ORIGINAL

Characterization of a gene cluster for sialoglycoconjugate utilization in *Bacteroides fragilis*

Haruyuki Nakayama-Imaohji[1,2], Minoru Ichimura[1], Tomoya Iwasa[2], Natsumi Okada[2], Yoshinari Ohnishi[2], and Tomomi Kuwahara[2,3]

[1]*Department of Immunology and Parasitology, and* [2]*Department of Molecular Bacteriology, Institute of Health Biosciences, the University of Tokushima Graduate School, Tokushima, Japan ; and* [3]*Department of Molecular Microbiology, Faculty of Medicine, Kagawa University, Kagawa, Japan*

Abstract : Recent analysis of the whole genome sequence of *Bacteroides fragilis* revealed extensive duplication of polysaccharide utilization genes in this anaerobe. Here we analyzed a unique 27-kb gene cluster (*sgu*) comprised of the 13 sialoglycoconjugates-utilization genes, which include the sialidase gene (*nanH1*) in *B. fragilis* strain YCH46. The genes were tightly organized and transcribed polycistronically. Comparative PCR scanning demonstrated that the *sgu* locus was conserved among the *Bacteroides* strains tested. Based on the transcriptional profiles generated by reverse transcriptase PCR, the *sgu* locus can be classified into at least three regulatory units : 1) sialic acid- or sialooligosaccharide-inducible genes, 2) constitutively expressed genes that can be down-regulated by catabolite repression, and 3) constitutively expressed genes. *In vitro* comparison of the growth of a *sgu* locus deletion mutant (SGUM172941) with a wild type strain indicates that this locus is necessary for *B. fragilis* to efficiently utilize mucin as a carbon source. Furthermore, SGUM172941 was defective in colonization of the intestines of germ-free mice under competitive conditions. These data indicate that the *sgu* locus in *B. fragilis* plays a crucial role in the utilization of host-derived sialoglycoconjugates and the stable colonization of this anaerobe in the human gut. J. Med. Invest. 59 : 79-94, February, 2012

Keywords **:** *Bacteroides fragilis, sialidase, sialoglycoconjugate, colonization, mucin*

INTRODUCTION

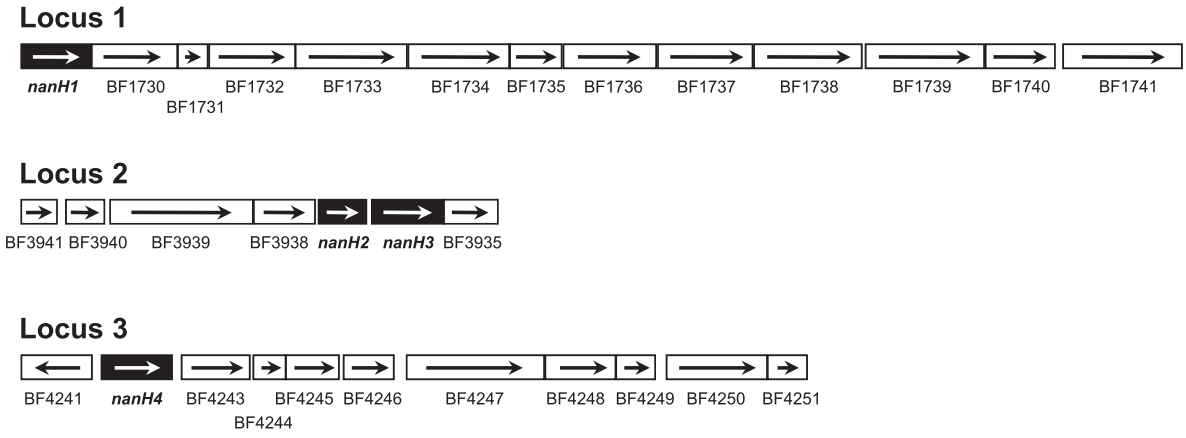
Bacteroides is the predominant member of human colonic microbiota. The environment of the lower intestine is poor in mono- and di-saccharides, most of which are absorbed by the host and the microbiota in the upper intestinal tract. Analysis of the three *Bacteroides* genomes *B. fragilis* (1), *B. thetaiotaomicron* (2) and *B. vulgatus* (3) revealed that *Bacteroides* is able to utilize a broad range of dietary and host-derived polysaccharides (4-8). These anaerobes have an exceptionally large number of genes involved in the utilization of a variety of complex polysaccharides (e.g., glycosylhydrolases, glycosyltransferases, and outer membrane polysaccharide-binding proteins). Although human colonic bacteria are commonly enriched for genes involved in polysaccharide utilization, the number of glycosylhydrolase genes in *Bacteroides* species is much higher than in other sequenced colonic bacteria such as *Bifidobacterium longum*, *Clostridium perfringens* and *Enterococcus faecalis* (9, 10). In

Received for publication October 26, 2011 ; accepted October 31, 2011.

Address correspondence and reprint requests to Haruyuki Nakayama-Imaohji, Department of Immunology and Parasitology, Institute of Health Biosciences, the University of Tokushima Graduate School, Kuramoto-cho, Tokushima 770-8503, Japan and Fax : +81-88-633-9229.

addition, many of the glycosylhydrolase genes colocalize with genes for the SusC/SusD family of proteins (outer membrane proteins involved in polysaccharide recognition and transport), extracellular function-type sigma factors and their cognate anti-sigma factors (3, 5, 11, 12), allowing *Bacteroides* to sense and utilize various types of polysaccharides. These adaptations in *Bacteroides* reflect the polysaccharide-rich environment in the human colon.

Sialidase (neuraminidase, EC 3.2.1.18) is an enzyme responsible for the cleavage of *N*-acetylneuraminic acid (sialic acid) from glycoproteins and glycolipids on host cell surfaces or in body fluids. It plays a key role in the sequential degradation of host-derived sialoglycoconjugates, and has been linked to the virulence of various pathogens (13, 14). Similarly, the most pathogenic *Bacteroides* species, *B. fragilis*, has a much higher level of sialidase activity than other *Bacteroides spp*. (15), an activity that is involved in the growth and survival of *B. fragilis* on tissue culture monolayers and in rat granuloma pouches *in vivo* (16). Four sialidase genes (designated *nanH1* to *nanH4*) have been identified in the genome of *B. fragilis* strain YCH46, while only two are found in *B. thetaiotaomicron* strain VPI-5482. These findings indicate that the host-derived polysaccharides are major nutritional sources for *B. fragilis* as well as dietary polysaccharides. As shown in Fig. 1, a number of genes for polysaccharide utilization such as glycosylhydrolases, outer membrane receptor proteins (SusC/SusD family), and transporters are clustered with sialidase genes in the *B. fragilis* YCH46 genome. Of these, the 27-kb region that includes *nanH1* is unique in that it contains (i) a much higher number of glycosylhydrolase genes than other loci and (ii) a sialate *O*-acetylesterase homologue, a rate-limiting enzyme of sialic acid catabolism that catalyzes removal of acetyl groups from *O*-acetylated sialic acid in sialoglycoconjugates (17, 18). *Streptococcus pneumoniae* sialidase and exoglycosidases sequentially deglycosylate human glycoconjugates, such as IgA1, to facilitate growth and colonize airways (19). In addition, this sequential deglycosylation of human serum increases resistance to complement deposition and subsequent phagocytic killing of *S. pneumoniae* (20). Thus, sialidase and other glycosylhydrolases may contribute to the ability of *B. fragilis* to establish a



Gene	Putative function	Gene	Putative function
BF3935, BF4245	major facilitator family transporter	BF4244	*N*-acetylglucosamine 2-epimerase
BF3938, BF4248	SusD homolog	BF4246	*N*-acetylneuraminic acid mutarotase
BF3939, BF4247	SusC homolog	BF4249	endo-β-*N*-galactosidase
BF3940	*N*-acetylneuramininate lyase	BF4250	β-hexosaminidase
BF3941	transcriptional regulator	BF4251	β-glucanase
BF4241, BF4243	transposase		

Fig. 1. Genomic region around the four sialidase paralogues (*nanH1* to *nanH4*, shown by closed boxes) in *B. fragilis* YCH46. Predicted coding regions are represented by open boxes with internal arrows showing transcriptional direction. Gene identifications are shown under the boxes. Putative functions of ORFs are listed in the table. Functions of the genes in locus-1 are summarized in Table 3.

stable colonization of the gut.

In this study, we characterized the *sgu* (sialoglycoconjugates utilization) locus, the 27-kb gene cluster that contains *nanH1*, and demonstrated its importance in the process of *B. fragilis* colonization of the host intestinal tract.

MATERIALS AND METHODS

Bacterial strains, plasmids and culture conditions

The bacterial strains and plasmids used in this study are listed in Table 1. All of the *Bacteroides* strains were grown anaerobically at 37°C in Gifu

Table 1. Bacterial strains and plasmids

Strain or plasmid	Relevant genotype or description	Reference or source
Strains		
Escherichia coli		
XLI-Blue	*end*Δ*1 hsdR17* ($r_K^- m_K^+$) *supE44 thi-1*λ*⁻ recA1 gyrA96* (Nal^R) *relA1* (Δ*lac*) [F′ *proAB*$^+$ *lacI*^{q}ZΔM15 : : Tn*10* (Tc^R)]	Stratagene
GMS407	*lacY1 galK2* (Oc) λ*⁻ manA4*, *uidA1*, *mtl-1*, *argE3* (Oc)	*E. coli* Genetic Stock Centre
Bacteroides fragilis		
ATCC25285	Type strain, appendix abscess	ATCC
ATCC43859	Laboratory strain, enterotoxigenic, *bft2*	36, ATCC
GAI97124	Clinical isolate, enterotoxigenic, *bft2*	37
YCH46	Clinical isolate	1
YCH2	Clinical isolate	I. Nakamura, Perianal abscess
YCH4	Clinical isolate	I. Nakamura, Abscess at vulva
YCH9	Clinical isolate	I. Nakamura, Peritoneal abscess
YCH21	Clinical isolate	I. Nakamura, Blood
YCH44	Clinical isolate	I. Nakamura, Abscess at abdominal wall
TDP101	Clinical isolate	38
B1	Clinical isolate	Peritoneal abscess
KMS5	Clinical isolate	Abscess
SGUM172941	YCH46 mutant with deletion of BF1729-BF1741	This study
B. uniformis		
BU1001	Laboratory strain, Gn^R	39
Plasmids		
pUC19	*E. coli* cloning vector ; Ap^R	Nippon gene
pBluescript II KS	*E. coli* cloning vector ; Ap^R	Stratagene
pGEM-T easy	*E. coli* cloning T-vector ; Ap^R	Promega
pT7blue	*E. coli* cloning T-vector ; Ap^R	Novagen
pVAL-1	*E. coli-Bacteroides* shutle vector ; Ap^R Tc^R in *E. coli*, Em^R in *Bacteroides* ; Mob$^+$ Rep$^+$	40
R751	Mobilizable plasmid used to move plasmids from *E. coli* to *B. uniformis* ; Tra$^+$ Tp^R	39
pNAND100	5.2-kb PCR fragment generated with primers APF1 and APR2 cloned into pT7blue, *estS*$^+$, *bmnA*$^+$	This study
pNAND170	3.2-kb *bmnA* fragment derived from pNAND100 cloned into pBluescript II KS, *bmnA*$^+$	This study
pNAND200	3.4-kb PCR fragment generated with primers KPF1 and H2R-3 cloned into pT7blue, *estA*$^+$, *estS*$^+$	This study
pNAND240	3.4-kb *estA/S* fragment derived from pNAND200 cloned into pBluescript II KS, *estA*$^+$, *estS*$^+$	This study
pNAND241	3.1-kb *estS* fragment derived from pNAND240 in pBluescript II KS, *estS*$^+$	This study
pNAND244	0.9-kb *estA* fragment derived from pNAND240 in pBluescript II KS, *estA*$^+$	This study
pVNAND241	3.1-kb *estS* fragment derived from pNAND241 cloned into pVAL-1, *estS*$^+$	This study
pNAND300	3.0-kb *bgaA*-PCR fragment generated with primers EIFZ4/new and ClaR8, cloned into pGEM-T *easy*, *bgaA*$^+$	This study
pNAND400	3.4-kb *bgaB*-PCR fragment generated with primers LONGsusD-F and LONGdnaJ-R cloned into pGEM-T easy, *bgaB*$^+$	This study
pOT3014	5.6-kb *nanH1-nahA-estA* fragment cloned into pHSG399, *nanH1*$^+$, *nahA*$^+$, *estA*$^+$	22
pNAKS	2.5-kb *nahA* fragment derived from pOT3014 cloned into pBluescript II KS, *nahA*$^+$	This study
pGN-1	6.2-kb *nahB-nanM-nahC*-PCR fragment generated with primers 3F-1 and KIW-R cloned into pGEM-T easy, *nahB*$^+$, *nanM*$^+$, *nahC*$^+$	This study
pNAP-1	2.8-kb *nahB* fragment derived from pGN-1 cloned into pUC19, *nahB*$^+$	This study
pNAP-3	3.4-kb *nahC* fragment derived from pGN-1 cloned into pUC19, *nahC*$^+$	This study

Abbreviations : Ap, ampicillin ; Em, erythromycin ; Gn, geneticin ; Nal, nalidixic acid ; Sm, streptomycin ; Tc, tetracycline ; Tp, trimethoprim ; Mob, ability to be mobilized by self-transmissible plasmid ; Rep, ability to replicate in *Bacteroides* ; Tra, ability to self-transfer.

anaerobic medium (GAM) broth (Nissui Pharmaceutical Co., Tokyo, Japan) or on GAM agar plates using the AnaeroPack System (Mitsubishi Gas Chemical Co., Inc., Tokyo, Japan) or an anaerobic chamber conditioned with mixed gases (N_2, 80% ; CO_2, 10% ; and H_2, 10%). *Escherichia coli* strains were routinely grown aerobically at 37°C in Luria-Bertani (LB) broth or on LB agar plates. If necessary, antibiotics were added to the media at the following final concentrations : ampicillin (Ap), 50 μg/ml ; erythromycin (Em), 10 μg/ml ; and gentamycin, 400 μg/ml.

In silico analysis of the sgu locus

The 27-kb *sgu* locus was identified in the complete genome sequence of *B. fragilis* strain YCH46 (Locus 1 in Fig. 1). Open reading frame (ORF) predictions and protein functional annotation were carried out as previously described (1). The similarity of deduced protein sequences to those in public databases was examined using BLASTP from the National Centre for Biotechnology Information (http : //www.ncbi.nlm.nih.gov/BLAST/). Alignments of nucleotide or amino acid sequences were performed using Clustal W (http : //clustalw.genome.ad.jp/). DNA sequencing was performed using an ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit, version 1.1 (Applied Biosystems, Foster City, CA) followed by analysis of the products on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems).

DNA manipulations

Standard molecular biology procedures were performed as described by Sambrook *et al.* (21). Restriction enzymes, DNA ligase, and DNA polymerase were used in accordance with the manufacturer's instructions. Synthetic oligonucleotide primers were purchased from Sigma-Aldrich Japan Co., Ltd. (Tokyo, Japan).

Construction of plasmids

As shown in Fig. 2A, regions within the *sgu* locus were separately amplified to subclone the putative glycosylhydrolase genes into *E. coli* cloning vectors. Amplification products of 5.2, 3.4, and 6.2 kb generated by primer sets APF1 and APR2, KPF1 and H2R-3, and 3F-1 and KIW-R were cloned into a pT7blue vector to construct pNAND100, pNAND200, and

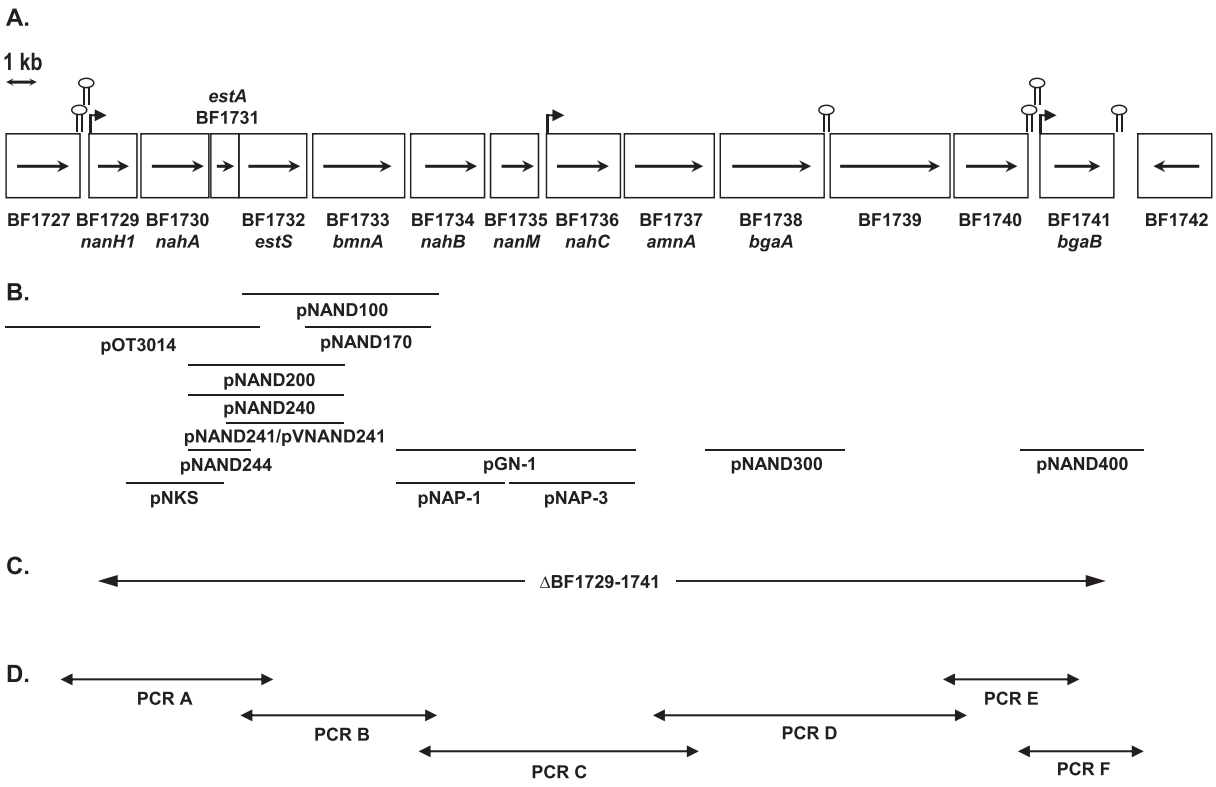


Fig. 2. Genetic organization of the 27-kb region (*sgu* locus) containing *nanH1* in *B. fragilis* YCH46 (A). Predicted coding regions are represented by open boxes with internal arrows showing transcriptional direction. Putative regulatory regions are indicated by bent arrows (promoter) and hairpin structures (transcriptional terminators). (B) Plasmids used in this study. Lines denote the DNA regions subcloned into individual plasmids. (C) ΔBF1729-1741 indicates the region that was deleted in the chromosome of mutant SGUM172941. (D) Diagrammatic representation of the PCRs performed to compare the genetic organization of *sgu* locus among *B. fragilis* strains.

pGN-1, respectively. The 3.2-kb *Hinc* II/*Cla* I-digested fragment containing *bmnA* from pNAND100 was subcloned into a pBluescript II KS vector to construct pNAND170. The 3.4-kb fragment containing the *estA/S* genes, from the *Xba* I/*Sma* I-digestion of pNAND200, was subcloned into a pBluescript II KS vector to construct pNAND240. The 0.9-kb *estA* and 3.1-kb *estS* fragments produced by *Pst* I- and *Sty* I/*Xba* I-digestion of pNAND240 were religated to construct pNAND244 and pNAND241, respectively. The 2.8-kb *nahB* and 3.4-kb *nahC* fragments obtained by *Sph* I/*Xho* I- and *Bam*H I/*Eco*R I-digestion of pGN-1 were subcloned into pUC19 to construct pNAP-1 and pNAP-3, respectively. The plasmid pNAKS was constructed by subcloning the 2.5-kb *Pvu* II/*Pst* I-digested fragment containing *nahA* from pOT3014 (22) into a pBluescript II KS vector. The 3.0-kb and 3.4-kb fragments containing *bgaA* and *bgaB* were amplified using the primer set EIFZ4/new and ClaR8 and the primer set LONGsusD-F and LONGdnaJ-R and cloned into a pGEM-T easy vector to construct pNAND300 and pNAND400, respectively. For the sialate *O*-acetylesterase assay in *B. uniformis* strain BU1001, mobilization of the *E. coli-Bacteroides* shuttle plasmid pVNAND241 carrying the *estS* gene from *E. coli* to *B. uniformis* was performed with a filter mating procedure using mobilizer plasmid R751 as previously described (22). The nucleotide sequences and annealing positions of oligonucleotide primers used for plasmid constructions are shown in Table 2.

Construction of deletion mutants

Deletion of the *sgu* locus (BF1729-BF1741) in

Table 2. Oligonucleotide primers

Primer name	Sequence (5' to 3')	Primer name	Sequence (5' to 3')
APF1	TAAGGGACATCGTCCTTACATGCAGCAAGA	KIW-F13	TAACTATCGATATGGGTGAG
APR1	TCTTATAATGATCGGTTACCGGAAGTACAC	KIW-R	AACGTTCAATTTTCCGGTTGT
APR2	GGCTTCTTCCTGAGTATAGAAACCACCGTA	KIW-R4	ATGGATCAGGTAACTCTGTA
ClaR1	CGAGTGACTGATTGCAAAGACTGACCACCT	KIW-R6	TCAGATTGGTATACAACAGG
ClaR6	TCCGGATTTCGGTGATTGGG	KIW-R8	TGTCTCATTCTCTTTACGCT
ClaR8	AGGGCAAAAGTCAATGTCGG	KIW-R9	TCGATACTGAGACTCCATAA
ClaR10	AATCGAATCCGGTCCATACG	KKF-1	TACATTCCGTCATGATAT
ClaR16	CTTTCATCCAGGTATCCACG	KKF-2	TTACGTAGCTAACGAACGCAAT
EIFZ	TGGCTATTGGAAGAAGATGCGTCCGATGAC	KKR-1	ATTGCGTTCGTTAGCTA
EIFZ4/new	TGCCGAATATTTCACTGACCG	KPF1	ATGCAGTGGACGGACTTATCTA
EIFZ7	AAGCTTGTGATGAAATGGGC	LONGdnaJ-R	TGCCAAGCGTAAACGGCTATGGTACAGGAG
EIFZ9	GTTATCGCCTGATGTGGATG	LONGsusD-F	CACTTATACGGGCAGACCTCAACTTTCGGG
EIFZ11	TTGGTGACCTCAGTATGCTC	nahB-F	GATTGACTTGCAACAACCTACCGAGATCTCC
EIFZ16	ATCTGGCTACTATCACAGGC	nanH-P11	CCTTTCAGAACAACCTGATTACCGGG
estA-1F	CGTGCAACTCTGTTTGAGGA	NF1	AGTACAGATGGTGGAAAAAC
H2F-3	TGCGCAACCATGCTTCATTG	NhAR1	CCAGTTTCACTTCAATACTTGCACGAGC
H2R-1	AATCGCGGTAAATCCGTGTC	NR5	TCTTTTGTAATAGCCACAGC
H2R-3	GATACCGCTTGTAACCATGC	PBF	ATTGAGTATACTAATGATGG
H3F-2	CAGTTGTCAAGTATCAACCG	SALF	GTCGACTACCGGATGTATTA
H3R-2	CTGCTGTTGCAGGTGAACAA	SauR1	GAACGGTCACAGATTGCCAC
HaeF2	CGATTGATGAAACTGCCACC	SUSInvF1	GTGGATAATGATGCCATTCAAGCCACGTGG
HaeF4	ATTGGCTTTAGAGGCCGATG	Δgal-R	ACTTCCCGACAGGTCTATCTTTGTGACCTGC
HaeR2	TGGTCTACCATCACCAATCG	3F-1	GTCCAGATTCCTGTTCAAGG
HaeR3	CTCTTATCGGATATATCGGG	3F-2	CGTAAGTCTATTCCGGTAGC
HaeR6	TACAGAGTCTTGCTCATCGG	3R-1	GTTGGCTTTGAGCAGTCGTA
InvRF	TCAACTAAGATTGGGGAAGTGGCTGAAGAA	nanH-1	GCTCTAGACCAGTGATGGTAACTGGAGTTTCAGC
IPMR1	CCAATACGGTTCACAATTGTATCTGC	nan-1	CCGGATACCATGTGCTGCTTGAACTACCGGTATAG
KIW-F5	TTTGGGTGTATCGGCATGTTAT	nan-2	AAGCAGCAGATGGTATCCGGCATTGATTTCTGGCA
KIW-F8	TTTCTACTGACGGATACTGTT	nanH-8	GCTCTAGAACCGTTAGCACCGATGATACCATAAC
KIW-F11	TATTCCTGAGATCGAAATGC	estAS-3	AATCCTCTTCACGTGCGTTACGGTT

B. fragilis YCH46 was accomplished by removing the 26.8-kb internal segment of the cluster (Fig. 2C). Briefly, DNA fragments upstream and downstream of the target region were separately PCR-amplified using primer sets nanH-1 and nan-1 (upstream fragment), and nan-2 and nanH-8 (downstream fragment). These amplicons were fused by a second PCR amplification using the primer set nanH-1 and nanH-8 *via* overlapping regions incorporated into the primer sequences. The resultant PCR products were ligated into pKK100 (23). The targeting plasmids were electroporated into *B. fragilis* YCH46 as described previously (23, 24). Diploids, in which the targeting plasmid integrated into the chromosome *via* a single genetic crossover, were selected on GAM agar plates containing Em. The diploids were grown in GAM broth, spread on nonselective GAM agar plates, and replica plated onto GAM agar plates containing Em to screen for mutants that resolved the diploid through a second homologous recombination. Em-sensitive colonies were selected, and the expected deletion was checked by PCR with primer pairs estAS-3 and LONGdnaJ-R, which flanked the deletion site (Table 2).

PCR scanning of the sgu locus among B. fragilis strains

As shown in Fig. 2D, six segments of the *sgu* locus were separately amplified to compare the genetic organization in 12 *B. fragilis* strains (PCRs A to F). Segments were amplified using primer sets InvRF and H3R-2 (PCR A), APF1 and APR2 (PCR B), 3F-2 and ClaR16 (PCR C), EIFZ and ClaR1 (PCR D), EIFZ16 and HaeR2 (PCR E), or LONGsusD-F and LONGdnaJ-R (PCR F). PCR amplification was performed with an iCycler Thermal Cycler (Bio-Rad) using *TaKaRa EX Taq* DNA polymerase (Takara Shuzo Co., Ltd., Otsu, Japan) using the following parameters : preheating at 94°C for 1 min, 30 cycles of 94°C for 30 s and 68°C for 10 min, and a final extension at 68°C for 10 min. The nucleotide sequence and annealing position of each primer are shown in Table 2.

Enzyme assays

Recombinant clones of *E. coli* or *B. uniformis* grown, in LB broth containing 50 μg/ml of Ap or in GAM broth containing 10 μg/ml of Em, were harvested, washed once with 50 mM sodium phosphate buffer (pH7.4), and resuspended in four volumes of buffer. The cells were disrupted by 50% pulsed sonication in an ice bath for 5 min and then centrifuged at 9,000×*g* for 30 min at 4°C. The supernatants were used for measurement of enzyme activities. Mucin purified from bovine submaxillary glands (Type I-S, Sigma), *p*-nitrophenyl acetate (*p*NPA), *p*NP-glycosides of β-D-galactopyranoside (*p*NPβGal), α-D-mannopyranoside (*p*NPαMan), β-D-mannopyranoside (*p*NPβMan), *N*-acetyl-β-D-glucosaminide (*p*NPβGlcNAc), and β-D-glucuronide (*p*NPβGlcU) were used as substrates in the enzyme assays. Enzymatic hydrolysis of the *p*NP-glycosides was carried out at 37°C in 500 μl of 50 mM sodium phosphate buffer (pH7.4) for up to 3 h at a substrate concentration of 1 mM before absorbance was measured at 400 nm. Sialate *O*-acetylesterase activity was assayed with mucin as described previously (17).

Reverse transcription (RT)-PCR analysis

B. fragilis YCH46 was grown in GAM broth, washed once with phosphate-buffered saline (pH7.4), and transferred to pre-reduced PY (polypeptone-yeast extract) liquid medium (25) containing one of the following mono- or oligo-saccharides as a carbon source : glucose, 0.5% ; sialic acid, 0.5% ; starch, 0.5% ; colominic acid, 0.5% ; mucin (type I-S), 0.5% ; fetuin, 0.5% ; asialofetuin, 0.5% ; or chitin, 0.5%. All carbohydrates were purchased from Sigma. Cell growth was monitored by measuring turbidity with a photoelectric colorimeter (Fuji Kogyo Co., Ltd., Tokyo, Japan) at 660 nm. The cells were harvested at mid-logarithmic phase, and total RNA was purified using an RNeasy Kit (QIAGEN) with the additional step of treatment with RNAprotect Bacteria Reagent (QIAGEN) and RQ1 RNase-Free DNase (Promega) according to the manufacturers' instructions. RT-PCR was performed using a OneStep RT-PCR Kit (QIAGEN) according to the supplier's instructions. Twenty-five nanograms of total RNA was used for cDNA synthesis at 50°C for 30 min. The RT reaction was terminated by heating the mixture at 95°C for 15 min and the cDNA products were amplified under the following conditions : 30 cycles of 30 s at 94°C, 30 s at 55°C and 1 min at 72°C with a final extension step of 10 min of 72°C. PCRs using the oligonucleotide primer sets NF1 and NR5, KKF-1 and KKR-1, estA-1F and IPMR1, APF1 and H3R-2, H2F-3 and H2R-1, 3F-2 and APR2, KIW-F5 and KIW-R8, KIW-F11 and KIW-R4, EIFZ and ClaR16, EIFZ7 and ClaR10, EIFZ11 and ClaR6, SUSInvF1 and HaeR6, and HaeF4 and HaeR2 were performed to detect *nanH1*,

nahA, *estA*, *estS*, *bmnA*, *nahB*, *nanM*, *nahC*, *amnA*, *bgaA*, BF1739, BF1740, and *bgaB* transcripts, respectively. In addition, PCRs using the primer sets SALF and nanH-P11, NF1 and NhAR1, KKF-2 and APR1, PBF and H3R-2, H3F-2 and H2R-3, H2F-3 and 3R-1, nahB-F and KIW-R9, KIW-F8 and KIW-R6, KIW-F13 and KIW-R, EIFZ4/new and SauR1, EIFZ9 and ClaR8, EIFZ16 and ClaR, HaeF2 and Δgal-R, LONGsusD-F and HaeR3, and LONGsusD-F and HaeR2 were performed to detect the transcripts that overlap BF1727-*nanH1*, *nanH1-nahA*, *nahA-estA*, *estA-estS*, *estS-bmnA*, *bmnA-nahB*, *nahB-nanM*, *nanM-nahC*, *nahC-amnA*, *amnA-bgaA*, *bgaA*-BF1739, BF1739-BF1740, and BF1740-*bgaB*, respectively. 16S rRNA was used as an internal control to compare the quantities of RNA in each reaction. The RT-PCR products were separated by electrophoresis in 0.7% agarose gels. The nucleotide sequences and annealing positions of the oligonucleotide primers are listed in Table 2.

In vitro growth comparisons

To compare the *in vitro* growth of the *sgu*-deletion mutant (SGUM172941) to that of wild type, each strain was grown separately in liquid media containing one of a variety of carbohydrates as the sole carbon source. Briefly, SGUM172941 and wild type strains were grown in GAM broth, washed once with phosphate-buffered saline (pH7.4), and transferred to pre-reduced PY liquid medium (25) containing 0.5% glucose or 0.5% mucin (type I-S) as a carbon source. Growth was monitored by optical density at 600 nm using a spectrophotometer (CO8000 Cell Density Meter, WPA, Cambridge, UK).

Competitive colonization assay in the gut of germ-free mice

Five-week-old male IQI/Jic germ-free mice (Clea Japan Inc., Tokyo, Japan) were individually housed in plastic cages under sterile conditions in a vinyl isolator (Natsume Seisakusho Co., Ltd., Tokyo, Japan). Three mice each were used per experimental group. Mice were fed standard laboratory chow sterilized by gamma irradiation (50 kGy) and received autoclaved tap water *ad libitum*. The gut colonization capacity of SGUM172941 was assessed by inoculating germ-free mice with a 0.4-ml cocktail containing 2.0×10^8 viable cells each of the wild type and the mutant strains. Ileal mucosa, ileal contents, caecal contents, colonic mucosa, and colonic contents were collected for one week after inoculation to compare the population levels of the wild type and SGUM172941 strains. Serial sample dilutions were spread on GAM agar plates, and 96 colonies were screened per sample. Wild type and SGUM172941 strains were distinguished by colony-PCR with primers flanking the *sgu* locus-deletion site. All animal experiments were performed according to the guidelines approved by the Animal Committee of the University of Tokushima.

Nucleotide sequence accession number

The nucleotide sequence of the *sgu* locus of *B. fragilis* YCH46 described in this study has been assigned DDBJ/EMBL/GenBank accession number AP006841.

RESULTS

General features of the region containing nanH1 in B. fragilis strain YCH46

In silico analysis of the complete genome sequence of *B. fragilis* YCH46 revealed a large cluster of genes encoding proteins involved in sialoglycoconjugate utilization. The first gene of the cluster was identified as a sialidase gene (*nanH1*) (Fig. 2). This gene cluster contained 13 unidirectionally transcribed open reading frames (ORFs), ten of them (BF1729-BF1734, BF1736, BF1737, BF1738 and BF1741) encoding putative glycosylhydrolases. Within the 27-kb chromosomal segment, BF1729 to BF1735, BF1736 to BF1738, and BF1739 to BF1740 were tightly organized without any obvious transcriptional initiation signals and transcriptional terminators within the intervening regions. *Bacteroides* promoter consensus sequences (TAnnTTTG) centered at -7 from the transcriptional initiation site (26) were found upstream of *nanH1* (BF1729), BF1736, and BF1741. Inverted repeat sequences that form stem-loop structures, which possibly work as a transcriptional terminator, were found upstream of *nanH1* and downstream of BF1738, BF1740, and BF1741. Analysis of the base composition of this cluster revealed an overall G+C content (43.43%) similar to that of the *B. fragilis* YCH46 chromosome. Putative functions of ORFs were assigned based on homology to proteins in databases and are summarized in Table 3.

This 27-kb region, which contains genes that appear to be involved in the binding, degradation and transport of sialated polysaccharides, was designated as the *sgu* (sialoglycoconjugate utilization) locus.

Table 3. Similarity of putative ORFs in the *sgu* locus to public database sequences

ORF	Size (no. of aa[a])	Best homology to functionally known protein (organism/% aa identity/accession no.)	Conserved domain structure[b]	Putative function of protein
BF1730 (*nahA*)	670	β-*N*-acetylhexosaminidase (*Porphyromonas gingivalis*/27%/P49008)	Catalytic domain (pfam00728) and domain 2 (pfam02838) for glycosylhydrolase family 20	β-*N*-acetylhexosaminidase
BF1731 (*estA*)	220	α2 subunit of platelet-activating facter acetylhydrolase Ib (rat/27%/O35264)	Conserved residues Ser-Gly-Asn-His of SGNH-hydrolase family, sialate *O*-acetylesterase (cd01828)	Sialate *O*-acetylesterase
BF1732 (*estS*)	690			Sialate *O*-acetylesterase
N-terminal domain	220	Arylesterase (*Vibrio mimicus*/25%/Q07792)	Conserved residues Ser-Gly-Asn-His of SGNH-hydrolase family, sialate *O*-acetylesterase (cd01828)	
C-terminal domain	470	Sialic acid-specific 9-*O*-acetylesterase (mouse/28%/P70665)	-	
BF1733 (*bmnA*)	856	β-mannosidase BT0458 (*Bacteroides thetaiotaomicron*/76%/NP_809371)	TIM barrel domain of glycosylhydrolase family 2 (pfam02836)	β-mannosidase
BF1734 (*nahB*)	747	*N*-acetyl-β-glucosaminidase (*Tannerella forsythensis*/39%/AAO33832)	Catalytic domain (pfam00728) and domain 2 (pfam02838) for glycosylhydrolase family 20. F5/8 type C domain (pfam00754).	β-*N*-acetylhexosaminidase
BF1735 (*nanM*)	404	*N*-acetylneuraminic acid mutarotase (*Escherichia coli*/24%/2UVK_A)	Kelch motifs (pfam01344), cyclically-permuted mutarotase family (TIGR03548)	*N*-acetylneuraminic acid mutarotase, the conversion of the α-sialic acid isomer to the β form.
BF1736 (*nahC*)	689	*N*-acetyl-β-glucosaminidase (*Tannerella forsythensis*/45%/ AAO33832)	Catalytic domain (pfam00728) and domain 2 (pfam02838) for glycosylhydrolase family 20	β-*N*-acetylhexosaminidase
BF1737 (*amnA*)	749	Immunoreactive 87 kD antigen PG92 (*Porphyromonas gingivalis*/59%/AAD51077) α-1,2-mannosidase (*Bacillus* sp./26%/BAA76709)	Glycosyl hydrolase family 92 (pfam 07971), α-1,2-mannosidase (TIGR 01180)	α-mannosidase
BF1738 (*bgaA*)	833	β-galactosidase (*Xanthomonas campestris pv. campestris*/40%/AAP86765)	Sugar-binding domain (pfam02837), TIM barrel domain (pfam02836), and immunoglobulin β-sandwich domain (pfam00703) of glycosylhydrolase family 2	β-galactosidase
BF1739	914	Outer membrane protein involved in starch binding, SusC (*Bacteroides thetaiotaomicron*/28%/NP_812613)	TonB-dependent receptor domain (pfam00593)	SusC homolog, polysaccharide binding outer membrane protein
BF1740	543	Outer membrane protein, SusD (*Bacteroides thetaiotaomicron*/26%/NP_812612)	SusD family (pfam07980)	SusD homolog, polysaccharide binding outer membrane protein
BF1741 (*bgaB*)	928	β-galactosidase (*Pseudoalteromonas haloplanktis*/24%/CAA10470)	Sugar-binding domain (pfam02837), TIM barrel domain (pfam02836), and immunoglobulin β-sandwich domain (pfam00703) of glycosylhydrolase family 2	β-galactosidase

[a] aa, amino acids.
[b] Conserved domains based on the Conserved Domain Database (http : //www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml).

Comparison of the sgu gene cluster in Bacteroides

The *sgu* locus was divided into six overlapping segments (A to F). These segments were separately amplified in 12 strains of *B. fragilis* (two laboratory strains and ten clinical isolates) using PCR primer sets that were synthesized on the basis of the nucleotide sequence of the *sgu* locus (Panel D in Fig. 2). As summarized in Table 4, all of the strains tested produced a PCR product from segments A to D, indicating that the region from *nanH1* to BF1740 in the *sgu* locus is highly conserved among

Table 4. Comparative PCR scanning of the *sgu* locus among *B. fragilis* strains

Strains	Results of PCR [a]					
	A	B	C	D	E	F (kb)[b]
ATCC25285	+	+	+	+		0.5
ATCC43859	+	+	+	+	+	3.4
YCH2	+	+	+	+	+	3.4
YCH4	+	+	+	+		0.5
YCH9	+	+	+	+		0.5
YCH21	+	+	+	+	+	3.4
YCH44	+	+	+	+		0.5
YCH46	+	+	+	+	+	3.4
TDP101	+	+	+	+	+	3.4
B1	+	+	+	+		0.5
KMS5	+	+	+	+		0.5
GAI97124	+	+	+	+	+	3.4

[a] Plus indicates that the product of expected size was successfully amplified.
[b] Size of PCR product.

B. fragilis strains. However, *bgaB* (β-galactosidase gene) was absent in six strains. Segment E-positive strains produced the expected 3.4-kb product, while segment E-negative strains generated a 0.5-kb PCR product (PCR F). Comparison of the nucleotide sequences of these 3.4-kb and 0.5-kb products revealed that *bgaB* is deleted *via* a 13-bp direct repeat sequence in segment E-negative strains (data not shown). To date, four *B. fragilis* strains have been sequenced. The structure of the *sgu* locus was highly conserved in these four strains, but the *bgaB* gene was absent in strains NCTC9343 and 638R (Fig. 3A).

To determine whether *sgu*-like regions are present in other bacterial species, we searched for gene clusters containing homologues of the sialidase, sialate *O*-acetylesterase and glycosylhydrolase genes in public databases. The *sgu*-like region was identified in many *Bacteroides* species : *B. caccae*, *B. dorei*, *B. fluxus*, *B. intestinalis*, *B. helcogenes*, *B. ovatus*, *B. thetaiotaomicron*, and *B. vulgatus*. Comparative analysis of the genetic organization of *sgu* loci among *B. fragilis*, *B. vulgatus*, and *B. thetaiotaomicron* demonstrated a high degree of synteny. However, BF1739, BF1740, and *bgaB* were absent in *B. vulgatus* and *B. thetaiotaomicron*. In addition, *estA*, *nanM* and *amnA* were absent in *B. thetaiotaomicron* (Fig. 3B).

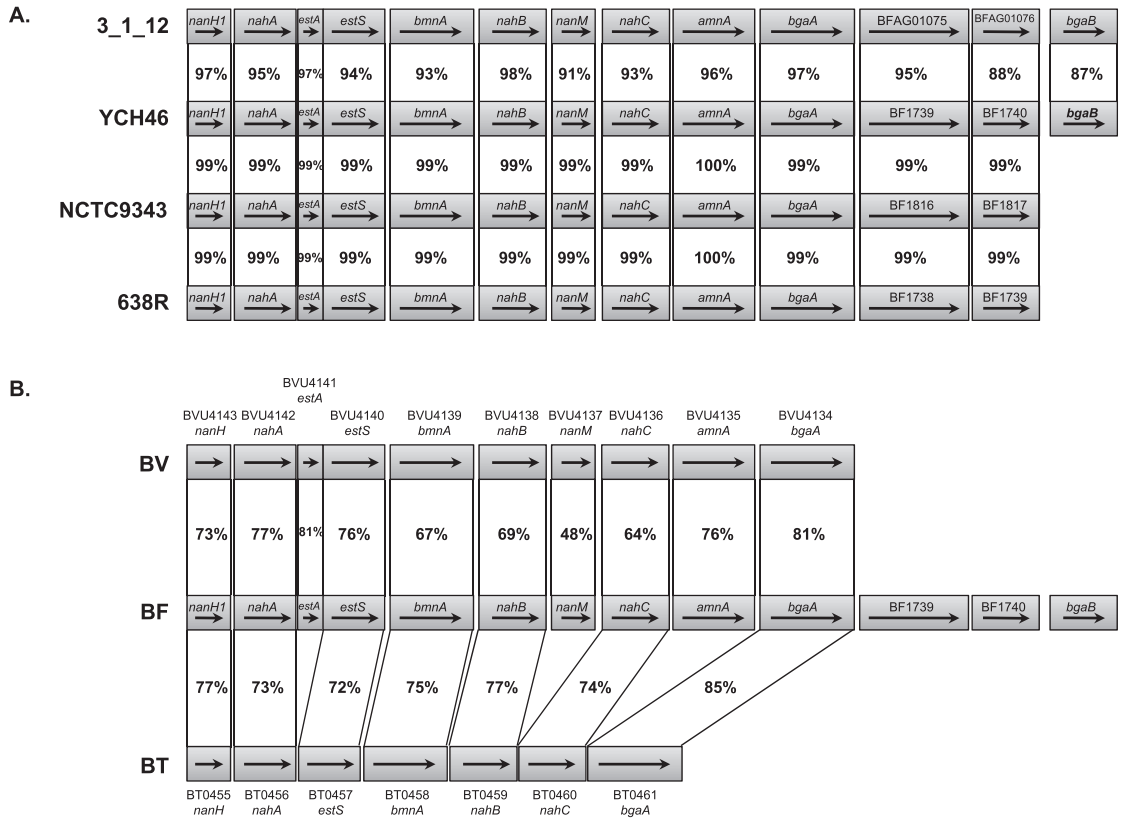


Fig. 3. Comparison of the *sgu* locus among *B. fragilis* strains (A) and *Bacteroides* species (B). BV : *Bacteroides vulgatus* ; BF : *Bacteroides fragilis* ; and BT : *Bacteroides thetaiotaomicron*. Homologous ORFs are linked. Amino acid sequence identities between homologous ORFs are indicated. The genetic structure of the region is conserved among *B. fragilis* strains and that the regions corresponding to regulatory units 1 and 3 are also conserved in *B. vulgatus* and *B. thetaiotaomicron*.

Functional analysis of the putative glycosylhydrolase genes in the sgu locus

Putative glycosylhydrolase genes (*nahA*, *estA*, *estS*, *bmnA*, *nahB*, *nahC*, *amnA*, *bgaA*, and *bgaB*) in the *sgu* locus were individually subcloned into a plasmid and introduced into *E. coli* (Fig. 2B). Glycosylhydrolase activities in crude extract of each recombinant *E. coli* strain were measured using various synthetic substrates (Fig. 4).

β-*N*-acetylhexosaminidase activity against *p*NPβGlcNAc was determined in crude extracts from *E. coli* cells carrying *nahA* (pNAKS), *nahB* (pNAP-1) or *nahC* (pNAP-3). β-*N*-acetylhexosaminidase activity of *E. coli* harboring pNAP-1 was 7.5-fold higher than that in *E. coli* cells with pNAP-3, while no β-*N*-acetylhexosaminidase activity was detected in *E. coli* harboring pNAKS.

Arylesterase and sialate *O*-acetylesterase activities were measured in *E. coli* strains carrying the plasmid pNAND244 (*estA*) or pNAND241 (*estS*) using *p*NPA and mucin, respectively. High levels of arylesterase and sialate *O*-acetylesterase activities were found in *E. coli* transformants harboring pNAND244, while no activity was detected in *E. coli* cells carrying pNAND241. It is presumed that the cloned *estS* gene was not expressed in *E. coli* due to differences in the transcription/translation systems or in posttranslational processing between *E. coli* and *Bacteroides* since both activities were detected in *B. uniformis* carrying pVNAND241. The *E. coli* strain harboring pNAND170 (*bmnA*) showed a high level of β-mannosidase activity when *p*NPβMan was used as a substrate.

The β-galactosidase activity of *E. coli* cells harboring the plasmid pNAND300 or pNAND400 was

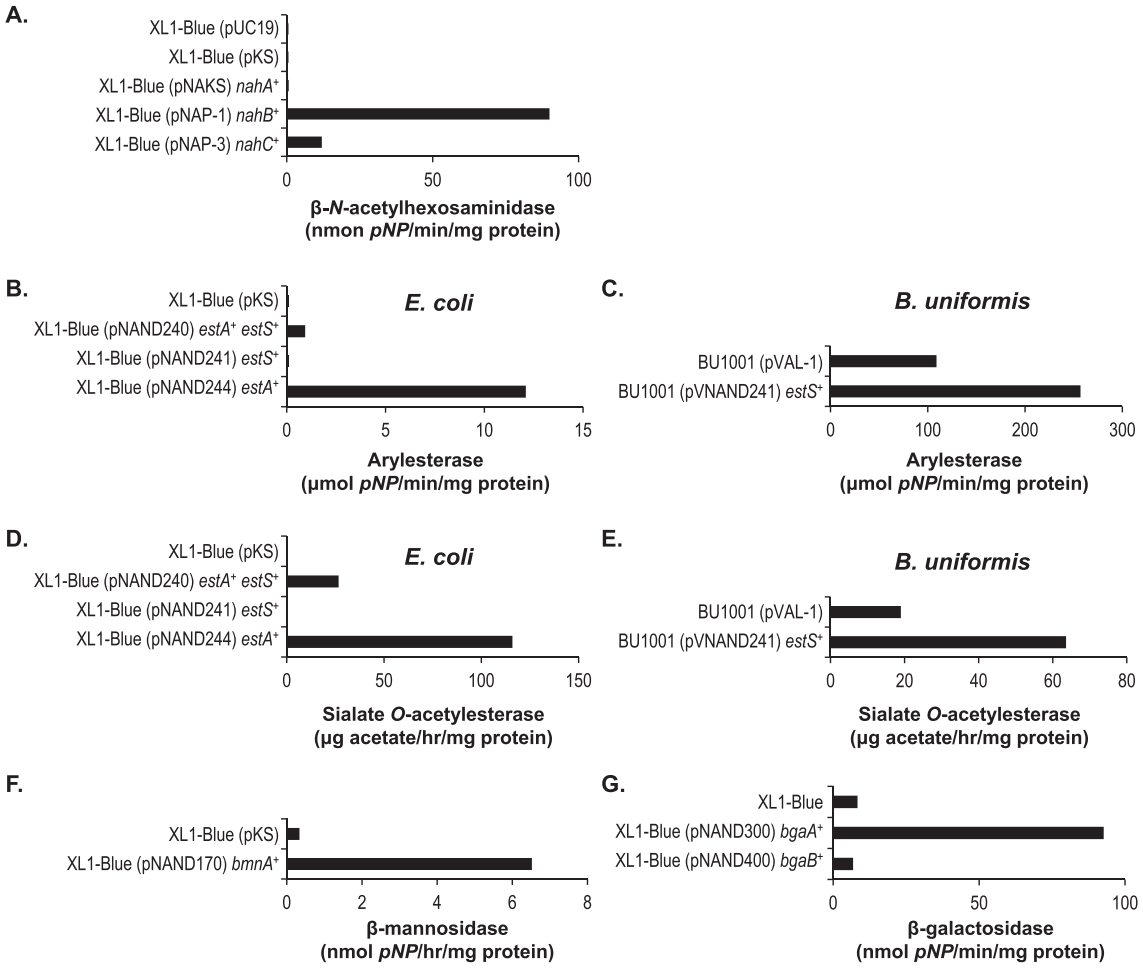


Fig. 4. Glycosylhydrolase activities of recombinant *E. coli* or *B. uniformis* clones. Glycosylhydrolase activities against various synthetic substrates were tested in each crude extract. (A) β-*N*-acetylhexosaminidase activity against *p*NPβGlcNAc in *E. coli* cells carrying *nahA* (pNAKS), *nahB* (pNAP-1) or *nahC* (pNAP-3). (B and D) Arylesterase and sialate *O*-acetylesterase activities measured in *E. coli* strains carrying *estA estS* (pNAND240), *estS* (pNAND241) or pNAND244 (*estA*) using *p*NPA and mucin, respectively. (C and E) Arylesterase and sialate *O*-acetylesterase activities of EstS expressed in *B. uniformis*. (F) β-mannosidase activity against *p*NPβMan in *E. coli* cells harboring pNAND170 (*bmnA*). (G) β-galactosidase activity produced by *E. coli* cells harboring pNAND300 (*bgaA*) or pNAND400 (*bgaB*) was measured using *p*NPβGal.

measured using *p*NPβGal. The pNAND300 plasmid produced a high level of β-galactosidase activity, while no activity was detected from pNAND400. β-glucuronidase activity against *p*NPβGlcU was also absent in the β-glucuronidase-negative *E.coli* mutant harboring the pNAND400 plasmid (data not shown). In addition, introduction of a shuttle plasmid carrying the *bgaB* gene into *B. uniformis* did not result in an increase of either β-galactosidase or β-glucuronidase activity when compared to *B. uniformis* carrying the control plasmid pVAL-1 (data not shown).

Although the *amnA* product showed a weak similarity to α-1,2-mannosidase of *Bacillus spp.*, α-mannosidase activity against *p*NPαMan was not observed in an *E. coli* strain carrying *amnA* (data not shown). Thus, the functional assignments found by a BLAST search for putative glycosylhydrolase genes in the *sgu* locus were confirmed experimentally, with the exceptions of *nahA*, *amnA*, and *bgaB*.

Transcriptional analysis of the sgu locus

To determine whether the genes in the *sgu* locus were co-transcribed, reverse transcription (RT)-PCR analysis using primer sets encompassing the neighboring ORFs was performed on total RNA extracted from a mid-logarithmic phase culture of *B. fragilis* YCH46 grown in GAM broth. As shown in Fig. 5A and 5B, the RT reaction produced no cDNA from the primer set spanning BF1727-*nanH1*, indicating that the transcription of BF1727 does not extend into *nanH1*. All of the other primer sets generated products of expected sizes, suggesting that the members of the *sgu* locus (*nanH1* to *bgaB*) are

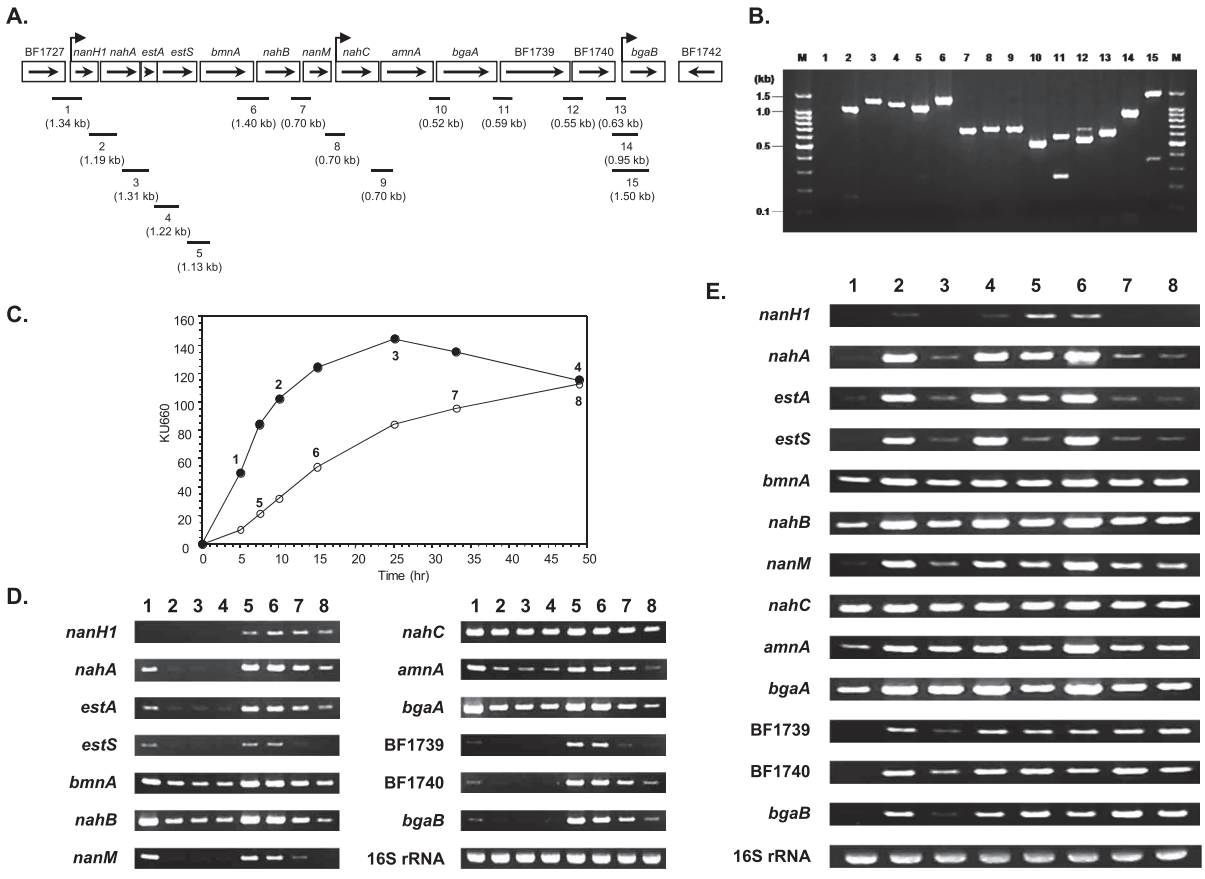


Fig. 5. Transcriptional analysis of the *sgu* locus. (A and B) Co-transcription with adjacent genes in the *sgu* locus assayed by RT-PCR. (A) The locations of the amplified fragments encompassing neighboring ORFs are indicated by black bars. The predicted sizes of amplification products are shown in the parentheses under the bars. ORFs are represented by open boxes as in Fig. 2. The bent arrows indicate putative *Bacteroides* promoters. (B) Agarose gel electrophoresis of the RT-PCR products. Lane numbers correspond to the amplified regions shown in panel A. M, 100-bp DNA ladder size markers. (C and D) Effects of growth phase and type of carbon source on the expression levels of *sgu* genes. (C) *B. fragilis* YCH46 was grown in PY medium supplemented with 0.5% glucose (closed circles) or 0.5% mucin (open circles) as the carbon source. Numbers around the growth curves indicate the points of cell harvest. (D) Expression levels of *sgu* genes assessed by RT-PCR analysis. Lane numbers indicate the harvest points shown in panel C. 16S rRNA was used as an internal control. (E) Effects of various types of sugars on the expression of *sgu* genes. Total RNA was isolated from *B. fragilis* YCH46 cells harvested at mid-logarithmic phase in PY medium supplemented with glucose (lane 1), sialic acid (lane2), starch (lane 3), colominic acid (lane 4), mucin (lane 5), fetuin (lane 6), asialofetuin (lane 7), or chitin (lane 8) as the carbon source. RT-PCR analysis was performed using 25 ng of each RNA sample. 16S rRNA was used as an internal control for RT-PCR.

organized in an operon, which is transcribed from the promoter upstream of *nanH1*. However, transcriptional analysis by RT-PCR of cultures at various phases of growth in PY medium supplemented with glucose or mucin as the carbon source showed complex expression patterns of the *sgu* genes (Fig. 5C and 5D). When *B. fragilis* YCH46 cells were grown in PY medium supplemented with mucin, the transcriptional levels of all of the *sgu* genes were elevated in the early- to mid-logarithmic phases, with a gradual decrease in early stationary phase. In contrast, the addition of glucose markedly repressed the expression of *nanH1*, *nahA*, *estA*, *estS*, *nanM*, BF1739, BF1740, and *bgaB* when the cells were at mid-logarithmic to stationary phase. The transcription of *nanH1* was completely repressed in the presence of glucose at all growth phases tested. *bmnA*, *nahB*, *nahC*, *amnA*, and *bgaA* were expressed constitutively and were not influenced by culture phase or carbohydrate source. These results indicated that the genes in the *sgu* locus could be divided into several regulatory units that change their expression levels depending on carbohydrate availability.

Effects of carbon source on the transcriptional expression of sgu genes

To further evaluate the effects of carbohydrate availability on the transcriptional expression of *sgu* genes, *B. fragilis* YCH46 cells were grown to mid-logarithmic phase in PY medium supplemented with 0.5% monosaccharide (glucose or sialic acid) or 0.5% polysaccharide (starch, colominic acid, mucin, fetuin, asialofetuin or chitin) as a carbon source (Fig. 5E). Glucose markedly repressed the expression of *nanH1*, *nahA*, *estA*, *estS*, *nanM*, BF1739, BF1740 and *bgaB*, but not the expression levels of of *bmnA*, *nahB*, *nahC*, *amnA* and *bgaA* (lane 1 in Fig. 5E). In contrast, sialic acid induced the expression of *nanH1*, *nahA*, *estA*, and *estS* (lane 2 in Fig. 5E). As shown in Fig. 5E, the expression of these sialic acid-inducible genes was also induced by sialoglycoconjugates such as colominic acid (lane 4), mucin (lane 5) and fetuin (lane 6), but not by asialopolysaccharides such as asialofetuin and chitin (lanes 7 and 8). In addition, when cells were grown on starch, very low levels of *nanH1*, *nahA*, *estA*, *estS*, *nanM*, BF1739, BF1740, and *bgaB* expression were observed (lane 3). Based on these results, the *sgu* locus can be divided into three regulatory units : unit 1 (*nanH1*, *nahA*, *estA*, and *estS*), which is induced by sialic acid or sialated polysaccharides ; unit 2 (*nanM*, BF1739, BF1740, and *bgaB*), which is expressed constitutively but suppressed by glucose or its polymer ; and unit 3 (*bmnA*, *nahB*, *nahC*, *amnA*, and *bgaA*), which is expressed constitutively.

Growth of SGUM172941 in vitro and in vivo

To assess the role of the sgu locus on *B. fragilis* growth, we compared the *in vitro* growth of wild type and SGUM172941 strains (Fig. 6A). SGUM172941

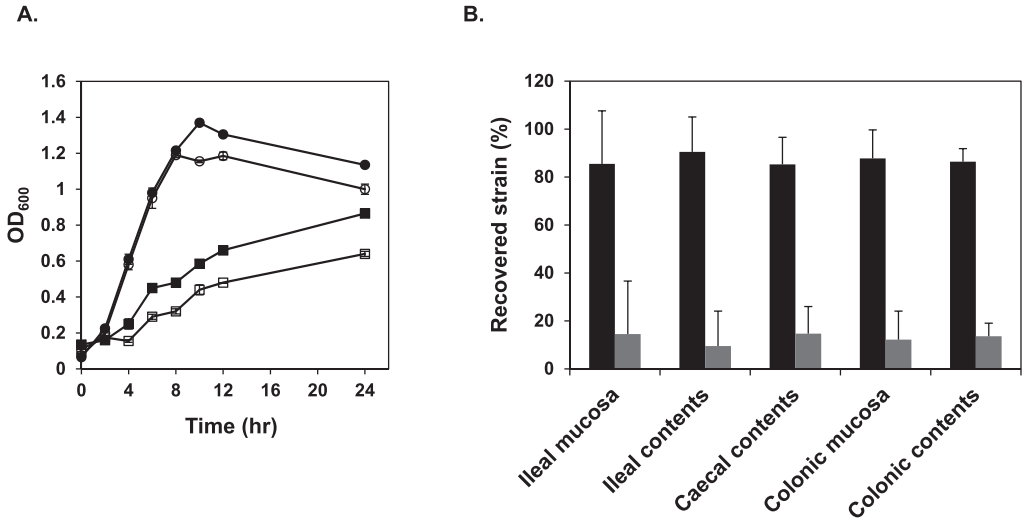


Fig. 6. Effect of *sgu* locus deletion on *in vitro* growth (A) and *in vivo* colonization (B) on *B. fragilis*. (A) Comparison of growth rates between wild type (black) and SGUM172941 mutant (white) strains inoculated into PY broth supplemented with glucose (circle) or mucin (square) as the sole carbon source. (B) A cocktail containing equal numbers of wild type and SGUM172941 mutant cells (2.0×10^8 viable cells each) was orally administered to IQI/Jic germ-free mice. One week after challenge, the ileal mucosa, ileal contents, caecal contents, colonic mucosa, and colonic contents were collected and diluted with phosphate-buffered saline. Appropriate dilutions were plated on GAM agar plates. Ratios of the strains in each sample were calculated from PCR analysis. Results are expressed as the mean ± standard deviation. Closed and gray bars indicate wild-type and SGUM172941 strains, respectively.

grew more slowly than wild type when they were cultured in PY medium supplemented with 0.5% mucin. However, both strains showed similar *in vitro* growth rates until late-logarithmic phase in PY medium containing glucose.

A gut colonization assay was performed to determine whether the *sgu* locus is essential for *B. fragilis* colonization in the intestine (Fig. 6B). A mixture containing an equivalent number of wild type and SGUM172941 strains was orally inoculated into germ-free mice. The wild type/SGUM172941 mutant ratio was assayed at 7 days after post-inoculation. The number of SGUM17294 mutant cells in all samples (ileal mucosa, ileal contents, caecal contents, colonic mucosa, and colonic contents) was reduced to nearly one-fifth the level of wild type cells.

These results indicate that the *sgu* locus plays an important role in the colonization process of *B. fragilis* in the host intestine by providing the capability to utilize complex host-derived polysaccharides such as mucin.

DISCUSSION

Sialic acid is a component of glycoproteins, glycolipids and various polysaccharides on mammalian cell surfaces. Sialylation is a common and functionally important modification of the glycoconjugates of vertebrates and some invertebrate species, and sialoglycoconjugates are involved in numerous biological and physiological processes, particularly intercellular and cell-molecule interactions (14, 27, 28). In the mammalian gut, sialoglycoconjugates are present on epithelial cell surfaces or in mucins. They show a great deal of structural complexity and diversity for evasion from pathogenic microbes, and the creation of mutually beneficial symbiotic relationships with nonpathogenic resident microflora (29). Intestinal residents have evolved a variety of gene repertoires for polysaccharide degradation through gene duplication or horizontal gene transfer to adapt to the structural heterogeneity of host-derived polysaccharides (3, 9, 30). The release of sialic acids from non-reducing ends by the action of sialidase is an initial step of sequential degradation of sialoglycoconjugates since sialic acid residues prevent the breakdown of sialoglycoconjugates by the action of other glycosylhydrolases. Another key enzyme for mucin degradation is sialate *O*-acetylesterase. Sialic acids in human rectal mucin are substituted by acetyl esters to a greater degree than in any other part of the gut (>40% of total sialic acids). The *O*-acetyl esters on the hydroxyl groups at positions C7, C8, and C9 (17, 31-33) inhibit the rate of sialic acid cleavage by sialidase. Thus, the cooperative action of sialidase and sialate *O*-acetylesterase plays a crucial role in the degradation of sialoglycoconjugate carbohydrate chains like mucin (17, 34). In addition, it has been shown that the relative content of sialic acid and *O*-acetylated sialic acids in human intestinal mucin increases proportionally from the ileum to the rectum (32), indicating that sialidase and sialate *O*-acetylesterase play key roles in mucin utilization of microbiota in the lower intestine. Most of the sequenced *Bacteroides* species possess sialidase and sialate *O*-acetylesterase homologues. Presumably, colocalization of *nanH* and *estA/S* in the *sgu* locus enables the cooperative action of both enzymes for efficient degradation of highly sialylated colonic mucin and contributes to the dominance of *Bacteroides* in the distal gut. Comparative PCR scanning analysis of the *sgu* locus revealed that the genetic structure of the region is conserved among *B. fragilis* strains and that the regions corresponding to regulatory units 1 and 3 are also conserved in other *Bacteroides* such as *B. vulgatus* and *B. thetaiotaomicron*. Furthermore, growth of the SGUM172941 mutant is inhibited when it is cultured in PY medium supplemented with mucin. Since SGUM172941 did not show a complete growth defect in mucin-supplemented medium, it is possible that mucin degradation can be attributed to other *nanH* paralogues (*nanH2*, *nanH3*, or *nanH4*). These observations indicate that the *sgu* locus plays an important role in *Bacteroides* nutrient acquisition in sialic acid-rich host environments like the human distal gut.

Transcriptional analysis of the *sgu* locus by RT-PCR using primer pairs that encompassed neighboring genes revealed that the region appeared to be transcribed polycistronically (Fig. 5). However, the transcription of this locus seemed to be regulated in a complex manner. The *sgu* genes can be classified into three regulatory units on the basis of transcriptional profiles in response to glucose, sialic acids, and sialo- (mucin and fetuin) and asialo-oligosaccharides (starch and asialofetuin). Genes categorized in the same regulatory unit also showed similar growth phase-dependent expression patterns. The genes categorized in regulatory unit 1 are predicted to play a key role in sialoglycoconjugate degradation since they are rate-limiting genes, such as *nanH1* and *estA/S*, which are induced by sialic

acid and sialoglycoconjugates. Genes in regulatory units 2 and 3 are expressed constitutively, but the expression of genes in regulatory unit 2 (*nanM*, BF1739, BF1740, and *bgaB*) are suppressed by glucose. The low levels of RT-PCR products for regulatory unit 2 genes when starch was used as the sole carbon source resulted from repressed expression by the increasing glucose concentration following the degradation of starch by α-glucosidase. Genes belonging to regulatory units 2 and 3 were not induced specifically by sialoglycoconjugates such as mucin and fetuin, suggesting that these genes are expressed constitutively for efficient degradation of de-sialylated glycoconjugates by the action of sialidase and sialate *O*-acetylesterase. Further analysis is needed to elucidate the complex regulatory mechanism(s) of the *sgu* locus.

The *in vivo* colonization competition assay demonstrated for the first time that the genes involved in degradation of sialoglycoconjugates have an essential role in *B. fragilis* colonization into the gut. The population level of SGUM172941 decreased to approximately 14% of the level of wild type cells one week after a mutant-wild type mixture of cells was orally administered to IQI/Jic germ-free mice. *In vivo* comparative transcriptome analysis of *B. thetaiotaomicron* demonstrated that the expression levels of *sgu* gene homologues (BT0455-BT0461) were elevated 2- to 10- fold in the ceca of monoassociated gnotobiotic mice compared to their expression in minimal medium plus glucose as the sole fermentable carbon source (12). Furthermore, these paralogs were expressed at higher levels in mice fed a simple-sugar diet than in mice fed a dietary polysaccharide-rich diet, indicating that the expression of this cluster is up-regulated to utilize host mucus glycans when polysaccharide availability from the diet is reduced (12). These findings indicate that the *sgu* locus is associated with the utilization of host-derived glycans such as mucin, and helps *B. fragilis* make close contact with the mucosal surface.

Genomic analysis of *Bacteroides* species revealed that they have evolved an exceptional ability to utilize polysaccharides through the extensive amplification of genes involved in polysaccharide binding, degradation and transport (1-3). Gene duplication is more extensive in *B. thetaiotaomicron*, which is likely correlated with its cell density in the colon, with a cell number that is 10- to 100-fold higher than that of *B. fragilis*. However, *B. fragilis* possesses a larger number of sialidase paralogues (four in *B. fragilis* vs. two in *B. thetaiotaomicron*). Considering the findings reported by Namavar *et al.* (35), in which *B. fragilis* is the predominant member of the *B. fragilis* group found on the gut mucosal surface but is in the minority in feces, it seems that host-derived sialoglycoconjugates, especially mucin, are major nutrient source for this anaerobe as well as host diet. Expansion of the sialidase gene would be beneficial not only for nutrient acquisition but also for infectious potential of *B. fragilis*. Therefore, exploration of the transcriptional regulation on *sgu* locus is needed to advance our understanding of the biology of *B. fragilis*.

CONFLICT OF INTEREST

None of the authors have any conflicts of interest to declare.

ACKNOWLEDGMENTS

We are grateful to Dr. Isao Nakamura and Dr. Naoki Kato for providing *B. fragilis* strains used in this study. This work was supported by grants-in-aid for Scientific Researches (C) (to T.K.) from the Ministry of Education, Science, Sports, Culture and Technology of Japan.

REFERENCES

1. Kuwahara T, Yamashita A, Hirakawa H, Nakayama H, Toh H, Okada N, Kuhara S, Hattori M, Hayashi T, Ohnishi Y : Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. Proc Natl Acad Sci USA 101 : 14919-14924, 2004
2. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI : A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. Science 299 : 2074-2076, 2003
3. Xu J, Mahowald MA, Ley RE, Lozupone CA, Hamady M, Martens EC, Henrissat B, Coutinho PM, Minx P, Latreille P, Cordum H, Van Brunt A, Kim K, Fulton RS, Fulton LA, Clifton SW, Wilson RK, Knight RD, Gordon JI : Evolution of Symbiotic Bacteria in the Distal Human Intestine. PLoS Biol 5 : 1574-1586, 2007

4. Holdeman LV, Kelley RW, Moore WEC : Family I. *Bacteroideaceae* Pribram 1933, 10[AL], p. 602-662. *In* : N. R. Kreig and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. Williams & Wilkins, Baltimore, Md, 1984
5. Martens EC, Chiang HC, Gordon JI : Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe 4 : 447-457, 2008
6. Roberton AM, Stanley RA : In vitro utilization of mucin by *Bacteroides fragilis*. Appl Environ Microbiol 43 : 325-330, 1982
7. Salyers AA, Vercellotti JR, West SE, Wilkins TD : Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. Appl Environ Microbiol 33 : 319-322, 1977
8. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, Sonnenburg JL : Specificity of polysa*ccharide use* in intestinal *Bacteroides* species determines diet-induced microbiota alterations. Cell 141 : 1241-1252, 2010
9. Thomas F, Hehemann JH, Rebuffet E, Czjzek M, Michel G : Environmental and gut *Bacteroidetes* : the food connection. Front Microbiol 2 (article 93), 2011
10. Xu J, Gordon JI : Inaugural Article : Honor thy symbionts. Proc Natl Acad Sci USA 100 : 10452-10459, 2003
11. Bjursell MK, Martens EC, Gordon JI : Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. J Biol Chem 281 : 36269-36279, 2006
12. Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, Buhler JD, Gordon JI : Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. Science 307 : 1955-1959, 2005
13. Corfield T : Bacterial sialidases--roles in pathogenicity and nutrition. Glycobiology 2 : 509-521, 1992
14. Vimr ER, Kalivoda KA, Deszo EL, Steenbergen SM : Diversity of microbial sialic acid metabolism. Microbiol Mol Biol Rev 681 : 132-153, 2004
15. Müller HE, Werner H : In vitro-Untersuchungen über das Vorkommen von Neuraminidase bei *Bacteroides*-arten. Pathol Microbiol 36 : 135-152, 1970
16. Godoy VG, Dallas MM, Russo TA, Malamy MH : A role for *Bacteroides fragilis* neuraminidase in bacterial growth in two model systems. Infect Immun 61 : 4415-4426, 1993
17. Corfield AP, Wagner SA, Clamp JR, Kriaris MS, Hoskins LC : Mucin degradation in the human colon : production of sialidase, sialate *O*-acetylesterase, *N*-acetylneuraminate lyase, arylesterase, and glycosulfatase activities by strains of fecal bacteria. Infect Immun 60 : 3971-3978, 1992
18. Steenbergen SM, Jirik JL, Vimr ER : YjhS (NanS) is required for *Escherichia coli* to grow on 9-*O*-acetylated *N*-acetylneuraminic acid. J Bacteriol 191 : 7134-7139, 2009
19. King SJ, Hippe KR, Weiser JN : Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by *Streptococcus pneumoniae*. Mol Microbiol 593 : 961-974, 2006
20. Dalia AB, Standish AJ, Weiser JN : Three surface exoglycosidases from *Streptococcus pneumoniae*, NanA, BgaA, and StrH, promote resistance to opsonophagocytic killing by human neutrophils. Infect Immun 78 : 2108-2116, 2010
21. Sambrook J, Frisch EF, Maniatis T : Molecular cloning : a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989
22. Ono T, Akimoto S, Kinouchi T, Kataoka K, Ohnishi Y : Cloning and expression of the *Bacteroides fragilis* YCH46 neuraminidase gene in *Escherichia coli* and *Bacteroides uniformis*. FEMS Microbiol Lett 121 : 153-158, 1994
23. Nakayama-Imaohji H, Hirakawa H, Ichimura M, Wakimoto S, Kuhara S, Hayashi T, Kuwahara T : Identification of the site-specific DNA invertase responsible for the phase variation of SusC/SusD family outer membrane proteins in *Bacteroides fragilis*. J Bacteriol 191 : 6003-6011, 2009
24. Ichimura M, Nakayama-Imaohji H, Wakimoto S, Morita H, Hayashi T, Kuwahara T : Efficient electrotransformation of *Bacteroides fragilis*. Appl Environ Microbiol 76 : 3325-3332, 2010
25. Frantz JC, McCallum RE : Growth yields and fermentation balance of *Bacteroides fragilis* cultured in glucose-enriched medium. J Bacteriol 137 : 1263-1270, 1979
26. Bayley DP, Rocha ER, Smith CJ : Analysis of *cepA* and other *Bacteroides fragilis* genes reveals a unique promoter structure. FEMS

Microbiol Lett 193 : 149-154, 2000
27. Traving C, Schauer R : Structure, function and metabolism of sialic acids. Cell Mol Life Sci 54 : 1330-1349, 1998
28. Varki A : Sialic acids in human health and disease. Trends Mol Med 14 : 351-360, 2008
29. Hooper LV, Gordon JI : Glycans as legislators of host-microbial interactions : spanning the spectrum from symbiosis to pathogenicity. Glycobiology 11 : 1R-10R, 2001
30. Lozupone CA, Hamady M, Cantarel BL, Coutinho PM, Henrissat B, Gordon JI, Knight R : The convergence of carbohydrate active gene repertoires in human gut microbes. Proc Natl Acad Sci USA 105 : 15076-15081, 2008
31. Rinninger A, Richet C, Pons A, Kohla G, Schauer R, Bauer HC, Zanetta JP, Vlasak R : Localisation and distribution of *O*-acetylated *N*-acetylneuraminic acids, the endogenous substrates of the hemagglutinin-esterases of murine coronaviruses, in mouse tissue. Glycoconj J 23 : 73-84, 2006
32. Robbe C, Capon C, Maes E, Rousset M, Zweibaum A, Zanetta JP, Michalski JC : Evidence of regio-specific glycosylation in human intestinal mucins : presence of an acidic gradient along the intestinal tract. J Biol Chem 278 : 46337-46348, 2003
33. Robbe-Masselot C, Maes E, Rousset M, Michalski JC, Capon C : Glycosylation of human fetal mucins : a similar repertoire of *O*-glycans along the intestinal tract. Glycoconj J 26 : 397-413, 2009
34. Schauer R, Reuter G, Stoll S : Sialate *O*-acetylesterases : key enzymes in sialic acid catabolism. Biochimie 70 : 1511-1519, 1988
35. Namavar F, Theunissen EB, Verweij-Van Vught AM, Peerbooms PG, Bal M, Hoitsma HF, MacLaren DM : Epidemiology of the *Bacteroides fragilis* group in the colonic flora in 10 patients with colonic cancer. J Med Microbiol 29 : 171-176, 1989
36. Myers LL, Shoop DS, Stackhouse LL, Newman FS, Flaherty RJ, Letson GW, Sack RB : Isolation of enterotoxigenic *Bacteroides fragilis* from humans with diarrhea. J Clin Microbiol 25 : 2330-2333, 1987
37. Kato N, Liu CX, Kato H, Watanabe K, Tanaka Y, Yamamoto T, Suzuki K, Ueno K : A new subtype of the metalloprotease toxin gene and the incidence of the three *bft* subtypes among *Bacteroides fragilis* isolates in Japan. FEMS Microbiol Lett 182 : 171-176, 2000
38. Kuhara T, Okada K, Kinouchi T, Tominaga T, Hirose M, Kagami S, Ohnishi Y : Hemolytic uremic syndrome in a patient with *Bacteroides fragilis* infection. Jpn J Infect Dis 72 : 162-165, 1998
39. Shoemaker NB, Getty C, Guthrie EP, Salyers AA : Regions in *Bacteroides* plasmids pBFTM10 and pB8-51 that allow *Escherichia coli-Bacteroides* shuttle vectors to be mobilized by IncP plasmids and by a conjugative *Bacteroides* tetracycline resistance element. J Bacteriol 166 : 959-965, 1986
40. Valentine PJ, Shoemaker NB, Salyers AA : Mobilization of *Bacteroides* plasmids by *Bacteroides* conjugal elements. J Bacteriol 170 : 1319-1324, 1988