemical tests indicated that the two strains could use sucrose, glucose, fructose, maltose, lactose, galactose, mannose, D-ribose, honey disaccharide, cellobiose and aesculin as carbon sources. The strains tested positive in the catalase test, and negative in the nitrate (reduction), gelatin test and nitrate (gas production) test (Table 2). In accordance with Berger's Manual of Bacterial Identification (eighth edition) [28], the physiological and biochemical characteristics of strains ZY0302 and ZY0804 were consistent with *B. fragilis* characteristics.

Phylogenetic analysis

The 16Sr RNAs of the isolates were sequenced and their lengths were 1423 and 1433 bases, respectively. Seventeen 16Sr RNA sequences that were closely related to the sequences of the isolates were selected as internal taxa from the results of BLAST homology matching analysis (The Per.Ident of the 16Sr RNA sequences of strains ZY0302 and ZY0804 with these sequences were 93.98–99.51% and 92.97–99.51%, respectively, and both had the highest Per.Ident with the sequence of *B. fragilis* strain W12014D), *Porphyromonas gingivalis* (ATCC 33277) as an outgroup. Maximum likelihood method concordance analysis based on 16Sr RNA gene sequence alignment showed that strains ZY0302 and ZY0804 had the highest homology with *B. fragilis* (Fig. 2A). In summary, strains ZY0302 and ZY0804 were identified to belong to *B. fragilis*. The PCR identification analysis of BFT and its specific typing primers identified strain ZY0302 as bft-1-producing ETBF and strain ZY0804 as NTBF (Fig. 2B).

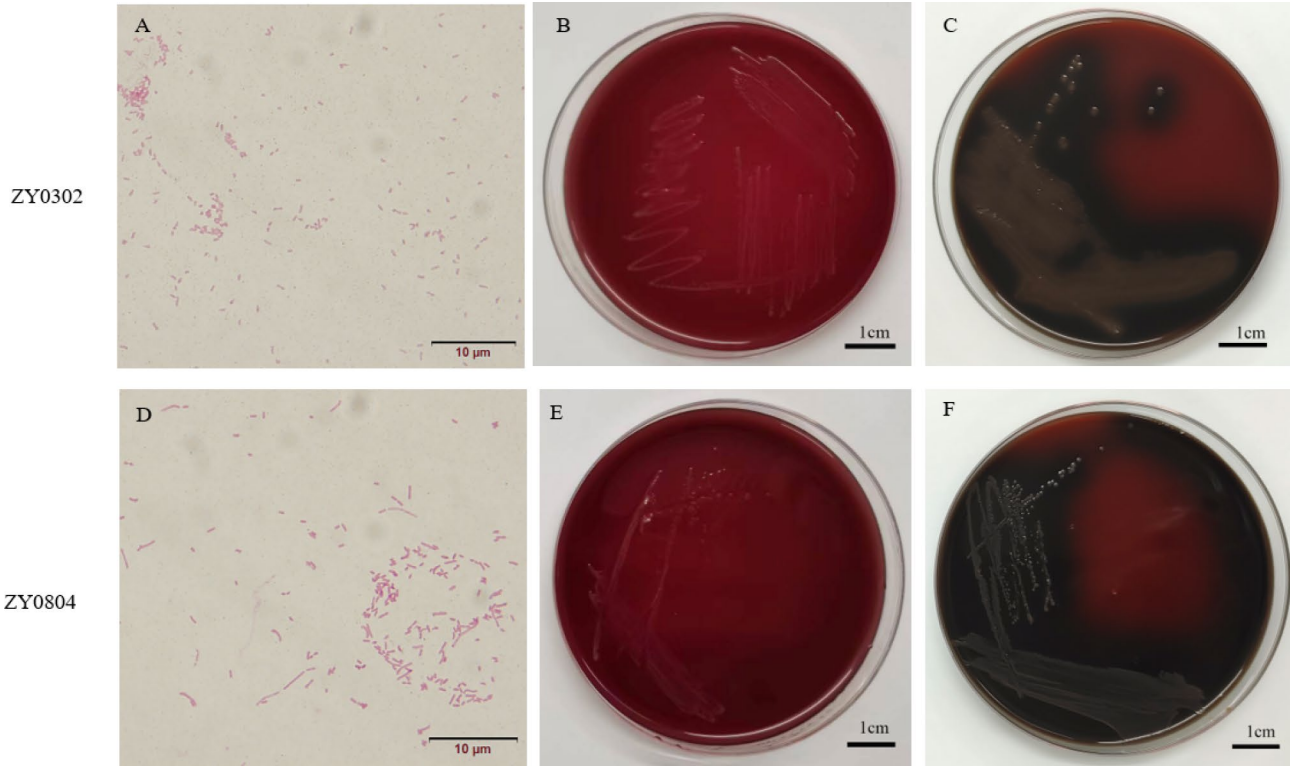

Fig. 1 Morphological characteristics of strains ZY0302 and ZY0804. (**A** and **D**) After Gram staining, observe the morphology of strains ZY0302 and ZY0804 under an optical microscope (100x) The bar represents a unit length of 10 μm. (**B** and **E**) The growth of this strain after 48 h at 37 °C on CDC anaerobic blood agar medium. The bar represents a unit length of 1 cm. (**C** and **F**) The growth of this strain after 48 h at 37 °C on BBE blood agar medium. The bar represents a unit length of 1 cm

Basic genomic features

ZY0302 (S2-BF-GZ-0302) and ZY0804 (S2-BF-GZ-0804) strains belonging to *B. fragilis* were isolated from tumor and paracancerous tissues of CRC patients, respectively. The chromosome of each strain was a single circular DNA with sizes of 5,141,346 bp and 5,174,916 bp, with an average G + C content of 43.43% and 43.19%, respectively (Fig. 3), these features were consistent with those of *B. fragilis* (41–44%) [29]. Both strains contained 18 ribosomal operons (5–23 S-tRNA^{Ala}-tRNA^{Ile}-16 S), and all amino acid specificity was provided by 73 tRNA genes. A total of 4,370 genes were predicted in the genome of strain ZY0302 and had an average gene length of 1,041.40 bp, which covered 88.52% of the whole genome sequence. Moreover, 4,369 genes were predicted in the genome of ZY0804, average gene length of 1,048.02 bp, and covered 88.48% of the whole genome sequence. Assessment of the GC skew results revealed that genomic rearrangements and horizontal gene acquisition may have occurred in the ZY0302 and ZY0804 genomes.

The two types of repetitive elements, DNA repeats and prophages, were present in the genomes of ZY0302 and ZY0804.The DNA repeats included tandem repeats (TR) (Table 3A) and regularly interspaced short palindromic repeats (CRISPR) (Table 3B). 9 and 12 prophages were predicted in both ZY0302 and ZY0804 strains, which accounted for about 4.87% and 6.14% of the total genome, respectively (Table 3C). Insertion sequences (Supplementary Table S1) and transposable elements (8 in ZY0302 and 1 in ZY0804) in the genomes of the two strains were found. However, the presence of plasmids were not detected in the complete genomic sequences of either strain.

DNA restriction and modification (R-M) system

The R-M system is regarded as ubiquitous type of defense mechanism that exists to protect the bacterial cell from invading foreign DNA, especially phage infection [30]. The analysis showed that the ZY0302 genome specified one type I and two type III R-M systems, while the ZY0804 genome specified three type I and one type III R-M systems (Supplementary Table S2). The type I R-M system consisted of three subunits including, HsdR, HsdM, and HsdS, where HsdS determined the DNA-binding specificity of methyltransferases and nucleic acid endonucleases [31]. The ZY0804 genome contained two loci (ZY0804 1855–1858; ZY0804 4192–4195) that were similar to a locus in the NCTC 9343 (BF1839-1842) genome. In the NCTC 9343 genome, this locus showed temporally variable expression of HsdS subunits with

Table 2 Physiological and biochemical properties of *B. fragilis* ZY0302 and ZY0804

Project	ZY0804	ZY0302
Cellobiose	W^+	W^+
Fructose	W^+	+
Melezitose	-	-
Rhamnose	-	-
Arabinose	-	-
Mannitol	-	-
D-ribose	W^+	W^+
Mannose	+	+
Sorbose	-	-
Melibiose	W^+	W^+
Trehalose	-	-
Sucrose	W^+	+
Salicin	-	-
Milk-mannitol	-	-
Glucose	+	+
Inositol	-	-
Purple milk	-	-
Lactose	+	W^+
Maltose	+	W^+
Galactose	W^+	W^+
Aesculin	+	+
3%H_2O_2	+	+
Gelatin	-	-
Nitrate (gas production)	-	-
Nitrate (reduction)	-	-

Key: + means the biochemical identification is positive, W^+ means weakly positive, - means negative

eight different possible polypeptide combinations. This in turn provided eight different DNA recognition specificities (Shufflon BB) [31]. In addition, the ZY0302 genome encodes the 5-methylcytosine-specific restriction endonuclease McrBC regulatory subunit McrC. McrBC consists of two polypeptides McrB and McrC. McrB is responsible for GTP binding and hydrolysis as well as DNA binding, while the other McrC is responsible for DNA cleavage. McrBC cleaves specific DNA sequences that have been modified with 5-methylcytosine, hence affording protection to the bacterial cell from foreign DNA [32].

Despite the presence of the R-M system in bacteria, phages have evolved a number of mechanisms to circumvent the limitations of the R-M system, including encoding restriction-resistant proteins and limiting the number of recognition sites in their genomes [33]. From our findings the presence of the gene encoding the restriction endonuclease-resistant protein *ArdA* in the phage gene of ZY0302 (ZY0302 1853) and the presence of the gene encoding the restriction endonuclease-resistant protein *ArdC* in the phage gene of ZY0804 (ZY0804 2152) may help elucidate the reason why these prophages can successfully be transferred into the bacterial genome.

Comparative genomics

The predicted proteome levels of ZY0302 and ZY0804 were compared with those of *B. fragilis* obtained through the Genebank and a phylogenetic evolutionary tree was

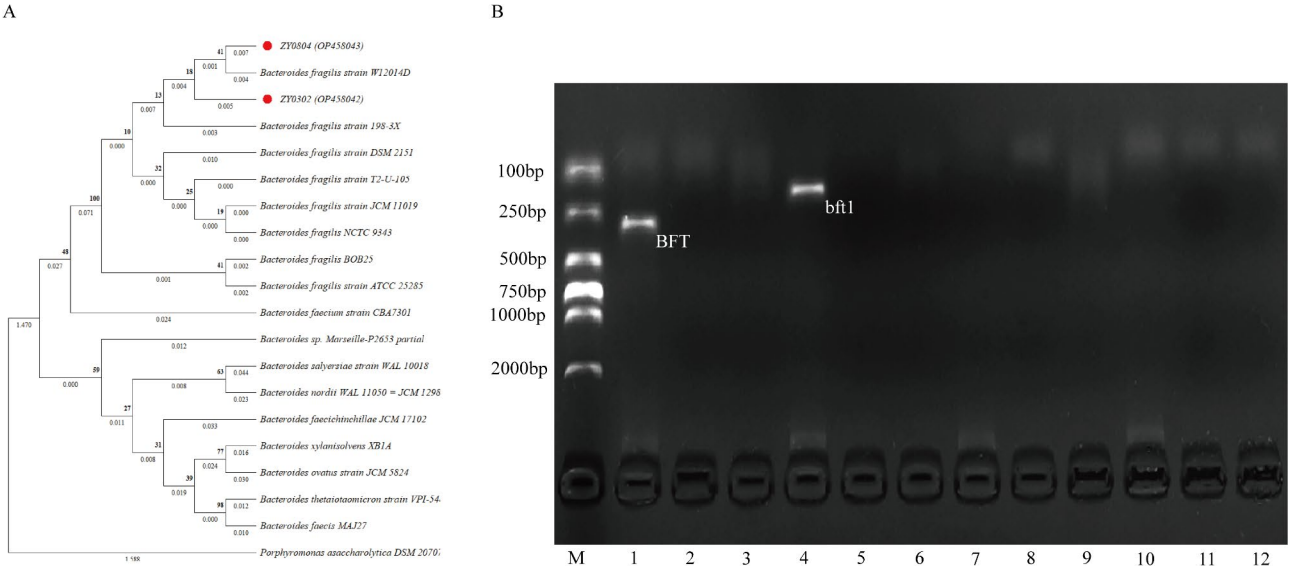

Fig. 2 Phylogenetic tree and typing identification of strains ZY0302 and ZY0804. (**A**) The evolutionary history was inferred by using the Maximum Likelihood method and Kimura 2-parameter model. Bootstrap values were based on 1,000 resamplings. Strain Porphyromonas gingivalis ATCC 33,277 was used as an outgrouping. Evolutionary analyses were conducted in MEGA11. (**B**) Gel electrophoresis images of PCR products amplified based on BFT and its typing primers. M: 2000 DL Marker; 1–3: BFT primer. 4–6: bft-1 primer. 7–9: bft-2 primer. 10–12: bft-3 primer. The products of 1,4,7,10 lanes were produced using DNA from strain ZY0302 as a template, 2,5,8,11 using DNA from strain ZY0804 as a template, and 3,6,9,12 using DEPC water as a template. Imaging was performed using a BIO-RAD gel imaging system

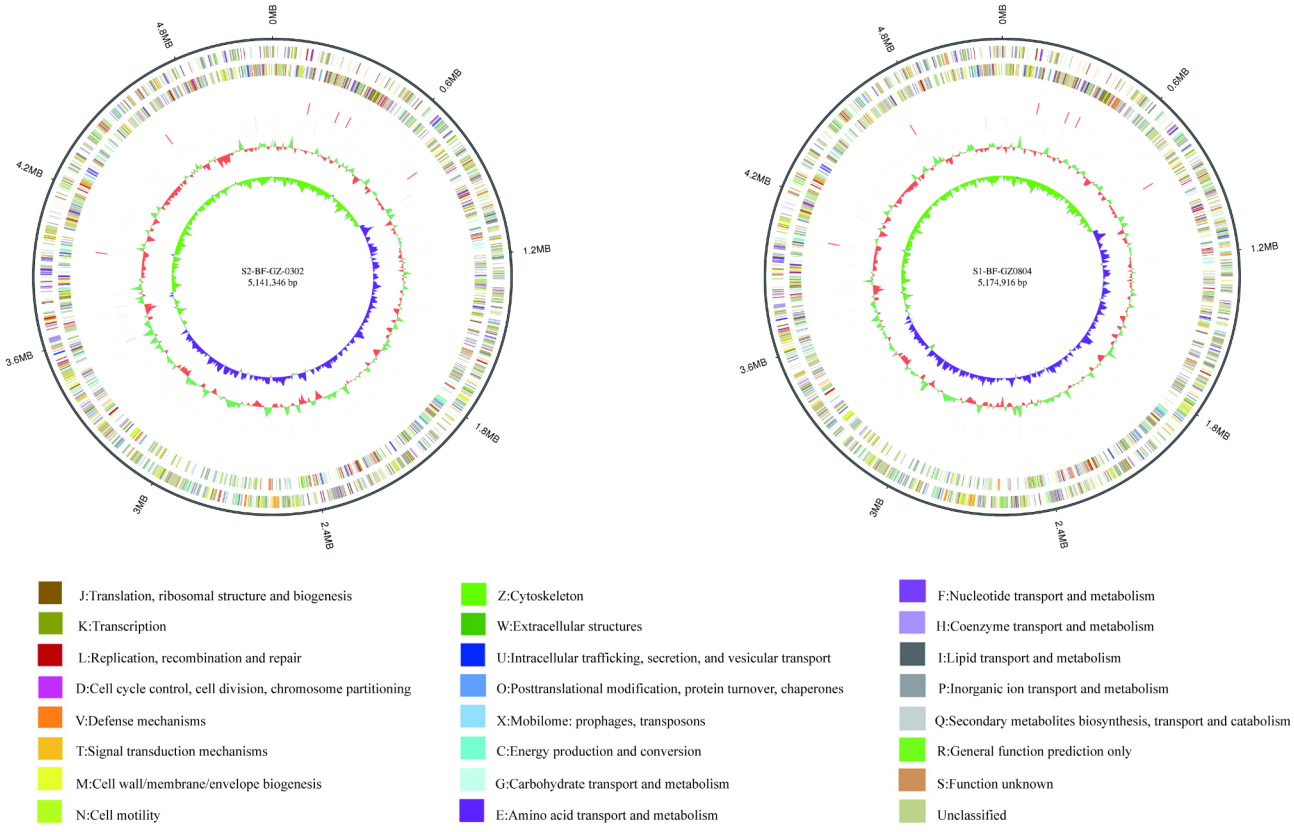

Fig. 3 Circular maps of the chromosome of *B. fragilis* strain ZY0302 and ZY0804. Each circle represents (From outer to inner): 1. Genome Size; 2–3. Forward strand gene and reverse strand gene, colored according to cluster of orthologous groups (COG) classification; 4–5. Forward and Reverse Strand ncRNA, including rRNA (Red), sRNA (Green), tRNA (Yellow); 6. Repeat; 7. G + C content; 8. GC skew (G-C/G + C)

then constructed. ZY0302 and ZY0804 shared 2,775 core genes with the reference strain and possessed 320 and 285 unique genes, respectively (Fig. 4A). Phylogenetic analysis of the core genes revealed that ZY0302 and ZY0804 were in the same evolutionary branch (Fig. 4B). However, the dispensable genes showed similar gene heat in different branches, which was unlike that of the reference strain (Fig. 4C).

When compared to the reference strain, gene families showed that in the ZY0302 predicted genome, 4,122 (94.32%) genes were clustered into 2,623 gene families, of which 17 unique paralogous genes belonged to 8 strain unique gene families. The 2,615 immediate homologous gene families included 1,962 multicopy immediate homologous genes. In the ZY0804 genome, 4,152 (95.03%) genes were clustered into 2,694 gene families, with 18 unique paralogous homologs in 9 strain unique gene families and 1,948 multicopy orthologs in 2,685 orthologous gene families. The largest lineal homologous family was composed of them while the reference strain consisted of 471 homologous genes, including genes such as *BF9343_RS01055* and *BF_RS10050*, which are *RagB/SusD* family vegetative uptake outer membrane protein genes. Nucleic acid collinearity results showed that the genomes between strains ZY0302 and ZY0804 had similar structural variability during the evolutionary process, with different degrees of genomic rearrangements, inversions, insertions, and deletions with different reference strains (Supplementary Fig. S1).

Carbohydrate metabolism

B. fragilis can utilize host-derived cell surface glycoproteins and glycolipids as an energy source, and these may include monosaccharides (e.g., glucose, galactose, mannose, etc.) and more complex compounds (for example, N-acetyl-d-glucosamine and N-acetylneuraminic acid.) [6]. Annotation using the CAZy database revealed the presence of various carbohydrate-associated enzymes in the genomes of ZY0302 and ZY0804, including Glycoside Hydrolases(GHs), GlycosylTransferases (GTs), Polysaccharide Lyases (PLs), Carbohydrate Esterases(CEs), Auxiliary Activities(AAs), and Carbohydrate-Binding Modules(CBMs) (Table 4).

Virulence factors

By comparing the ZY0302 and ZY0804 genomes with the Virulence Factor Database (VFDB), 123 and 121 virulence genes were predicted, respectively (Supplementary

Table 3 Repeat sequence and prophage statistics

A Sample Name	Type	Number	Repeat Size(bp)	Total Length (bp)	In Genome (%)
ZY0804	TRF	161	2-102	8,311	0.1606
	Minisatellite DNA	122	15–63	6,460	0.1248
	Microsatellite DNA	15	2–10	559	0.0108
ZY0302	TRF	160	2-327	11,125	0.2164
	Minisatellite DNA	112	15–63	6,546	0.1273
	Microsatellite DNA	7	2–10	285	0.0055

B Sample Name	CRISPR Start	CRISPR End	CRISPR Length (bp)
ZY0804	3,830,423	3,831,030	607
	4,080,330	4,080,424	94
ZY0302	3,510,355	3,512,565	2210
	4,011,706	4,011,800	94

C Sample Name	Phage id	Phage Start	Phage End	attL Start	attL End	AttR Start	attR End
ZY0804	pp1	103,007	116,343	101,436	101,448	114,520	114,532
	Pp2	266,809	292,542	265,539	265,551	292,493	292,505
	Pp3	398,224	443,144	397,567	397,582	440,854	440,869
	Pp4	2,500,368	2,557,616	2,503,491	2,503,503	2,555,926	2,555,938
	Pp5	2,611,537	2,637,461	2,613,907	2,613,933	2,636,547	2,636,572
	Pp6	2,849,395	2,857,581	2,853,383	2,853,510	2,863,233	2,863,360
	Pp7	3,038,638	3,049,143	3,042,626	3,042,753	3,052,476	3,052,603
	Pp8	3,549,249	3,565,329	3,553,237	3,553,364	3,563,087	3,563,214
	Pp9	3,832,133	3,862,386	3,831,476	3,831,491	3,874,763	3,874,778
	Pp10	3,954,193	3,976,415	3,956,563	3,956,589	3,979,203	3,979,228
	Pp11	4,091,813	4,118,593	4,094,183	4,094,209	4,116,823	4,116,848
	Pp12	4,442,053	4,478,683	4,441,396	4,441,411	4,484,683	4,484,698
ZY0302	Pp1	376,457	429,664	376,774	376,787	429,946	429,959
	Pp2	2,264,276	2,295,477	2,265,599	2,265,613	2,301,807	2,301,821
	Pp3	2,368,935	2,411,515	2,369,252	2,369,265	2,422,424	2,422,437
	Pp4	2,477,232	2,498,103	2,480,935	2,480,954	2,502,396	2,502,415
	Pp5	3,105,863	3,130,983	3,109,306	3,109,330	3,144,154	3,144,178
	Pp6	3,652,381	3,659,000	3,656,160	3,656,304	3,657,883	3,658,027
	Pp7	4,023,720	4,053,207	4,027,163	4,027,187	4,062,011	4,062,035
	Pp8	4,785,891	4,802,425	4,789,334	4,789,358	4,824,182	4,824,206
	Pp9	4,846,935	4,871,861	4,850,378	4,850,402	4,885,226	4,885,250

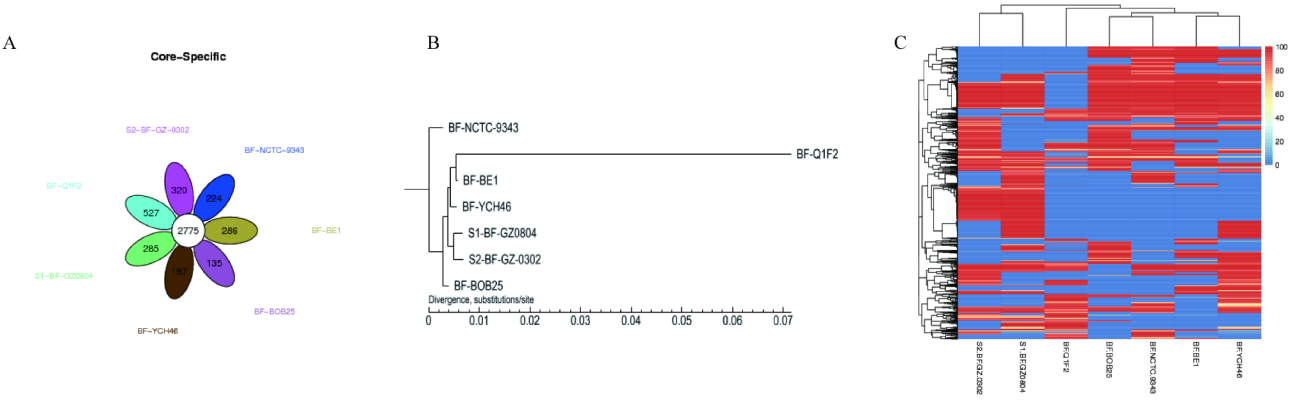

Fig. 4 Common and unique genes of ZY0302 and ZY0804 with reference strains. (**A**) Venn diagram of the Pan gene set. (**B**) Phylogenetic trees were constructed based on core genes of the sample and reference strains. The NJ algorithm of TreeBeST was used to construct the phylogenetic tree. (**C**) Dispensable gene heatmap

Table 4 Taxonomic annotation of carbohydrate enzymes of ZY0302 and ZY0804

Sample Name	AAs	CBMs	CEs	GHs	GTs	PLs
ZY0302	1	35	10	185	129	2
ZY0804	1	34	9	185	121	2

Table S3). Consistent with the PCR results, the 16 virulence factors specific to ZY0302 included zinc-dependent metalloproteinase toxin (bft1), the major virulence factor of ETBF. Among the 11 specific virulence factors of ZY0804, α-1, 3-fucosyltransferase was identified, this enzyme is a key to the last step of the formation of LPS Lewis antigen. Fucose-containing LPS are commonly expressed by pathogenic bacteria such as Helicobacter pylori and *E. coli*, and they are thought to be involved in molecular mimicry, adhesion, colonization, and modulation of the host immune response [34].

Although *B. fragilis* is exclusively anaerobic, it is also highly tolerant to oxygen and can survive exposure to an aerobic environment for 48–72 h [35]. The high tolerance of *B. fragilis* to oxidative stress is thought to be an important factor in its escape from the anaerobic environment of the gastrointestinal tract to aerobic tissues such as the peritoneal cavity as well as its tolerance to the respiratory burst induced by immune cells [35]. We found the presence of catalase, superoxide dismutase, aspartate 1-decarboxylase and peptide methionine sulfoxide reductase msrA/ msb in the genomes of ZY0302 and ZY0804, as well as alkyl hydroperoxide reductase AHPC and the heat shock protein CLPB. Among them, CLPB is a member of the AAA family (ATPases associated with a wide range of cellular activities), which has been shown to be important in a variety of pathogenic bacteria and is an important virulence factor, playing a key role in the cellular response to stressful conditions such as heat stress, antibiotic resistance, and oxidative stress [36]. In addition, CLPB proteins regulate the secretion of bacterial effector molecules associated with the type VI secretion system (T6SS), which is critical for the survival and infectivity of clinically relevant microorganisms [36]. However, there are no studies that directly refer to the specific functions or roles of CLPB proteins in *B. fragilis*.

Among the ZY0302 and ZY0804 virulence genes, we identified genes related to type IV bacteriophage biosynthesis, the hmu direct heme uptake system (two in ZY0302 and one in ZY0804) and PDTC (pyridine-2,6-dithiocarboxylic acid). PDTC is thought to act as an iron carrier in *Pseudomonas aeruginosa* and is associated with the transport of iron and other metals [37]. Iron is removed from heme in the periplasmic space and the released iron needs to be transported to the cytoplasm of *B. fragilis* via the FeoAB system and the CobN-like proteins BtuS1 and BtuS2 [38]. It is therefore believed that *B. fragilis* employs similar systems in its virulence mechanism. Through COG database gene function annotation, we found the presence of FEOB (ZY0302 2050; ZY0804 1995) and COBN proteins (ZY0302 3039, 3304; ZY0804 3338) in the genomes of ZY0302 and ZY0804. In addition, five VirB4 subunits of the type IV secretion system (1867, 1969, 3147, 4060, 4109), six VirD4 subunits (1054, 1876, 1979, 3134, 4023, 4071), and three TrbF fractions were identified in the genome of ZY0302 by functional annotation of genes in the COG database (1864, 1966, 4056), and the genome of ZY0804 identified one VirB4 subunit (2067) and one VirD4 subunit (2085) of the type IV secretion system. Further analysis showed that most of these subunits were not present on the bacterial genome chromosome but originated from prophages, except for one VirB4 subunit (3147) and three VirD4 subunits (1054, 3134, 4023) which were identified on the bacterial chromosome of the ZY0302 genome.

B. fragilis has been shown to induce abscess formation and its potential to induce abscess formation is significantly related to its capsule [39]. Virulence gene analysis showed that there were 74 (74/121) virulence genes were involved in capsule formation in ZY0804 and 76 (76/123) in ZY0302. At least eight and nine capsular biosynthesis gene clusters in their genomes were found in the ZY0804 (107 to 112meme 1442 to 1447 meme 1456 to 1465, 2038 to 2046, 2253 to 22268, 3228 to 3232, 3516 to 3518, 4170 to 4172) and ZY0302 (95 to 99991436 to 1447, 1701 to 1701 to 1776 to 2087, 2183 to 2185, 2499 to 2502, 3192 to 3194, 3488 to 3488 to 4215), respectively. ZY0804 had two capsular biosynthetic gene clusters (2038 to 2046 and 3228 to 3232) derived from prophages and ZY0302 had one (2076 to 2087).

B. fragilis can produce a variety of extracellular and physiological enzymes to attack and penetrate the host's extracellular matrix. Extracellular enzyme-related genes such as neuraminidase, lipase, hemolysin (Table 5), protease, and phosphatase (not listed) were predicted in the genomes of ZY0302 and ZY0804 using COG database gene function annotation. The products synthesized by these extracellular enzyme genes play important roles in the adhesion, invasion and growth of *B. fragilis*.

Discussion

In this study, we isolated ZY0302 and ZY0804 from tumor tissues and paracancerous tissues of CRC patients. They were confirmed to be *B. fragilis* by morphological analysis, physiological and biochemical characteristics, and 16 S rRNA sequencing according to Berger's Manual of Bacterial Identification. Furthermore, we detected the expression of BFT gene in the two bacterial strains

Table 5 Extracellular enzymes of ZY0302 and ZY0804

Locus of virulence genes		Extracellular enzymes and related genes
ZY0302	ZY0804	
		Neuraminidase
121, 1240	126, 1257	Neuraminidase (sialidase) NanH, contains C-terminal autotransporter domain
2391	2448	Predicted neuraminidase (sialidase)
		lipase
172, 173, 318, 432, 1029, 2393, 2394, 2805	179, 180, 324, 453, 1061, 2450, 2451, 2865	Lysophospholipase L1 or related esterase. Includes spore coat protein LipC/YcsK
704, 2677, 2678, 2778	738, 2737, 2738, 2837	Acetyl esterase/lipase
2771, 2773	2830, 2832	Predicted ABC-type transport system involved in lysophospholipase L1 biosynthesis, permease component
2858, 2974, 3549, 4351	2970, 3073, 3565, 4350	Predicted acylesterase/phospholipase RssA, containd patatin domain
3510	3530	Lysophospholipase, alpha-beta hydrolase superfamily
3800	3865	Predicted phospholipase, patatin/cPLA2 family
		Hemolysin
961, 962	994, 995	Putative hemolysin
1684, 2031, 2207	1704, 1977, 2278	Hemolysin-related protein, contains CBS domains, UPF0053 family
4231	4231	Predicted membrane channel-forming protein YqfA, hemolysin III family

by applying PCR. The final strains, ZY0302, were identified as a bft-1-producing ETBF, and ZY0804 as an NTBF. Whole genome sequencing (WGS) is able to identify taxa at a higher taxonomic resolution relative to 16 S rRNA and provides more information about gene function. We characterized the genomes of the isolates and identified and analyzed their putative virulence factors.

Pan-genomic analysis revealed that the isolates shared multiple immediate homologous gene families with the *B. fragilis* reference strain. RagB/SusD family nutrition uptake outer membrane protein gene was observed to be their largest gene family. SusD-like proteins have α-helix folding, bind polysaccharides on the cell surface, and may promote polysaccharides to shuttle to related TonB-dependent porins for transport to periplasm for degradation [40]. Also, the SusC family of outer membrane proteins is the largest paralogous family of *B. fragilis* and *Bacteroides pleomorphic* [19]. The multiple copies of these polysaccharide utilization genes, as well as the presence of large numbers of GHs, GTs, and CBMs in the genome, may have been designed to assist in the effective utilization of nutrients in the distal gut. The distal gut is deficient in monosaccharides and disaccharides, and these amenities could help in the efficient utilization of polysaccharides and a variety of dietary carbohydrates. Phylogenetic analysis of core and dispensable genome sequences showed that *B. fragilis* isolated from cancerous and paracancerous tissues belonged to the same evolutionary branch but was in a different evolutionary branch from *B. fragilis* from other tissue sources (BF-BE1 was isolated from a patient with a wound infection, BF-BOB25 was isolated from a fecal sample from a patient with intestinal ecological dysbiosis, BF-NCTC – 9343 was isolated from patients with abdominal infections, BF-Q1F2 was isolated from patients with skin and soft tissue infections, and BF-YCH46 was isolated from patients with sepsis). From these results, we speculated that ZY0302 and ZY0804 co-evolved the core and non-essential genomes as a means to adapt to the changing microenvironment of the tumor and paracancerous tissues, thus enabling them to colonize the ever-changing microenvironment long-term. In addition, ZY0302 and ZY0804 have undergone structural variation in gene rearrangements (inversions, insertions, and deletions) during evolution. This crucial evolutionary driver regulates gene expression and advances ecological niche adaptation [41]. Through a process known as horizontal gene transfer (HGT), some virulence factors and antibiotic resistance genes that were not originally present in the bacterial genome may be acquired. Consequently, horizontally transferred genes in a recipient bacterium may incite a new adaptation thereby affording the bacterium new ecological niches that were inaccessible through mutation alone [42]. The presence of phages, a class of mobile genetic elements that can be transmitted from cell to cell, may also allow some bacteria to acquire antibiotic resistance, enhance adaptation to the environment, improve adhesion and enable bacteria to become pathogenic [43]. Insertion Sequences (ISs), transposases, and mobile elements from phages were identified within the genomes of ZY0302 and ZY0804. The phages contained several genes associated with capsular biosynthesis and the type IV secretion system, while they lacked other virulence genes and antibiotic resistance genes. The presence of these phages enables better adaptation of the strain to new ecological niches and colonization.

B. fragilis makes up the gut microbiota and remains commensal with the host gut, however under unfavorable conditions, this bacteria can turn into an opportunistic pathogen leading to abscess formation and bacteremia

in several parts of the body (abdomen, brain, liver, pelvis, and lungs) [5]. Several virulence factors contribute to its conversion from a symbiont into a pathogenic bacterium, including complex capsular polysaccharides, BFT, LPS, and virulence-related enzymes [6]. Furthermore, recent studies have shown that the *B. fragilis* correlation with the development of CRC [18, 44]. The development of CRC is accompanied by the generation of an inflammatory microenvironment and excessive oxidative stress [45]. ZY0302 and ZY0804 encode multiple genes involved in the regulation of oxidative stress, which is favorable for their survival in the excessively oxidatively stressed tumor microenvironment.

B. fragilis effectively evades the host's immune response due to its unique cellular structure. It possesses the ability to modulate the surface composition of its capsular polysaccharides, which contributes to its capsule's role in increasing the bacteria's resistance to complement-mediated lysis as well as to phagocytosis and subsequent destruction [19, 20, 46, 47]. In addition, its capsule and LPS also act as adhesins, allowing the bacteria to attach and colonize the infection site [6]. ZY0302 and ZY0804 from CRC and paracancerous tissues are rich in capsular biosynthesis genes which is consistent with the characteristics of *B. fragilis*. Among the predicted virulence factors, 61.79% of them for ZY0302 and 61.16% for ZY0804 are related to capsular biosynthesis, and have multiple capsular biosynthesis gene clusters. Certain gene clusters originate from phages via horizontal gene transfer, encoding proteins associated with cell structure derived from phage sequences. This process can further augment the adaptability of bacteria to their environment and enhance their adhesion capabilities. We also identified some genes in the genomes of ZY0302 and ZY0804 that are related to the synthesis of type IV bacterial hairs. *B. fragilis* can use bacterial hairs to adhere to other bacterial species and host tissues, which is important for bacterial mucosal colonization [48]. In addition, the genomes of ZY0302 and ZY0804 contain the *kdsA* gene encoding 2-dehydro3-deoxyphosphate octanoate aldolase. *KdsA* genes have been implicated in LPS synthesis with adhesion and endotoxin properties [49]. *B. fragilis* strains isolated from patients with polyps are enriched in the LPS gene and have a greater ability to activate TLR4 and induce a proinflammatory cytokine response [18]. ZY0302 and ZY0804 encode genes for glycosyltransferases involved in LPS biosynthesis, and 18 and 16 of their virulence genes, respectively, are associated with LPS synthesis (Supplementary Table S3).

Iron and heme are essential nutrients for the growth and survival of *Bacteroides*, and key determinants of bacterial virulence as well. The hmu loci identified in the ZY0302 and ZY0804 genomes play a major role in heme acquisition in *Porphyromonas gingivalis*, and a similar locus was found in *B. fragilis* NCTC9343 [50]. Alluding to its vital potential role in heme acquisition of *B. fragilis*. Bacterial type IV secretion systems (T4SSs) are a functionally diverse translocation superfamily that not only mediate the horizontal transfer of DNA, but also transport macromolecules such as proteins and ribonucleoprotein complexes, which allows the bacteria to encode resistance to heavy metals and antibiotics [51]. However, previous studies have shown that there are no traces of III type, IV type, autotransporter or two-partner secretion systems in the genome of *B. fragilis* [52]. In this study, we discovered that the loci linked to the type IV secretion system were identified within the genomes of ZY0302 and ZY0804; however, they were predominantly located within the phage genes, suggesting that these genes were likely obtained through horizontal gene transfer.

B. fragilis can produce a variety of extracellular and physiological enzymes, with neuraminidase being the most prevalent. This enzyme cleaves mucin polysaccharides, thereby promoting microbial growth, and serves as a significant virulence factor that impairs the immune function of the host [6]. Lipase is a secreted virulence factor, and the putative roles of microbial secretion of lipase during infection include host cell adhesion, lipid digestion for nutrients and triggering of the inflammatory cascade [53]. The production of hemolysins may facilitate the in vivo access of *B. fragilis* to iron and heme through the destruction of host cells and erythrocytes. Two hemolysins, namely HlyA and HlyB, have been identified in *B. fragilis*, exhibiting synergistic functions in the process of erythrocyte hemolysis [54]. The proteases attack a range of host proteins, including extracellular matrix proteins and cell adhesion molecules, and their disruption leads to the loss of cell surface receptors and tissue integrity [55, 56]. It is evident that these enzymes not only play a key role in their pathogenicity, but are also involved in bacterial nutrient metabolism and immune escape to enhance their own survival.

The genomic diversity and genetic differences between strains of *B. fragilis* may be related to its adaptations in different hosts and environments. Considering the dual functions of *B. fragilis* in both intestinal health and disease, it is imperative that future research delves deeper into the ecological implications of these genomic variations and their impact on host health and disease. Utilizing high-throughput sequencing technology, researchers should perform extensive genomic comparisons of *B. fragilis* strains derived from various sources to uncover their evolutionary connections and functional distinctions. This approach will lay the groundwork for the formulation of targeted intervention strategies. However, the limited sample size of *B. fragilis* examined in this study may have led to findings that do not provide robust evidence regarding their epidemiological significance. The primary

factor contributing to this limitation is that, in the present study, we isolated and identified *B. fragilis* from a single tissue section. This approach effectively addresses issues of contamination and pollution, thereby ensuring the specificity of the strain source. However, it also heightens the risk of leakage, which may be a significant reason for our limited positive sample size. Future studies could enhance this sample size by either homogenizing the tissue or developing a more systematic protocol to improve the isolation efficiency of *B. fragilis*. In addition, although we observed the correlation between genomic alterations in *B. fragilis* and adaptation to the tumor microenvironment, we did not directly confirm that these alterations promote *B. fragilis* adaptation to the tumor microenvironment through corresponding experiments. In future studies, we will also further validate the role and significance of these genomic alterations in the adaptation of *B. fragilis* to the tumor microenvironment in combination with modern genetic engineering techniques such as CRISPR-Cas9.

Conclusions

The genomic genetic characteristics of *B. fragilis* strains ZY0302 and ZY0804 isolated from tumor and paracancerous tissues of CRC patients are different from those isolated from non-tumor tissues, the selection pressure from the tumor microenvironment may be the important driver of pan-genomic variability in *B. fragilis*. Changes in the genomic characteristics of ZY0302 and ZY0804 may facilitate bacterial adaptation to the changing tumor microenvironment, where gene rearrangement and horizontal gene transfer offer the possibility of better adaptation to changes in the microenvironment. Of course, These alterations may also elevate the infectious potential of the bacteria; however, further research is necessary to elucidate their roles in the progression of CRC.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11421-3.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
Not applicable.

Author contributions
HY and YG: Writing–review & editing, Writing–original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization; SJ and XZ: Methodology, Investigation, Validation; YX and DG: Investigation, Data curation; XX, YG and YZ: Software; QL: Data curation; MW: Resources, Writing–review & editing, Supervision; JL: Writing–review & editing, Supervision, Project administration, Funding acquisition. All authors have read and approved the final manuscript.

Funding
Natural Science Research Project of Guizhou Education Department in 2024 (Guizhou Education Technology [2024] No. 335).
Zunyi city Science and Technology Program project (Zunyi City, Kehe HZ character (2024) No. 303).
Zunyi Medical University 2021 Special Project for Academic New Seedling Cultivation and Innovative Exploration (Guizhou Science and Technology Platform Talents [2021]1350-038).
Science and Technology Fund Project of Guizhou Health Care Commission (No. gzwjkj2019-1-123).
Governor's Special Fund for Outstanding Scientific and Technological Education Talents in Guizhou Province (No. [2011]57).
Scientific Research Program of Guizhou Provincial Department of Education (QJJ [2023] 019).

Data availability
Data will be made available on request. The 16 S rRNA sequences of ZY0302 and ZY0804 has been assigned the GenBank accession no. OP458042 and OP458043. Whole generation sequencing data of ZY0302 and ZY0804 has been assigned the GenBank accession no. CP139162 and CP139161.

Declarations

Ethics approval and consent to participate
The study was conducted in accordance with the principles of the Declaration of Helsinki and approved by the Medical Ethics Review Committee (ZunYi ethical examination No. [2023] 1–014). Written Informed Consent was obtained from all subjects involved in the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Institute of Zoonosis, College of Public Health, Zunyi Medical University, Zunyi, Guizhou, China
[2]Key Laboratory of Maternal & Child Health and Exposure Science of Guizhou Higher Education Institutes, Zunyi, Guizhou, China
[3]Institute of Gastroenterology, Affiliate Hospital of Zunyi Medical University, Zunyi, Guizhou, China
[4]Southwest Guizhou Vocational and Technical College, Xingyi, Guizhou, China
[5]No. 149, Dalian Road,, Zunyi City 563003, Guizhou Province, China
[6]No. 6, Xuefu West Road, Xinpu New District, Zunyi City 563000, Guizhou Province, China

Received: 11 November 2024 / Accepted: 28 February 2025
Published online: 18 March 2025

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49. https://doi.org/10.3322/caac.21660.
2. Song M, Chan AT, Environmental, Factors. Gut microbiota, and colorectal Cancer prevention. Clin Gastroenterol Hepatol. 2019;17(2):275–89. https://doi.org/10.1016/j.cgh.2018.07.012.
3. Leung PHM, Subramanya R, Mou Q, Lee KT, Islam F, Gopalan V, et al. Characterization of mucosa-Associated microbiota in matched Cancer and Non-neoplastic mucosa from patients with colorectal Cancer. Front Microbiol. 2019;101317. https://doi.org/10.3389/fmicb.2019.01317.
4. Liu W, Zhang X, Xu H, Li S, Lau HC, Chen Q, et al. Microbial community heterogeneity within colorectal neoplasia and its correlation with colorectal carcinogenesis. Gastroenterology. 2021;160(7):2395–408. https://doi.org/10.1053/j.gastro.2021.02.020.

5. Sun F, Zhang Q, Zhao J, Zhang H, Zhai Q, Chen W. A potential species of next-generation probiotics? The dark and light sides of *Bacteroides fragilis* in health. Food Res Int. 2019;126108590. https://doi.org/10.1016/j.foodres.2019.108590.
6. Wexler HM. Bacteroides: the good, the bad, and the nitty-gritty. Clin Microbiol Rev. 2007;20(4):593–621. https://doi.org/10.1128/CMR.00008-07.
7. Dougherty MW, Jobin C. Intestinal bacteria and colorectal cancer: etiology and treatment. Gut Microbes. 2023;15(1):2185028. https://doi.org/10.1080/19490976.2023.2185028.
8. Chung L, Orberg ET, Geis AL, Chan JL, Fu K, DeStefano Shields CE, et al. *Bacteroides fragilis* toxin coordinates a Pro-carcinogenic inflammatory cascade via targeting of colonic epithelial cells. Cell Host Microbe. 2018;23(3):421. https://doi.org/10.1016/j.chom.2018.02.004.
9. Wu S, Morin PJ, Maouyo D, Sears CL. *Bacteroides fragilis* enterotoxin induces *c-Myc* expression and cellular proliferation. Gastroenterology. 2003;124(2):392–400. https://doi.org/10.1053/gast.2003.50047.
10. Wu S, Rhee KJ, Albesiano E, Rabizadeh S, Wu X, Yen HR, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. Nat Med. 2009;15(9):1016–22. https://doi.org/10.1038/nm.2015.
11. Liu QQ, Li CM, Fu LN, Wang HL, Tan J, Wang YQ, et al. Enterotoxigenic *Bacteroides fragilis* induces the stemness in colorectal cancer via upregulating histone demethylase JMJD2B. Gut Microbes. 2020;12(1):1788900. https://doi.org/10.1080/19490976.2020.1788900.
12. Takashima Y, Kawamura H, Okadome K, Ugai S, Haruki K, Arima K, et al. Enrichment of *Bacteroides fragilis* and enterotoxigenic *Bacteroides fragilis* in CpG Island methylator phenotype-high colorectal carcinoma. Clin Microbiol Infect. 2024;30(5):630–6. https://doi.org/10.1016/j.cmi.2024.01.013.
13. Park PH, Keith K, Calendo G, Jelinek J, Madzo J, Gharaibeh RZ, et al. Association between gut microbiota and CpG Island methylator phenotype in colorectal cancer. Gut Microbes. 2024;16(1):2363012. https://doi.org/10.1080/19490976.2024.2363012.
14. Shao X, Sun S, Zhou Y, Wang H, Yu Y, Hu T, et al. *Bacteroides fragilis* restricts colitis-associated cancer via negative regulation of the NLRP3 axis. Cancer Lett. 2021;523170–81. https://doi.org/10.1016/j.canlet.2021.10.002.
15. Deng H, Yang S, Zhang Y, Qian K, Zhang Z, Liu Y, et al. *Bacteroides fragilis* prevents Clostridium difficile infection in a mouse model by restoring gut barrier and Microbiome regulation. Front Microbiol. 2018;92976. https://doi.org/10.3389/fmicb.2018.02976.
16. Kazmierczak-Siedlecka K, Skonieczna-Zydecka K, Hupp T, Duchnowska R, Marek-Trzonkowska N, Polom K. Next-generation probiotics - do they open new therapeutic strategies for cancer patients? Gut Microbes. 2022;14(1):2035659. https://doi.org/10.1080/19490976.2022.2035659.
17. Huang Y, Cao J, Zhu M, Wang Z, Jin Z, Xiong Z. Nontoxigenic *Bacteroides fragilis*: A double-edged sword. Microbiol Res. 2024;286127796. https://doi.org/10.1016/j.micres.2024.127796.
18. Kordahi MC, Stanaway IB, Avril M, Chac D, Blanc MP, Ross B, et al. Genomic and functional characterization of a mucosal symbiont involved in early-stage colorectal cancer. Cell Host Microbe. 2021;29(10):1589–. https://doi.org/10.1016/j.chom.2021.08.013.
19. Kuwahara T, Yamashita A, Hirakawa H, Nakayama H, Toh H, Okada N, et al. Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. Proc Natl Acad Sci U S A. 2004;101(41):14919–24. https://doi.org/10.1073/pnas.0404172101.
20. Cerdeno-Tarraga AM, Patrick S, Crossman LC, Blakely G, Abratt V, Lennard N, et al. Extensive DNA inversions in the B. fragilis genome control variable gene expression. Science. 2005;307(5714):1463–5. https://doi.org/10.1126/science.1107008.
21. Sears CL, Garrett WS. Microbes, microbiota, and Colon cancer. Cell Host Microbe. 2014;15(3):317–28. https://doi.org/10.1016/j.chom.2014.02.007.
22. Jean S, Wallace MJ, Dantas G, Burnham CAD. Time for some group therapy: update on identification, antimicrobial resistance, taxonomy, and clinical significance of the *Bacteroides fragilis* group. J Clin Microbiol. 2022;60(9). https://doi.org/10.1128/jcm.02361-20.
23. Fathi P, Wu S. Isolation, detection, and characterization of enterotoxigenic *Bacteroides fragilis* in clinical samples. Open Microbiol J. 2016;1057–63. https://doi.org/10.2174/1874285801610010057.
24. L.Leber A. Clinical Microbiology Procedures Handbook, Volume 1-3, 4th Edition. John Wiley & Sons, Inc;2016.
25. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64. https://doi.org/10.1093/nar/25.5.955.
26. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006;34(Database issue):D32–6. https://doi.org/10.1093/nar/gkj014.
27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25(1):25–9. https://doi.org/10.1038/75556.
28. Buchanan RE, Gibbons NE. Berger's Bacterial Identification Manual. 8th ed. Science Publishing Company; 1984.
29. Wang Y, Deng H, Li Z, Tan Y, Han Y, Wang X, et al. Safety evaluation of a novel strain of *Bacteroides fragilis*. Front Microbiol. 2017;8435. https://doi.org/10.3389/fmicb.2017.00435.
30. Kobayashi I. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. Nucleic Acids Res. 2001;29(18):3742–56. https://doi.org/10.1093/nar/29.18.3742.
31. Patrick S, Blakely GW, Houston S, Moore J, Abratt VR, Bertalan M, et al. Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of *Bacteroides fragilis*. Microbiol (Reading). 2010;156(Pt 11):3255–69. https://doi.org/10.1099/mic.0.042978-0.
32. Sukackaite R, Grazulis S, Tamulaitis G, Siksnys V. The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. Nucleic Acids Res. 2012;40(15):7552–62. https://doi.org/10.1093/nar/gks332.
33. Ben-Assa N, Coyne MJ, Fomenkov A, Livny J, Robins WP, Muniesa M, et al. Analysis of a phase-variable restriction modification system of the human gut symbiont *Bacteroides fragilis*. Nucleic Acids Res. 2020;48(19):11040–53. https://doi.org/10.1093/nar/gkaa824.
34. Zhang L, Lau K, Cheng J, Yu H, Li Y, Sugiarto G, et al. *Helicobacter hepaticus Hh0072* gene encodes a novel alpha1-3-fucosyltransferase belonging to CAZy GT11 family. Glycobiology. 2010;20(9):1077–88. https://doi.org/10.1093/glycob/cwq068.
35. Yekani M, Baghi HB, Vahed SZ, Ghanbari H, Hosseinpur R, Azargun R, et al. Tightly controlled response to oxidative stress; an important factor in the tolerance of *Bacteroides fragilis*. Res Microbiol. 2021;172(2):103798. https://doi.org/10.1016/j.resmic.2021.103798.
36. Alam A, Broms JE, Kumar R, Sjostedt A. The role of ClpB in bacterial stress responses and virulence. Front Mol Biosci. 2021;8668910. https://doi.org/10.3389/fmolb.2021.668910.
37. Lewis TA, Leach L, Morales S, Austin PR, Hartwell HJ, Kaplan B, et al. Physiological and molecular genetic evaluation of the dechlorination agent, pyridine-2,6-bis(monothiocarboxylic acid) (PDTC) as a secondary siderophore of *Pseudomonas*. Environ Microbiol. 2004;6(2):159–69. https://doi.org/10.1046/j.1462-2920.2003.00558.x.
38. Rocha ER, Bergonia HA, Gerdes S, Jeffrey Smith C. *Bacteroides fragilis* requires the ferrous-iron transporter FeoAB and the CobN-like proteins BtuS1 and BtuS2 for assimilation of iron released from Heme. Microbiologyopen. 2019;8(4):e00669. https://doi.org/10.1002/mbo3.669.
39. Tzianabos AO, Onderdonk AB, Rosner B, Cisneros RL, Kasper DL. Structural features of polysaccharides that induce intra-abdominal abscesses. Science. 1993;262(5132):416–9. https://doi.org/10.1126/science.8211161.
40. Bolam DN, Koropatkin NM. Glycan recognition by the Bacteroidetes Sus-like systems. Curr Opin Struct Biol. 2012;22(5):563–9. https://doi.org/10.1016/j.sbi.2012.06.006.
41. West PT, Chanin RB, Bhatt AS. From genome structure to function: insights into structural variation in microbiology. Curr Opin Microbiol. 2022;69102192. https://doi.org/10.1016/j.mib.2022.102192.
42. Lee IPA, Eldakar OT, Gogarten JP, Andam CP. Bacterial Cooperation through horizontal gene transfer. Trends Ecol Evol. 2022;37(3):223–32. https://doi.org/10.1016/j.tree.2021.11.006.
43. Borodovich T, Shkoporov AN, Ross RP, Hill C. Phage-mediated horizontal gene transfer and its implications for the human gut Microbiome. Gastroenterol Rep (Oxf). 2022. 10goac012.
44. Dejea CM, Fathi P, Craig JM, Boleij A, Taddese R, Geis AL, et al. Patients with Familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. Science. 2018;359(6375):592–7. https://doi.org/10.1126/science.aah3648.
45. Morgillo F, Dallio M, Della Corte CM, Gravina AG, Viscardi G, Loguercio C, et al. Carcinogenesis as a result of multiple inflammatory and oxidative hits: a comprehensive review from tumor microenvironment to gut microbiota. Neoplasia. 2018;20(7):721–33. https://doi.org/10.1016/j.neo.2018.05.002.
46. Gemmell CG, Peterson PK, Schmeling D, Mathews J, Quie PG. Antibiotic-induced modification of *Bacteroides fragilis* and its susceptibility to phagocytosis by human polymorphonuclear leukocytes. Eur J Clin Microbiol. 1983;2(4):327–34. https://doi.org/10.1007/BF02019462.

47. Reid JH, Patrick S. Phagocytic and serum killing of capsulate and non-capsulate *Bacteroides fragilis*. J Med Microbiol. 1984;17(3):247–57. https://doi.org/10.1099/00222615-17-3-247.
48. Yekani M, Baghi HB, Naghili B, Vahed SZ, Soki J, Memar MY. To resist and persist: important factors in the pathogenesis of *Bacteroides fragilis*. Microb Pathog. 2020;149104506. https://doi.org/10.1016/j.micpath.2020.104506.
49. Huang Z, Yu K, Xiao Y, Wang Y, Xiao D, Wang D. Comparative genomic analysis reveals potential pathogenicity and Slow-Growth characteristics of genus Brevundimonas and description of *Brevundimonas pishanensis Sp. nov*. Microbiol Spectr. 2022;10(2):e0246821. https://doi.org/10.1128/spectrum.02468-21.
50. Lewis JP, Plata K, Yu F, Rosato A, Anaya C. Transcriptional organization, regulation and role of the *Porphyromonas gingivalis W83* Hmu haemin-uptake locus. Microbiol (Reading). 2006;152(Pt 11):3367–82. https://doi.org/10.1099/mic.0.29011-0.
51. Costa TRD, Harb L, Khara P, Zeng L, Hu B, Christie PJ. Type IV secretion systems: advances in structure, function, and activation. Mol Microbiol. 2021;115(3):436–52. https://doi.org/10.1111/mmi.14670.
52. Wilson MM, Anderson DE, Bernstein HD. Analysis of the outer membrane proteome and secretome of *Bacteroides fragilis* reveals a multiplicity of secretion mechanisms. PLoS ONE. 2015;10(2):e0117732. https://doi.org/10.1371/journal.pone.0117732.
53. Toth R, Alonso MF, Bain JM, Vagvolgyi C, Erwig LP, Gacser A. Different *Candida parapsilosis* clinical isolates and lipase deficient strain trigger an altered cellular immune response. Front Microbiol. 2015;61102. https://doi.org/10.3389/fmicb.2015.01102.
54. Robertson KP, Smith CJ, Gough AM, Rocha ER. Characterization of *Bacteroides fragilis* hemolysins and regulation and synergistic interactions of HlyA and HlyB. Infect Immun. 2006;74(4):2304–16. https://doi.org/10.1128/IAI.74.4.2304-2316.2006.
55. Chen Y, Kinouchi T, Kataoka K, Akimoto S, Ohnishi Y. Purification and characterization of a fibrinogen-degrading protease in *Bacteroides fragilis* strain YCH46. Microbiol Immunol. 1995;39(12):967–77. https://doi.org/10.1111/j.1348-0421.1995.tb03300.x.
56. Rudek W, Haque RU. Extracellular enzymes of the genus Bacteroides. J Clin Microbiol. 1976;4(5):458–60. https://doi.org/10.1128/jcm.4.5.458-460.1976.

Publisher's note